Signal, Image and Video Processing (2024) 18:8251–8266
https://doi.org/10.1007/s11760-024-03468-8

ORIGINAL PAPER

Robust coverless image steganography based on DenseUNet with multi-scale feature fusion and attention mechanism

Xiaopeng Li[1] · Qiuyu Zhang[1] · Zhe Li[1]

Received: 7 May 2024 / Revised: 9 July 2024 / Accepted: 26 July 2024 / Published online: 6 August 2024

Abstract

Coverless image steganography (CIS) have attracted considerable attention due to their ability to resist steganalysis detection completely. However, most of the existing CIS methods are weak in robustness to geometric attacks, and it is difficult to achieve a balance between geometric attacks and non-geometric attacks. So, a robust coverless image steganography method based on DenseUNet with multi-scale feature fusion attention mechanism is proposed in this paper. At the sender, an end-to-end hash sequence generation model is designed by combining the DenseUNet network with the multi-scale feature fusion attention mechanism to extract the multi-scale CNN features of the images, and as serve matching benchmarks. Secondly, a hybrid loss function is introduced into the network model for network training to generate hash sequences with robust features. Finally, the secret information is segmented into equal-length segments, and the image whose hash sequence matches the secret information segment is selected as a stego-images using the inverted index. At the receiver, the secret information was recovered from the stego-images using the constructed network model. Experimental results show that the proposed method has stronger performance in terms of robustness and security compared with existing CIS schemes, and achieves enhanced robustness against both the geometric attacks and non-geometric attacks at four different datasets.

Keywords Coverless image steganography · DenseUNet · Multi-scale feature fusion · Attention mechanism · Hybrid loss function

1 Introduction

Steganography, a covert communication technique, is employed to conceal communication channels from any third party aside from the intended sender and receiver. However, with the advancement of steganalysis technology, traditional image steganography methods face challenges. The embedding of secret messages disrupts the statistical characteristics of the carrier images, leading to weakened robustness and an increased risk of detection by steganalysis tools [1–3]. In response to the increasing sophistication of steganalysis tools, the CIS [4] has emerged as a crucial solution. CIS effectively counters existing steganalysis methods, providing heightened security in covert communication. Its importance is particularly notable in sensitive contexts such as military intelligence, national defense, and private information transmission. Consequently, CIS has gained considerable attention and emerged as a prominent topic in the realm of covert communication [2, 5, 6].

Currently, the majority of existing CIS algorithms utilize images and videos as their primary mediums or carriers [6]. Most CIS algorithms currently in use employ images as their principal mediums or carriers and can be categorized into two distinct groups depending on the methods used to acquire the cover images: coverless steganography based on image generation and coverless steganography based on mapping rules. Scholars have proposed robust CIS to ensure the safe transmission of stego-images in public channels [1, 5–17]. Among these robust steganography methods for coverless images, some rely on robust image features to encode secret information. Examples include Scale-Invariant Feature Transform (SIFT) features [7, 8], Discrete Cosine Transform (DCT)

Xiaopeng Li, Qiuyu Zhang and Zhe Li contributed equally to this work

✉ Qiuyu Zhang
zhangqy@lut.edu.cn

Xiaopeng Li
rain0237@163.com

Zhe Li
lzjslbb@163.com

[1] School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, Gansu, China

features [9], Discrete Wavelet Transform (DWT) features [10, 11], and Convolutional Neural Networks (CNN) features [12–15]. Additionally, some methods achieve high robustness by transmitting disguised images [16, 17]. The aforementioned CIS methods have demonstrated high robustness, enabling them to effectively resist interference from various image processing algorithms. However, in the process of extracting image features to generate hash sequences, achieving optimal image representation becomes challenging due to the complex structure of the CNN models employed. Furthermore, striking a balance between robustness against geometric and non-geometric attacks poses additional difficulties. At the same time, coverless steganography based on mapping rules often relies on pre-existing image databases, resulting in a limited hiding capacity. Consequently, a large number of unrelated stego-images and auxiliary information must be transmitted during communication, thereby reducing the security of the hiding algorithm. Additionally, CIS may require the use of side channels, especially when dealing with large capacities, which necessitates the sender to specify additional information for proper communication, rendering the process vulnerable to attackers [2]. Furthermore, other CIS methods [1–8] directly generate stego-images by constructing Generative Adversarial Network (GAN) models with improved performance, such as Deep Convolutional Generative Adversarial Networks (DCGAN). However, these methods often entail training multiple generative models, leading to increased computational complexity and steganography difficulty. Moreover, directly generated stego-images semantics may lack naturalness, making them susceptible to loss of details after being subjected to attacks. As a result, receivers may encounter difficulty in extracting the correct secret information. Most existing CIS currently employ low-level features of entire images for information mapping, a technique widely employed in computer vision due to its inherent resistance to steganalysis. However, several challenges persist, including limited capacity, dependency on extensive image databases, poor quality of generated images, and vulnerability to loss of image content and low-level features under geometric attacks, rendering them less robust. Furthermore, the limited steganography capacity and the issue of auxiliary information compromise their security to some extent.

To enhance the generalization, robustness, and security of the CIS, this paper proposes an end-to-end CNN network based on the multi-scale feature fusion and attention mechanism of DenseUNet. Through the utilization of the meticulously crafted DenseUNet model, multi-scale feature fusion and attention mechanism, and hybrid loss function, the network markedly enhances the robustness and security of the algorithm. In summary, the key contributions of this paper can be outlined as follows.

(1) DenseUNet is introduced into the feature extraction task of the steganography model, and an end-to-end CNN network model based on DenseUNet is designed. Through the joint optimization facilitated by the hybrid feature fusion layer, the network model adeptly handles various image changes, thereby endowing the extracted image features with robustness.
(2) Through the integration of an improved channel attention mechanism and spatial attention mechanism, a multi-scale feature fusion and attention mechanism module is designed. This module is deeply integrated into the network model, enabling precise extraction of multi-scale CNN features. Utilizing these features as matching benchmarks for images, the approach demonstrates robustness against various image attacks.
(3) Through the fusion of similarity loss, reconstruction loss, and quantization loss, a hybrid loss function is devised to optimize the performance of the network model across multiple dimensions. This function is utilized during network training to achieve a balance among similarity, reconstruction ability, and quantization characteristics. The objective is to generate a hash sequence with robust features, effectively enhancing the robustness of image features.

The rest of the paper is arranged as follows. The related research work of CIS is described in Sect. 2. The specific implementation steps of the proposed CIS method are detailed in Sect. 3. Section 4 outlines the experimental results and analysis, while Sect. 5 draws conclusions based on these findings.

2 Related works

2.1 CIS based on image generation

CIS based on image generation eliminates the need for carrier selection, instead directly generating secret carriers from secret information using specified rules or algorithms. These methods leverage state-of-the-art conditional generative adversarial networks (CGANs) to produce images by inputting secret information, thereby enhancing resistance to steganalysis attacks compared to traditional information hiding schemes. For instance, Hu et al. [18] proposed a DCGAN-based steganography scheme, wherein bit segments are transformed into noise vectors with multiple subintervals, and noise values are selected based on secret information. The steganographic image is then generated by DCGAN using input noise vectors, and secret data is extracted from these vectors according to inverse mapping rules. However, this scheme may struggle to correctly extract secret data even if the steganographic image is not under attack. Building on Hu's method, Li et al. [19] introduced a generative steganography method based on WGAN-GP, transforming secret data into noise vectors and simultaneously training the generator and extractor to achieve high extraction accuracy. Cao et al. [20] proposed a steganography scheme based on anime character synthesis, enhancing steganographic capacity by converting secret information into attribute tag sets of anime characters and using them for direct synthesis. To optimize steganographic image quality, Qin et al. [21] and Li et al. [22] incorporated adversarial loss and extraction modules into the image synthesis stage. The method introduced by Peng et al. [23] involves updating the noise vector in an iterative manner during the data extraction process. To prioritize generated image security, Liu et al. [24] improved secret information extraction using structural representation stability, while Wei et al. [25] employed hierarchical gradient decay to enhance the robustness of the generated steganographic images against steganalysis methods. Zhou et al. [26] introduced a novel approach using the Glow model to establish a one-to-one correspondence between secret messages and stego-images. This method provides a high hiding capacity while ensuring accurate extraction of the secret messages. Sun et al. [27] proposed a generative steganography model based on image style transfer and deep fusion of features. They designed a channel reduction network to extract image structural features using the VGG network as the base model fused secret matrices and extracted structural features through a fusion network to improve hiding capacity and enhance transmission security.

2.2 CIS based on mapping rules

CIS based on mapping rules usually requires the establishment of an image database in advance. They utilize the hash sequence generation method to produce a unique hash sequence for each image contained within the database. Subsequently, the secret information is segmented, with the length of each segment adjusted to match the length of the hash sequence of the image. In secret communication, the sender and receiver select a relevant image from the database based on the content of the hidden information to transmit secret information. For instance, Zhang et al. [28] proposed a CIS scheme based on the Discrete Cosine Transform (DCT) and the Latent Dirichlet Allocation (LDA) topic classification method. LDA is used to screen the image set for candidate images with similar topics, and DCT is then applied to obtain robust features of each candidate image. Zou et al. [29] proposed a coverless information-hiding method based on the average pixel value of sub-images, leveraging the stability of the mean feature to enhance robustness against attacks. Liu et al. [11] calculated low-frequency component coefficients after block transformation and performed zigzag scanning of DWT coefficients between blocks to generate a robust feature sequence. Govindasamy et al. [30] proposed an image steganography method based on Haar integer wavelet transform. The image was divided into 1024 non-overlapping sub-blocks, and wavelet transform was applied to each sub-block. Image features were then constructed based on the relationship between wavelet coefficients of sub-blocks. Compared to Liu's method [11], Govindasamy's method significantly improves the hiding ability of each image while ensuring robustness. However, all the above steganography methods for coverless images use low-level semantic features to represent secret information. Although they can effectively resist steganalysis, their robustness is still limited, especially when facing geometric attacks.

In recent years, researchers have integrated deep learning into coverless steganography based on mapping rules, introducing numerous deep learning-based techniques to address issues in traditional mapping-based schemes. To enhance robustness and security further, Luo et al. [31] introduced a novel CIS scheme based on multi-target recognition. This approach constructs object labels alphabetically to create a label mapping table. However, its hiding capacity is constrained by the types of objects in the image. Liu et al. [32] proposed a CIS scheme leveraging DenseNet feature mapping. By extracting high-dimensional CNN features and mapping them into hash sequences, this scheme enhances robustness against geometric attacks. Additionally, Luo et al. [33] presented a CIS scheme based on image segmentation. It utilizes ResNet to extract semantic features and employs Mask-RCNN to segment the target area from the image, thereby improving robustness against geometric attacks. Liu et al. [17] proposed a robust coverless steganography scheme based on camouflage images. This scheme clusters and regroups the image database, retrieving camouflage images with characteristics related to the cover image for transmission to the receiver instead of the stego-images, thus enhancing robustness and security. Furthermore, Meng et al. [34] introduced a robust CIS scheme based on an end-to-end hash generation model. This approach enhances the robustness of the method by integrating attention mechanisms and adversarial training into the design of the model. Lastly, Zou et al. [35] proposed a robust CIS scheme by disregarding the construction of a coverless image dataset. They utilized an unsupervised clustering algorithm to construct the coverless image dataset, enhancing dataset construction efficiency and improving the robustness of the proposed steganography method.

Based on the aforementioned analysis, CIS based on mapping rules offers fundamental resistance to steganalysis detection and ensures high security in image perception. However, due to their reliance on the size of pre-existing image databases, these methods exhibit limited hiding capacity. Moreover, a significant amount of unrelated stego-images and auxiliary information must be transmitted during com-

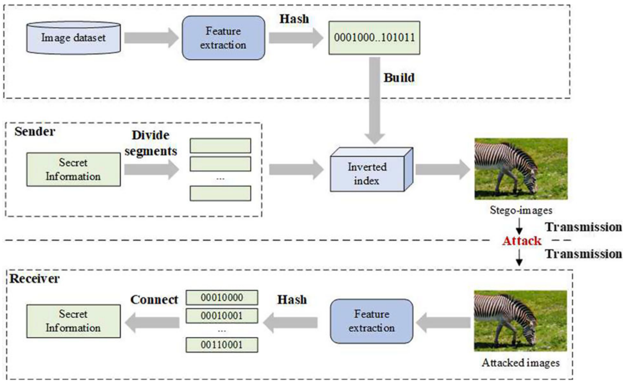

Fig. 1 The framework of the proposed method

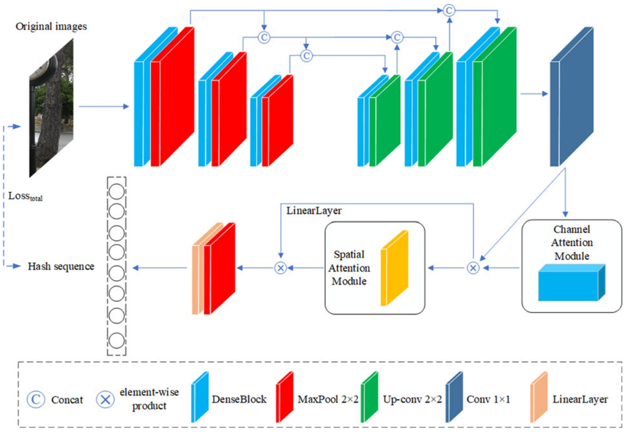

Fig. 2 The architecture of the network

munication, thereby reducing the security of the algorithm and weakening its robustness. Consequently, a robust CIS method based on DenseUNet multi-scale feature fusion and attention mechanism is proposed in this paper. This approach enhances anti-interference capabilities against both geometric and non-geometric attacks, preserves excellent image perceptual quality, and demonstrates strong performance in terms of robustness and security.

3 Proposed scheme

Figure 1 illustrates the proposed framework for CIS, comprising two main components: information hiding at the sender and information extraction at the receiver. At the receiver, each received stego-image is processed by a pre-trained feature extraction model to produce the corresponding binary hash sequence. Subsequently, these binary sequences are concatenated in the order they were transmitted by the sender, ultimately resulting in the successful retrieval of the original secret information.

Figure 1 depicts the proposed steganography process tailored for coverless images, which can be broadly delineated into three modules: image feature extraction, inverted index construction, and secret information steganography. At the sender, the process begins with training the proposed end-to-end hash generation model. Subsequently, an image index database is constructed based on this model. During the information mapping stage, the secret messages are partitioned into num segments, each of length l. For every segment of the secret message, the sender queries the established image index database to identify the image that precisely matches the hash sequence of that segment. Finally, the entire collection of chosen stego-images is transmitted to the receiver., ensuring covert transmission of the secret messages.

3.1 Feature extraction

Feature extraction stands as a pivotal component in image processing, enabling the accurate identification and retrieval of images, crucially so in the detection of stego-images. The feature extraction algorithm must possess the capability to uphold the stability of the feature sequence amidst various image attacks, ensuring that the extracted features authentically and precisely reflect the intrinsic characteristics of the image. To address this challenge, this paper introduces a novel network architecture founded on the DenseUNet multi-scale feature fusion and attention mechanism. This approach aims to extract robust features from images, thereby enhancing the efficacy of steganalysis and image recognition tasks.

(1)***Network structure*** The network architecture adopted in this paper builds upon DenseUNet, which amalgamates two prominent networks, U-Net and DenseNet while integrating two enhanced attention mechanism modules to establish an image feature extraction network. Figure 2 illustrates the specific design of this network structure. In the encoder phase, the Dense Block structure from three DenseNet variants, denoted as Dense Block [36], is utilized, along with the downward Transition layers, known as Transition Down, from these networks. Employing dense blocks and downward transition layers instead of conventional convolutional and pooling layers further enhances the feature extraction performance of U-Net, facilitating feature reuse across the entire network and rendering the model more compact. Furthermore, a multi-level attention mechanism module comprising a channel attention module and a spatial attention module is introduced within this network framework. Through the integration of these improved attention modules, precise capturing and enhancement of image features are achieved.

(2) *Attention mechanism module* To achieve comprehensive extraction and enhancement of image features, a multi-scale feature fusion and attention module is introduced. By integrating this hybrid attention mechanism, the performance of the model in image feature extraction is enhanced, while also improving its efficiency and robustness in handling complex image tasks. Specifically, the convolutional block attention module structure [37] is adopted in this paper, which comprises both a channel attention module and a spatial attention module.

(1) Channel attention module Because the channel attention module usually only focuses on the dependencies between channels, but ignores the importance of different scale features, the pyramid pooling structure can capture the feature information of different scales through multi-scale pooling operation, and fuse it to enrich the feature expression. Therefore, the combination of channel attention module and Spatial Pyramid Pooling (SPP) can more effectively focus on the dependencies between feature channels at different scales and improve the performance of the model. The specific implementation steps of the improved channel attention mechanism are as follows: the spatial context descriptor is extracted from the original feature map by using the global average pooling and global max pooling operations. At the same time, the original feature map is fed into the pyramid pooling layer. This layer captures multi-scale feature information from local to global through pooling operations at different scales. This enables the network to more comprehensively understand the spatial structure of the input data and form a richer and more comprehensive feature representation. Assume that the input of the channel attention module is F and the output is F_C, and the implementation process is shown in Fig. 3. This process is defined in Eqs. (1) and (2) as follows.

$$F_{\text{chamel}} = \sigma\left(Concat\left(f_N\left(AvgPool\left(F\right)\right),\; f_N\left(\text{MaxPool}\left(F\right)\right)\right)\right) \tag{1}$$

$$F_C = Concat\left(F_{channel}, SPPool\left(F\right)\right) \tag{2}$$

where $F_{channel}$ denotes the output of the channel attention module; *AvgPool*, *MaxPool*, and *SPPool* denote the average pooling, maximum pooling, and pyramid pooling processes, respectively; σ and F_N denote the sigmoid function and the shared network, respectively; and *Concat* denotes the splicing operation.

(2) Spatial attention module The traditional spatial attention module often struggles to process global context effectively, particularly in complex scenes, as it primarily focuses on local information extraction. To address this limitation, this paper introduces the fusion of Local Importance-based Pooling (LIP) with the spatial attention module. This fusion strategy retains the ability of the spatial attention module to capture key regions while further enhancing feature robustness through Local Importance pooling. By incorporating LIP, the model becomes more comprehensive and accurate in the feature extraction process, capable of simultaneously considering global context and local details. This integration overcomes the limitations of traditional spatial attention modules in extracting robust features, leading to improved model performance, generalization ability, and enhanced capacity to accurately capture key information in complex environments. Assuming the input of the spatial attention module is denoted as F and the output as F_S, the implementation process is depicted in Fig. 4. This process is defined by Eq. (3).

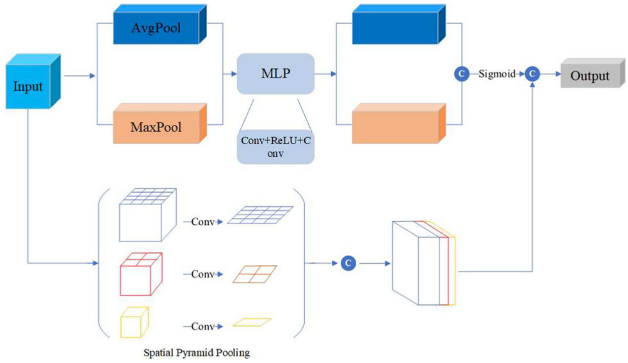

Fig. 3 Improved channel attention module

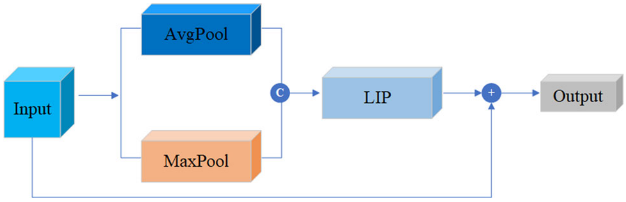

Fig. 4 Improved spatial attention module

$$F_S = LIPool\left(Concat\left(AvgPool\left(F\right),\; MaxPool\left(F\right)\right)\right) + F \tag{3}$$

where *LIPool* and *Concat* denote the pooling operation and splicing operation of LIPs, respectively.

3) Loss function To enhance the robustness of the model, preserve original information, and achieve discretized output, the proposed method incorporates similarity loss, reconstruction loss, and quantization loss.

(1) Similarity loss The similarity loss is primarily utilized to gauge the similarity between the hash sequence generated by the attacked image and that of the original image, with the aim of enhancing the robustness of the model. In order to ensure the ability of the model to produce reliable output even when subjected to noise, data perturbation, or adversarial attacks, this paper introduces the concept of similarity loss. By minimizing this loss, our objective is to maintain the coherence and consistency of the output in the presence of minor alterations. Additionally, we aim to ensure that the features extracted by the model remain resilient to changes in the input, thereby contributing to improved performance in subsequent tasks. The loss function is defined as shown in Eq. (4).

$$SimilarityLoss = \frac{1}{N}\sum_{i=1}^{N}\left(h_i - h_{attack,j}\right)^2 \tag{4}$$

where N represents the dimension of the hash code, h_i denotes the ith element of the original hash code, and $h_{attack,i}$ represents the ith element of the hash code after the attack.

(2) Reconstruction loss The reconstruction loss is designed to account for the ability of the model to retain information during feature encoding and decoding. By minimizing this loss, the objective of this paper is to ensure that the model preserves the original information to the greatest extent possible while extracting features, thus reducing information loss. Additionally, the reconstruction loss aids the model in learning more effective feature representations. The loss function is defined by

$$\mathrm{Re}constructionLoss = \frac{1}{M}\sum_{j=1}^{M}\left(f_j - {f'}_j\right)^2 \tag{5}$$

where M represents the dimension of the feature, f_j denotes the jth element of the original hash code, and f'_j represents the jth element of the reconstructed feature.

(3) Quantization loss. In hashing algorithms, it is typically desirable for the output to consist of a discrete, finite set of hash values. Optimizing the quantization loss aligns the output of the model with predefined quantization levels, streamlining subsequent processing steps and minimizing storage demands. Additionally, the quantization loss helps improve the robustness of the model to noise, as the discretized output is more resistant to small perturbations. The loss function is defined by

$$\begin{aligned}\mathrm{QuantizationLoss} = \frac{1}{N}\sum_{i=1}^{N}|h_i \\ - (2\cdot \mathrm{sigmoid}\,(h_i) - 1)|\end{aligned} \tag{6}$$

where N represents the dimension of the hash code, and h_i denotes the ith element of the hash code.

By combining these three losses, the overall loss function enables the model to simultaneously consider similarity, reconstruction accuracy, and quantization accuracy during the training process, thereby enhancing the model's robustness and generalization ability to the input data. The overall hybrid loss function is defined by

$$\begin{aligned}Loss_{\mathrm{total}} &= \alpha\cdot \mathrm{SimilarityLoss} + \beta\cdot \mathrm{ReconstructionLoss} \\ &+ \gamma\cdot \mathrm{QuantizationLoss}\end{aligned} \tag{7}$$

where α, β and γ are hyperparameters.

3.2 The process of establishing inverted index

To rapidly index the secret information within the stego-images, the hash sequence must first be extracted from the image database. Following this, the binarized secret message is segmented into portions of equal length as the hash sequence. Then, by establishing the mapping relationship between the secret information and the stego-images, an efficient inverted index can be generated. Here's how it works.

1) Hash sequence generation process Hash sequence generation is a key step in constructing an inverted index, which lays the foundation for efficient retrieval. The hash sequence, extracted according to specific hashing rules based on image features, is a fixed-length sequence. Since the output of feature extraction yields a continuous numerical sequence, it must be discretized to form the required hash sequence. Discretization ensures that the continuous numerical sequence is converted into a discrete sequence with a specific format, typically comprising binary digits.

(1) To extract global feature information, a global average pooling operation is applied, which reduces the spatial dimensions of features to 1 × 1.
(2) A fully connected layer is then utilized to map the pooled features to the same dimension as the expected hash sequence length. To confine the output range, the tanh activation function is applied, ensuring that the output value falls within the range of -1 to 1.
(3) To convert these continuous values into discrete hash sequences, a bi-half layer [38] is used in this paper. The bi-half layer generates a binary hash sequence by sorting the input features and assigning the top half of the sorted features to 1 and the bottom half to -1.
(4) To ensure the hash sequence is mapped between 0 and 1, a simple linear transformation is applied. By adding 1 to the output of the bi-half layer and then dividing by 2, the values originally ranging from -1 to 1 are successfully mapped to the range between 0 and 1. As a result, the final hash sequence, consisting of 0 s and 1 s, is obtained.

The entire process not only achieves the discretization of features but also ensures that the hash sequence adheres to a specific format. This aspect furnishes a convenient and efficient foundation for subsequent tasks such as feature matching and retrieval. Building upon the preceding analysis, the entire process can be succinctly described by the formulation outlined in Eq. (8)- Eq. (11).

$$F_{avg} = AvgPool\,(F) \in R^{B\times C\times 1\times 1} \tag{8}$$

$$F_{hash} = \tanh\left(Linear\left(F_{avg}\right)\right) \in R^{B\times l} \tag{9}$$

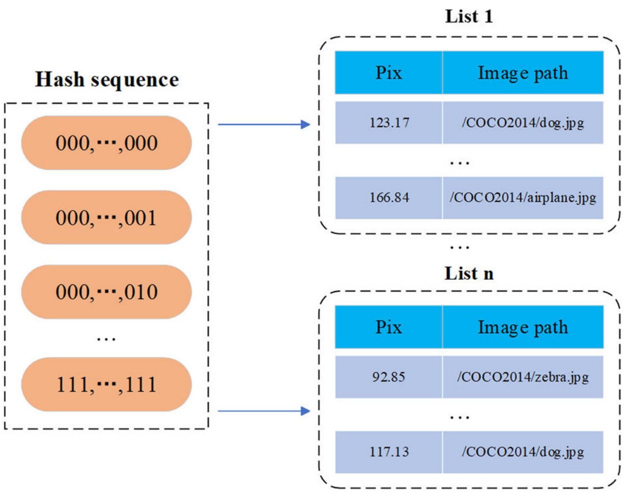

Fig. 5 The established image index database

$$F_{bihalf} = BiHalfLayer\,(F_{hash}) \in R^{B\times l} \tag{10}$$

$$F_{final} = \frac{F_{bihalf}+1}{2} \in R^{B\times l} \tag{11}$$

where F denotes the input features; F_{avg} denotes the features obtained after the global average pooling operation; F_{hash} denotes the hash sequence obtained after the hash function processing; F_{bihalf} denotes the features obtained after the bi-half layer operation; F_{final} denotes the final features obtained after the mapping operation; B denotes the batch size; C denotes the number of channels; and l denotes the length of the hash sequence.

(2) Construction of inverted index In CIS schemes, the secret message typically requires careful segmentation into multiple segments to ensure secure transmission. However, in practice, manually matching these segments with carriers in a vast image database is prohibitively time-consuming and inefficient. This exhaustive search becomes especially impractical when dealing with large image databases, consuming significant time and computational resources. To substantially enhance retrieval efficiency, an image index database based on inverted indexing is established. Figure 5 illustrates the index structure.

As shown in Fig. 5, this database comprises three crucial components: hash sequences, average pixel values, and image paths. Among these, the hash sequence not only serves as the retrieval gateway for the entire database but also ensures the uniqueness of each image, thereby enhancing the accuracy and efficiency of the retrieval process. Through the design of this indexing structure, it becomes feasible to swiftly locate the image matching the secret information fragment, thus significantly enhancing the effectiveness and efficiency of CIS. In the first part, the hash sequence serves as the main access point for the complete image index database as a whole, acting as the primary identifier for each image. The sequence is arranged in an ascending binary format, commencing from ’000,..., 000’ and concluding with ’111,..., 111’. Given the potential variability in the order of received images, an intermediate component, ’pix’, stores the average pixel value of the images to maintain the actual sequence order. The average pixel value is computed as follow:

$$\text{pix} = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N} f(i,j) \tag{12}$$

where pix represents the average pixel value of the image, M and N are the width and height of the image (i.e., the total number of pixels in the image), respectively, and $f(i,j)$ denotes the value of the pixel in the image located at point (i,j).

The final part is the image path, indicating the storage location of the image. During the process of information concealment, it is imperative that the pix value of the chosen image exceeds that of the preceding image, and the pix value of the initially selected image must be the lowest value among the entire list. This method guarantees that regardless of any changes in the order of stego-images resulting from network routing at the receiver, the original order can be accurately reconstituted by utilizing this parameter. As an example, consider a scenario where the current segment of the secret message is “00000000”, its associated pix value is 123.17, and the path to the image is “/COCO2014/dog.jpg”. When the next secret message segment is “11111111”, “/COCO2014/airplane.jpg” can be selected as the stego-image because its pix value is 166.84, which is greater than 123.17. However, “/COCO2014/zebra.jpg” and “/COCO2014/dog.jpg” cannot be chosen as carriers because their pix values are 92.85 and 117.13, respectively, failing to meet the aforementioned conditions. This design ensures the sequence and accuracy of secret information during the transmission process.

3.3 Secret information hiding

After establishing the repository of image indices, the next step is to segment the secret information and match the image through the index structure to realize the steganography of the secret information. Algorithm 1 is the pseudo-code of the secret information steganography process.

The specific processing steps of secret information steganography are as follows.

Step 1: Suppose the secret message S has a length of L. It is divided into num segments, each of length l. If L is not divisible by l, zeros are appended to the final segment to complete the sequence, and noting the number of appended

Algorithm 1 Secret information hiding

```
Require: Image database: $I = \{I_1, I_2, \ldots, I_{\text{Num}}\}$, Secret information $S$
Ensure: Stego images: $ST = \{st_1, st_2, st_3, \ldots, st_{\text{num}}\}$ or $ST = \{st_1, st_2, st_3, \ldots, st_{\text{num}}, st_{\text{padding}}\}$
for $j = 1$ to INum do
    Feature extract: $feature = \text{DenseUNet}(I)$
    Calculate $pix_j$ by Equation (11)
    Update index database: Index item→ $\{pix, \text{Image path}\}$
end for
Divide S: $segment(S) = \{S_1, S_2, S_3, \ldots, S_{\text{num}}\}$
for $i = 1$ to num do
    Select stego-image for $S$: $sti = I_j$ , if $HS_j = Si$ and $pix_i > pix_{i-1}$
end for
if $L\%l = 0$ then
    Get Steo images: $ST = \{st_1, st_2, st_3, \ldots, st_{\text{num}}\}$
else
    Get Steo images: $ST = \{st_1, st_2, st_3, \ldots, st_{\text{num}}, st_{\text{padding}}\}$ And Record the number of padding zeros and map it to the last image $st_{\text{padding}}$
end if
```

zeros. The calculation method is provided by

$$num = \begin{cases} L/l, & if L\%l = 0 \\ \lfloor L/l \rfloor + 1, & Otherwise \end{cases} \tag{13}$$

where $\lfloor . \rfloor$ denotes the downward rounding function.

Step 2: For a secret message segment Si, where $i = 1, 2, \ldots,$ *num*, select the corresponding stego-image, where the hash sequence of the stego-image st_i is equal to S_i.

$$\begin{aligned} &\text{st}_i = I_j \text{ if } HS_j = S_i \text{ and } pix_j > pix_{j-1}, \\ &\text{for } i = 1, 2, \ldots, \text{num}, \quad j = 1, 2, \ldots, \text{Num} \end{aligned} \tag{14}$$

where HS_j represents the hash sequence of the selected image. I_j denotes the j-th image in the image database, and *Num* denotes the number of images.

Step 3: Repeat **Step 2** until all stego-images are obtained to obtain a set of stego-images ST. If filler zeros are present, their count is translated into binary format and mapped to the final stego-image.

$$ST = \begin{cases} st_1, st_2, st_3, \ldots, st_{\text{num}}, & \text{if } L\%l = 0 \\ st_1, st_2, st_3, \ldots, st_{\text{num}}, & \\ st_{\text{padding}}, & \text{Otherwise} \end{cases} \tag{15}$$

Step 4: Send the successfully matched stego-images to the receiver.

3.4 Secret information extraction

After successfully receiving the stego-images, the receiver utilizes a predetermined feature extraction model to generate the corresponding hash sequence. Following the specific order in which the sender transmitted them, the receiver accurately assembles these hash sequences to extract the secret information. This step enables the receiver to efficiently and accurately reconstruct the original secret information, ensuring its integrity and confidentiality are preserved. Algorithm 2 outlines the secret information extraction process.

Algorithm 2 Secret information extraction

```
Require: Steo images: $ST = \{st_1, st_2, st_3, \ldots, st_{\text{num}}\}$ or $ST = \{st_1, st_2, st_3, \ldots, st_{\text{num}}, st_{\text{padding}}\}$
Ensure: Secret information $S$
for $j = 1$ to INum do
    Feature extract: $feature = \text{DenseUNet}(I)$
    Calculate $pix$ by Equation (11)
    Sort the images by $pix$
end for
if $L\%l = 0$ then
    Connect all of the hash sequences in order
else
    Subtract padding zeros from the last sequence. Connect all of the hash sequences in order
end if
```

The specific processing steps of secret information extraction are as follows:

Step 1: In the process of sending stego-images group *ST*, the image order may be changed. The receiver first checks the received stego-images group *ST* with the parameter *pix*, and reorganizes the image order according to the inspection results.

Step 2: The received stego-images group *ST* is input into the designed feature extraction network to extract the hash sequence of image features. In the case of zero padding, the hash sequence of the last image is adjusted by subtracting the recorded number of zero padding.

Step 3: By connecting the hash sequences in order, one can obtain the desired result *S*.

4 Experimental results and analysis

4.1 Configuration

The CNN model adopted in this paper was trained on an NVIDIA GTX3080Ti GPU using the PyTorch framework. Other experiments were conducted on a personal computer equipped with an Intel(R) Core i7-8750 H @ 2.20GHz processor and 24GB of memory. The PyTorch framework was used for deep learning, and image attacks were implemented using functions from the OpenCV library. All experiments were conducted using PyCharm.

Four widely used datasets are adopted, namely PASCAL VOC 2012 [39], INRIA Holidays [40], MS COCO 2014 [41], and Caltech-256 [42], which are described as follows.

(1) The PASCAL VOC2012 dataset. The PASCAL VOC2012 dataset can be divided into 20 subcategories and contains a total of 11,540 images.
(2) The INRIA Holidays dataset. The INRIA Holidays dataset consists of 500 image groups, each representing a different scene or object, and contains a total of 1,491 high-resolution images.
(3) The MS COCO2014 dataset. The MS COCO 2014 dataset contains 82,783 training images, 40,504 validation images, and 40,775 test images divided into 80 classes.
(4) The Caltech-256 dataset. The Caltech-256 dataset contains 257 object images of different classes, totaling 30,608 images. The number of images in each class varies, but a minimum of 80 and a maximum of 827 images per class are ensured to ensure diversity and richness of the data.

In the experiments, this paper utilizes the PASCAL VOC 2012 and MS COCO 2014 training sets to train the CNN model. For the testing phase, 500 images are selected from the PASCAL VOC 2012, MS COCO 2014, Caltech-256, and INRIA Holidays datasets. The parameters for the comparison experiments are set as follows: the length l of the hash sequence used for the experiment is set to 8, consistent with the comparison requirements. For the experiment, all images have been resized to a dimension of 256 × 256. Additionally, the hyperparameters α, β and γ are set to 1.0, 0.1, and 1.0, respectively. To assess the superiority of the proposed method, Liu's [17], Zhang's [28], Luo's [31], and Meng's [34] schemes are used as reference benchmarks for comparison.

4.2 Capacity

In this section, we analyze and compare the capacity of the proposed method, specifically the number of binary bits required to hide secret information. The capacity of most CIS methods based on mapping rules is determined by the image feature mapping rules, which in turn is limited by the length of the hash sequence, denoted as l, and is proportional to the capacity. In theory, there is no fixed upper limit to the length of the hash sequence l, as it can take any value. However, longer hash sequences may increase the capacity but also pose challenges during image retrieval. Longer sequences lead to increased complexity and difficulty in matching, potentially making the retrieval process more cumbersome and inefficient. Therefore, steganography methods must make trade-offs and carefully choose the length l of the hash sequence.

In the proposed method, the length of the hash sequence can be flexibly selected between 1 and 15. This approach ensures that the scheme maintains a certain capacity while avoiding retrieval difficulties caused by excessively long sequences. It achieves a balance between capacity and operational convenience. The evaluation index for capacity is determined by the number of images capable of hiding the secret message L of a fixed length. This evaluation is closely related to the length of the hash sequence l. Specifically, a longer length of the hash sequence l results in a smaller number of sent images, indicating a larger capacity. This metric, denoted as N_P, can be determined by the following calculation:

Table 1 The capacity comparison

Method	Np				Capacity
	1B	10B	100B	1KB	
Zhang's [28]	2∼9	7∼81	55∼801	548∼8193	1∼15
Luo's [31]	1∼2	4∼14	34∼134	342∼1366	6n
Liu's [17]	2∼9	7∼81	55∼801	548∼8193	1∼15
Meng's [34]	2∼9	7∼81	55∼801	548∼8193	1∼15
Ours	2∼9	7∼81	55∼801	548∼8193	1∼15

$$N_p = \frac{L}{l} \tag{16}$$

To verify the capacity of the proposed method, comparisons are made with existing methods, including Zhang's [28], Luo's [31], Liu's [17], and Meng's [34]. The steganographic capacity comparison results are presented in Table 1.

Table 1 illustrates that Luo's [31] selects objects for information hiding, with its capacity being closely related to the number n of selected objects. Each object represents 6 bits of information, with a maximum value set to 4. While its maximum capacity is large, its randomness is significantly high, resulting in unstable capacity. In contrast, in Liu's [17], Zhang's [28] and Meng's [34] and the proposed method, the hidden capacity is more controllable and stable.

4.3 Robustness

In image processing or transmission, images inevitably face various geometric and non-geometric attacks, resulting in content loss and feature destruction. Therefore, in CIS, the ability to effectively extract secret messages from attacked images serves as the primary robustness evaluation criterion. Secret information extraction accuracy (ACC) aims to precisely gauge the capability of the algorithm to recover information from damaged images, thus offering a comprehensive assessment of its robust performance. A higher ACC value indicates greater robustness, with a maximum ACC of 100%. The extraction accuracy ACC is defined as

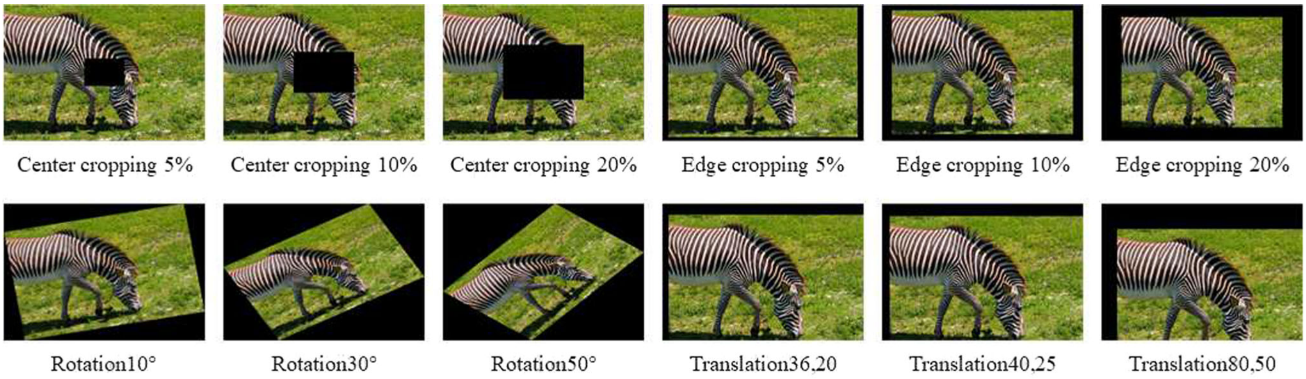

Fig. 6 The example of the geometric attacked images

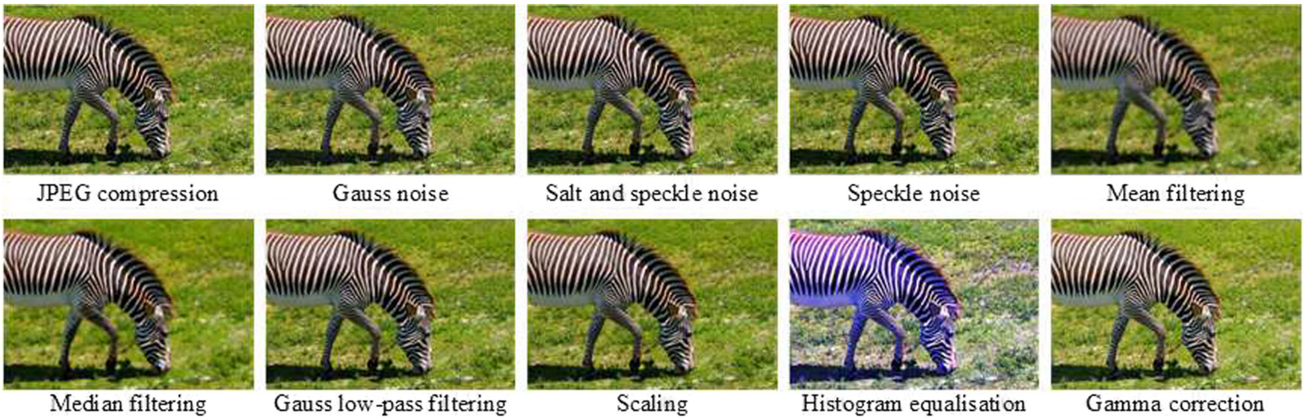

Fig. 7 The example of the geometric attacked images

$$ACC = \frac{\sum_{i=1}^{M} F(i)}{M} \times 100\%, \quad F(i) = \begin{cases} 1, & \text{if } hs_i = hs' \\ 0, & \text{otherwise} \end{cases}. \tag{17}$$

where M represents the total number of pieces of secret message in the stego-images, hs_i denotes the hidden secret message of the ith stego-image, and hs_i' represents the extracted secret message of the ith stego-image.

In this paper, the robustness analysis is conducted using the PASCAL VOC2012, INRIA Holidays, MS COCO 2014, and Caltech-256 datasets. The attack forms are categorized into geometric and non-geometric attacks. Examples of images after being attacked are illustrated in Figs. 6 and 7, respectively. Their parameters are detailed in Tables 2 and 3, respectively. Notably, the original images depicted in Figs. 6 and 7 are selected from the MS COCO 2014 dataset.

4.3.1 Ablation study

(1) ***Analysis of baseline networks:*** This section aims to explore the robustness of three CNN models – Unet,

Table 2 Kind of geometric attacks

Attack	Parameters
Centered cropping	Ratios: 5%, 10%, 20%
Edge cropping	Ratios: 5%, 10%, 20%
Translation	(36,20), (40,25), (80,50)
Rotation	Ratios angles: 10°, 30°, 50°

Table 3 Kind of non-geometric attacks

Attack	Parameters
Gaussian noise	The mean μ: 0, the variance σ: 0.001
Salt & pepper noise	The mean μ: 0, the variance σ: 0.001
Speckle noise	The mean μ: 0, the variance σ: 0.01
Median filter	The filter size: 3 × 3
Mean filter	The filter size: 3 × 3
Gaussian filter	The filter size: 3 × 3
JPEG compression	The quality factor Q: 10% and 90%
Scaling	The scaling ratio: 3.0
Color histogram equalization	None
Gamma correction	Factor: 0.8

Table 4 Robustness(%) comparison with network models

Attack	Parameters	U-Net	Dense-Net121	Dense-UNet
Gaussian noise	σ(0.001)	93.29	89.92	**97.28**
Salt & pepper noise	σ(0.001)	91.90	89.60	**96.07**
Speckle noise	σ(0.01)	93.62	85.36	**96.11**
Median filter	3×3	68.71	37.50	**97.07**
Mean filter	3×3	75.04	37.50	**96.82**
Gaussian filter	3×3	87.50	50.00	**98.53**
JPEG compression	Q(10)	62.19	48.63	**98.87**
	Q(90)	56.20	39.56	**97.62**
Scaling	3.0	86.20	98.75	**98.87**
Color histogram equalization	–	62.82	83.20	**85.40**
Gamma correction	0.8	83.60	93.40	**97.90**
Centered cropping	5%	62.35	81.2	**93.04**
	10%	58.39	62.5	**89.30**
	20%	45.28	18.7	**88.13**
Edge cropping	5%	50.25	6.2	**83.81**
	10%	25.30	6.2	**82.38**
	20%	6.25	18.7	**78.48**
Translation	(36,20)	70.36	72.60	**89.39**
	(40,25)	68.29	71.36	**88.59**
	(80,50)	58.30	62.58	**88.97**
Rotation	10°	70.25	69.82	**84.98**
	30°	53.62	64.50	**79.28**
	50°	49.87	62.89	**78.23**

The bold indicates the best robustness performance among the comparison methods

Table 5 Robustness(%) of adding attemtion module

Attack	Parameters	Without attention	With attention
Gaussian noise	σ(0.001)	97.28	**99.04**
Salt & pepper noise	σ(0.001)	96.07	**98.95**
Speckle noise	σ(0.01)	96.11	**98.99**
Median filter	3×3	97.07	**99.04**
Mean filter	3×3	96.82	**98.74**
Gaussian filter	3×3	98.53	**99.04**
JPEG compression	Q(10)	98.87	**99.04**
	Q(90)	97.62	**99.12**
Scaling	3.0	98.87	**99.41**
Color histogram equalization	—	85.40	**89.47**
Gamma correction	0.8	97.90	**98.90**
Centered cropping	5%	93.04	**95.76**
	10%	89.30	**94.00**
	20%	88.13	**94.44**
Edge cropping	5%	83.81	**88.88**
	10%	82.3	**86.28**
	20%	78.48	**80.70**
Translation	(36,20)	89.39	**91.90**
	(40,25)	88.59	**91.36**
	(80,50)	88.97	**90.69**
Rotation	10°	84.98	**88.84**
	30°	79.28	**85.74**
	50°	78.23	**84.82**

The bold indicates the best robustness performance among the comparison methods

DenseNet, and DenseUNet – in image feature extraction. These models are trained and tested using INRIA Holidays. The experimental results are presented in Table 4.

Table 4 illustrates that DenseUNet excels in extracting robust image features. This is because DenseUNet combines the structural characteristics of dense connection and skip connection of Unet and DenseNet. Dense connections facilitate full propagation and reuse of features within the network, thereby enhancing feature utilization. Meanwhile, skip connections aid in the fusion of feature information across different levels, enabling the network to effectively handle complex backgrounds and noise interference.

(2) ***Analysis of attention mechanisms:*** To further enhance the robustness of the DenseUNet network, the proposed method incorporates a multi-scale feature fusion and attention module into its architecture. This design not only integrates feature information from different scales but also leverages the benefits of the attention mechanism to bolster the ability of the network to extract image features. Details regarding the design of the attention mechanism module are provided in section 3.1. This section aims to investigate the impact of adding an attention module to the network on image feature extraction robustness. Training and testing are conducted on the public image dataset INRIA Holidays, and the results of the robustness experiments are presented in Table 5.

Table 5 demonstrates that the addition of multi-scale feature fusion and attention mechanism to the network results in improved robustness, particularly against geometric attacks. This enhancement stems from the ability of multi-scale feature fusion to capture both detailed information and the global context of the image simultaneously. By integrating features at various levels, the network gains a more comprehensive understanding of the image content, thereby enhancing feature robustness and discrimination. This design enables the network to extract key features more reliably, even in the presence of challenges such as scale changes and variations in object size.

4.3.2 Robustness on different datasets

To assess the robustness of the proposed method across four different datasets (PASCAL VOC 2012, MS COCO 2014, INRIA Holidays, and Caltech-256), comparisons are made with existing schemes, including Zhang's [28], Luo's [31],

Table 6 Robustness(%) comparison with four CIS methods in PASCAL VOC 2012

Attack	Parameters	Zhang's [28]	Luo's [31]	Meng's [34]	Ours
Gaussian noise	σ(0.001)	92.1	72.4	92.6	**97.2**
Salt & pepper noise	σ(0.001)	94.6	74.0	92.6	**95.5**
Speckle noise	σ(0.01)	90.0	67.1	91.4	**97.1**
Median filter	3×3	93.2	71.0	84.2	**94.0**
Mean filter	3×3	**96.4**	67.3	82.4	92.7
Gaussian filter	3×3	**97.0**	72.2	84.2	94.2
JPEG compression	Q(90)	**97.7**	82.3	96.4	97.1
Scaling	3.0	**99.1**	92.1	98.8	98.1
Color histogram equalization	—	77.0	70.2	86.6	**91.3**
Gamma correction	0.8	92.4	81.7	93.8	**96.6**
Centered cropping	5%	85.3	62.3	75.0	**95.8**
	10%	77.4	37.9	65.4	**94.3**
	20%	75.2	25.6	56.4	**91.8**
Edge cropping	5%	68.0	82.9	**98.8**	91.0
	10%	60.4	80.0	**98.8**	89.6
	20%	55.3	70.6	**99.2**	84.0
Translation	(36,20)	7.3	70.6	80.8	**90.8**
	(40,25)	5.1	70.8	80.6	**90.2**
	(80,50)	3.9	60.5	76.6	**89.7**
Rotation	10°	12.0	63.1	80.6	**90.2**
	30°	4.3	48.0	67.4	**86.9**
	50°	3.5	35.9	62.0	**84.8**
Average	—	63.1	66.3	83.8	**92.4**

The bold indicates the best robustness performance among the comparison methods

Liu's [17], and Meng's [34]. Since PASCAL VOC 2012 and MS COCO 2014 are more suitable for evaluating object detection algorithms, Luo's [31] proposed a CIS method based on multi-object recognition. Therefore, considering the dataset content, PASCAL VOC 2012 and MS COCO 2014 are used for comparison with Luo's [31]. However, INRIA Holidays and Caltech-256 are used to compare with Liu's [17]. Nonetheless, Zhang's [28], Meng's [34], and the proposed method are not affected by the dataset content and can be applied to all four datasets. The detailed comparison results are shown in Table 6 through Table 7. Among them, the types of attacks in Tables 8 and 9 that are not considered in Liu's [17] are represented by '—'.

As can be seen from Tables 6, 7, 8, 9, compared with Zhang's [28] scheme, the proposed method has better robustness, especially in terms of geometric attacks, such as cropping, rotation and translation, etc. This can be attributed to the block-wise processing of images required by the Discrete Cosine Transform (DCT) for feature extraction, which renders the relationships between adjacent blocks fragile. Consequently, the stability of DCT features is compromised in the face of geometric attacks, potentially resulting in information loss or distortion. In contrast, deep hashing technology can encapsulate richer semantic information of images, which makes it more robust in steganography, more effective in resisting various forms of attacks, and maintaining the integrity and security of information. Luo's [31] analyzed multiple objects in the image, extracted their feature information, and used these feature information to encode and hide secret information. This approach relied on the feature extraction of the object in the image. However, Luo's [31] performs poorly in terms of non-geometric attacks because noise can affect the detection accuracy of Faster RCNN. Liu's [17] transmits secret information by disguised images, exhibiting robustness against most attacks, particularly geometric ones. However, its performance significantly deteriorates when confronted with noise attacks. This decline can be attributed to the disruptive nature of noise, which has the ability to disrupt the visual features of images, making it challenging to accurately identify the disguised image during retrieval. Consequently, this leads to errors in secret information extraction. Meng's [34] adopts an end-to-end hash generation model, which has strong robustness against geometric attacks and non-geometric attacks. However, when

Table 7 Robustness(%) comparison with four CIS methods in MS COCO 2014

Attack	Parameters	Zhang's [28]	Luo's [31]	Meng's [34]	Ours
Gaussian noise	$\sigma(0.001)$	93.2	68.20	97.0	**99.8**
Salt & pepper noise	$\sigma(0.001)$	95.1	88.18	97.0	**99.6**
Speckle noise	$\sigma(0.01)$	90.5	84.65	96.8	**98.1**
Median filter	3×3	92.7	74.68	95.6	**97.8**
Mean filter	3×3	**96.5**	73.24	95.0	96.2
Gaussian filter	3×3	97.0	71.30	95.6	**97.9**
JPEG compression	Q(90)	97.4	81.60	98.2	**98.9**
Scaling	3.0	98.3	88.90	98.2	**98.9**
Color histogram equalization	—	76.1	70.50	**94.8**	90.0
Gamma correction	0.8	92.3	82.30	97.4	**98.3**
Centered cropping	5%	84.6	56.3	90.0	**95.5**
	10%	76.7	32.0	84.8	**93.4**
	20%	74.3	23.4	70.0	**91.0**
Edge cropping	5%	66.4	74.5	**98.2**	90.4
	10%	59.0	72.1	**98.2**	89.0
	20%	54.3	64.3	**98.2**	86.0
Translation	(36,20)	7.1	65.6	**93.4**	89.3
	(40,25)	6.5	66.2	**93.2**	89.1
	(80,50)	3.4	56.1	**91.6**	88.3
Rotation	10°	9.7	55.4	92.2	**94.9**
	30°	4.3	33.8	81.2	**86.1**
	50°	3.2	22.9	67.8	**84.4**
Average	—	62.7	59.6	92.1	**93.3**

The bold indicates the best robustness performance among the comparison methods

dealing with complex and variable image content, the model may struggle to identify all key features accurately. This limitation can impact the generation of hash sequences and consequently affect the overall robustness of the steganography. In addition, Meng's [34] method demonstrates a higher average robustness than the proposed method in the Caltech-256. This discrepancy can be attributed to the lower robustness performance of the proposed method against geometric attacks compared to Meng's [34]. The images in the Caltech-256 dataset typically exhibit low resolution and lack detailed information, which hinders accurate and comprehensive feature extraction by the proposed method. Specifically, under geometric attacks such as rotation and scaling, the object shapes and structures in the images may become blurred and challenging to identify due to the low resolution and lack of detail. Consequently, significant changes occur in the extracted features, diminishing the robustness of feature extraction.Conversely, for non-geometric attacks like noise addition, feature extraction prioritizes the overall structure and semantic information of the image rather than fine details. Consequently, the robustness against non-geometric attacks is higher. Overall, the proposed method demonstrates superior robustness and generalization performance against both geometric and non-geometric attacks across four different datasets.

4.4 Security

(1) ***Anti-steganalysis security*** In the information hiding stage, the proposed method adopts a CIS, which does not modify the images, but transmits a set of unmodified natural images that have nothing to do with the secret information. Since the existing steganalysis tools can detect the existence of secret information by the modification traces left in the images. Therefore, this enables the proposed method to completely evade the detection of any steganalysis algorithm, thus ensuring the concealment of information transmission.

(2) ***Transmission security*** While CIS algorithms can effectively evade detection by steganalysis methods, it's essential to recognize the potential security implications associated with auxiliary information during transmission. Liu's [17], Zhang's [28] and Luo's [31] and other methods will send additional information of variable length when transmitting information, and this uncertainty increases the

Table 8 Robustness(%) comparison with four CIS methods in INRIA HOLIDAYS

Attack	Parameters	Zhang's [28]	Liu's [17]	Meng's [34]	Ours
Gaussian noise	$\sigma(0.001)$	93.2	46.0	98.0	**99.1**
Salt & pepper noise	$\sigma(0.001)$	94.4	46.0	97.8	**99.0**
Speckle noise	$\sigma(0.01)$	91.7	46.0	96.6	**99.0**
Median filter	3×3	94.9	**100.0**	98.0	99.1
Mean filter	3×3	95.7	**100.0**	97.8	98.7
Gaussian filter	3×3	96.0	**100.0**	98.2	99.1
JPEG compression	Q(10)	81.0	**100.0**	95.8	99.1
Scaling	3.0	98.65	**100.0**	98.2	99.4
Color histogram equalization	—	75.9	**95.0**	92.2	91.5
Gamma correction	0.8	91.6	**100.0**	96.8	98.9
Centered cropping	5%	81.8	—	89.0	**95.8**
	10%	74.0	—	82.4	**94.0**
	20%	72.3	93.0	67.6	**94.4**
Edge cropping	5%	63.4	—	**98.2**	88.9
	10%	57.8	—	**98.2**	86.3
	20%	54.1	**100.0**	**98.2**	82.7
Translation	(36,20)	25.6	—	**94.4**	91.9
	(40,25)	22.9	—	**94.6**	91.4
	(80,50)	12.2	**100.0**	**94.2**	90.7
Rotation	10°	10.7	92.0	**93.4**	88.8
	30°	4.7	—	76.4	**85.7**
	50°	4.3	—	63.4	**84.8**
Average	—	63.4	87.0	91.8	**93.6**

The bold indicates the best robustness performance among the comparison methods

risk of secret information being intercepted or tampered with. In contrast, the proposed method and Meng's [34] scheme do not involve additional auxiliary information transmission in the transmission process, thus greatly reducing the possibility of secret information being exposed or destroyed in the transmission process, and significantly improving the security of steganography. Therefore, while pursuing efficient steganography, the proposed method attaches great importance to the security of information transmission to ensure that the secret information can be transmitted safely and reliably to the receiver.

5 Conclusion

In this paper, a robust CIS method based on DenseUNet with multi-scale feature fusion and attention mechanism is proposed. The proposed method can fully extract the feature information from images by combining the DenseUNet network model with the multi-scale feature fusion and attention mechanism. This integration significantly enhances the accuracy of detecting and extracting robust features from images. This integration significantly enhances the accuracy of detecting and extracting robust features from images. Utilizing the end-to-end hash sequence generation model enables us to directly generate a robust hash sequence, thereby avoiding interference from intermediate steps and minimizing the risk of information leakage. Consequently, our method exhibits improved robustness and anti-interference capabilities in steganography applications. To ensure high robustness in feature extraction, we have designed three loss functions to compose a hybrid loss function. This approach contributes to the overall effectiveness of our network model. From a security perspective, our proposed method eliminates the need for auxiliary information transmission, leading to a significant enhancement in security. The experimental results validate the exceptional performance of our method across various metrics, including anti-steganalysis, security of secret information transmission, robustness, and capacity. The future work will continue to deepen the proposed method and strive to further expand the steganography capacity while maintaining a high degree of security and robustness.

Table 9 Robustness(%) comparison with four CIS methods in CALTECH-256

Attack	Parameters	Zhang's [28]	Liu's [17]	Meng's [34]	Ours
Gaussian noise	$\sigma(0.001)$	80.3	73.0	97.8	**98.3**
Salt & pepper noise	$\sigma(0.001)$	83.9	71.0	98.6	**99.5**
Speckle noise	$\sigma(0.01)$	77.8	73.0	98.8	**99.1**
Median filter	3×3	91.6	93.0	97.8	**98.5**
Mean filter	3×3	93.6	93.0	97.2	**98.5**
Gaussian filter	3×3	94.6	93.0	**97.8**	96.1
JPEG compression	Q(10)	76.9	93.0	95.4	**98.2**
Scaling	3.0	98.5	**100.0**	99.4	98.9
Color histogram equalization	—	73.7	94.0	**95.8**	94.6
Gamma correction	0.8	91.6	94.0	98.8	**99.0**
Centered cropping	5%	84.5	—	96.4	**97.7**
	10%	76.5	—	92.8	**95.8**
	20%	74.6	91.0	90.8	**93.7**
Edge cropping	5%	63.3	—	**99.4**	92.7
	10%	56.0	—	**99.4**	91.1
	20%	50.7	94.0	**99.4**	87.7
Translation	(36,20)	6.3	—	**97.0**	95.2
	(40,25)	5.6	—	**96.8**	91.5
	(80,50)	4.5	94.0	**94.2**	91.6
Rotation	10°	13.8	94.0	**97.4**	94.4
	30°	5.1	—	91.6	**93.0**
	50°	4.6	—	89.0	**90.5**
Average	—	59.5	89.3	**96.4**	95.3

The bold indicates the best robustness performance among the comparison methods

Acknowledgements This work is supported by the National Natural Science Foundation of China (No. 61862041).

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Xiaopeng Li and Zhe Li. The first draft of the manuscript was written by Xiaopeng Li, Qiuyu Zhang commented on previous versions of the manuscript and critically revised the work. All authors read and approved the final manuscript.

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

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Li, G., Feng, B., He, M., Weng, J., Lu, W.: High-capacity coverless image steganographic scheme based on image synthesis. Signal Process.: Image Commun. **111**, 116894 (2023)
2. Wen, W., Huang, H., Qi, S., Zhang, Y., Fang, Y.: Joint coverless steganography and image transformation for covert communication of secret messages. IEEE Trans. Netw. Sci. Eng. **11**(3), 2951–2962 (2024)
3. Mandal, P.C., Mukherjee, I., Paul, G., Chatterji, B.: Digital image steganography: A literature survey. Inf. Sci. **609**, 1451–1488 (2022)
4. Zhou, Z.L., Cao, Y., Sun, X.M.: Coverless information hiding based on bag-of-words model of image. J. Appl. Sci. **34**(5), 527–536 (2016)
5. Anggriani, K., Chiou, S.-F., Wu, N.-I., Hwang, M.-S.: A robust and high-capacity coverless information hiding based on combination theory. Informatica **34**(3), 449–464 (2023)
6. Meng, L., Jiang, X., Zhang, Z., Li, Z., Sun, T.: A robust coverless video steganography based on maximum dc coefficients against video attacks. Multimed. Tools Appl. **83**(5), 13427–13461 (2024)
7. Meng, L., Jiang, X., Sun, T.: A review of coverless steganography. Neurocomputing **566**, 126945 (2023)
8. Chen, X., Zhang, Z., Qiu, A., Xia, Z., Xiong, N.N.: Novel coverless steganography method based on image selection and stargan. IEEE Trans. Netw. Sci. Eng. **9**(1), 219–230 (2020)
9. Kulkarni, T., Debnath, S., Kumar, J., Mohapatra, R.K.: Dct based robust coverless information hiding scheme with high capacity. In: 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 358–364 (2023). IEEE
10. Biswas, S., Debnath, S., Mohapatra, R.K.: Coverless image steganography based on dwt approximation and pixel intensity averaging. In: 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1554–1561 (2023). IEEE
11. Liu, Q., Xiang, X., Qin, J., Tan, Y., Tan, J., Luo, Y.: Coverless steganography based on image retrieval of densenet features and dwt sequence mapping. Knowl.-Based Syst. **192**, 105375 (2020)
12. Zhou, Z., Mu, Y., Wu, Q.J.: Coverless image steganography using partial-duplicate image retrieval. Soft. Comput. **23**(13), 4927–4938 (2019)
13. Qin, J., Wang, J., Sun, J., Xiang, X., Xiang, L.: Coverless image information hiding based on deep convolution features. In: National

Conference of Theoretical Computer Science, pp. 15–30 (2021). Springer
14. Liu, J., Tan, L., Zhou, Z., Li, Y., Chen, P.: A dynamic yolo-based sequence-matching model for efficient coverless image steganography. arXiv preprint arXiv:2401.11946 (2024)
15. Zhou, Z., Ding, Y.: A coverless image steganography method based on invertible neural network. In: Third International Symposium on Computer Engineering and Intelligent Communications (ISCEIC 2022), vol. 12462, pp. 657–665 (2023). SPIE
16. Liu, Q., Xiang, X., Qin, J., Tan, Y., Zhang, Q.: Reversible subfeature retrieval: Toward robust coverless image steganography for geometric attacks resistance. KSII Trans. Internet Inf. Syst. (TIIS) **15**(3), 1078–1099 (2021)
17. Liu, Q., Xiang, X., Qin, J., Tan, Y., Zhang, Q.: A robust coverless steganography scheme using camouflage image. IEEE Trans. Circuits Syst. Video Technol. **32**(6), 4038–4051 (2021)
18. Hu, D., Wang, L., Jiang, W., Zheng, S., Li, B.: A novel image steganography method via deep convolutional generative adversarial networks. IEEE Access **6**, 38303–38314 (2018)
19. Li, J., Niu, K., Liao, L., Wang, L., Liu, J., Lei, Y., Zhang, M.: A generative steganography method based on wgan-gp. In: Artificial Intelligence and Security: 6th International Conference, ICAIS 2020, Hohhot, China, July 17–20, 2020, Proceedings, Part I 6, pp. 386–397 (2020). Springer
20. Cao, Y., Zhou, Z., Wu, Q.J., Yuan, C., Sun, X.: Coverless information hiding based on the generation of anime characters. EURASIP J. Image Video Process. **2020**, 1–15 (2020)
21. Qin, J., Wang, J., Tan, Y., Huang, H., Xiang, X., He, Z.: Coverless image steganography based on generative adversarial network. Mathematics **8**(9), 1394 (2020)
22. Li, Q., Wang, X., Wang, X., Ma, B., Wang, C., Shi, Y.: An encrypted coverless information hiding method based on generative models. Inf. Sci. **553**, 19–30 (2021)
23. Peng, F., Chen, G., Long, M.: A robust coverless steganography based on generative adversarial networks and gradient descent approximation. IEEE Trans. Circuits Syst. Video Technol. **32**(9), 5817–5829 (2022)
24. Liu, X., Ma, Z., Ma, J., Zhang, J., Schaefer, G., Fang, H.: Image disentanglement autoencoder for steganography without embedding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2303–2312 (2022)
25. Wei, P., Li, S., Zhang, X., Luo, G., Qian, Z., Zhou, Q.: Generative steganography network. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 1621–1629 (2022)
26. Zhou, Z., Su, Y., Li, J., Yu, K., Wu, Q.J., Fu, Z., Shi, Y.: Secret-to-image reversible transformation for generative steganography. IEEE Trans. Dependable Secure Comput. **20**(5), 4118–4134 (2022)
27. Sun, Y., Liu, J., Zhang, R.: Large capacity generative image steganography via image style transfer and feature-wise deep fusion. Appl. Intell. **53**(23), 28675–28693 (2023)
28. Zhang, X., Peng, F., Long, M.: Robust coverless image steganography based on DCT and LDA topic classification. IEEE Trans. Multimed. **20**(12), 3223–3238 (2018)
29. Zou, L., Sun, J., Gao, M., Wan, W., Gupta, B.B.: A novel coverless information hiding method based on the average pixel value of the sub-images. Multimed. Tools Appl. **78**, 7965–7980 (2019)
30. Govindasamy, V., Sharma, A., Thanikaiselvan, V.: Coverless image steganography using haar integer wavelet transform. In: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), pp. 885–890 (2020). IEEE
31. Luo, Y., Qin, J., Xiang, X., Tan, Y.: Coverless image steganography based on multi-object recognition. IEEE Trans. Circuits Syst. Video Technol. **31**(7), 2779–2791 (2020)
32. Liu, Q., Xiang, X., Qin, J., Tan, Y., Qiu, Y.: Coverless image steganography based on densenet feature mapping. EURASIP J. Image Video Process. **2020**, 1–18 (2020)
33. Luo, Y., Qin, J., Xiang, X., Tan, Y., He, Z., Xiong, N.N.: Coverless image steganography based on image segmentation. Comput., Mater. Continua **64**(2), 1281–1295 (2020)
34. Meng, L., Jiang, X., Zhang, Z., Li, Z., Sun, T.: A robust coverless image steganography based on an end-to-end hash generation model. IEEE Trans. Circuits Syst. Video Technol. **33**(7), 3542–3558 (2022)
35. Zou, L., Li, J., Wan, W., Wu, Q.J., Sun, J.: Robust coverless image steganography based on neglected coverless image dataset construction. IEEE Trans. Multimedia **25**, 5552–5564 (2022)
36. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
37. 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), pp. 3–19 (2018)
38. Li, Y., Gemert, J.: Deep unsupervised image hashing by maximizing bit entropy. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 2002–2010 (2021)
39. Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes challenge: A retrospective. Int. J. Comput. Vision **111**, 98–136 (2015)
40. Jegou, H., Douze, M., Schmid, C.: Hamming embedding and weak geometric consistency for large scale image search. In: Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I 10, pp. 304–317 (2008). Springer
41. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, pp. 740–755 (2014). Springer
42. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset (2007)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.