



ARTICLE

A Cover-Independent Deep Image Hiding Method Based on Domain Attention Mechanism

Nannan Wu¹, Xianyi Chen^{1,*}, James Msughter Adeke² and Junjie Zhao²

¹School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China

²School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, 210044, China

*Corresponding Author: Xianyi Chen. Email: xianyi_chen@nuist.edu.cn

Received: 23 August 2023 Accepted: 21 November 2023 Published: 26 March 2024

ABSTRACT

Recently, deep image-hiding techniques have attracted considerable attention in covert communication and high-capacity information hiding. However, these approaches have some limitations. For example, a cover image lacks self-adaptability, information leakage, or weak concealment. To address these issues, this study proposes a universal and adaptable image-hiding method. First, a domain attention mechanism is designed by combining the Atrous convolution, which makes better use of the relationship between the secret image domain and the cover image domain. Second, to improve perceived human similarity, perceptual loss is incorporated into the training process. The experimental results are promising, with the proposed method achieving an average pixel discrepancy (APD) of 1.83 and a peak signal-to-noise ratio (PSNR) value of 40.72 dB between the cover and stego images, indicative of its high-quality output. Furthermore, the structural similarity index measure (SSIM) reaches 0.985 while the learned perceptual image patch similarity (LPIPS) remarkably registers at 0.0001. Moreover, self-testing and cross-experiments demonstrate the model's adaptability and generalization in unknown hidden spaces, making it suitable for diverse computer vision tasks.

KEYWORDS

Deep image hiding; attention mechanism; privacy protection; data security; visual quality

1 Introduction

In recent years, image hiding based on deep learning has gained significant attention owing to its potential applications in various fields, such as secure communications [1], digital watermarking [2], and steganography [3]. The prevailing approaches in this field involve embedding a secret image within a specific region [4–6], which generally refers to texture regions or high-frequency regions of the cover image. However, these conventional techniques have several limitations that hinder their practicality and effectiveness. A key drawback of the current approach is the lack of self-adaptability in the cover image [7]. This deficiency often renders the cover image incompatible with specific application requirements, thereby limiting its use in real-world scenarios. For example, in applications



where the cover image is expected to seamlessly blend into a specific environment, such as digital advertising or camouflage systems, the lack of self-adaptability hinders concealment effectiveness and increases detection risk. This limitation necessitates the development of novel techniques that can be adapted for various application contexts. Furthermore, existing methods do not adequately consider the correlation between cover and secret images during the embedding process [4,6,8]. As a result, there is a risk of information leakage and weak concealment, jeopardizing the security and robustness of the hidden image. The correlation between the cover and secret images can provide valuable contextual information, which improves the effectiveness of the hiding process [9,10]. Ignoring this correlation limits the ability to achieve high levels of concealment and increases the vulnerability of hidden information to attacks and unauthorized extraction.

To address the pressing need to achieve an optimal balance between image quality preservation and hiding capacity enhancement. By analyzing the relationship between the secret image domain and the cover image domain, the proposed approach combines the attention mechanism of the mixed domain and the Atrous convolution. To create a highly adaptable method that does not rely on a cover image, the attention mechanism selects and manipulates regions within the cover image to enhance concealment and improve robustness. Our method builds on the learning technique of secret images in the existing general depth-hiding model [11] and incorporates the attention mechanism of mixed domain and Atrous convolution. This combined approach significantly improves the reconstruction performance of the cover secret information independent of the image-to-image network. This was achieved by improving the secret feature reconstruction module and incorporating Atrous convolution to enhance the sensitivity field of the hidden image recovery network. By leveraging the sparse sampling of secret information from images containing hidden information, we successfully preserved the spatial structure of the secret image while extracting high-level semantic features.

Moreover, recent advancements in deep learning have demonstrated the effectiveness of incorporating perceptual loss into image-generation tasks [12]. Perceptual loss, which is derived from deep neural networks, captures high-level semantic information and encourages the embedded image to align with human visual perception, resulting in more visually plausible and imperceptible hidden images. By adding perceptual loss to the training process, the proposed method improves the similarity perceived by humans between the embedded secret image and the cover image, further improving the overall concealment effectiveness and ensuring a visually convincing output.

To evaluate the feasibility and effectiveness of the proposed algorithm, a comprehensive set of experiments was carried out. The experimental results demonstrate promising outcomes, with the proposed method achieving a peak signal-to-noise ratio (PSNR) value of 40.72 dB between the stego and cover images. Moreover, self-testing and cross-experiments were performed to validate the adaptability and generalization of the proposed model in the presence of unknown hidden spaces, thereby confirming its applicability to diverse computer vision tasks.

In summary, the major contributions of this study are as follows:

- Proposed a cover-independent image-to-image hiding method based on a mixed-domain attention mechanism, it incorporates a secret redundant noise filtering module, which improves the feature extraction ability and adaptability of secret images.
- Introduced the application of the mixed-domain attention mechanism to efficiently compress essential secret information within images, thereby enhancing hiding efficiency and improving the concealment of secret data in various cover images.

- Introduced Atrous convolution enhances sensitivity, preserves the spatial structure, and captures high-level semantics, significantly improving the capability of cover-independent image-to-image networks to reconstruct hidden information.

The remainder of this paper is organized as follows. [Section 2](#) provides related works. In [Section 3](#), we present a cover-independent framework based on a mixed-domain attention mechanism with detailed components. [Section 4](#) describes the experimental setup, including the parameter settings with software components, dataset selection, performance evaluation metrics, and details of the experimental results and analysis. Finally, [Section 5](#) presents the conclusion of the work presented in this paper.

2 Related Works

The field of image hiding has witnessed significant advances in recent years, driven by the application of deep learning techniques. These techniques offer promising solutions in various domains, including secure communication, digital watermarking, and steganography. Traditional image-hiding methods, which rely on manual design distortion cost functions, were extensively studied before 2017. However, this review focuses on the emerging trend of using deep learning for image mapping and storage.

2.1 *The Method of Embedding Images-in-Images Based on Cover*

To expand the transmission capacity of secret images in image-hiding tasks, researchers initially explored the manual design of distortion cost functions to improve the hiding performance [13–17]. In a groundbreaking study in 2017, Baluja [18] introduced deep learning to image-hiding tasks for the first time. By leveraging convolutional neural networks, the author developed a paired neural network structure capable of embedding secret color images into color cover images of the same size. The experimental findings revealed that the secret image information was distributed across all pixel bits of the cover image, rather than being confined to the least significant bit (LSB). Moreover, modifying a single pixel affects the number of bits in the seven surrounding pixels. This breakthrough has opened new possibilities for image information hiding, garnering significant attention, and driving further research in this domain.

Subsequent studies [4,5] explored secret images hiding within different spatial dimensions of cover images, such as the RGB space and the YUV space. These approaches aim to preserve the quality of secret images while maximizing the hiding capacity. This is in contrast to earlier methods, which relied on manually designed feature functions. For example, Filler et al. [19], and Holub et al. [20] achieved a substantial improvement in the hiding capacity, which was nearly 60 times greater. However, it is crucial to note that higher hiding capacities often occur at the expense of security [18]. As the number of hidden secret images increases, the risk of being detected by potential attackers also increases. Therefore, this study primarily focuses on the case of hiding a single secret image to strike a balance between the hiding capacity and security.

2.2 *Methods Based on Cover Generation to Hide Secret Images*

Methods based on cover generation were introduced to generate cover images using generative adversarial networks (GANs) [21,22]. Since the pioneering work of Goodfellow et al. [23] in 2014, GANs have emerged as powerful tools in various computer vision domains. Volkhonskiy et al. [24] were the first to combine image hiding with GANs, using the generator component to simulate random noise suitable for the input cover image in the image hiding task. Attention-Driven Binary Hiding

(ADBH) [8] is another method that leverages GANs and attention masks to achieve directional concealment of secret information in specific regions of the cover images. ADBH first employs an attention mechanism to generate a cover image attention mask. Subsequently, by adding the pixel values of the original secret image and the attention mask, the authors synthesized a secret image that resembled the cover image. Finally, a separate generator was utilized to recover the secret image directly from the synthesized image. The recovery loss function constrains the discrepancy between the generated secret image and the recovered secret image, ensuring that the secret image can be regenerated from the target image. The integration of GAN image-hiding tasks offers a convenient pathway, making image hiding akin to image translation, and improving the concealment of secret images.

Although researchers have made significant contributions to image-hiding tasks using cover-based embedding or cover-generated methods, these approaches typically require encoding secret images based on specific cover images for practical applications [25,26]. Furthermore, most researchers have focused on designing efficient network modules and complex network structures to enhance the transmission capacity and concealment of secret information in covert communication [27]. However, they often overlook the underlying reason for the success of image hiding, which lies in the distinct frequency differences between the features of the secret image (after removing redundant information) and the cover image [11]. These frequency differences enable cover-independent image hiding.

3 Methodology

3.1 Overall Framework

The key distinction between cover-independent and conventional approaches lies in the fact that the cover image is unknown and independent. In the conventional approach, the cover image is predetermined, whereas in the cover-independent approach, the identity of the cover image remains undisclosed. This fundamental difference introduces complexities in adaptively concealing crucial secret information within the cover image, thereby limiting the exploration of the hidden feature representation space of the secret image. On the other hand, the task of hiding the image, independent of the cover image, operates under a different paradigm. It involves encoding the cover and secret images without requiring a conjugated attribute, which is often referred to as conjugated [28]. Fundamentally, the encoding of a secret image is performed independently of any specific cover image. This pioneering approach allows for the decoupling of the hidden secret image from its reliance on a particular cover image.

As a result, this decoupling opens new possibilities for the development and advancement of adaptive coding techniques. It provides flexibility to adapt the coding process to different cover images, leading to improved efficiency and enhanced adaptability in the encoding and decoding of secret images. Fig. 1 illustrates the overall framework of the cover-independent image hiding network (CIIHN) based on the mixed domain attention mechanism. The CIIHN framework consists of two key modules: the secret image self-coding network and the secret image feature recovery network. These interconnected modules synergistically facilitate efficient image-hiding and recovery processes. In the following subsections, we provide an in-depth explanation of each module.

3.1.1 Secret Image Self-Coding Network

In information theory, information entropy quantifies the reduction in the uncertainty. The law of entropy increase states that isolated systems tend to increase their entropy [29]. In image processing, the process of hiding a secret image in a cover image can be considered as an increase in entropy,

where the information entropy of the secret image surpasses that of the cover and hidden images. The challenge is to manage this entropy increase effectively. To address this challenge, we focus on improving the adaptability of secret image hiding. Redundancy exists in feature representations within the image space, including spatial, temporal, visual, information entropy, and structural redundancies. By filtering redundant features in the channel and spatial domains, the information volume of the secret image is reduced, allowing accurate extraction of essential information and enhancing the adaptability to different cover images. Secret image feature extraction involves a subnetwork structure with five-stage convolutional modules as shown in Fig. 2. The instance normalization and leakyReLU [30] activation function described in Eq. (1) is applied to preserve the image details, resulting in a 512-dimensional feature tensor in the image space of the input sample.

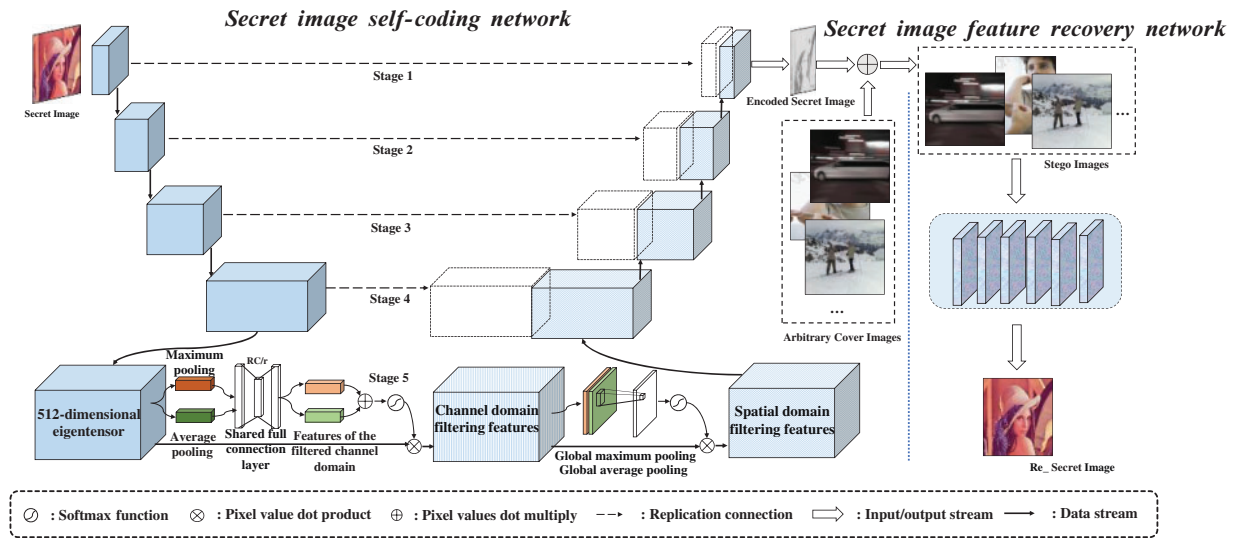


Figure 1: The overall framework of the cover independent image hiding network (CIIHN) based on the mixed domain attention mechanism

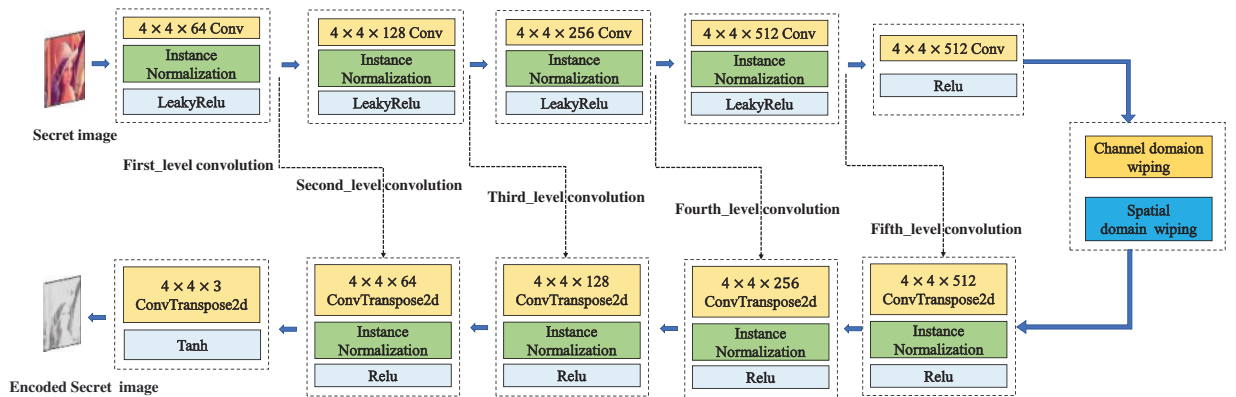


Figure 2: The detailed workflow of secret image self-coding network

$$\text{Prep}_{\text{down}}(s) = \text{leakyRelu}(\text{IN}(\text{conv}(s))) \quad (1)$$

where $\text{Prep}_{\text{down}}(s)$ refers to the preprocessed secret feature of the secret image s after down-sampling, and the feature tensor has a dimensionality of 512. conv represents a convolutional operation, IN denotes instance normalization, which operates on a single channel of the image and normalizes the mean of that channel for pixel-level feature processing, and leakyRelu denotes the activation function employed in our approach.

To enhance adaptability, redundant feature filtering is performed in the channel and spatial domains. Channel redundancy features are learned using a feed-forward neural network with a hidden layer, whereas spatial domain redundancy is suppressed using pooling operations and convolutions. The resulting hybrid domain feature filtering module effectively characterizes the essential information of the secret image. Residual connections are utilized to preserve the underlying feature information. The output of the secret image self-coding network represents a strongly encoded image suitable for tasks where the cover image is independent of the image used for hiding. A specific training process was implemented to allow training iterations that were independent of the cover images. The sampling operations for the cover and secret images in each image space domain are mutually exclusive. These efforts aim to enhance the adaptability and efficiency of secret image-hiding techniques in various scenarios. The filter coefficient of the channel redundancy features is calculated as shown in Eq. (2).

$$1 - M_c(F) = 1 - \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (2)$$

where $1 - M_c(F)$ denotes the redundant feature information after removing redundant information, σ represents the activation function, and F is the 512-dimensional feature tensor after downsampling, which is the weight coefficient of the vector dimension $1 \times 1 \times C$. This coefficient is the dot product of a 512-dimensional feature tensor obtained from the feature extraction part to remove secret redundant features in the channel domain.

3.1.2 Secret Image Feature Recovery Network

The process of eliminating redundant features from secret images is carried out to mitigate the increase in entropy [29]. During extensive training, the neural network may tend to compensate secret images with lower information content from the secret image domain. Thereby reducing their dependence on the macro input for the image-hiding task. However, as the macro-input undergoes processing through various components of the secret image self-coding network, the secret information contained within it undergoes further reduction. As a result, the task content involving secret images containing confidential information is diminished, making it more challenging to reconstruct these secret images using the secret image recovery network.

To address these challenges, we employed Atrous convolution to enhance the receptive field of the hidden image recovery network, enabling it to extract multiscale depth features from hidden images. Moreover, to preserve the spatial structure of the recovered secret image, we introduce Atrous ratios of 1, 2, 5, 1, 2, and 5 in the secret image recovery network distributed in a saw tooth pattern [31]. This form of Atrous convolution significantly expands the receptive field of the secret image recovery network, resulting in an exponential improvement in each pixel acquisition degree within the same feature tensor. Consequently, the recovered secret images demonstrate a higher precision and superior quality. The detailed workflow of the secret image feature recovery network is shown in Fig. 3.

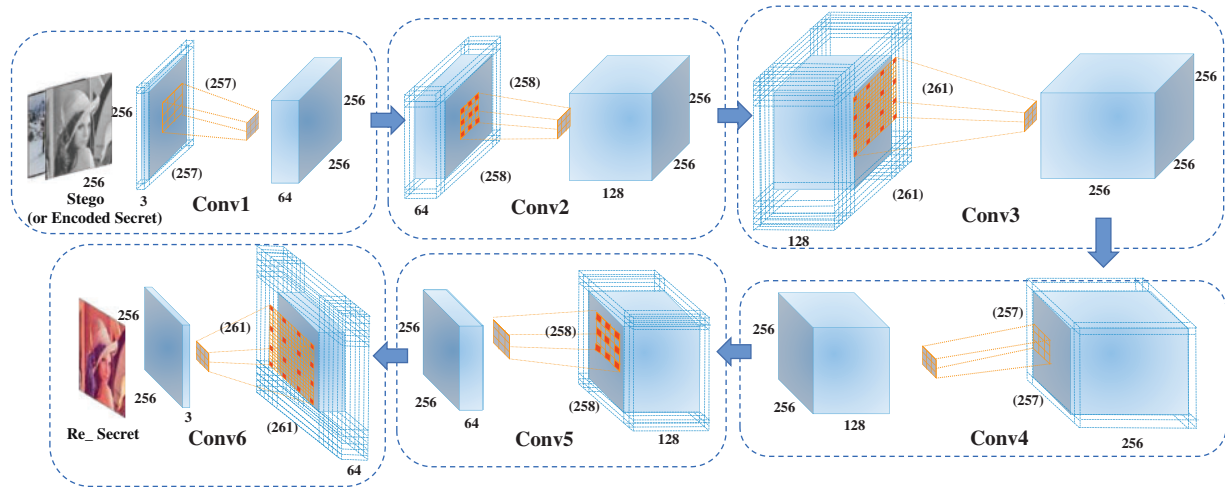


Figure 3: The detailed workflow of the secret image feature recovery network

3.2 Error Propagation and Loss Calculation

The model utilizes the output features from the secret image self-encoding network to obtain a secret image that closely resembles a cover image. This includes the secret coded image, the sum of the tensor obtained by adding pixel values with a random cover image, and the error signal between the feature tensor of the cover image. The error signal is propagated through the neural network layer-by-layer, allowing for error signal adjustment and weight optimization in each layer to minimize pixel-level differences between the secret and cover images. This process is governed by the hidden loss, as expressed in Eq. (3).

$$\text{encoded}_{\text{secret}} = \sum_{i=0}^N (\text{SA} \cdot (\text{M}_c(\text{F}) \cdot \text{F})_i \oplus (\text{leakyRelu}(\text{IN}(\text{conv}(\text{s}))))_i) \quad (3)$$

where i represents the coding phase, whose range is $[0,5]$. The SA denotes the spatial domain redundancy feature filtering coefficient, $(\text{M}_c(\text{F}) \cdot \text{F})$ corresponds to the image features after removing redundant information, and \oplus represents the pixel value addition operation.

In the image space U , which comprises the cover image domain $R^{H \times W \times C_s}$ and secret image domain $R^{H \times W \times C_s}$, two methods are used to recover the secret image without relying on the cover image.

$$\mathcal{L}_{\text{con}(\theta)} = \sum_{i=1}^N \|\mathbf{H}^{(i)} - (\mathcal{H}(\mathbf{S}^{(i)}) \oplus \mathbf{H}^{(i)})\|_2 \quad (4)$$

where \mathbf{S} represents the secret image, i denotes the secret image feature at the i -th convolutional layer, \mathbf{H} represents the carrier image, and $\mathcal{H}(\cdot)$ denotes the encoding process of the secret image. $\mathcal{L}_{\text{con}(\theta)}$ denotes the value of the loss function that needs to be calculated during the hiding process, \mathbf{H} corresponds to the coding characteristics of the cover image and $\mathcal{H}(\mathbf{S}^{(i)})$ represents the secret coding features.

The first method involves reconstructing the secret image directly from its representation, which is denoted as $\mathbf{S}' = \mathcal{R}(\mathcal{H}(\mathbf{S}))$ as shown in Eq. (4), and the second method involves recovering the secret image from the secret encoded image expressed as $\mathbf{S}' = \mathcal{R}((\mathcal{H}(\mathbf{S}) \oplus \mathbf{H}))$. To ensure accurate recovery, the hidden image recovery network is meticulously trained to minimize recovery loss. This

loss function effectively measures the discrepancy between the recovered secret image and the original secret image, as shown in Eq. (5).

$$\mathcal{L}_{rev(\theta)} = \begin{cases} \sum_{i=1}^N \|S^{(i)} - (\mathcal{H}(S^{(i)}) \oplus H^{(i)})\|_2, & \text{recovered from } \mathcal{H}(S) \\ \sum_{i=1}^N \|S^{(i)} - \mathcal{H}(S^{(i)} \oplus H^{(i)})\|_2, & \text{recovered from } \mathcal{H}(S^{(i)} \oplus H^{(i)}) \end{cases} \quad (5)$$

where $\mathcal{L}_{rev(\theta)}$ denotes the value of the loss function calculated during the reconstruction phase, $S^{(i)}$ corresponds to the i^{th} secret image selected from sets and $\mathcal{H}(S^{(i)})$ represents the secret coding features after the secret image coding.

The weight of the secret image recovery network is not influenced by the error term between the original cover image and the secret image. This is because the objective of the secret image restoration network is to enhance the similarity between the secret images before and after they are hidden in the cover image rather than directly restoring the original cover image. However, all the weight parameters of the network components are affected by the error term between the original secret image and the recovered secret image, as each component plays a role in preserving and transmitting information about the secret image.

Traditional loss functions, such as the L_2 norm, only consider the pixel-level similarity between the secret images and cover image, as well as the secret image and recovered secret image, without addressing the similarity at the feature level. To overcome this limitation, we introduced perceptual loss, which is a technique commonly used in image style transfer algorithms. This constraint function can be formally expressed as shown in Eqs. (6) and (7) for the hidden perceptual loss and reconstructed perceptual loss, respectively.

$$L_{pips}(C, C') = \sum_l \frac{1}{H_l \times W_l} \sum_{h,w} \|w_l \odot (C - C')\|_2^2 \quad (6)$$

where (C, C') denotes the cover image and the stego image, $H_l \times W_l$ corresponds to the layer of l , and $\|w_l \odot (C - C')\|_2^2$ is the feature tensor after the perception weight and difference image dot multiplication.

By leveraging the feature values extracted by a convolutional neural network, perceptual loss measures the difference between the generated image and the target image at the feature level, thus improving visual similarity.

$$L_{pips}(S, S') = \sum_l \frac{1}{H_l \times W_l} \sum_{h,w} \|w_l \odot (S - S')\|_2^2 \quad (7)$$

where $H_l \times W_l$ is the height of the feature tensor in the l^{th} feature layer and W_l is the adaptive weight of the image feature values, respectively. \odot represents the XOR operation. By calculating the distance between the corresponding feature tensors layer by layer and continuously reducing the distance, the image perception similarity was improved.

$$\mathcal{L}_{loss} = \lambda \mathcal{L}_{con(\theta)} + \alpha \mathcal{L}_{rev(\theta)} + \gamma (L_{pips}(C, C') + L_{pips}(S, S')) \quad (8)$$

where $\lambda \mathcal{L}_{con(\theta)}$ denotes the hidden loss with a tradeoff factor, $\alpha \mathcal{L}_{rev(\theta)}$ correspond to the reconfiguration losses with trade-offs, and $\gamma (L_{pips}(C, C') + L_{pips}(S, S'))$ is the Pixel level perceptual loss.

Therefore, our model is trained using hidden loss (λ), recovery loss (α), and perceived loss (γ) as total compatibility measures and hyperparameters. The setting of the hyperparameter α was inspired

by Baluja [18] and is set to 0.75. Finally, the gradient descent algorithm is used to update the weights until the model converges.

4 Experimental Results and Discussion

4.1 Experimental Setup

In the experiment, we employed the Adam optimizer [32] for the cover-independent image-hiding model. The initial learning rate was set to 1×10^{-3} , and a reduction of 20% was applied to the learning rate when the loss value did not decrease for 8 consecutive epochs. The batch size for the cover image and secret image was set to 4 samples each. In a batch, 4 cover image samples were randomly selected from the cover image domain, and 4 secret image samples were chosen from the secret image domain. The selection of secret and cover images was mutually exclusive for each batch. During training, the dataset was shuffled, and four samples were chosen sequentially until the entire training set was traversed. Detailed hyper-parameter settings employed in the experiments are listed in Table 1.

Table 1: Hyper-parameter settings

Parameter	Value
Epoch	80
Learning rate	0.001
Decay	10
Normalization	Instance
Iteration per epoch	2000
Batch size	4

The experimental platform consisted of a Linux system with an NVIDIA GeForce RTX3080 graphics card. The IDE used was VS-code, and the programming framework used was PyTorch 1.7, CUDA version 10.1.

4.2 Datasets and Performance Measurements

4.2.1 Datasets

During the experiments, the proposed method was trained and validated using the ImageNet dataset [33]. The ImageNet dataset contains over 1500 high-resolution images with labels for 22000 different image categories. Each image in the dataset underwent meticulous manual screening and labeling. To create the training and validation sets, we employed a PyTorch data sampler, which randomly selects samples from the ImageNet dataset. Before training, all the sample images were resized uniformly to dimensions of $256 \times 256 \times 3$ to ensure compatibility with the vector image-independent model. Furthermore, we employed the DIV2K [34] and COCO [35] datasets to verify the generalization ability of the proposed algorithm.

4.2.2 Performance Measurements

To evaluate the efficacy of the proposed method, the following performance measurement metrics were used to appraise image quality: peak signal-to-noise ratio (PSNR) [36], structural similarity index

(SSIM) [37], average pixel discrepancy (APD) [11], and learned perceptual image patch similarity (LPIPS) [38].

- **PSNR:** The PSNR quantifies the difference in the peak signal-to-noise ratio between two images, i.e., the cover image and stego image, as well as the secret image and recovered secret image. The higher the ratio, the better the image quality. This is mathematically expressed by Eq. (9).

$$\text{PSNR} = 10 \log_{10} \frac{255^2}{MSE} \quad (9)$$

where $10 \log_{10}$ denotes the correlation coefficient and MSE represents the mean square error of the pixels between images.

- **SSIM:** The SSIM measures the similarity between two images, that is, the stego image and cover image, and is formulated by Eq. (10).

$$\text{SSIM} = \frac{(2\mu_x\mu_y + c1)(2\sigma_{xy} + c2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c2)} \quad (10)$$

where μ_x, σ_x^2 is the average and variance of x , μ_y, σ_y^2 is the average and variance of y , σ_{xy} is the covariance of x and y , $c1$, and $c2$ are constant with values $c1 = m_1L^2$ and $c2 = m_2L^2$ with m_1 and m_2 having values of 0.001 and 0.003, respectively and L is the range of pixels.

- **APD:** This metric concisely measures the prediction error of the regression model. Because the image-hiding task can be viewed as a regression task, the cover and secret images are treated as the ground truth, while the secret image and the secret recovery image serve as the predicted values. This is mathematically formulated as shown in Eq. (11).

$$\text{APD} = \frac{1}{n} \sum_{i=1}^N |f(x_i) - y_i| \quad (11)$$

where $\frac{1}{n}$ denotes the coefficient of pixel difference, and $|f(x_i) - y_i|$ represents the absolute value between the input and output terms.

- **LPIPS:** Because our method incorporates perceptual loss in the training to optimize the model, to assess the similarity of human visual perception between images, we introduced a novel performance measurement metric called the LPIPS to evaluate the effectiveness of the image hiding task by considering the perception of hidden images-in-images. It is defined by Eq. (12).

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\hat{y}'_{hw} - \hat{y}'_{0hw})\|_2^2 \quad (12)$$

where d denotes the spatial distance between points x_0 and x . The model procures a feature stack from the L-layer, subsequently applying unit normalization within the channel dimension. Vector w_l serves the purpose of diminishing the count of active channels, enabling the computation of the L_2 distance. Ultimately, spatial averaging is performed in conjunction with channel summation.

4.3 Results and Analysis

In this section, we conduct comparison experiments with the conventional approaches in four respects: (1) effectiveness of redundancy removal, (2) effectiveness of perceived loss, (3) validity of Atrous convolution, and (4) generalization ability analysis.

4.3.1 Effectiveness of Redundancy Removal

We conducted several attempts to remove redundant information from the secret image (Table 2). Initially, we encoded and transformed the secret image into channel dimensions to reduce the required number of bits. By utilizing the local cross-channel interaction strategy (ECA) without dimensionality reduction [39] and fast one-dimensional convolution, we effectively selected essential information from the channel dimension, resulting in reduced redundancy.

Table 2: Evaluating the effectiveness of mixed attention mechanisms on overall performance

	Errors	APD	PSNR	SSIM	LPIPS	Δ PSNR
Baseline network	UDH C	2.51	36.22	0.982	0.0001	–
	UDH S (from_C)	4.53	32.49	0.938	0.0850	–
	UDH S (from_Se)	2.64	36.49	0.983	0.0150	–
Local cross-channel interaction	UDH C	2.78	37.51	0.978	0.0004	+1.29
	UDH S (from_C)	2.88	36.21	0.977	0.0160	+3.72
	UDH S (from_Se)	2.63	37.04	0.982	0.0101	+0.55
CIHN (Ours)	UDH C	1.83	40.72	0.985	0.0002	+4.50
	UDH S (from_C)	2.54	36.73	0.983	0.0159	+4.24
	UDH S (from_Se)	2.06	38.54	0.988	0.0074	+2.05

The experimental data in Table 2 validate the effectiveness of the proposed method, where ‘UDH C,’ ‘UDH S (from_C),’ and ‘UDH S (from_Se)’ respectively represent: the metrics (APD, PSNR, SSIM, LPIPS) between the carrier image and the hidden image; the metrics (APD, PSNR, SSIM, LPIPS) between the secret image and the reconstructed secret image from the carrier image; and the metrics (APD, PSNR, SSIM, LPIPS) between the secret image and the reconstructed secret image from the secret encoded image. The second row of Table 2 demonstrates the positive impact of these attempts, with increased numerical indices between the secret image and the recovered secret image, as well as between the cover image and the stego image by 1.29 and 3.72 dB, respectively. However, the local cross-channel interaction strategy focuses only on the channel dimension and fails to fully explore the spatial representation ability of the secret information. To overcome this limitation, our algorithm removes redundant features from both the channel and spatial domains, enabling a more convenient hiding of secret information in any cover image. The effectiveness of this approach is confirmed by the experimental results in the third row of Table 2.

4.3.2 Effectiveness of Perceptual Loss

The primary objective of hiding images-in-images is to minimize pixel-level differences using a similarity measurement function. The basic loss function achieves this by constraining pixel-level Euclidean distance. However, it only considers local pixel differences and overlooks overall image

dissimilarity. To address this, we incorporated perceptual loss into the model training, enhancing the perceptual similarity between the secret image and the recovered secret image, as well as between the cover image and the secret image. By adding perceptual loss, we achieved a significant reduction in perceptual similarity between the secret image and recovered secret image (60% reduction) and between the secret image recovered from the encoded image and the original secret image (21.3% reduction). These results shown in Table 3, highlight the effectiveness of incorporating perceptual loss. This approach strikes a balance between the cover image and the secret image without relying solely on metrics such as PSNR and LPIPS. By improving perceptual similarity, we enhanced the overall performance of the image-hiding task.

Table 3: Evaluation of the effectiveness of perceptual loss overall performance

	Errors	APD	PSNR	SSIM	LPIPS
Base loss	UDH C	2.51	36.22	0.982	0.0001
	UDH S (from_C')	4.53	32.49	0.938	0.0850
	UDH S (from_Se)	2.64	36.49	0.983	0.0150
Base loss ++ LPIPS (S,S)	UDH C	2.69	37.88	0.975	0.0054
	UDH S (from_C')	2.92	35.87	0.974	0.0340
	UDH S (from_Se)	2.36	37.63	0.986	0.0118
Base loss ++ LPIPS (C,C')	UDH C	2.78	37.38	0.972	0.0006
	UDH S (from_C')	2.72	36.29	0.986	0.0207
	UDH S (from_Se)	2.54	36.70	0.992	0.0090
CIHN (Ours)	UDH C	1.83	40.72	0.985	0.0002
	UDH S (from_C')	2.54	36.73	0.983	0.0159
	UDH S (from_Se)	2.06	38.54	0.988	0.0074

4.3.3 The Validity of Atrous Convolution

In this experiment, we aimed to validate the efficacy of Atrous convolution [31] in the secret image recovery network. To achieve this, we incorporate Atrous convolution into the deep secret image recovery module of the deep hidden network, which we refer to as the basic model. The results of these experiments are listed in Table 4. The performance of the model deteriorates when Atrous convolution is applied solely to the secret image recovery network of the baseline model.

Table 4: Evaluating the effectiveness of the Atrous convolution on the overall performances

	Errors	APD	PSNR	SSIM	LPIPS
Baseline	UDH C	2.51	36.22	0.982	0.0001
	UDH S (from_C')	4.53	32.49	0.938	0.0850
	UDH S (from_Se)	2.64	36.49	0.983	0.0150

(Continued)

Table 4 (continued)

	Errors	APD	PSNR	SSIM	LPIPS
Baseline + Atrous	UDH C	2.78	37.38	0.972	0.0006
CIHN (Ours)	UDH S (from_C')	2.72	36.29	0.986	0.0207
	UDH S (from_Se)	2.54	36.70	0.992	0.0090
Baseline + Cbam +	UDH C	1.83	40.72	0.985	0.0002
LPIPS ₁ + LPIPS ₂ +	UDH S (from_C')	2.54	36.73	0.983	0.0159
Atrous CIHN (Ours)	UDH S (from_Se)	2.06	38.54	0.988	0.0074

This declination can be attributed to the inherent nature of Atrous convolution, which employs sparse samples of image pixels at a specified dilation rate, as opposed to utilizing all pixels for conventional convolution operations. Consequently, when Atrous convolution is applied solely to the secret image recovery network without reinforcing the secret information, it leads to information loss during the recovery process. Moreover, the accuracy of the final model was lower than that of the baseline. Therefore, as indicated by the results in row 3 of Table 4, the model's performance reaches optimal levels only when the Atrous convolution is utilized in conjunction with the mixed-domain attention mechanism. However, it is important to note that Atrous convolution is not necessarily the optimal solution for enhancing the performance of a cover-independent secret image recovery network within this domain.

4.3.4 Generalization Ability Analysis

To evaluate the generalization ability of the cover-independent image-hiding method, we conducted self-tests and cross-tests on two datasets, DIV2K and COCO. The experimental results presented in Table 5 highlight the performance of the model. By analyzing the first and third lines of the experimental data, we observed that the model effectively accomplished image hiding without relying on specific cover images in both the DIV2K and COCO datasets. The numerical indicators obtained from these experiments were highly compelling, further reinforcing the strong generalization ability of the model. In particular, the model exhibited exceptional adaptability even within the unfamiliar hidden space of the DIV2K dataset, which was originally intended for super-resolution tasks. It successfully conceals secret images within various cover images, including those with a high resolution, vibrant colors, and complex textures. Moreover, the model demonstrated an excellent performance on the COCO dataset.

Table 5: Evaluation of the generalization ability of the proposed method for different datasets

Train	Validation	Errors	APD	PSNR	SSIM	LPIPS
ImageNet	DIV2K	UDH C	2.08	39.50	0.988	0.0002
		UDH S (from_C')	3.85	33.55	0.971	0.0281
		UDH S (from_Se)	2.84	35.81	0.984	0.0088
DIV2K	DIV2K	UDH C	6.44	31.06	0.930	0.0599
		UDH S (from_C')	8.89	26.57	0.858	0.2360

(Continued)

Table 5 (continued)

Train	Validation	Errors	APD	PSNR	SSIM	LPIPS
		UDH S(from_Se)	7.35	27.91	0.897	0.1887
ImageNet	COCO	UDH C	3.21	36.56	0.087	0.0789
		UDH S (from_C')	4.16	34.78	0.945	0.0987
		UDH S (from_Se)	3.89	35.67	0.956	0.0060
COCO	COCO	UDH C	2.62	38.19	0.977	0.0014
		UDH S (from_C')	4.00	32.97	0.968	0.0326
		UDH S (from_Se)	3.34	34.53	0.977	0.0140

An important finding lies in the last line of the experimental data, which reveals a peak signal-to-noise ratio of approximately 38.19 dB between the cover and secret images. This achievement represents a remarkable breakthrough in image hiding without relying on the cover image itself. In other words, the model based on this scheme exhibits notable adaptability when confronted with unknown domain image spaces. Through careful analysis, it is fair to state that the proposed model exhibits robust generalization abilities, demonstrating its potential for real-world applications in various image-hiding scenarios.

One subset of hiding images, denoted as (a), (b), (c), and (d), was discreetly employed during the course of the experiment. After the removal of the hidden redundant attribute through the implementation of the secret self-coding network, cover images (a), (b), (c), and (d) featured in the uppermost row of Fig. 4, undergo a process of concealment. This procedure culminates in the emergence of hidden images, as presented in the second row of Figs. 4a–4d. The visual analysis reveals a strikingly high degree of similarity between the hidden images and their corresponding cover images. This outcome can be attributed to the intrinsic adaptability of the hidden image within the cover-independent image-hiding model, allowing proficient integration of essential hidden information into diverse cover images. Upon traversing the secret image reconstruction network, the hidden image was subjected to restoration.

The discernible divergence between the restored hidden image and its original counterpart, in terms of visual appearance, is minimal, thereby validating the effectiveness of the algorithm. Additionally, the provided summary elucidates pixel disparity images, which elucidate variations between the cover image and the hidden image, as well as between the hidden image and its restored form. These images, featured in the third row of both Figs. 4 and 5, afford insights into the absence of substantial reliance on the image hiding model by the cover image, thus ensuring the preservation of confidential information.

Furthermore, in Fig. 5, there is a significant resemblance between the secret-containing image and the cover image. This high similarity arises from the model design, wherein the cover image remains independent of the secret image, exhibiting a notable adaptability to conceal the secret information more efficiently within any cover image. In the first row, one can observe the presence of secret images (a), (b), (c), and (d), which become recoverable after undergoing the secret image recovery network. The resulting recovered secret image is presented in the second row of the figure.

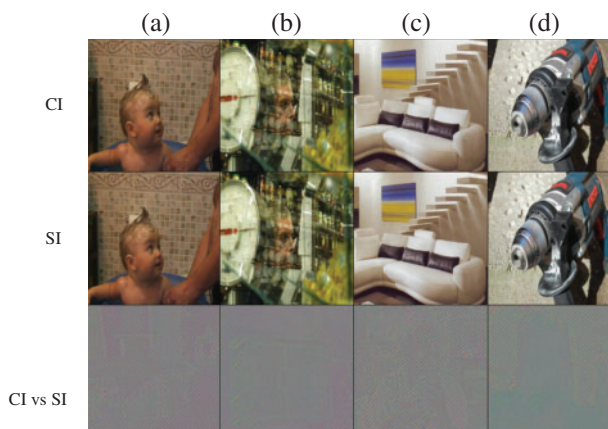


Figure 4: Visual effect obtained by applying the proposed method for secret image hiding. The elements depicted in the figure included the cover image (CI), stego image (SI), and CI vs. SI

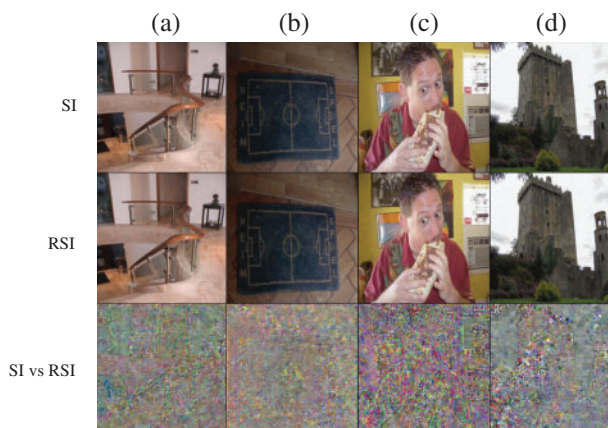


Figure 5: Visual effectiveness of the proposed method for secret image hiding recovery: similarity analysis secret image (SI), recovered secret image (RSI), and SI vs. RSI

In addition, [Table 6](#) offers a rigorous comparative analysis of the experimental findings derived from the evaluation of deep learning image-hiding models using the renowned ImageNet dataset. The primary objective of this analysis was to meticulously assess the performance and efficacy of established image-hiding techniques in concealing sensitive information within images while upholding the overall image quality. The table comprises essential performance metrics that were judiciously selected to evaluate the image-hiding models in terms of precision and comprehensiveness. These critical metrics encompass but are not limited to APD, PSNR, SSIM, and LPIPS. Each row in the table corresponds to specific image-hiding models that were subjected to meticulous evaluation, whereas the columns represent the corresponding results for the aforementioned performance metrics. To provide an insightful overview of the model’s comparative performance, the table also includes aggregated scores for each metric.

Table 6: Comparative analysis of different methods

Method	Errors	APD	PSNR	SSIM	LPIPS
UDH [11]	UDH C	2.35	39.13	0.985	0.0001
	UDH S (from_C')	3.56	35.00	0.976	0.0136
	UDH S (from_Se)	1.98	39.18	0.992	0.0022
UDH [11]	(Cover,Stego)	2.51	36.22	0.982	0.0001
	(Secert,re_secert)(C')	4.53	32.49	0.938	0.0850
	(Secert,Se)(Se)	2.64	36.49	0.983	0.0150
Baluja [18]	(Cover,Stego)	–	39.60	0.980	–
	(Secert,re_secert)	–	32.05	0.910	–
Baluja [40]	(Cover,Stego)	–	36.59	0.952	–
	(Secert,re_secert)	–	34.13	0.925	–
Rehman [41]	(Cover,Stego)	29.60	0.880	31.30	0.880
	(Secert,re_secert)	32.90	0.960	36.60	0.960
HiDDeN [42]	(Cover,Stego)	–	34.79	0.938	–
	(Secert,re_secert)	–	35.70	0.960	–
ISGAN [43]	(Cover,Stego)	–	34.89	0.968	–
	(Secert,re_secert)	–	33.42	0.947	–
ABDH [8]	(Cover,Stego)	–	33.91	0.958	–
	(Secert,re_secert)	–	30.21	0.948	–
HiNet [44]	(Cover,Stego)	–	38.50	0.954	–
	(Secert,re_secert)	–	35.38	0.955	–
CIHN (Ours)	(Cover,Stego)	1.83	40.72	0.985	0.0001
	(Secert,re_secert) (C')	2.54	36.73	0.983	0.0159
	(Secert,re_secert) (Se)	2.06	38.54	0.988	0.0074

This diligently conducted analysis serves as a foundational reference for researchers, practitioners, and stakeholders involved in the domains of image hiding and data security. The methodological findings obtained from this study will contribute significantly to the advancement of image-hiding techniques and aid in the informed selection of suitable models for diverse use cases, including secure communication, data watermarking, and confidential information protection. To further reinforce the robustness and generalizability of the findings, supplementary details concerning the experimental setup, dataset specifications, and statistical significance are deemed essential.

5 Conclusion and Future Work

In conclusion, our research advances hidden image embedding for secure and efficient data transmission via steganography. However, a significant limitation is the disparity in PSNR values between the secret and reconstructed images caused by the inherent convolution characteristics during encoding, which results in information loss in the image features. To address this limitation, our future research will take a multidisciplinary approach, including the investigation of advanced neural architectures, such as attention-based models and GANs, to improve security and minimize information loss while optimizing temporal efficiency for real-time embedding. We also aim to develop robust countermeasures against steganalysis techniques to bolster security, while expanding our approach to encompass diverse multimedia data types, including audio and video, for comprehensive secure communication. By overcoming these challenges, we aim to advance secure data embedding, enhancing both security and efficiency, while extending its applicability across various multimedia platforms, underscoring our dedication to advancing the field of steganography and secure data transmission.

Acknowledgement: The authors would like to express our sincere gratitude and appreciation to each other for our combined efforts and contributions throughout the course of this research paper.

Funding Statement: This work was supported by the National Key R&D Program of China (Grant Number 2021YFB2700900), the National Natural Science Foundation of China (Grant Numbers 62172232, 62172233), and the Jiangsu Basic Research Program Natural Science Foundation (Grant Number BK20200039).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: N. Wu, X. Chen; data collection: J. Zhao; analysis and interpretation of results: N. Wu, X. Chen, J. M. Adeke; draft manuscript preparation: J. M. Adeke, N. Wu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The datasets used in the experiments are cited in the article.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. Rahman, F. Masood, W. U. Khan, N. Ullah, F. Q. Khan *et al.*, “A novel approach of image steganography for secure communication based on LSB substitution technique,” *Computers, Materials & Continua*, vol. 64, no. 1, pp. 31–61, 2020.
- [2] O. Byrnes, W. La, H. Wang, C. Ma, M. Xue *et al.*, “Data hiding with deep learning: A survey unifying digital watermarking and steganography,” arXiv preprint arXiv:2107.09287, 2021.
- [3] P. C. Mandal, I. Mukherjee, G. Paul and B. Chatterji, “Digital image steganography: A literature survey,” *Information Sciences*, vol. 609, pp. 1451–1488, 2022.
- [4] B. Chen, J. Wang, Y. Chen, Z. Jin, H. J. Shim *et al.*, “High-capacity robust image steganography via adversarial network,” *KSI Transactions on Internet & Information Systems*, vol. 14, pp. 366–381, 2020.
- [5] X. Duan, K. Jia, B. Li, D. Guo, E. Zhang *et al.*, “Reversible image steganography scheme based on a U-Net structure,” *IEEE Access*, vol. 7, pp. 9314–9323, 2019.
- [6] X. Liao, Y. Yu, B. Li, Z. Li and Z. Qin, “A new payload partition strategy in color image steganography,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, pp. 685–696, 2019.

- [7] C. Xu, C. Zhang, M. Ma and J. Zhang, "Blind image deconvolution via an adaptive weighted TV regularization," *Journal of Intelligent & Fuzzy Systems*, vol. 44, pp. 1–15, 2023.
- [8] C. Yu, "Attention based data hiding with generative adversarial networks," in *Proc. of the AAAI Conf. on Artificial Intelligence*, New York, HM, USA, pp. 1120–1128, 2020.
- [9] R. Anushiadevi, P. Praveenkumar, J. B. B. Rayappan and R. Amirtharajan, "Uncover the cover to recover the hidden secret—a separable reversible data hiding framework," *Multimedia Tools and Applications*, vol. 80, pp. 19695–19714, 2021.
- [10] J. Qin, Y. Luo, X. Xiang, Y. Tan and H. Huang, "Coverless image steganography: A survey," *IEEE Access*, vol. 7, pp. 171372–171394, 2019.
- [11] C. Zhang, P. Benz, A. Karjauv, G. Sun and I. S. Kweon, "UDH: Universal deep hiding for steganography, watermarking, and light field messaging," *Advances in Neural Information Processing Systems*, vol. 33, pp. 10223–10234, 2020.
- [12] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao *et al.*, "IFCNN: A general image fusion framework based on convolutional neural network," *Information Fusion*, vol. 54, pp. 99–118, 2020.
- [13] T. Pevný, T. Filler and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *The 12th Int. Conf. on Information Hiding*, Calgary, AB, Canada, pp. 161–177, 2010.
- [14] V. Holub and J. Fridrich, "Digital image steganography using universal distortion," in *Proc. of the First ACM Workshop on Information Hiding and Multimedia Security*, Montpellier, France, pp. 59–68, 2013.
- [15] B. Li, S. Tan, M. Wang and J. Huang, "Investigation on cost assignment in spatial image steganography," *IEEE Transactions on Information Forensics and Security*, vol. 9, pp. 1264–1277, 2014.
- [16] T. Denemark and J. Fridrich, "Side-informed steganography with additive distortion," in *2015 IEEE Int. Workshop on Information Forensics and Security (WIFS)*, Rome, Italy, pp. 1–6, 2015.
- [17] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 868–882, 2012.
- [18] S. Baluja, "Deep steganography," *Advances in Neural Information Processing Systems*, vol. 30, pp. 2069–2079, 2017.
- [19] T. Filler and J. Fridrich, "Gibbs construction in steganography," *IEEE Transactions on Information Forensics and Security*, vol. 5, pp. 705–720, 2010.
- [20] V. Holub, J. Fridrich and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, pp. 1–13, 2014.
- [21] J. Liu, Y. Ke, Z. Zhang, Y. Lei, J. Li *et al.*, "Recent advances of image steganography with generative adversarial networks," *IEEE Access*, vol. 8, pp. 60575–60597, 2020.
- [22] Q. Cui, Z. Zhou, Z. Fu, R. Meng, X. Sun *et al.*, "Image steganography based on foreground object generation by generative adversarial networks in mobile edge computing with Internet of Things," *IEEE Access*, vol. 7, pp. 90815–90824, 2019.
- [23] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley *et al.*, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, pp. 53–65, 2014.
- [24] D. Volkhonskiy, B. Borisenko and E. Burnaev, "Generative adversarial networks for image steganography," in *ICLR 2017 Conf.*, pp. 1–8, 2017.
- [25] M. Shyla, K. S. Kumar and R. K. Das, "Image steganography using genetic algorithm for cover image selection and embedding," *Soft Computing Letters*, vol. 3, pp. 100021, 2021.
- [26] C. Li, Y. Jiang and M. Cheslyar, "Embedding image through generated intermediate medium using deep convolutional generative adversarial network," *Computers, Materials & Continua*, vol. 56, no. 2, pp. 313–324, 2018.
- [27] W. Mazurczyk, S. Wendzel, M. Chourib and J. Keller, "Countering adaptive network covert communication with dynamic wardens," *Future Generation Computer Systems*, vol. 94, pp. 712–725, 2019.
- [28] Y. Zhao, S. Wang, L. Jiao and K. Li, "SAR image despeckling based on sparse representation," in *MIPPR 2007: Multispectral Image Processing*, Wuhan, China, pp. 657–664, 2007.
- [29] A. Nikulov, "The law of entropy increase and the Meissner effect," *Entropy*, vol. 24, pp. 83, 2022.

- [30] D. Ulyanov, A. Vedaldi and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," arxiv preprint arXiv:1607.08022, 2016.
- [31] F. Yu, V. Koltun and T. Funkhouser, "Dilated residual networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 472–480, 2017.
- [32] K. Diederik and B. Jimmy, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [33] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [34] B. Lim, S. Son, H. Kim, S. Nah and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, pp. 136–144, 2017.
- [35] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona *et al.*, "Microsoft COCO: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conf.*, Zurich, Switzerland, pp. 740–755, 2014.
- [36] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th Int. Conf. on Pattern Recognition*, Istanbul, Turkey, pp. 2366–2369, 2010.
- [37] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.
- [38] R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 586–595, 2018.
- [39] H. Xue, M. Sun and Y. Liang, "ECANet: Explicit cyclic attention-based network for video saliency prediction," *Neurocomputing*, vol. 468, pp. 233–244, 2022.
- [40] S. Baluja, "Hiding images within images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 1685–1697, 2019.
- [41] A. ur Rehman, R. Rahim, S. Nadeem and S. ul Hussain, "End-to-end trained CNN encoder-decoder networks for image steganography," in *Computer Vision–ECCV 2018 Workshops*: Munich, Germany, pp. 723–729, 2019.
- [42] J. Zhu, R. Kaplan, J. Johnson and F. F. Li, "Hidden: Hiding data with deep networks," in *Proc. of the European Conf. on Computer Vision (ECCV)*, Munich, Germany, pp. 657–672, 2018.
- [43] R. Zhang, S. Dong and J. Liu, "Invisible steganography via generative adversarial networks," *Multimedia Tools and Applications*, vol. 78, pp. 8559–8575, 2019.
- [44] J. Jing, X. Deng, M. Xu, J. Wang and Z. Guan, "HiNet: Deep image hiding by invertible network," in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, Montreal, QC, Canada, pp. 4733–4742, 2021.