IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received January 3, 2020, accepted January 17, 2020, date of publication January 27, 2020, date of current version February 5, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2969524*

# Technology of Hiding and Protecting the Secret Image Based on Two-Channel Deep Hiding Network

**FENG CHEN, QINGHUA XING, AND FUXIAN LIU**
Department of Air Defense and Anti-Missile, Air Force Engineering University, Xi'an 710051, China

Corresponding author: Feng Chen (cf1904812819@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71771216 and Grant 71701209.

**ABSTRACT** The development of new media technology brings serious security problems to the transmission of secret remote sensing or military images. It is a new and challenging task to study the technology of protecting these secret images. In this paper, based on the powerful spatial feature extraction capability of the convolutional neural network, a novel two-channel deep hiding network (TDHN) is designed by introducing advanced ideas such as skip connection, feature fusion, etc., and the two channels are respectively used to input the cover image and the secret image simultaneously. This network consists of two parts: the hiding network and the extraction network. The sender uses the hiding network to hide a secret image in a common cover image and generates a hybrid image called the hidden image. The receiver uses the extraction network to extract and reconstruct the secret image from the hidden image. Meanwhile, an innovative loss function is constructed by introducing two metrics called MSE and SSIM. Experimental results show that the TDHN optimized by the loss function can generate the hidden image and extracted image in high quality. The SSIM value between the hidden image and the original cover image is up to around 0.99, and the SSIM value between the extracted image and the original secret image is up to around 0.98. Through testing on different datasets, it is verified that the designed and optimized TDHN has excellent generalization capability, and thus it has important theoretical significance and engineering value.

**INDEX TERMS** Convolutional neural network, steganography technology, two-channel deep hiding network, skip connection, feature fusion.

## I. INTRODUCTION

With the rapid development of information technology, digital media has become an important carrier for military and commercial organizations to transmit information. Some important remote sensing or secret military images are easy to be intercepted and abused by criminals in the transmission process. At present, the research on remote sensing or military images focuses on image classification, image recognition and detection, high resolution reconstruction, etc., but the information security of these images has not aroused people's attention. How to ensure the security of these important images has become an imminent challenge. Currently, information protection technologies in the field of digital communication are mainly divided into two types. One is the traditional encryption technology [1], [2], encryption technology can protect the integrity and security of the secret content by converting the secret content into cipher-text. However, this kind of meaningless and garbled cipher-text is easy to attract the attention and interest of the monitor, so as to be attacked and intercepted. It is possible that the sender adopts very complex encryption algorithms, which will make it impossible for the attacker to crack the secret contents in a short time. But at the same time, the transmission process of the secret information is also cut off by the attacker, so that the secret information can't be effectively and timely delivered to the recipient, what's worse, the secret message also faces the risk of being deciphered at any time. Another type of information protection technology is digital steganography [3], [4], which has become popular in recent years. Steganography can be regarded as a disguised encryption technology, but it is different from the traditional encryption technology.

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

Digital steganography utilizes the human visual perception system's insensitivity to the redundant information in the digital carrier to embed the text or picture containing secret information into the digital carrier and ensures the carrier embedded secret information looks very similar to the original carrier visually, so as not to attract the attention and suspicion of the monitor, and avoid interception and attack by the monitor. Therefore, compared with the traditional encryption technology for the purpose of protecting secret content, the purpose of digital steganography is to conceal the transmission of secret information and the existence of the process. In practice, images are usually chosen as the carrier of steganography technology, because the image carrier has a large capacity to hold more secret information. Besides, the image carrier contains rich color and texture features, high frequency noise and other redundant information, which is easy to confuse the human visual perception system (HVPS) for the reason that the HVPS is not sensitive to the slight changes of image content, so selecting image as carrier has stronger security.

In recent years, deep learning based on neural network has been widely used in computer vision [5], [6], natural language processing [7] and other tasks due to its superior ability of feature extraction and feature representation, and has achieved remarkable success, especially its ability of self-learning and high-dimensional feature extraction, which are similar to the functions of human brain, has attracted extensive attention in various fields. In the field of remote sensing, deep learning methods have been successively applied to remote sensing scene classification [8], [9] and remote sensing target recognition [10], [11]. Generally, these deep learning methods will generate high-dimensional feature vectors, which can more objectively reflect the features of objects in essence [12], [13]. While extracting and analyzing features by hand is time-consuming, laborious and subjective, and the features and methods of artificial design usually only targets at specific objects or tasks, which obviously does not meet the actual requirements [14]. Deep convolutional neural network (CNN) has been widely used in the field of computer vision due to its ability to fully learn local complex structure information of images and extract high-dimensional features [15], [16]. The purpose of this paper is to hide an important remote sensing or military image in a common cover image, which requires automatic recognition and extraction of high-dimensional features of the two images, and then gets a high degree of feature fusion. Most of the image steganography methods [17]–[19] need to analyze and design the statistical characteristics of the image manually, the image features and texture synthesis through artificial analysis and design are generally limited to one-dimensional or low-dimensional statistical features, while the image contains complex high-dimensional features that can't be accurately described by establishing mathematical functions. Therefore, the CNN which can automatically extract high-dimensional features naturally becomes our first choice.

Based on the deep CNN, a novel two-channel deep hiding network (TDHN) is designed in this paper, which is composed of two parts: hiding network and extraction network. The first hiding network can fully extract the high-dimensional features of the cover image and the secret image, and then highly integrate the information of these two images, so as to embed the secret image into the cover image and produce a hidden image embedded with secret information that is visually similar to the original cover image; The second extraction network can automatically extract feature information from the hidden image embedded with secret information, and then reconstruct and restore the secret image.

The main contributions of this study are summarized as follows:

A novel TDHN is designed to protect secret remote sensing and military image information. This network can automatically learn the high-dimensional feature information of images based on data-driven, and realize end-to-end mapping between the original cover image and hidden image, and between the hidden image and extracted image.

Tune skills such as skip connection, residual connection, feature fusion, feature dimension reduction and multi-scale feature extraction are introduced into the model structure of TDHN, and multi-level features are integrated to optimize the structure and performance of TDHN.

An innovative loss function is constructed by introducing the metrics to measure the similarity of image content and the metrics to measure the similarity of image structure to optimize the parameters of the model iteratively. Finally, an amazing image hiding effect is achieved by the optimized model.

The remainder of this article is arranged as follows. Section 2 briefly reviews and analyzes the previous works related to the steganography technology, and gives a detailed introduction of the least significant bit (LSB) method. Section 3 gives a detailed description of our proposed framework of TDHN and then makes a description of three datasets. The experimental results and discussion are presented in Section 4. Finally, Section 5 summarizes this paper and puts forward the future work ideas.

## II. RELATED WORKS

### A. TRADITIONAL STEGANOGRAPHY METHODS

The most representative steganography method of traditional steganography techniques is the least significant bit (LSB) steganography method based on pixel level information [17]. The motivation for this approach is that the visual appearance of an image is mainly determined by the highest bit of each pixel instead of the lowest bit, and the lowest bit of each pixel is statistically similar to randomly generated noise. Therefore, by changing the LSB of a cover image and embedding a small amount of secret information into it, the overall visual appearance of the cover image generally remains unchanged. But this method has obvious and simple statistical rule, which makes itself easy to be attacked and deciphered. Based on more complex statistical rule, Holub and Fridrich [18] proposed a wavelet obtained weight (WOW) method to embed the secret information into the original image according to the texture complexity of an image, the more complex the image region, the more the pixel value was modified in that region. By introducing a general distortion function independent of the embedded domain, Holub and Fridrich [19] further designed a spatial-universal wavelet relative distortion (S-UNIWARD) method, which performed multidirectional filtering of the pixels in the spatial domain, gave smaller distortion to the elements with smaller residuals, and inserted more information into the region with more noise or complex texture in the cover image. However, the steganography capability of above methods is limited and the cover image will leave traces of modification, which may attract the attention of the attacker. To further improve the security of the steganography technique, scholars have proposed the novel concept of steganography without embedding (SWE). One kind of SWE methods were designed by mapping the secret information from the semantic features of natural images with the help of bag of words [20] or image hashing [21], [22]. This type of method needed to construct an image library collecting a set of natural images and then established the relationship between the secret information and the images in the library. The other kind of SWE methods were proposed to map the secret information from a class of texture synthesis[23], [24] with the help of carefully designed reversible mathematical functions. But the steganography capacity of the second type of SWE method is also limited.

It can be concluded that the traditional steganography methods usually have two obvious shortcomings. The first defect is that, for most studies, the amount of hidden information is limited and the size of the information to be hidden is quite small compared with the size of cover image, that means the relative capability of the cover image is very small. However, the research of this paper is to hide an important remote sensing or secret military image (in the form of N × N × R GB) in a common cover image of the same size. Due to the large amount of the secret information, obviously, the above methods can't meet the requirement, which will be a huge challenge for these traditional steganography methods.

The second defect is that traditional image steganography needs to analyze and design the statistical characteristics of the image manually, so as to design a reasonable mathematical functions. Therefore, the effect of traditional steganography methods depends on whether the feature model designed manually is reasonable or not, and whether it can accurately describe the statistical characteristics of the image. Meanwhile, extracting and designing feature manually are very laborious and heuristic, the selection of effective feature depends more on human experience, and establishing the relationship between the secret information and the images in the library takes a lot of time and energy.

### B. STEGANOGRAPHY METHODS BASED ON DEEP LEARNING

Traditional steganography methods need to design rules manually and analyze the characteristics of images artificially, which will takes a lot of time and energy. In recent years, some scholars have tried to apply the theory of deep learning to steganography by using the powerful feature extraction ability of deep learning.

Tang *et al.* [25] proposed an automatic steganography distortion learning framework with GAN, which can automatically learn embedding change probabilities for every pixel in a given spatial cover image and directly learn and analyze the distortion function. Zhu *et al.* [26] pointed out that neural networks can learn to use invisible perturbations to encode a rich amount of useful information. and can exploit this capability for the task of data hiding. Then, a pair of encoder and decoder networks are designed, where given an input message and cover image, and it can learn to reconstruct hidden information from the hidden image. However, the amount of the secret information embedded through the methods in literature [25], [26] is small. Both methods just can hide textual messages or specific form of data into the cover image. To make the designed model hide image message, Rehman *et al.* [27] proposed a deep learning based generic encoder-decoder architecture for image steganography, and it is a completely automatic steganography method for hiding one image to another. However, this method can only ensure that a gray image is embedded into a color image, and the capacity of the hidden information is also limited. To solve this problem, Baluja [28] proposed a deep neural networks and attempted to place a full size color image within another image of the same size. The deep neural networks are simultaneously trained to create the hiding and revealing processes. This method is to our knowledge the first try to use neural networks to visually hide a full N × N × RGB pixel secret image into another N × N × RGB cover image. However, the quality of the image produced by this method is not high, the extracted secret image and the original secret image have obvious distortion in chrominance and content details of images.

### C. LSB STEGANOGRAPHY METHOD

In order to show the manual rules and small capacity characteristics of the traditional steganography methods, this section will briefly introduce the LSB steganography method, which is the most representative method of traditional steganography methods. The basic idea of LSB steganography method can be explained vividly in Fig. 1.

The LSB steganography method presented in Fig. 1 can be explained as follows. For an RGB image with three color channels, the value of each channel of a pixel is determined by an 8-bit binary, with a value range of 0-255. The color of each pixel in the image depends on the value of its three channels, while the overall visual appearance of the image

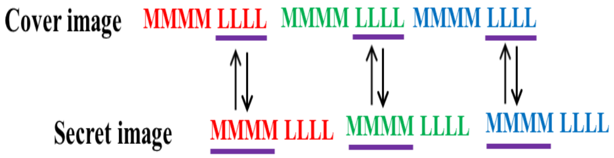


**FIGURE 1.** Explanation of the LSB steganography method.

mainly depends on the high bits of 8-bit binary, and the influence of the lowest bit on the visual appearance of the image is negligible. Assume that the byte of a channel of a pixel is 11111111 (equivalent to decimal 255), if the value 1 of the least significant bit is replaced by 0 (equivalent to decimal 254), and this kind of tiny change can't be detected in the visual appearance of the image. Therefore, most LSB methods aim to change the least significant bit of the image so as not to change the visual appearance of the image, but this method is usually used to hide a small amount of text information. The purpose of this paper is to hide a full N × N × RGB secret remote sensing image into another N × N × RGB common cover image. In order to hide all the secret information in the cover image, the 4 least significant bits of the cover image is replaced by the 4 most significant bits of the secret image, generating a hybrid image called the hidden image. In order to reconstruct and restore the secret image, the receiver copies the 4 least significant bits from the hidden image as the 4 most significant bits of the reconstructed secret image, and sets the remaining bits to 0.

## III. THE PROPOSED TDHN

### A. TDHN ARCHITECTURE

Our method is inspired by the traditional auto-encoding networks [29], which compresses and codes the input image through its compression network, and then reconstructs and restores the original input image from the compressed information through the corresponding decoding network. Its process can be described by mathematical formula, encoding process $y = encode(x)$, corresponding decoding process $\tilde{x} = decode(y)$. Under the constraint of $\tilde{x} = x$, the parameters of the model are trained and optimized by back propagation until the probability distribution of $\tilde{x}$ is closest to that of $x$, at this time, the decoded image is very similar to the original input image. Based on the powerful spatial feature extraction capability of the CNN, a novel TDHN is designed and its architecture is presented in Fig. 2.

Fig. 2 shows that the TDHN consists of two parts: The first hiding network has two input channels and one output channel. The cover image and the secret image to be hidden are input from channel 1 and channel 2 respectively. Through a series of convolution operations, complex and high-level feature information is extracted from massive low-level pixel information. Then the abstract and complex features of two channels are highly fused. Finally, the fused high-dimensional feature maps are reduced to three-channel dimensions (RGB) by convolution operation, and produce a hybrid image called the hidden image that is visually similar to the original cover image. The hidden image contains all the information of the imperceptible secret remote sensing image.

For the second extraction network, the input is a hidden image embedded with the secret image. The network also extracts abstract high-level feature information from massive low-level pixel information through a series of convolution operations, and uses these high-dimensional features to recover and reconstruct the secret image embedded in the hidden image.

The sender covertly uses the hiding network to hide the secret remote sensing or military image in the ordinary cover image, producing a hidden image with imperceptible secret information, and send the hidden image to the receiver. The receiver recovers the secret remote sensing or military image from the hidden image by the extraction network. Theoretically, the receiver hopes that the reconstructed secret image will be exactly the same as the original one, but it is difficult to achieve this goal in practice. However, for a secret image, there will be a lot of redundant information irrelevant to visual perception, even if it can't be reconstructed without loss, we can still judge the secret content. As long as the reconstructed secret image is as similar as possible to the original

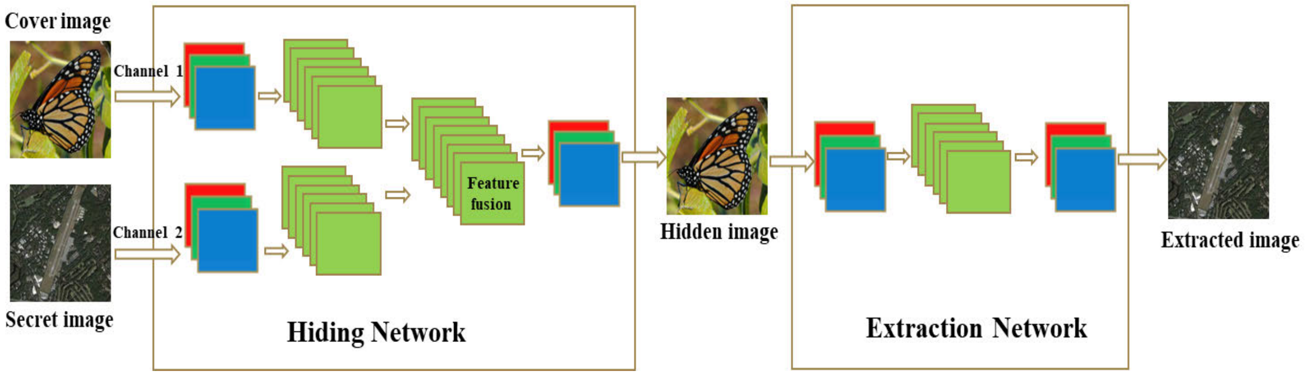


**FIGURE 2.** Architecture of TDHN. The TDHN consists of two parts: hiding network and extraction network.

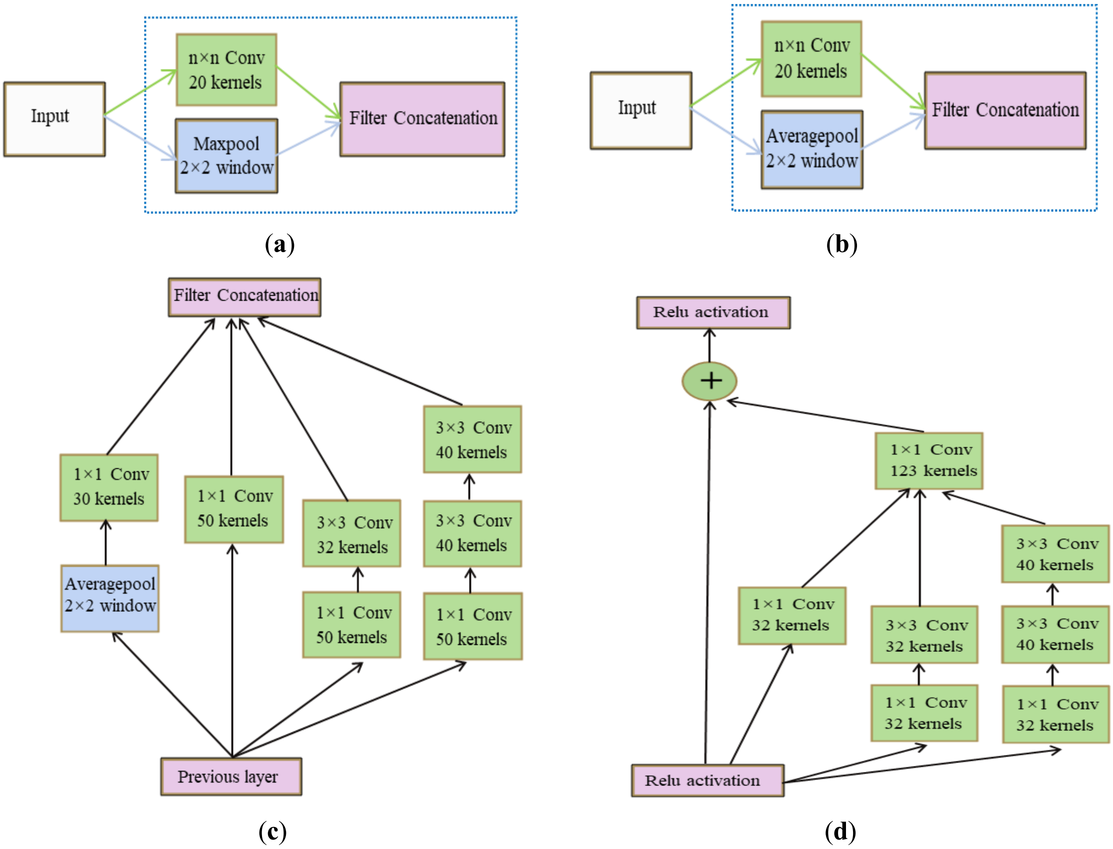


**FIGURE 3.** **Four main modules in TDHN. (a) structure of the Maxmod_n module; (b) structure of the Avgmod_n module; (c) structure of the inception module; (d) structure of the inception_Resnet module.**

secret image, this will not affect the receiver's understanding of the secret image information.

### B. MODULES IN TDHN

The previous section shows the overall architecture of TDHN, next, special convolution network structures are required to design the TDHN with excellent performance. Looking back on the development of AlexNet [30], VGGNet [31], GoogLeNet [32], andResNet [33], we can draw the following conclusions: The improvement of model performance mainly depends on three aspects, The first is to have a large and diverse training dataset, which can fully train and optimize the model; The second aspect is to increase the depth and width of the network to further improve the ability of through extracting complex high-dimensional features; The third is to improve the structure of the network, for example, by combining multi-scale convolution operations to get more abundant feature information, or by using identity mapping in the network structure to form a residual network to eliminate the gradient vanishing problem caused by the deep network. In this paper, we will fully absorb the advantages of various convolution network structures to design the TDHN for hiding secret remote sensing or military image. There will be four main modules in the TDHN, which is shown in Fig. 3.

The Maxmod_n and Avgmod_n module shown in Fig. 3 (a) and (b) are specially designed for the TDHN. The difference between them lies in the different types of pooling operations. Maximum pooling is to take the maximum value of feature points in the neighborhood, for digital images, high frequency or texture features can be fully preserved, while average pooling is to take the average value of feature points in the neighborhood, mainly to retain the overall data features and information integrity [34]. The purpose of this paper is to embed the important secret image into the common cover image and ensure that the synthesized hidden image is the same as the original image visually. In order to achieve the visual imperceptibility, the secret information is usually embedded in the high frequency region of the cover image with rich texture features. Therefore, Maxmod_n module is used to extract high-frequency texture features of the cover image, and Avgmod_n module is used to maintain the integrity of secret image information. The Inception module [35] in Fig. 3 (c) is a network with excellent topology designed by GoogLeNet team, and this module performs well in the task of ImageNet classification. It contains several convolution cores of different sizes as shown in Fig. 3 (c), that is, multi-scale convolution operations or pooling operations are performed on the input image in parallel to obtain different information of input image. This topology structure can well

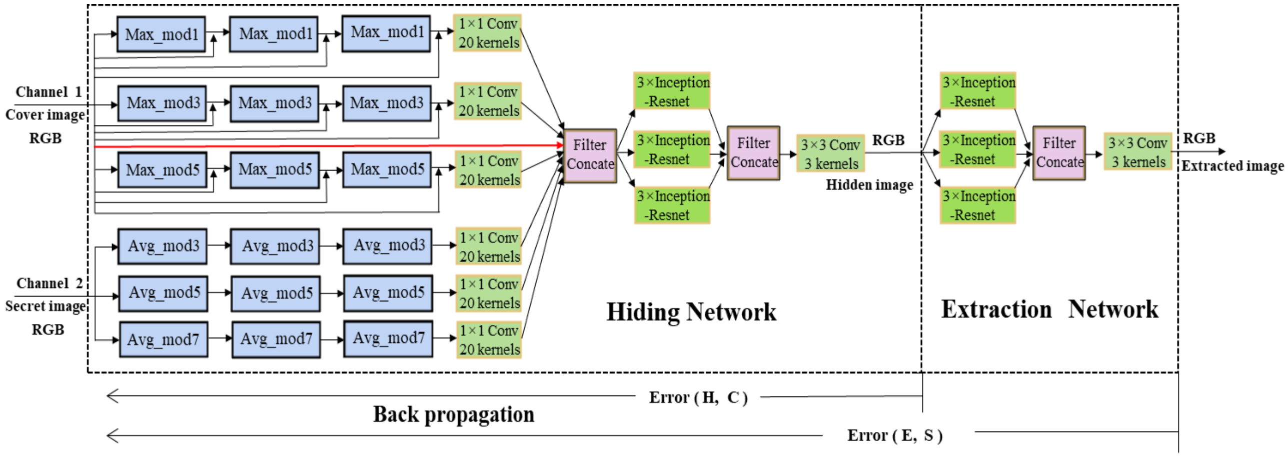


**FIGURE 4.** Implementation details of TDHN. H and C are the abbreviations of the hidden image generated by the hiding network and the original cover image respectively; Error (H, C) indicates the difference between the hidden image and the cover image; E and S are abbreviations of the extracted secret image and the original secret image respectively; Error (E, S) indicates the difference between the extracted secret image and the original secret image.

integrate the feature mappings of different sizes of receptive field. Compared with VGG-Net which overuses full connection network, it greatly reduces the scale of parameters. The Inception_Resnet [36] module shown in Fig. 3 (d) combines Inception and ResNet, which can effectively eliminate the gradient vanishing problem caused by the deep network through residual connections. And this module has achieved excellent performance in the reduction of training time and the improvement of image recognition accuracy.

### C. IMPLEMENTATION DETAILS OF TDHN

In this part, we mainly discuss and analyze the specific implementation details and ideas of the TDHN. As is illustrated in Fig. 4, there are two input channels in the hiding network. Channel 1 inputs an ordinary cover image in the form of RGB, it can obtain different information of the cover image through three parallel operations of different scales: Maxmod_1, Maxmod_3 and Maxmod_5. Parallel extraction of these feature information and combination of all high-level features will get better image representation. The maximum pooling operation in Maxmod_n can extract and retain the high frequency texture features of the cover image as much as possible. It should be noted that after each Maxmod_n, there will be a skip connection to connect the original cover image in the form of RGB. This is mainly inspired by the advanced feature fusion strategy in U-Net structured convolutional neural network [37] and Dense Convolutional Network (DenseNet) [38]. In the U-net structure, each upsampling operation is cascaded with the corresponding feature map from the downsampling layer by skip connection, so as to make full use of the features of the front layer, but it does not effectively use the information of the original image. In the DenseNet structure, each layer is cascaded with the features of all previous layers through a dense skip connection, for each layer, its input is the output features of all previous layers, and its own output feature map is the input of all subsequent layers. By reusing the features of each layer, it achieves a high degree of feature fusion. However, this kind of dense connection consumes a lot of memory during training, and the current deep learning framework does not support DenseNet's dense connection very well. The final output of the hiding network is to get a hidden image similar to the original cover image as much as possible. Therefore, we hope to retain the information of the original cover image as much as possible in Channel 1. Through skip connections, the low-level pixel information of the original cover image can be transmitted to each layer behind, so that the original information can be fused with the high-level features of each later layer. And the skip connection of red line can ensure that all the original information of the cover image can be transmitted to the last layer of the hiding network through the shortcut connection (identity mapping) in the Inception_Resnet module, which aids to generate a hidden image similar to the cover image visually. Moreover, in the process of back propagation, the gradient vanishing problem can be effectively alleviated by these skip connections.

For the secret image, we only want to extract its high-level feature information for hiding, and do not want its original image information to interfere with the visual appearance of the hidden image. Therefore, skip connection is not used in Channel 2 and advanced features of the secret image are extracted through three parallel operations of different scales: Avgmod_3, Avgmod_5 and Avgmod_7. The Avgmod_7 with larger convolution kernel size is used to extract the global information of the secret image in a larger range of receptive field. And the influence of the high frequency part of the secret image on the visual appearance of the hidden image is weakened through the average pooling operation in Avgmod_n module. Besides, it can retain the overall data characteristics and information integrity of the secret image, which is convenient for the later extraction network to reconstruct and restore the secret image. After each branch of Channel 1 and Channel 2, there will be a convolution operation of conv1 × 1, which mainly reduces the dimension and further enhances the representation of higher-order nonlinear features.

The middle part of the hiding network uses three parallel processing branches, each of which consists of three consecutive Inception_Resnet modules, by using the Inception_Resnet module in parallel, not only the width of the network is widened, but also the gradient vanishing problem can be effectively eliminated. In the last layer of the hiding network, the fused high-dimensional feature mappings are reduced to three-channel dimensions (RGB) and produce a hidden image by a convolution operation with three kernels of 3 × 3 Conv.

The extraction network uses three parallel processing branches, each of which is composed of three consecutive Inception modules. According to the requirements of this paper, the maximum pooling operation in the original Inception module designed by GoogLeNet is replaced by the average pooling operation. The main reason is that the input of the extraction network is a hidden image which is similar to the cover image visually. Therefore, the high-frequency region (texture and edge) of the hidden image contains more information of the original cover image, rather than secret information. Through the average pooling operation, the influence of the high-frequency region of the hidden image is weakened, and the global information in the hidden image is extracted as much as possible, which is conducive to the reconstruction and restoration of the secret image. In the last layer of the extraction network, a convolution operation with three kernels of 3 × 3 Conv is also used to reduce dimensions, and reconstruct the secret image through the fused high-dimensional feature mappings.

Fig. 4 shows the TDHN that realizes complete end-to-end hiding and end-to-end extraction functions. The hiding network and extraction network are trained together rather than separately. Each convolution operation in the TDHN follows a nonlinear activation function (Rectified Linear Unit, ReLU) [39] to extract nonlinear high-dimensional features except the last convolution operation with three kernels of 3 × 3 Conv that is used to generate RGB image. The TDHN calculates and adjusts the weight parameters of the network through the back propagation of errors. Through the back propagation process shown in Fig. 4, we can see that Error (H, C) and Error (E, S) affect the updating of weight parameters in different locations. Error (H, C) only affects the updating of weight parameters in the hiding network, but does not affect the weight parameters in the later extraction network. While Error (E, S) not only affects the updating of weight parameters in the extraction network, but also forces the previous hiding network to extract valuable high-dimensional features that are conducive to the reconstruction of the secret image by influencing the weight parameters of the hiding network.

### D. SIMILARITY AND LOSS FUNCTION

In order to evaluate the effect and performance of hiding, not only the size of the payload of the hidden image and the visual effect of the hidden image need to be considered, but also the quality of the generated image needs to be evaluated with quantitative metrics. In most previous work, the mean square error (MSE) between the original image pixel and the generated image pixel is usually used as metrics to evaluate the similarity between the two images. The MSE is computed as follows:

$$MSE = \frac{1}{N \times N} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (I(i,j) - G(i_w, j_w))^2 \tag{1}$$

where $I(i,j)$ and $G(i_w, j_w)$ represent the original image and the generated image respectively; $(i_w, j_w)$ indicates that the pixels of the generated image depend on the weight parameters of the network.

In addition to MSE, peak signal-to-noise ratio (PSNR) [40] is also a widely used metrics for evaluating image quality. This metrics is a standard to measure image distortion or noise level, and has been proved to be related to the average opinion score of human experts. Calculated by using:

$$PSNR = 10 \log_{10}(\frac{MAX_I^2}{MSE}) \tag{2}$$

where $MSE$ represents the mean square error between the original image and the generated image; $MAX_I$ represents the maximum value of the image pixels. If each pixel is represented by 8-bit binary, the maximum value is 255. The PSNR unit is dB, the smaller the value of MSE is, the larger the value of PSNR is; and the larger the value of PSNR is, the smaller the image distortion is, which means the higher the image quality. The general benchmark is 30dB. If the value of PSNR is lower than 30dB, the image distortion is obvious.

However, MSE only represents the pixel level error of two images, but ignores the correlation between the pixels, and the potential structural features of two images. PSNR is essentially the same as MSE, and it is the logarithmic representation of MSE. Actually, the natural image has a very high structure, which shows that there are strong correlations between the pixels of the image. These correlations carry important information about the structure of objects in the visual scene. Usually, the HVPS is more sensitive to the change of brightness and color in the non-texture area and it can sense the degree of image distortion by detecting whether the structural information changes, and so, structural distortion is an important consideration when measuring image quality. SSIM [41], [42] is introduced to compare the structural similarity between the generated image and the original image, which can be computed as follows:

$$SSIM(x, y) = [l(x, y)]^{\alpha} [c(x, y)]^{\beta} [s(x, y)]^{\gamma} \tag{3}$$

$$l(x, y) = \frac{2u_x u_y + C_1}{u_x^2 + u_y^2 + C_1} \tag{4}$$

$$c(x, y) = \frac{2\delta_x \delta_y + C_2}{\delta_x^2 + \delta_y^2 + C_2} \tag{5}$$

$$s(x, y) = \frac{\delta_{xy} + C_3}{\delta_x \delta_y + C_3} \tag{6}$$

In all of the above formulas, $x$ and $y$ represent the original image and the generated image respectively. SSIM metrics divides the task of similarity measurement into three comparisons: luminance, contrast and structural similarity. The luminance, contrast and structural similarity of the two images are measured by formula (4), formula (5) and formula (6) respectively. $u_x$ and $u_y$ represent the pixel mean of image $x$ and image $y$ respectively; $\delta_x$ and $\delta_y$ are the standard deviation of image $x$ and image $y$ respectively; $\delta_{xy}$ represents the covariance of image $x$ and image $y$. $\alpha > 0$, $\beta > 0$, $\gamma > 0$ are the parameters to adjust the relative importance of $l(x, y)$, $c(x, y)$ and $s(x, y)$; $C_1$, $C_2$ and $C_3$ are constants, which are used to maintain the stability of $l(x, y)$, $c(x, y)$ and $s(x, y)$, avoid denominator being set to zero.

In practical application, it is usually set $\alpha = \beta = \gamma = 1$, and then, combine formula (3), formula (4), formula (5) and formula (6) to obtain the following formula.

$$SSIM(x, y) = \frac{(2u_x u_y + C_1)(\delta_{xy} + C_2)}{(u_x^2 + u_y^2 + C_1)(\delta_x^2 + \delta_y^2 + C_2)} \tag{7}$$

where $C_1 = (k_1 L)^2$, $C_2 = (k_2 L)^2$; $L$ is the maximum value of pixel value;$k_1$ and $k_2$ are set as $k_1 = 0.01$, $k_2 = 0.03$ by default [43]; The value range of SSIM is [0, 1], the closer the SSIM value is to 1, the less distortion the two images have, when two images are completely identical, SSIM is equal to 1.

Consider the similarity of image content and image structure simultaneously, the mixed loss function is constructed by merging MSE and SSIM. PSNR is not introduced into the loss function, because it can't well reflect the perception relationship between HVPS and image quality and it is essentially equivalent to MSE [47], but PSNR is often used as an important metrics to measure the degree of image distortion. The loss function in this paper is constructed as follows:

$$\begin{aligned} Loss\_sum &= Loss(H_W, C) + Loss(E_W, S) \\ &= \eta MSE(H_W, C) + \mu\,[1 - SSIM(H_W, C)] \\ &\quad + \lambda MSE(E_W, S) + \kappa\,[1 - SSIM(E_W, S)] \end{aligned} \tag{8}$$

where $C$ and $H_W$ represent the original cover image and the generated hidden image respectively; $S$ and $E_W$ represent the original secret image and the extracted secret image respectively; $\eta$ and $\lambda$ are a pair of hyper parameters to trade off the content quality of hidden image and extracted image; $\mu$ and $\kappa$ are a pair of hyper parameters to trade off the structure quality of hidden image and extracted image.

### E. DESCRIPTION OF THREE DATASETS

Pascal VOC [44] is an excellent standardized dataset for image recognition and classification. From 2005 to 2012, an image recognition challenge will be held every year. The main purpose of this challenge is to recognize some kinds of objects in real scenes. VOC2007 dataset is the benchmark to measure the ability of image classification and recognition. There are 20 classifications in this dataset, including 9963 labeled pictures. The dataset includes five files called Annotation, ImageSets, JPEGImages, SegmentationClass and SegmentationObject. We only focus on the file called JPEGImages. Fig. 5 shows some class representatives from this dataset. More information about this dataset can be obtained at http://host.robots.ox.ac.uk/pascal/VOC/voc2007/.

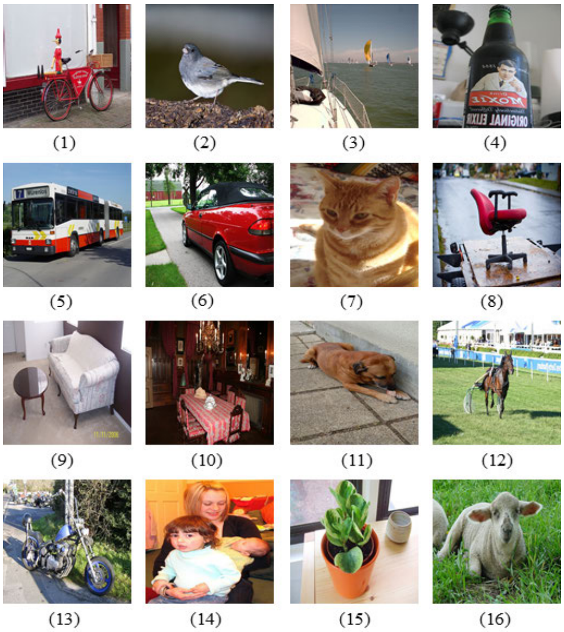


**FIGURE 5. Class representatives of the VOC2007 dataset: (1) Bicycle; (2) Bird; (3) Boat; (4) Bottle; (5) Bus; (6) Car; (7) Cat; (8) Chair; (9) Sofa; (10) Diningtable; (11) Dog; (12) Horse; (13) Motorbike; (14) Person; (15) Pottedplant; (16) Sheep.**

NWPU-RESISC45 dataset [13] is a new large remote sensing dataset, which contains 31500 aerial images and 45 scene categories. Each category contains 700 images with a size of 256 × 256 pixels. The dataset has the characteristics of large scale and rich image variations. Fig. 6 shows some class representatives from this dataset. More information about this dataset can be obtained at http://www.escience.cn/people/JunweiHan/NWPU-RESISC45.html

The AID dataset [8] is obtained from Google Earth using different remote sensing imaging sensors. There are 10000 images with a size of 600 × 600 pixels in this dataset, which are divided into 30 scene categories. The images in this dataset come from different countries in different seasons. Therefore, this dataset has rich inter-class variations. Fig. 7 shows some class representatives from this dataset. More information about this dataset can be obtained at https://pan.baidu.com/s/1mifOBv6#list/path =%2F

## IV. EXPERIMENTAL SETUP AND RESULTS

### A. EXPERIMENTAL PLATFORM AND SETUP

In our experiment, we use a workstation equipped with Intel Xeon(R) Bronze 3104 CPU running at 1.7 GHz,

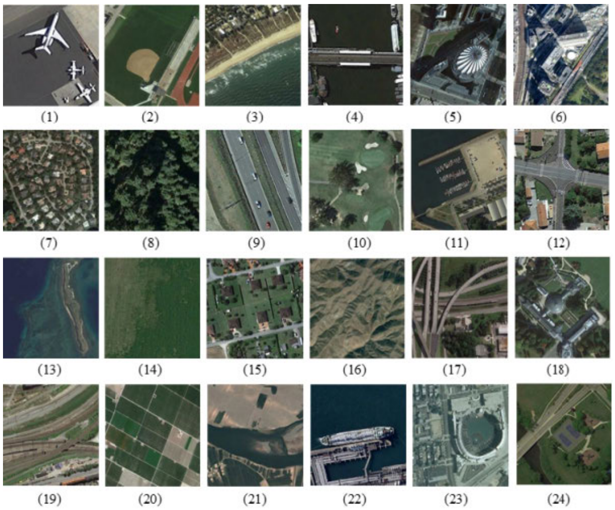


**FIGURE 6. Class representatives of the NWPU-RESISC45 dataset: (1) Airplane; (2) Baseball diamond; (3) Beach; (4) Bridge; (5) Church; (6) Commercial area; (7) Dense residential; (8) Forest; (9) Freeway; (10) Golf course; (11) Harbor; (12) Intersection; (13) Island; (14) Meadow; (15) Mediumresidential; (16) Mountain; (17) Overpass; (18) Palace; (19) Railway; (20) Rectangular farmland; (21) River; (22) Ship; (23) Stadium; (24) Tenniscourt.**

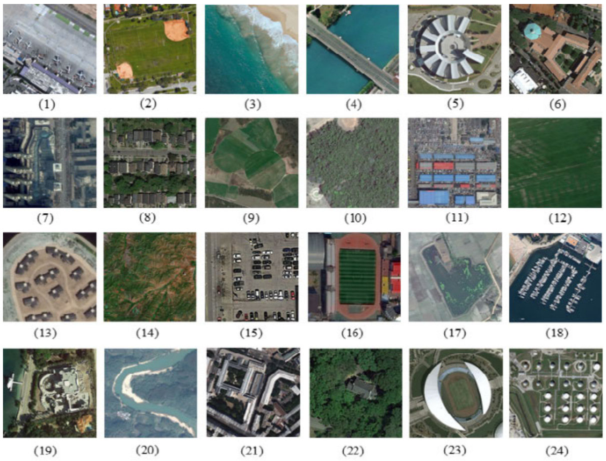


**FIGURE 7. Class representatives of the AID dataset: (1) Airport; (2) BaseballField; (3) Beach; (4) Bridge; (5) Center; (6) Church; (7) Commercial; (8) Dense residential; (9) Farmland; (10) Forest; (11) Industrial; (12) Meadow; (13) Mediumresidential; (14) Mountain; (15) Parking; (16) Playground; (17) Pond; (18) Port; (19) Resort; (20) River; (21) School; (22) SparseResidential; (23) Stadium; (24) StorageTanks.**

32GB DDR4 memory, a graphics processing unit (GPU) NVIDIA GeForce RTX2080Ti with a 11 GB memory. The deep learning framework in this paper was open source Tensorflow, which is based on the Ubuntu18.04 operating system and a Python3.6 interface.

In order to optimize the loss function better, this paper adopts Adam optimizer, which is computationally efficient and requires less memory. In order to show the large capacity embedding characteristics of our method, this paper embeds a secret image into a common cover image of the same size, instead of embedding a small secret image into a much larger image. Therefore, the cover image should be limited to the same size as the secret image. Due to the limitation of memory, two input images of the network are resized to 224 × 224. The VOC2007 dataset is selected as the dataset of common cover images, and the NWPU-RESISC45 dataset is selected as the dataset of secret images. Since the cover image and secret image are input and trained in pairs, we do not need to use all the pictures in the NWPU-RESISC45 dataset. According to the scale of VOC2007 dataset, we randomly select 9963 images from NWPU-RESISC45 dataset as the dataset of secret images. In this way, the dataset size will become smaller, but we adopt the way of random pairing between secret image and cover image to enhance the diversity of data, and the combined patterns can reach $9963^2$, which is enough to meet the training needs. Then 90% of the two datasets are selected as the training set of cover image and secret image respectively, and the remaining images are used for testing to verify the generalization capability of our method. We use the training samples with a mini batch size of 4 because of the limitation of memory, that is to say, eight pictures are fed to the model at one time, four ordinary cover images and four secret remote sensing images.

In the process of hiding the secret image, we pay more attention to the quality of the generated hidden image to ensure the security of the transmission process. As for the extraction and reconstruction of the secret image, it can't completely restore the original secret image, but as long as the extracted image does not affect the receiver's understanding of the information and content of the original secret image, it can meet the requirements, which is the trade-off principle when setting parameters, thus, the hyper parameters can be set as Table 1.

**TABLE 1. Values of hyper parameters.**

| Hyper parameters | $\eta$ | $\lambda$ | $\mu$ | $\kappa$ | Learning rate |
|---|---|---|---|---|---|
| Value | 1 | 0.6 | 0.2 | 0.1 | $5\times10^{-4}$ |

## B. OPTIMIZATION PROCESS

In order to intuitively show the training and optimization process of the model, this section sets breakpoint and continuation training in the training process to show the hidden effect of the secret image under different training epochs. The hidden effect of the secret image reflects the optimization process of the model. To ensure the objectivity of the comparison, we select the same pair of images and the experimental results are shown in Fig. 8 and Table 2.

Fig. 8 vividly shows that, in the early stage of model training, the outline information of the secret image is clearly visible on the hidden image, and the extracted image reconstructed by the extraction network and the original secret image are quite different in luminance, color and contrast. As the training epoch increases, the performance of the model is continuously optimized. When the training epoch

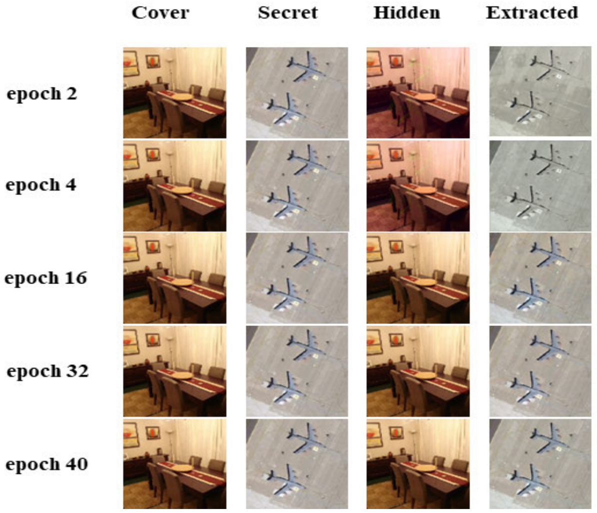


**FIGURE 8.** Optimization process on the NWPU-RESISC45 training set.

**TABLE 2.** Optimization process on the NWPU-RESISC45 training set.

| Epoch | SSIM<br>(H, C) | PSNR<br>(H, C) | SSIM<br>(E, S) | PSNR<br>(E, S) |
|---|---|---|---|---|
| 2 | 0.92 | 27.90 | 0.95 | 27.58 |
| 4 | 0.95 | 28.92 | 0.96 | 29.67 |
| 16 | 0.97 | 39.32 | 0.97 | 34.14 |
| 32 | 0.99 | 44.28 | 0.98 | 38.92 |
| 40 | 0.99 | 44.42 | 0.99 | 38.71 |

reaches 32, the optimized model can completely hide the secret image in the cover image, and the generated hidden image is quite similar to the original cover image visually. It is difficult to distinguish whether there are any traces of modification, so that it has enough security. And the reconstructed secret image is also very similar to the original secret image in vision.

Table 2 reflects from the perspective of specific quantitative metrics that with the increase of training epochs, the similarity between the hidden image generated by the hiding network and the original cover image increases gradually, and the similarity between the secret image extracted from the extraction network and the original secret image also increases gradually, which indicates that the quality of the generated image is getting higher and higher until it tends to be stable.

### C. COMPARISON OF HIDING PERFORMANCE

To fully verify the superiority and generalization capability of our designed TDHN, in the test set, we compare it with two methods, the LSB method which is the most representative method of embedding steganography methods, and the state-of-the-art image steganography method [28] based on deep learning. Comparison of their performance will be carried out from two aspects: one is subjective visual effect, the other is objective quantitative metrics. Next, we will define two other quantitative metrics, one is the changing rate, the other is the extracting rate.

In the process of embedding the secret image into the common cover image, it is necessary to modify the cover image to some extent, and the degree of modification can be quantified by the changing rate, which is defined as follows:

$$changing\ rate = \frac{1}{N \times N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{|H_W(i,j) - C(i,j)|}{C(i,j) + \varepsilon} \quad (9)$$

where $C(i,j)$ and $H_W(i,j)$ represent the original cover image and the hidden image embedded with secret information respectively; $\varepsilon$ is a small constant to avoid a certain pixel value in the denominator being zero. For color image, the value range of the pixel value is [0, 255], $\varepsilon = 1$ can be set.

When extracting the secret image from a hidden image, it is usually impossible to completely extract the secret image without loss. The extracting rate can be used to measure the effect and ability of extracting the secret image. It can be defined as follows:

$$extracting\ rate = 1 - \frac{1}{N \times N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{|E_W(i,j) - S(i,j)|}{S(i,j) + \varepsilon} \quad (10)$$

where $S(i,j)$ and $E_W(i,j)$ represent the original secret image and the extracted secret image respectively, and $\varepsilon = 1$ is also selected.

Here, three common cover images and three secret images to be hidden are selected from the test sets of VOC2007 and NWPU-RESISC45 datasets respectively. Then, three groups of test experiments are conducted with the above optimized model respectively, and the test results are shown in Fig. 9 and Table 3.

As is shown in Fig. 9, Column 1-2 show the original cover image and the original secret image; Column 3-4 show the hidden image and the extracted secret image; Column 5-6 show the residual image between the hidden image and the original cover image with different magnification.

From the perspective of subjective visual effect, it can be seen that hiding the secret image by LSB method will leave obvious modification traces on the hidden image, which is shown by the red box in Column 3 and it's easy to attract the attention or attack of the monitor. While the state-of-the-art image steganography method [28] based on deep learning and our proposed method have better hiding performance compared with the LSB method, no trace of modification can been found on the hidden image. However, Column 4 shows that the quality of the extracted secret image extracted by the state-of-the-art image steganography method [28] is poorer compared with our method, because the chrominance of the extracted secret image extracted by the method [28] is quite different from that of the original secret image. While our proposed method can completely embed the secret image into the ordinary cover image. The hidden image embedded with secret information is almost the same as the original

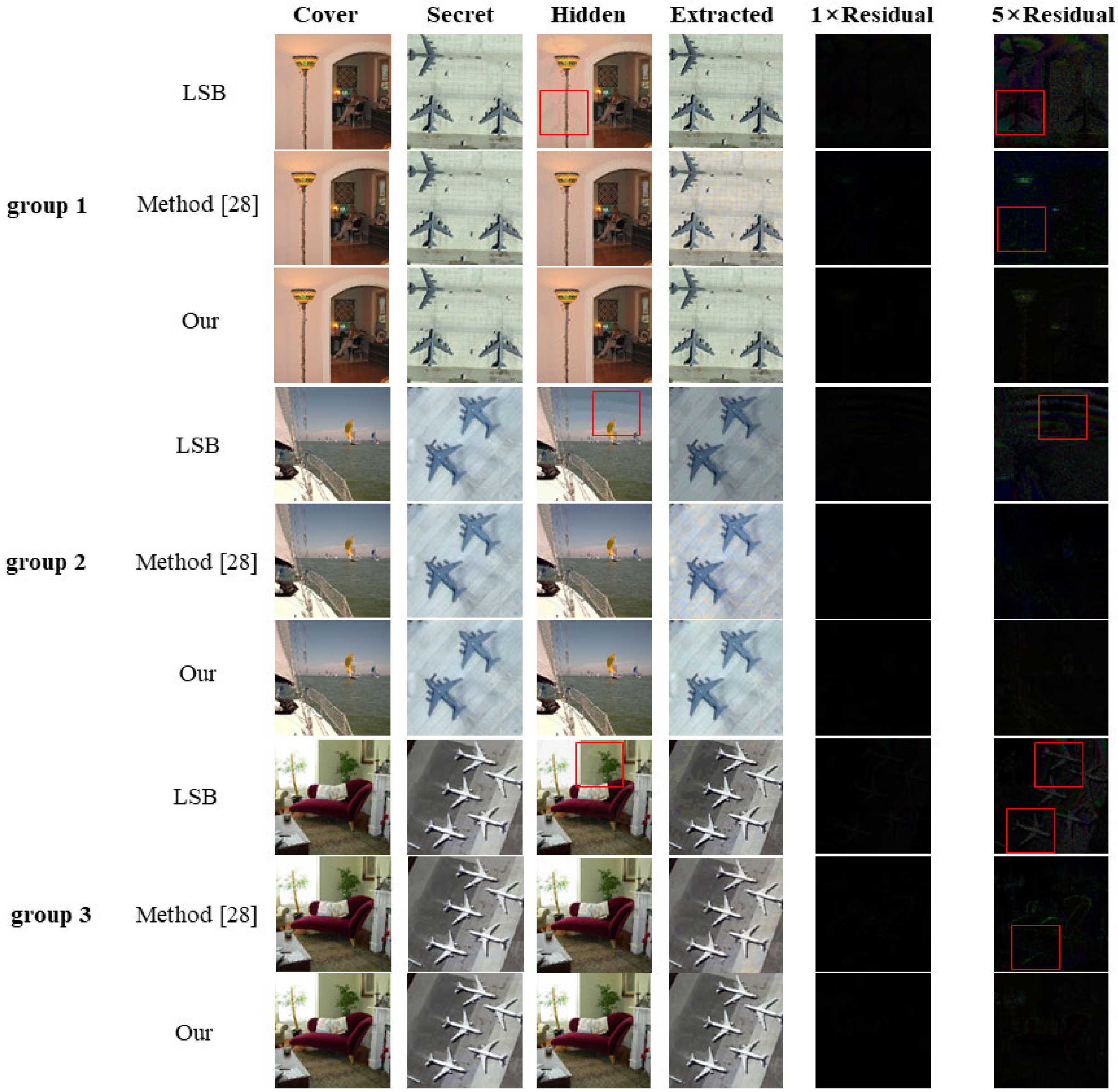


**FIGURE 9.** Comparison of hiding performance on the NWPU-RESISC45 test set.

cover image in vision. Moreover, without knowing any information about the secret image, the receiver can reconstruct and extract the high-quality secret image by the extraction network, which means our method has a stronger hiding reversibility compared with the traditional LSB method and method [28].

From the perspective of objective quantitative metrics, Table 3 shows that our method has a minimum value of changing rate compared with LSB method and the state-of-the-art image steganography method, and the changing rate is only 1.3-2.7%, which means our proposed TDHN only just needs a slight amount of modification on the hidden image to achieve an excellent hiding performance. The extracting rate of our method is largest compared with other two methods, which means our method has strongest hiding reversibility. Moreover, in terms of our method, the SSIM between the hidden image and the original cover image is surprisingly about 0.99 and the PSNR is also as high as 40-43. The SSIM and PSNR between the extracted secret image and the original secret image are as high as about 0.99 and 38 respectively. Obviously, The SSIM and PSNR values of our method are largest compared with other two methods, which shows that the hidden image and the extracted secret image generated by our method have highest quality. And thus, it can be concluded that the method proposed in this paper can achieve the best hiding performance compared with the LSB method and the state-of-the-art image steganography method.

To further illustrate the security of our designed TDHN in transmitting secret images, the last two columns of Fig. 9 show an extreme case. Suppose that our secret transmission process is discovered and intercepted by the attacker, and the attacker also has the original cover image (this case

**TABLE 3.** Comparison of hiding performance on the NWPU-RESISC45 test set.

| Group | Method | Changing rate (H, C) (%) | Extracting rate (E, S) (%) | SSIM (H, C) | PSNR (H, C) | SSIM (E, S) | PSNR (E, S) |
|---|---|---|---|---|---|---|---|
| 1 | LSB | 7.58 | 94.95 | 0.89 | 32.80 | 0.90 | 29.28 |
|  | Method [28] | 4.06 | 95.68 | 0.96 | 35.47 | 0.94 | 30.60 |
|  | **Our** | **2.36** | **98.04** | **0.98** | **40.30** | **0.98** | **37.81** |
| 2 | LSB | 7.01 | 94.58 | 0.92 | 34.19 | 0.87 | 29.26 |
|  | Method [28] | 2.27 | 97.49 | 0.98 | 39.60 | 0.91 | 32.05 |
|  | **Our** | **1.37** | **98.47** | **0.99** | **43.18** | **0.99** | **37.93** |
| 3 | LSB | 11.14 | 92.31 | 0.93 | 32.45 | 0.90 | 28.99 |
|  | Method [28] | 5.15 | 96.85 | 0.98 | 38.83 | 0.97 | 34.82 |
|  | **Our** | **2.7** | **98.05** | **0.99** | **42.87** | **0.99** | **39.74** |

usually does not exist in practice). If the attacker wants to obtain the secret image, he can extract the secret image through the residual image between the captured hidden image and the original cover image. Column 5 in Fig. 9 shows that no valuable information can be found through direct residual image for these three methods. However, when we enhance the residual image by 5 times, which is shown in the last column of Fig. 9, the LSB method obviously exposes the outline information of the secret image in the enhanced residual image. The method [28] exposes the outline information of the hidden image and a small amount of secret information, which is presented by the red box. While our method doesn't expose any secret information so that the attacker can't find any secret information, which verifies the strong security and engineering applicability of our designed TDHN.

### D. HIDING CAPACITY

To further illustrate the large hiding capacity of our method, we compare the hiding capacity with other steganography methods. At present, the hiding capacity of the traditional embedding steganography methods is relatively low. Since our method is a new hiding method, in order to make a more intuitive and convincing comparison, in this paper, we compare our method with the SWE methods proposed in recent years and the state-of-the-art method based on deep learning. The comparison results are shown in Table 4, where the second column is the absolute hiding capacity (hiding capacity per image), the third column is the size of the cover image, and the last column is the relative hiding capacity (hiding capacity per pixel), the relative hiding capacity can be calculated as follows:

$$Relative\ capacity = \frac{Absolute\ capacity}{The\ size\ of\ the\ image} \quad (11)$$

Table 4 show that the relative hiding capacity of the state-of-the-art method based on deep learning is much greater than those of SWE methods. The relative hiding capacity of the state-of-the-art method based on deep learning is 2.5e-2 bytes/pixel, which is shown in Row 6. However, the relative hiding capacity of our method is improved significantly and it is much greater than that of the state-of-the-art method based on deep learning. The relative hiding capacity of our method is 1 bytes/pixel, which is shown in the last row. Rows 1-5 show the capacities of SWE methods. In other words, our method outperforms these state-of-the-art methods.

**TABLE 4.** Comparison of hiding capacities.

| Methods | Absolute capacity (bytes/image) | Image size | Relative capacity (bytes/pixel) |
|---|---|---|---|
| [21] | 1.125 | 512×512 | 4.77e-6 |
| [22] | 2.25 | 512×512 | 8.58e-6 |
| [20] | 3.72 | ≥512×512 | 1.42e-5 |
| [24] | 1535~4300 | 1024×1024 | 1.46e-3~4.10e-3 |
| [23] | 64×64 | 800×800 | 6.40e-3 |
| [26] | 6.5 | 16×16 | 2.5e-2 |
| 0ur | **224×224** | **224×224** | **1** |

### E. WIDE APPLICATION AND HIGH SECURITY

To further illustrate the wide application and high security of our designed model, the proposed TDHN model was applied to a new dataset called AID dataset which is obtained from Google Earth by using different remote sensing imaging sensors, and the images in the dataset come from different countries in different seasons. Therefore, this dataset has rich inter-class variations and it is totally different from the NWPU-RESISC45 dataset in terms of image structure and form, which is a huge challenge for the TDHN model. The specific results are shown in Fig. 10 and Table 5.

It can be seen from Fig. 10 that the TDHN designed in this paper can still achieve excellent performance on a new AID dataset, and can hide the secret image in the cover image without changing the visual appearance of the cover image. Moreover, the extraction network can highly restore the original secret image and its specific details. Table 5 shows that the structure similarity metrics SSIM between the hidden

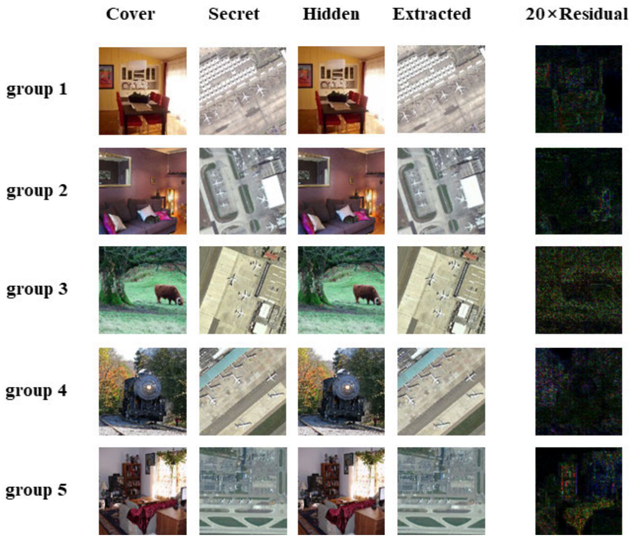


**FIGURE 10.** Performance of TDHN on the new AID dataset.

**TABLE 5.** Performance of TDHN on the new AID dataset.

| Group | SSIM (H, C) | PSNR (H, C) | SSIM (E, S) | PSNR (E, S) |
|---|---|---|---|---|
| 1 | 0.98 | 40.19 | 0.98 | 33.75 |
| 2 | 0.99 | 41.52 | 0.98 | 36.80 |
| 3 | 0.98 | 36.03 | 0.96 | 32.76 |
| 4 | 0.99 | 38.22 | 0.96 | 32.05 |
| 5 | 0.99 | 38.20 | 0.98 | 36.25 |

image generated by the hiding network and the original cover image is maintained above 0.98, and the content similarity metrics PSNR between them is around 38. The structure similarity metrics SSIM between the extracted secret image and the original secret image is maintained above 0.96, and the content similarity metrics PSNR is around 34. Both the generated hidden image and the extracted secret image are of high quality, which verifies the excellent migration capability of our designed TDHN.

In this part, we dramatically enhance the residual image between the hidden image and the original cover image by 20 times as shown in the last column of Fig. 10. Through the enhanced residual image, we still can't find any secret information, only the outline of the cover image can be seen, which further verifies the excellent performance and high security of our designed TDHN.

## V. CONCLUSION

In this paper, a novel end-to-end TDHN is designed based on the strong feature extraction ability of CNN and the structure of TDHN is optimized by using the tune skills such as skip connection, feature dimension reduction, feature fusion, etc. The simulation results show that the innovation of architecture and loss function makes our designed TDHN have large hiding capacity and sound hiding performance. The relative hiding capacity of our method is 1 bytes/pixel, which is much greater than that of the state-of-the-art method. Moreover, the hidden image embedded with secret information and the original cover image have quite high similarity in both visual appearance and statistical characteristics. The SSIM value between the hidden image and the original cover image is up to around 0.99. Meanwhile, the receiver can extract the secret image from the hidden image with high quality by using the extraction network of TDHN. And the SSIM value between the extracted image and the original secret image is up to around 0.98. The TDHN designed in this paper can completely realize end-to-end automatic hiding and extraction. Moreover, the model of TDHN has excellent generalization and migration capability, and has a wide prospect of engineering application.

It is a challenging and meaningful task to hide and protect secret or important military remote sensing images. Inspired by GANs, we plan to introduce GANs into our framework in the future to reduce dependence on data-driven. There is a lot of redundant information in secret image for the HVPS and the HVPS mainly focuses on the important information of secret image. Next, we will introduce the attention theory into the model design to simplify and refine the model, and further optimize and improve the hiding performance.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable suggestions.

## REFERENCES

[1] B. Koziel, R. Azarderakhsh, M. Mozaffari Kermani, and D. Jao, "Post-quantum cryptography on FPGA based on isogenies on elliptic curves," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 1, pp. 86–99, Jan. 2017.

[2] J. Howe, A. Khalid, C. Rafferty, F. Regazzoni, and M. O'Neill, "On practical discrete Gaussian samplers for lattice-based cryptography," *IEEE Trans. Comput.*, vol. 67, no. 3, pp. 322–334, Mar. 2018.

[3] H. Zhou, K. Chen, W. Zhang, and N. Yu, "Comments on 'steganography using reversible texture synthesis,'" *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1623–1625, Apr. 2017.

[4] K. Chen, H. Zhou, W. Zhou, W. Zhang, and N. Yu, "Defining cost functions for adaptive JPEG steganography at the microscale," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 4, pp. 1052–1066, Apr. 2019.

[5] C. Huang, C. C. Loy, and X. Tang, "Unsupervised learning of discriminative attributes and visual representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 5175–5184.

[6] Y. Wang, L. Zhang, X. Tong, L. Zhang, Z. Zhang, H. Liu, X. Xing, and P. T. Mathiopoulos, "A three-layered graph-based learning approach for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6020–6034, Oct. 2016.

[7] H. Strobelt, S. Gehrmann, M. Behrisch, A. Perer, H. Pfister, and A. M. Rush, "Seq2seq-Vis: A visual debugging tool for sequence-to-sequence models," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 353–363, Jan. 2019.

[8] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[9] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land–cover classification of high-resolution imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, Apr. 2017.

[10] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[11] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.

[12] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5585–5599, Oct. 2017.

[13] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.

[14] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approachh," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[15] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, and D. Li, "Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 4224–4231, Sep. 2018.

[16] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, Sep. 2018.

[17] J. Mielikainen, "LSB matching revisited," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 285–287, May 2006.

[18] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2012, pp. 234–239.

[19] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP J. Inf. Secur.*, vol. 2014, no. 1, pp. 1–13, 2014.

[20] Z. L. Zhou, Y. Cao, and X. M. Sun, "Coverless information hiding based on bag-of-words model of image," *J. Appl. Sci.*, vol. 34, no. 5, pp. 527–536, 2016.

[21] Z. Zhou, H. Sun, R. Harit, X. Chen, and X. Sun, "Coverless image steganography without embedding," in *Proc. Int. Conf. Cloud Comput. Secur.*, 2015, pp. 123–132.

[22] S. Zheng, L. Wang, B. Ling, and D. Hu, "Coverless information hiding based on robust image hashing," in *Intelligent Computing Methodologies*. Cham, Switzerland: Springer, 2017, pp. 536–547, doi: 10.1007/978-3-319-63315-2_47.

[23] J. Xu, X. Mao, X. Jin, A. Jaffer, S. Lu, L. Li, and M. Toyoura, "Hidden message in a deformation-based texture," *Vis. Comput., Int. J. Comput. Graph.*, vol. 31, no. 12, pp. 1653–1669, 2015.

[24] K.-C. Wu and C.-M. Wang, "Steganography using reversible texture synthesis," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 130–139, Jan. 2015.

[25] W. Tang, S. Tan, B. Li, and J. Huang, "Automatic steganographic distortion learning using a generative adversarial network," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1547–1551, Oct. 2017.

[26] J. Zhu, R. Kaplan, J. Johnson, and L. Fei-Fel, "HiDDeN: Hiding data with deep networks," in *Proc. ECCV*, 2018, pp. 682–697.

[27] A. U. Rehman, R. Rahim, M. S. Nadeem, and S. U. Hussain, "End-to-end trained CNN encoder-decoder networks for image steganography," in *Proc. ECCV*, 2018, pp. 723–729.

[28] S. Baluja, "Hiding images in plain sight: Deep steganography," in *Proc. NIPS*, 2017, pp. 2069–2079.

[29] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst.*, La Jolla, CA, USA, 2012, pp. 1097–1105.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas Valley, NV, USA, Jun./Jul. 2016, pp. 770–778.

[34] Y. L. Boureau, F. Bach, Y. Lecun, and J. Ponce, "Learning mid-level features for recognition," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2559–2566.

[35] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015, *arXiv:1512.00567*. [Online]. Available: https://arxiv.org/abs/1512.00567

[36] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-Resnet and the impact of residual connections on learning," 2016, *arXiv:1602.07261*. [Online]. Available: https://arxiv.org/abs/1602.07261

[37] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, 2015, pp. 234–241.

[38] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 21–26.

[39] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist. (ICAIS)*, Klagenfurt, Austria, Sep. 2011, pp. 315–323.

[40] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.

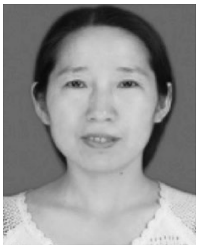
[41] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[42] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2941–2958, May 2017.

[43] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 1477–1480.

[44] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

**FENG CHEN** received the bachelor's degree from Hunan University, Hunan, China, in 2016, and the master's degree from Air Force Engineering University, Xi'an, in 2019, where he is currently pursuing the Ph.D. degree. His research interests include deep learning, information security, and computer vision.



**QINGHUA XING** received the bachelor's degree from Shanxi University, Shanxi, China, in 1989, and the master's and Ph.D. degrees from Air Force Engineering University, Xi'an, in 1992 and 2003, respectively. She is currently a Professor with the Air Force Engineering University. Her research interests include computer vision and information security.

**FUXIAN LIU** received the bachelor's degree from Lanzhou University, Lanzhou, China, in 1994, and the master's and Ph.D. degrees from Air Force Engineering University, Xi'an, in 1998 and 2001, respectively. He is currently a Professor with the Air Force Engineering University. His research interests include deep learning and pattern recognition.

• • •