A Universal Framework for Improving the Robustness of Coverless Image Steganography Based on Image Restoration

Laijin Meng, Fan Li, Xinghao Jiang, *Senior Member, IEEE*, and Qiang Xu, *Member, IEEE*

***Abstract*— Compared with traditional modification image steganography, coverless image steganography can resist the detection of steganalysis algorithms relying on no modification to the carriers. Previous works have made great efforts to improve the robustness against image attacks. However, the robustness of resisting geometric attacks performs not that well. After studying the general flow of the coverless image steganography, we find out that the receiver always needs to generate or map the hash sequences directly from the received images, which causes a significantly negative impact for extracting correct secret information because these received images might be attacked. Inspired by this finding, we surprisingly explore a common way to solve the problem by proposing a universal restoration framework for the attacked images. The most important module of the framework, the restoration module, contains two main parts, i.e., the classification sub-module and the attack restoration sub-module. The attacked images at the receiving end are first sent to a classification sub-module to estimate the type of the attack. Then, the corresponding attack restoration sub-module is utilized to repair the attacked images to improve the robustness. Experimental results show that the robustness of the existing coverless image steganography methods have been greatly improved after using the proposed framework without introducing extra security issues.**

***Index Terms*— Coverless image steganography, universal framework, high robustness, image restoration, image attacks.**

I. Introduction

IN THE field of information security, people have started to pay attention to their personal information security due to the rapid advancement of social networks and computer technology. Different from cryptography, steganography is devoted to embedding the secret information into carriers without raising suspicion. As for carriers, researchers tend to choose text [1], [2], images [3], [4], [5], [6], [7], and videos [8], [9], [10], [11] to embed the secret information. Among them, images attract more researchers due to their wide application scenarios [12]. Previous image steganography methods accomplished embedding the secret information by modifying the parameters of carriers, which would inevitably leave a trace of modification. Although these modifications are too minor to be noticed by human eyes, they can be detected by specially designed steganalysis tools [13], [14], [15], [16], [17]. Therefore, how to fundamentally solve this security issue has become a hot spot for researchers.

In 2015, the concept of coverless steganography was proposed by Zhou et al. [18]. It does not mean that no carriers are needed when hiding the secret information, but no modification to the carriers. Different from the previous modification steganography, it embeds the secret information by constructing a mapping rule between the carrier and the secret bits. Since there is no trace of modification left on the carrier, the threat of steganalysis methods has disappeared. Recently, more coverless image steganography methods have been proposed. According to how to obtain the cover images, the existing coverless image steganography can be divided into two types, i.e., construction-based methods and selection-based methods. The construction-based methods use generative models, i.e., the generative adversarial network (GAN) and glow model to generate stego-images from the secret information. For example, different kinds of GANs, i.e., DCGAN, WG-GAN, and StarGAN, were utilized to generate images in the previous works [19], [20], [21], [22], [23], [24], [25], [26], respectively. However, these methods cannot resist steganalysis detection fundamentally.

As for the selection-based methods, Zhou et al. [18], as the pioneer of coverless image steganography, utilized the average pixel of the image to generate hash sequences. Then, many other works have been proposed recently. For this type of coverless image steganography, it can be further divided into two types of methods according to whether the hash sequence is generated directly from the image, i.e., generation-based methods [18], [27], [28], [29], [30], [31], [32] and mapping-based methods [33], [34], [35], [36]. The former type utilized all kinds of hash generation methods to extract hash sequences from images. However, the latter one constructed a mapping dictionary between the fixed hash sequences and the images by some rules, i.e., they artificially assigned a hash sequence to each image.

Although the existing methods have greatly promoted the research on the coverless image steganography, many issues still need to be discussed. First, as a lossy channel, social networks contain various types of attacks. The images may be

Manuscript received 24 May 2024; revised 8 August 2024 and 22 August 2024; accepted 2 September 2024. Date of publication 4 September 2024; date of current version 30 January 2025. This work was supported by the National Natural Science Foundation of China under Grant 62272299. This article was recommended by Associate Editor Z. Xiang. *(Corresponding author: Xinghao Jiang.)*

Laijin Meng, Xinghao Jiang, and Qiang Xu are with the National Engineering Laboratory on Information Content Analysis Techniques, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: menglaijin@sjtu.edu.cn; xhjiang@sjtu.edu.cn; xuqiangwhu@sjtu.edu.cn).

Fan Li is with the School of Information and Communication Engineering, Hainan University, Haikou 570228, China (e-mail: fanli@hainanu.edu.cn).

Digital Object Identifier 10.1109/TCSVT.2024.3454457

attacked during transmission. Therefore, the ability of resisting various image attacks is very essential for coverless image steganography methods. Although some existing methods have been specially designed for geometric attacks, most cannot resist them well. Geometric attacks are image processing attacks that change the pixel position or geometry of the image. Compared with non-geometric attacks, these attacks usually have a greater impact on the image content. Second, in terms of security, most of the existing methods need to transmit some other auxiliary information along with the stego-images, which would arouse the suspicion of the monitors. So, how to improve the robustness of coverless image steganography against all kinds of image attacks on the premise of ensuring security is still an important issue. Last, the existing research on the selection-based coverless image steganography is scattered and there is no research work to integrate the existing methods and give a general improved method.

From the above analysis, we find that the common characteristic for both generation-based methods and mapping-based methods is that their robustness can be improved by repairing attacked images. Especially for the generation-based methods, there is still much room for promotion. In response to this discovery, we have an innovative idea of proposing a universal image restoration framework that is universal and suitable for all existing algorithms. However, it is still a challenge to work for so many algorithms and resisting so many attacks. The main module of the proposed framework is the restoration module, which contains two main parts, i.e., the classification sub-module and the attack restoration sub-module. The attacked images at the receiving end are first sent to a classification module to determine the type of attacks. Then, the corresponding restoration sub-module is utilized to repair the attacked images to improve the robustness. After utilizing this framework for different types of coverless image steganography methods, the robustness of resisting common image attacks, including geometric attacks (e.g., rotation, cropping, and translation) and non-geometric attacks (e.g., noise, filter, and JPEG compression) can be improved significantly without introducing additional security issues.

In summary, the main contributions of this paper are concluded as follows.

- A universal framework for improving the robustness of coverless image steganography is proposed. This framework contains two main parts, i.e., the classification sub-module and the attack restoration sub-module. The restoration module adopts the blind image restoration, which can complete the process of image restoration without additional given specific parameters of the attacks, such as rotation angle and noise intensity, etc.
- The proposed framework is not limited by the type of coverless image steganography and can be used for most existing methods. To the best of our knowledge, it is the first universal framework to improve the robustness of the coverless image steganography.
- Since the restoration operation is accomplished at the receiving end, there is no need for the communication parties to transmit additional information, and the original security of coverless image steganography is not affected.
- Experimental results show that the proposed framework can effectively improve the robustness of the existing typical coverless image steganography methods without affecting the capacity and security of the algorithms.

The rest of this paper is organized as follows. Related work is introduced in Section II. The proposed universal framework is described in Section III. The experimental results and analysis are discussed in Section IV and Section V. Conclusions are drawn in Section VI.

II. Related Work

A. Selection-Based Coverless Image Steganography

Since the construction-based coverless image steganography methods cannot fundamentally resist steganalysis [36], we focus on the selection-based coverless image steganography in this paper. As the pioneer of coverless image steganography methods, Zhou et al. [18] proposed a novel coverless image steganography by generating hash sequences from the average pixel. After that, researchers have started to focus on improving the robustness of resisting all kinds image attacks. Zheng et al. [27] introduced Scale Invariant Feature Transform (SIFT) in generating hash sequences. Zhang et al. [28] and Liu et al. [29] obtained the low-frequency part of the image by using discrete cosine transform (DCT) and discrete wavelet transform (DWT), respectively. Liu et al. [30] provided a new idea by transmitting camouflage images instead of stego-images to the receiver. Similar to Liu et al.'s [29] method, Karim et al. [31] extracted the average value of the LL subband to generate hash sequences. Meng et al. [32] combined the process of extracting features and generating hash sequences by an end-to-end network. Zou et al. [33] generated hash sequences by the average pixel value of each sub-image and realized the secret information hiding through the mapping relationship. Zhou et al. [34] and Luo et al. [35] tried to improve the robustness of resisting geometric attacks by the object detection network. Zou et al. [36] attempted to improve the construction efficiency of the coverless image database.

According to whether the hash sequence is generated directly from the image, the selection-based methods can be divided into generation-based and mapping-based methods. In coverless image steganography, the robustness represents whether different algorithms can still correctly extract secret information after various image attacks. The robustness performance of these two types of selection-based methods against geometric attacks and non-geometric attacks is shown in Table I. From this table, it can be found that the robustness of different types of algorithms performs various against geometric attacks and non-geometric attacks, and even the same type of algorithms would also have different performances against different types of attacks. For example, Zhou et al.'s [18] and Liu et al.'s [30] methods both belong to the generation-based methods. However, the former method performs well in resisting non-geometric attacks, while the latter method shows the opposite.

To further improve the robustness of these two types of methods, we concluded the general flow of these two types of methods, as shown in Fig. 1. We find that the received images in both types of methods may be attacked during transmission. The attacked images are utilized to generate hash sequences for the generation-based methods by hash generation methods. For the mapping-based methods, the attacked images are also utilized to generate objects or retrieve original images in the coverless image database. In other words, a common feature of these two types of methods is that both the receiver and the

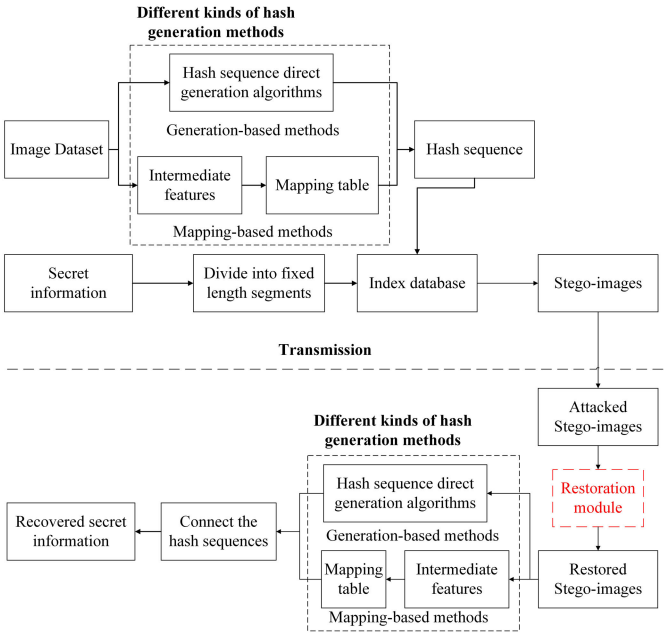


Fig. 1. The general flow of generation-based and mapping-based methods.

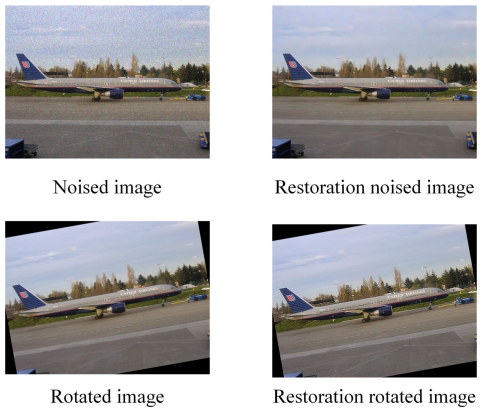


Fig. 2. The example of an image restoration network for repairing different attacks.

sender use the same hash generation method to extract the hash sequence from the attacked image. Obviously, if the image is not attacked, i.e., the received image is the same as the original stego-image, the extracted hash sequences will be the same as the sender's. Therefore, if we can repair the received images to make them close to the original ones, it will be of great help to improve the robustness of both generation-based and mapping-based methods, which is highlighted in red in Fig. 1. However, due to different kinds of methods utilizing different image features, it is still a challenge to propose a universal restoration framework. The example of an image restoration network for repairing different kinds of attacked images is shown in Fig. 2. The four images from left to right and from top to bottom are: the image with noise, the restoration noised image by using some kind of restoration network, the image with rotation, and the restoration rotated image by using the same kind of restoration network as used in repairing noised image. We can find that this network can repair the noised image well but has no effect on the rotated image. Therefore, we realize that it would be impossible to apply only one kind of restoration network to repair all kinds of attacked images.

TABLE I

THE ROBUSTNESS OF TWO TYPES OF SELECTION-BASED COVERLESS IMAGE STEGANOGRAPHY METHODS AGAINST GEOMETRIC AND NON-GEOMETRIC ATTACKS

		Geometric attacks	Non-geometric attacks
Generation -based methods	Zhou's method [18]	Poor	Good
	Zheng's method [27]	Poor	Poor
	Zhang's method [28]	Poor	Middle
	Liu's method [29]	Poor	Middle
	Liu's method [30]	Good	Poor
	Karim's method [31]	Poor	Good
	Meng's method [32]	Middle	Good
Mapping -based methods	Luo's method [35]	Middle	Poor
	Zou's method [33]	Poor	Middle
	Zhou's method [34]	Poor	Poor
	Zou's method [36]	Middle	Good

In summary, there are many selection-based coverless image steganography methods and different kinds of attacks that may be encountered in image transmission. Even we know that if the attacked images could be repaired as close as the original images, the robustness to the attacks would be improved eventually, but how to implement an image restoration module suitable for various algorithms and attacks is still a big challenge.

B. Image Restoration

According to the previous analysis, restoring the attacked images is the core technique in our framework. Therefore, it is important to figure out the existing image restoration works. Especially in the coverless image steganography, the type of attacks is unknown in advance. Therefore, we focus on concluding the blind image restoration methods without masks as follows.

Compared with image restoration with masks, blind image restoration is a challenge due to the unknown and multi-property complexity of contamination (e.g., diverse contents with various shapes and positions) in different contaminated images. The image blind restoration works mainly contain blind image denoising and blind image inpainting. For the former type, most existing methods contain two steps, i.e., noise estimation and non-blind denoising. For additive white Gaussian noise (AWGN), Chen et al. [37] developed a method to estimate noise standard deviation. Nam et al. [38] modeled the cross-channel noise as a multivariate Gaussian and performed denoising by the Bayesian nonlocal means filter. Guo et al. [39] developed a convolutional blind denoising network for the real-word images. Xu et al. [40] suggested a multi-channel weighted nuclear norm minimization model to exploit channel redundancy.

For the other type, i.e., blind image inpainting, Cai et al. [41] proposed a blind inpainting convolutional neural network to directly learn an end-to-end mapping between a pre-acquired dataset of corrupted/ground truth subimage pairs. After that, Zhang et al. [42] proposed a novel feature-oriented blind face

inpainting framework, which is implemented by an end-to-end network. Wang et al. [43] proposed a two-stage visual consistency network, which predicted semantically inconsistent regions first and then it repaired these estimated missing regions using a new spatial normalization. Zhao et al. [44] proposed a one-stage network, which combined the Transformer and CNN to repair contaminated regions.

In a word, the existing blind image restoration methods mainly focus on image denoising and image inpainting. There are limited types of image restoration works for common image attacks. In coverless image steganography, common image attacks including geometric and non-geometric attacks are considered. Some of the structures of the restoration network can be directly used for repairing the attacked images in coverless image steganography, i.e., blind image denoising network. However, for other attacks, e.g., the image translation attack, it needs to make a new design to repair. Therefore, we focus on designing a universal framework to eliminate the common image attacks, including geometric and non-geometric attacks that would be considered in coverless image steganography.

C. Motivation

As discussed in the last subsection, during the transmission of stego-images from the sender to the receiver, they would be inevitably damaged by a lot of attacks, i.e., geometric and non-geometric attacks. Considering that different types of attacks will destroy the image from different aspects, it is difficult to repair the image with a single restoration module. Fortunately, we develop a novel thought to solve this problem. First, an image classification network is used to classify the received images into several groups, and then different restoration sub-modules are designed for different groups of received images. The detail of the thought will be illustrated in the next section.

III. The Proposed Universal Framework

A. The Proposed Universal Framework

The whole process of the proposed framework is shown in Fig. 3. In this figure, a black dotted line partitions the whole framework into two parts, i.e., the part above this line is the process of hiding secret information and the below part shows extraction secret information. Specifically, for a given image database, different kinds of hash generation methods, including generation-based and mapping-based methods, are selected to generate hash sequences. Then, the hash sequences are used to construct the index database to accelerate the efficiency of searching. For the sender, the secret information is divided into fixed-length segments. After that, the images in the index database, corresponding to the secret segments are selected as stego-images. Finally, the stego-images are sent to the receiver. During the transmission, the stego-images would be damaged by image attacks.

At the receiving end, the attacked stego-images are put into the restoration module. The restoration module contains a classification sub-module and five attack restoration sub-modules. The attacked stego-images are classified by the classification module first to obtain the type of attacks. Then, they are repaired by the corresponding restoration sub-module. For example, if the attacked stego-images are classified to

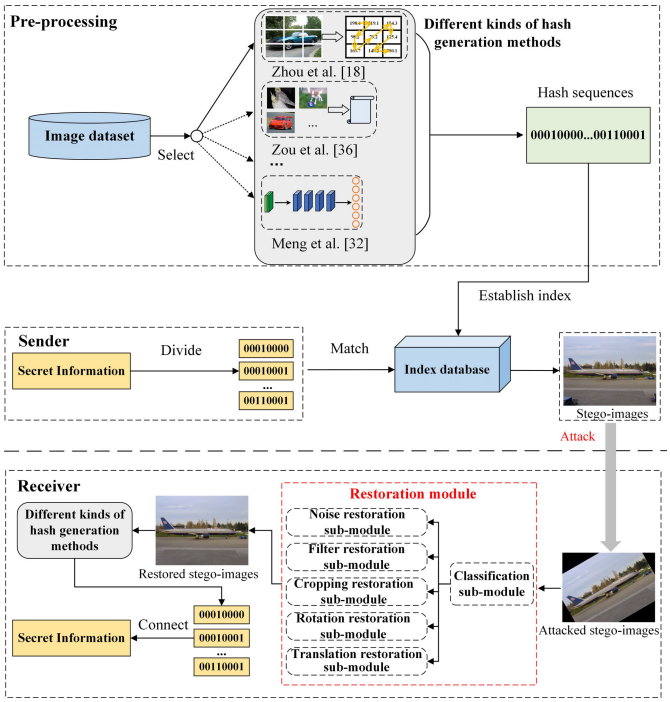


Fig. 3. The process of the proposed framework.

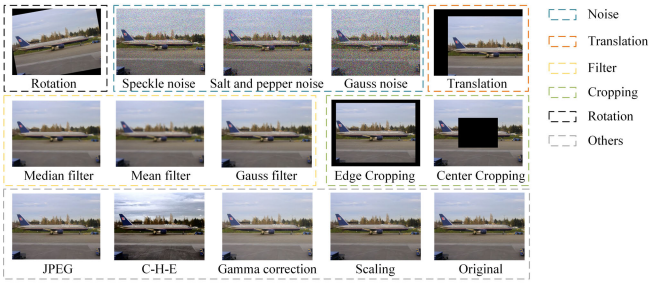


Fig. 4. The grouped categories of the classification module.

"Rotation", the rotation restoration sub-module is used to repair them. After obtaining the restored stego-images, the same kind of hash generation method is used to extract hash sequences. Finally, the segmented hash sequences are connected to recover the secret information. The specific introduction to the restoration module will be expanded in the next section.

B. Restoration Module

1) Classification Sub-Module: To further repair the attacked images precisely, a classification sub-module is used to obtain the category of the attacked stego-images first. The common attacks for the images are shown in Fig. 4. From this figure, we find that compared with the original image, the attacked images with noise, filtering, cropping, rotation and translation are changed as different characteristics. For geometric attacks, including cropping, translation, and rotation, the content of the image itself is missing, i.e., forming a black area. However, the location of the black areas varies greatly for different attacks, so they are each grouped into their own categories. For the noise and filtering in non-geometric attacks, the visual changes of the image as a whole are obvious compared with the original image. For example, the filtering images become significantly more blurry in general. So, the noise and filtering

TABLE II
THE STRUCTURE OF THE CLASSIFICATION MODULE

Layer	Configuration
Conv1	64×3×3, padding=1, ReLU
Conv2	64×3×3, padding=1, ReLU
MaxPooling3	kernel size=2×2, stride=2
Conv4	128×3×3, padding=1, ReLU
Conv5	128×3×3, padding=1, ReLU
MaxPooling6	kernel size=2×2, stride=2
Conv7	256×3×3, padding=1, ReLU
Conv8	256×3×3, padding=1, ReLU
MaxPooling9	kernel size=2×2, stride=2
Conv10	512×3×3, padding=1, ReLU
Conv11	512×3×3, padding=1, ReLU
MaxPooling12	kernel size=2×2, stride=2
Conv13	512×3×3, padding=1, ReLU
Conv14	512×3×3, padding=1, ReLU
MaxPooling15	kernel size=2×2, stride=2
Fully connected16	4096
Fully connected17	4096
Fully connected18	6
BatchNormalization	-
Softmax	-

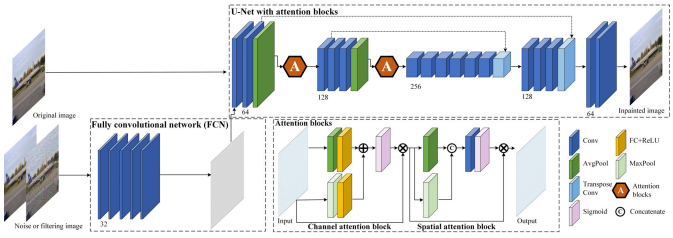


Fig. 5. The structure of the noise and filter restoration sub-module.

are treated as separate categories. For JPEG compression, scaling, color histogram equalization (C-H-E), gamma correction and original images, human vision changes are small, so we group these attacks into one category and label them as "others". Therefore, when we design the classification network, we divide the received attacked stego-images into six categories, i.e., noise, filter, cropping, translation, rotation, and others. For the first five categories, the corresponding restoration sub-module is designed for each category. For the category of "others", as most of the existing methods can resist these attacks well, we do not further design a restoration sub-module and the original stego-images are preserved.

In the proposed framework, a VGG-13 [45] network is selected as the backbone of this module, and the last fully connected (FC) layer is changed from FC-1000 to FC-6. In addition, to speed up the network convergence, we add a batch normalization (BN) layer after the last fully connected layer. The structure of the classification module is shown in Table. II. In short, after obtaining the category of the attacked stego-images, the corresponding attack restoration sub-modules, i.e., noise, filter, cropping, rotation, and translation restoration sub-modules are selected to repair the attacked stego-images.

2) Noise and Filter Restoration Sub-Module: For the noise and filtering attacks, they can be regarded as adding perturbations on the images. Therefore, these two restoration sub-modules use the same network structure. The overall structure of the proposed module is shown in Fig. 5. Inspired by [39], a plain five-layer fully convolutional network (FCN) is first used to predict perturbations applied to images. In each convolution layer, the number of feature channels is set to 32, and the kernel size is 3×3. Moreover, the ReLU function is followed after each convolutional layer. Assume that the original image is $\boldsymbol{I} \in R^{W \times H}$, and the attacked image with noise or filtering is $\boldsymbol{I}_n$. The process can be described as follows.

$$\hat{\boldsymbol{n}} = f_{FCN}(\boldsymbol{I}_n), \tag{1}$$

where $f_{FCN}(.)$ represents the fully convolutional network and $\hat{\boldsymbol{n}}$ devotes the estimated perturbation.

Then, the U-Net [46] with attention blocks is utilized to obtain the final restored image $\boldsymbol{I}'_n$. The 16-layer U-Net architecture contains symmetric skip connections, strided convolutions and transpose convolutions, which is actually an encoder-decoder structure. All the kernel size is 3×3, and the ReLU function is applied after each convolution layer except the last one. After adding with the input $\boldsymbol{I}_n$, the restored image $\boldsymbol{I}'_n$ can be obtained.

$$\boldsymbol{I}'_n = U\text{-}Net_{att}[\boldsymbol{I}_n, \hat{\boldsymbol{n}}] + \boldsymbol{I}_n, \tag{2}$$

where [.] represents the operation of concatenating.

Specifically, the attention blocks contain two parts, i.e., the channel attention block and spatial attention block. This process can be represented as

$$\begin{aligned} \boldsymbol{I}_m &= \boldsymbol{I}_{in} \otimes f_{ca}(\boldsymbol{I}_{in}), \\ \boldsymbol{I}_{out} &= \boldsymbol{I}_m \otimes f_{sa}(\boldsymbol{I}_m), \end{aligned} \tag{3}$$

where $\otimes$ represents element-wise multiplication and $\boldsymbol{I}_m$ represents the intermediate feature map. f_{ca} and f_{sa} are the channel attention block and spatial attention block, respectively. For the channel attention block, it utilizes the inter-channel relationship of the features. The input feature map is first sent to an average-pooling and a max-pooling layer to generate spatial context descriptors. After that, both descriptors are forwarded to a shared network to produce the channel attention map. The shared network is composed of a convolution layer, a ReLU function, and another convolution layer. Finally, the output feature vectors are merged by the element-wise summation operation.

$$\boldsymbol{I}_{out} = \sigma(f_N(AvgPool(\boldsymbol{I}_{in})) + f_N(MaxPool(\boldsymbol{I}_{in}))), \tag{4}$$

where $\sigma(.)$ denotes the sigmoid function and f_N represents the shared network. On the other hand, the spatial attention block exploits two pooling operations first, i.e., average-pooling and max-pooling. Then the obtained features are concatenated and convolved by a convolution layer. The process can be described as follows.

$$\boldsymbol{I}_{out} = \sigma(conv([AvgPool(\boldsymbol{I}_{in}); MaxPool(\boldsymbol{I}_{in})])), \tag{5}$$

where $conv(.)$ denotes the convolution layer.

Finally, the constraints of the loss function are described, Except for the original loss constraints in [39], i.e., the reconstruction loss L_{rec} and total variation loss L_{TV}, the perceptual loss L_{per} is added. The definition of these three constraints is as follows.

The reconstruction loss minimizes the L2 distance between the output restored image $\boldsymbol{I}'_n$ and the original image $\boldsymbol{I}$.

$$L_{rec} = \left\| \boldsymbol{I} - \boldsymbol{I}'_n \right\|_2, \tag{6}$$

where $\|.\|_2$ means the Euclidean norm.

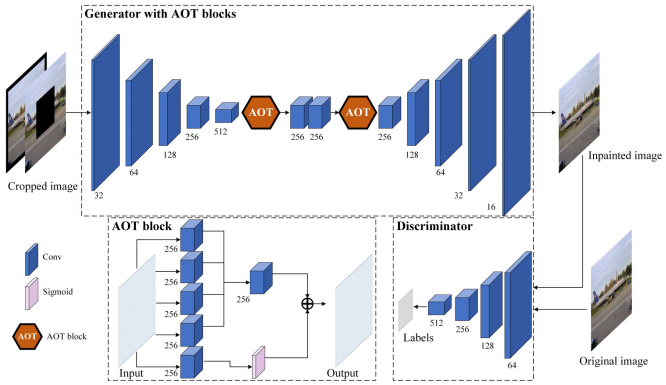


Fig. 6. The structure of the cropping restoration sub-module.

Furthermore, the total variation (TV) loss is introduced to constrain the smoothness of $\hat{\boldsymbol{n}}$ as follows.

$$L_{TV} = \left\|\nabla_h \hat{\boldsymbol{n}}\right\|_2^2 + \left\|\nabla_v \hat{\boldsymbol{n}}\right\|_2^2, \tag{7}$$

where ∇_h and ∇_v denote the gradient operator along the horizontal and vertical direction, respectively.

For the perceptual loss, it aims at minimizing the L1 distance between $\boldsymbol{I}_n'$ and $\boldsymbol{I}$, which is denoted as

$$P = \left\|\boldsymbol{I} - \boldsymbol{I}_n'\right\|_1, \tag{8}$$

where $\|.\|_1$ represents the L1 norm.

Finally, the overall optimization loss is concluded as

$$L_{n+f} = L_{rec} + \lambda_{TV} \times L_{TV} + \lambda_{per} \times P, \tag{9}$$

where λ_{TV} and λ_{per} represent hyperparameters. The effects of the attention blocks and the loss P are discussed in the ablation study part.

3) Cropping Restoration Sub-Module: In the proposed module, two kinds of cropping attacks are considered, i.e., edge cropping and center cropping. As it is difficult to obtain the specific parameters of cropping, a blind image restoration GAN, which means that the shape of the mask is unknown in advance, is designed to repair the cropped stego-images. The structure of the cropping restoration sub-module is shown in Fig. 6. It contains two important parts, i.e., the generator G and the discriminator D. For the generator, it utilizes the structure of encoder-decoder. Moreover, two aggregated contextual-transformation (AOT) blocks [47] are used to connect the encoder and decoder. Compared with the residual block, the stack of AOT blocks largely enriches the combinations of different pathways in the generator network, allowing the generator to capture as many as possible patterns of interest for context reasoning. As for the discriminator, it consists of several convolutional layers, each of which reduces the spatial size of feature maps by two. The discriminator takes as input an image from restored results or original images, and outputs a prediction label.

The loss constraint of this module is as follows. For the discriminator D, a least squares discriminator loss is used to evaluate the similarity between the restored image and the original image, which is denoted as

$$L_D = \frac{(D(\boldsymbol{I}) - b)^2 + (D(G(\boldsymbol{I}_c) - a))^2}{2}, \tag{10}$$

where $\boldsymbol{I}_c$ represents the cropped image and $a = 0$ and $b = 1$ are the targeted labels. On the other hand, the loss constraint of the generator contains four parts, i.e., the least squares generator loss, the mean squared error (MSE) loss, the mean absolute error (MAE) loss, and the structural similarity (SSIM) loss. First, the least squares generator loss, which corresponds to the loss of the discriminator, is formulated as

$$L_{lsg} = \frac{(D(G(\boldsymbol{I}_c)) - c)^2}{2}, \tag{11}$$

where $c = 1$ is the label. Then, two types of distance losses, i.e., MSE and MAE, are used to minimize the difference between the restored image $G(\boldsymbol{I}_c)$ and the original image $\boldsymbol{I}$

$$L_{MSE} = \|G(\boldsymbol{I}_c) - \boldsymbol{I}\|_2, \tag{12}$$

$$L_{MAE} = \|G(\boldsymbol{I}_c) - \boldsymbol{I}\|_1. \tag{13}$$

Last, the SSIM loss is used to evaluate the similarity as follows

$$L_{SSIM} = 1 - SSIM(G(\boldsymbol{I}_c), \boldsymbol{I}), \tag{14}$$

where $SSIM(.)$ means the calculation of structural similarity.

Finally, the overall constraint for the generator is denoted as

$$L_G = \lambda_{lsg} \times L_{lsg} + \lambda_{MSE} \times L_{MSE} + \lambda_{MAE} \times L_{MAE} + L_{SSIM}, \tag{15}$$

where λ_{lsg}, λ_{MSE}, and λ_{MAE} are the hyperparameters.

4) Rotation Restoration Sub-Module: Compared with cropping attacks, rotation not only leads to losing the content in the image, but also the shift of pixel position. A straightforward restoration idea is to predict the angle of the rotated image, and then rotate this image in the opposite direction. However, we find that it is very difficult to predict the rotation angle of the image with high accuracy. Once the forecast angle is wrong, it will cause more serious rotation errors in the image. Therefore, a rotation correction network is used in the rotation restoration sub-module to repair the rotated image directly. Recently, Nie et al. [48] proposed a deep rotation correction method. However, it can repair the rotated images well with a small rotation angle. For a big angle, the performance of this method is degraded and the training convergence is slow. In our module, the overall network structure follows the structure in Nie's method, but the constraints are improved. In Nie's method, the loss constraint is denoted as

$$L_{flow} = L_{content} + \omega \times L_{symmetry}, \tag{16}$$

where $L_{content}$ is a content term that uses the features extracted after the conv4-3 layer of VGG19 [45] as an effective perceptual representation. $L_{symmetry}$ is a symmetry-equivariant term to further facilitate the network the capability of horizon perception. ω is the hyperparameter.To further improve the performance of this restoration sub-module, another loss is added, which is used to minimize the distance between the original image $\boldsymbol{I}$ and the rotation correction image $\boldsymbol{I}_r'$

$$R = \left\|\boldsymbol{I} - \boldsymbol{I}_r'\right\|_2. \tag{17}$$

The final loss of the rotation restoration sub-module is formulated as

$$L_{rot} = L_{flow} + \lambda_{rot} \times R, \tag{18}$$

where λ_{rot} is the hyperparameter. The discussion of this loss is illustrated in the ablation experiments.

Algorithm 1 Translation Restoration Algorithm

Input: Translated image $\boldsymbol{I}_t \in R^{W\times H}$.
Output: Translation restored image $\boldsymbol{I}'_t$.

For $x = 1{:}H$ do
  Calculate $ave_{row} = \frac{1}{W}\sum_{y=1}^{W} \boldsymbol{I}(x, y)$
  If $ave_{row} < \epsilon$
    $T_y += 1$
End for
For $y = 1{:}W$ do
  Calculate $ave_{col} = \frac{1}{H}\sum_{x=1}^{H} \boldsymbol{I}(x, y)$
  If $ave_{col} < \epsilon$
    $T_x += 1$
End for
Translate the image as $\boldsymbol{I}'_t(x, y) = \boldsymbol{I}_t(x + T_x, y + T_y)$
Return the translation restored image $\boldsymbol{I}'$
End

5) Translation Restoration Sub-Module: For the translation attacked images, we notice that the position coordinates of the pixels that contain efficient information in these images are shifted. If we can restore the position coordinates of these pixels, the similarity between the attacked image and the original image can be significantly improved. Therefore, a simple but efficient image Translation Restoration Algorithm (TRA) based on pixel scanning is proposed in this sub-module. In this algorithm, the received attacked stego-image is scanned from top to bottom row-by-row, and the average pixel value of each row is calculated. If the calculated result is less than a given threshold, i.e., ϵ, this row is considered as the "translated row". Similarly, the image is scanned column by column from left to right to get the corresponding "translated columns". Then, the estimated result of the translated coordinates is (T_x, T_y). Finally, the received stego-image is translated backwards, that is, the image is translated with the coordinate $(-T_x, -T_y)$. Therefore, the position of the part containing efficient information in the image is repaired. The pseudo-code is shown in Algorithm 1.

In summary, based on the reasonable classification of the attack type and the specific design of the image restoration module, a universal framework to improve the robustness of existing coverless image steganography methods against various types of image attacks is finally achieved. In the next section, the robustness experiments are conducted on almost all the existing coverless image steganography methods to evaluate the effectiveness of the proposed universal framework.

IV. Experimental Results and Analysis

A. Configuration

All the experiments are implemented using the following configurations. The restoration module of the framework is trained on two NVIDIA RTX3090 GPUs with the Pytorch framework. Other experiments are implemented on a personal computer with an Intel (R) Core (TM) i5-10400F CPU @ 2.90GHz, 32GB memory. The attacks such as noise and filtering are implemented by the library function of OpenCV.

For the datasets, three widely used image datasets, including PASCAL VOC2012 [49], MS COCO2014 [50], and ImageNet [51] are adopted. The former two datasets are consistent with Meng et al.'s [32] method and the remaining dataset is consistent with Zou et al. et al.'s [36] method.

(1) The PASCAL VOC2012 dataset [49]. PASCAL VOC is a world-class computer visual challenge that provides standardized image datasets for object class recognition. The PASCAL VOC2012 dataset contains 11540 images.

(2) The MS COCO2014 dataset [50]. The MS COCO dataset is a large-scale object detection, segmentation, and captioning dataset. The 2014 version has 82783 training images.

(3) The ImageNet dataset [51]. The ImageNet dataset is a large visual database for research on visual object recognition software. It contains more than 15 million images.

In the experiments, the PASCAL VOC2012 and MS COCO2014 training datasets are utilized to train the restoration module. In the testing stage, 500 images from the testing dataset of PASCAL VOC2012, 500 images from the testing dataset of MS COCO2014 are randomly selected, which is consistent with Meng et al.'s [32] work. Besides, a total of 2588 images from the testing dataset of ImageNet are selected in files "n01440764" and "n01514668" for experiments, which is consistent with Zou et al.'s [36] work. It is worth noting that the images in the test set did not appear in the training set. All the images are resized to 256×256 in the comparison experiments.

Specifically, the hyperparameters of the restoration module are set as follows. $\lambda_{per} = \lambda_{TV} = 5 \times 10^{-2}$, $\lambda_{lsg} = 10^{-3}$, $\lambda_{MSE} = \lambda_{MAE} = 1$, $\omega = 10^{-1}$ and $\lambda_{rot} = 10^{4}$. In the training phase, the Adam optimizer is used to determine the optimal settings of noise, filter, cropping, and rotation restoration sub-modules.

In order to verify the effectiveness of the proposed universal framework, six typical coverless image steganography methods, including generation-based and mapping-based methods, are used to test our universal framework, i.e., Zhou et al.'s [18], Zhang et al.'s [28], Liu et al.'s [29], Karim et al.'s [31], Meng et al.'s [32], and Zou et al.'s [33] methods. In robustness experiments, the length of the hash sequence for Zhou et al.'s [18], Zhang et al.'s [28], Liu et al.'s [29], Karim et al.'s [31], Meng et al.'s [32], Zou et al.'s [33] are set to 8, which is consistent with their settings of the papers.

B. Robustness

As the originality of this paper is to improve the robustness of the existing coverless image steganography methods, we choose six typical generation-based methods, i.e., Zhou et al.'s [18], Zhang et al.'s [28], Karim et al.'s [31], Liu et al.'s [29], and Meng et al.'s [32] methods and mapping-based methods, i.e., Zou et al.'s [36] method to compare the robustness to image attacks before and after using the proposed universal framework.

In our experiments, the robustness of different kinds of coverless image steganography is evaluated by the extraction accuracy. Assume that each hash sequence generated by each image is hs_k, and the number of hash sequences generated by the entire dataset is K. The extracted hash sequence from the attacked image is denoted as hs'_k and the extraction accuracy Acc can be defined as

$$Acc = \frac{\sum_{k=1}^{K} f(k)}{K} \times 100\%,$$

TABLE III
THE CLASSIFICATION ACCURACY OF THE SUB-MODULE ON THREE DATASETS

	PASCAL VOC2012	MS COCO2014	ImageNet
Acc	99.2%	99.3%	99.0%

TABLE IV
THE SPECIFIC PARAMETERS FOR THE IMAGE ATTACKS

Attack	Parameters
Center cropping	Ratios: 5%, 10%, 20%
Edge cropping	Ratios: 5%, 10%, 20%
Rotation	Rotation angles: 10°, 30°, 50°
Translation	(36,20), (40,25), (80,50)
JPEG compression	The quality factor: Q(90)
Gauss noise	The mean μ:0, the variance σ:0.001
Salt and pepper (S&P) 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
Gauss filter	The filter size: 3×3
Scaling	The scaling ratio: 3.0
C-H-E	-
Gamma correction	Factor: 0.8

$$f(k) = \begin{cases} 1, & if\ hs_k = hs'_k \\ 0, & if\ hs_k \neq hs'_k. \end{cases} \tag{19}$$

To further verify the effectiveness of the proposed framework, three widely used image datasets with different content are considered, i.e., the PASCAL VOC2012 dataset, the MS COCO2014 dataset, and the ImageNet dataset. The classification accuracy on these three datasets is shown in Table III. From this table, we can find that the classification accuracy of the sub-module on three datasets can be achieved over 99.0%, which ensures the subsequent restoration sub-modules repair the attacked stego-images correctly.

1) The Performance of Resisting the Single Attack: In this subsection, the performance of resisting the single attack is discussed. The specific parameters of widely used image attacks are shown in Table IV. The comparison results of different kinds of coverless image steganography using the proposed framework on three datasets are shown in Table V to Table X, respectively. Each table is composed of 6 columns of data, and each of the two columns represents the extraction accuracy before and after using our framework for one coverless image steganography method, respectively. Moreover, the values in parentheses represent the increased percent after using our framework. The result in bold represents the better one. From these tables, it can be found that for the existing typical coverless image steganography methods, the overall robustness of using our framework is improved. For the VOC2012 dataset, the average increase is 28.9% for geometric attacks, and 1.0% for non-geometric attacks. In the COCO2014 dataset, the average accuracy increases 23.2% in resisting geometric attacks, and 0.3% in non-geometric attacks. Last is the ImageNet dataset, the average increase for geometric and non-geometric attacks is 23.9% and 1.2%, respectively. The specific results and analysis are as follows.

For the PASCAL VOC2012 dataset, the experimental results are shown in Table V and Table VI. It can be seen that the average accuracy of Zhou et al.'s [18], Zhang et al.'s [28], Karim et al.'s [31] and Liu et al.'s [29] is improved by over 20 percent. For Zhou et al.'s [18] and Karim et al.'s [31] methods, they both divide an image into nine blocks first. Then, they generate hash sequences by comparing the pixel intensity and calculating the average value of the LL subband, respectively. Therefore, geometric attacks such as cropping and translation will have a huge impact on the image blocks used to generate the hash sequence for each image, resulting in poor robustness. After using our framework, the content in each block has been restored and the extraction accuracy is greatly improved. As for Zhang et al.'s [28] and Liu et al.'s [29] methods, they divide the image into blocks twice, and each image can generate many hash sequences. The cropping operation would only affects partial sub-blocks. So, the robustness of resisting cropping attacks is higher. After using our framework, the improvement for the cropping attacks is not that much. However, the rotation and translation attacks would still cause big damage to Zhang et al.'s [28] and Liu et al.'s [29] methods. Owing to the great performance of the restoration modules, the robustness of resisting rotation and translation is improved efficiently, i.e., by more than 50 percent. In Meng's [32] method, an end-to-end hash generation model is utilized to generate hash sequences, which contains a specific design for improving the robustness of resisting geometric attacks. However, the robustness of this method against non-geometric attacks such as noise is not that good, and the result is significantly improved after using our framework. Especially for filter attacks, the increase is more than 10 percent. Zou et al.'s [18] method constructs a coverless image database (CID) for secret information hiding and extraction. As it extracts the secret information by comparing stego-images with the existing image database, it achieves strong robustness against most attacks except rotation. Our rotation restoration sub-module effectively compensates for the lack of resistance to this attack. As illustrated in Section III-*A*, the attacks of "others", including JPEG compression, scaling, C-H-E, and gamma correction, no more restoration sub-module is used to repair. Due to the limitation of the classification accuracy in the classification module, the extraction accuracy of resisting these four attacks is slightly reduced.

Table VII and Table VIII demonstrate the comparison results in the MS COCO2014 dataset. For Zhou et al.'s [18], Zhang's [28], Karim et al.'s [31] and Liu et al.'s [29] methods, the original average accuracy is slightly lower than the performance of PASCAL VOC2012 dataset. After using our framework, the average increase for extraction accuracy can still exceed 15.0%. Differently, the performance of Meng et al.'s [32] method in the MS COCO2014 dataset is better than that of the PASCAL VOC2012 dataset. However, our framework can still improve by more than 10 percent for rotation attacks mostly. For Zou et al.'s [34] method, similar to the results of the PASCAL VOC2012 dataset, it has poor robustness in resisting rotation attacks. Our framework can significantly improve the extraction accuracy to more than 93.0%.

For the ImageNet dataset, from Table IX and Table X, the performance of Zhou's [18] and Karim's [31] methods can also be improved by about 20 percent overall with our framework. For the noise and filter attacks, the original result of these two methods is high enough, i.e., most results can be achieved more than 98.0%. Therefore, it is difficult to

TABLE V

THE ROBUSTNESS COMPARISON OF ZHOU ET AL.'S [18], ZHANG ET AL.'S [28] AND KARIM ET AL.'S [31] METHODS USING OUR FRAMEWORK IN PASCAL VOC2012 DATASET

Attack	Parameter	Zhou's [18]	Zhou's [18] with our framework(Δ)	Zhang's [28]	Zhang's [28] with our framework(Δ)	Karim's [31]	Karim's [31] with our framework(Δ)
CenterCropping	5%	34.8%	**89.8%**(+55.0%)	83.8%	**88.9%**(+5.1%)	34.8%	**89.2%**(+54.4%)
	10%	26.8%	**78.2%**(+51.4%)	76.0%	**82.1%**(+6.1%)	26.8%	**76.0%**(+49.2%)
	20%	9.8%	**57.6%**(+47.8%)	73.9%	**75.5%**(+1.6%)	10.0%	**55.4%**(+45.4%)
EdgeCropping	5%	69.0%	**94.2%**(+25.2%)	66.9%	**84.9%**(+18.0%)	79.8%	**95.2%**(+15.4%)
	10%	49.2%	**89.8%**(+40.6%)	59.3%	**76.6%**(+17.3%)	58.8%	**91.0%**(+32.2%)
	20%	26.2%	**75.2%**(+49.0%)	54.1%	**62.1%**(+8.0%)	30.0%	**78.6%**(+48.6%)
Rotation	10°	23.4%	**90.0%**(+66.6%)	12.0%	**84.7%**(+72.7%)	24.2%	**89.6%**(+65.4%)
	30°	3.6%	**68.6%**(+65.0%)	4.3%	**73.9%**(+69.6%)	3.8%	**71.0%**(+67.2%)
	50°	1.8%	**57.2%**(+55.4%)	3.5%	**69.2%**(+65.7%)	2.2%	**59.8%**(+57.6%)
Translation	(36,20)	13.8%	**33.4%**(+19.6%)	7.3%	**64.6%**(+57.3%)	15.0%	**36.2%**(+21.2%)
	(40,25)	11.4%	**30.6%**(+19.2%)	5.1%	**66.4%**(+61.3%)	12.4%	**34.0%**(+21.6%)
	(80,50)	4.4%	**16.8%**(+12.4%)	3.9%	**56.9%**(+53.0%)	4.4%	**17.0%**(+12.6%)
JPEGcompression	Q(90)	100.0%	**100.0%**(+0.0%)	**97.7%**	97.6%(-0.1%)	100.0%	**100.0%**(+0.0%)
Gauss noise	σ(0.001)	99.2%	**99.4%**(+0.2%)	92.1%	**92.4%**(+0.3%)	98.4%	**99.2%**(+0.8%)
S&P noise	σ(0.001)	**99.6%**	99.6%(-0.0%)	94.6%	**97.0%**(+2.4%)	99.6%	**99.8%**(+0.2%)
Speckle noise	σ(0.01)	98.0%	**99.2%**(+1.2%)	90.0%	**90.9%**(+0.9%)	98.0%	**98.8%**(+0.8%)
Median filter	3×3	98.2%	**99.0%**(+0.8%)	93.2%	**94.7%**(+1.5%)	96.0%	**98.0%**(+2.0%)
Mean filter	3×3	**99.8%**	99.6%(-0.2%)	96.4%	**97.0%**(+0.6%)	**99.6%**	99.4%(-0.2%)
Gauss filter	3×3	100.0%	**100.0%**(+0.0%)	97.0%	**97.5%**(+0.5%)	99.6%	**99.8%**(+0.2%)
Scaling	3.0	**99.8%**	99.8%(-0.0%)	**99.1%**	99.0%(-0.1%)	100.0%	**100.0%**(+0.0%)
C-H-E	-	**69.6%**	69.6%(-0.0%)	**77.0%**	77.0%(-0.0%)	**68.8%**	68.8%(-0.0%)
Gamma correction	0.8	**90.6%**	90.6%(-0.0%)	**92.4%**	92.3%(-0.1%)	**91.0%**	91.0%(-0.0%)
Average	-	55.9%	**79.0%**(+23.1%)	62.7%	**82.8%**(+20.1%)	57.0%	**79.4%**(+22.4%)

TABLE VI

THE ROBUSTNESS COMPARISON OF LIU ET AL.'S [29], MENG ET AL.'S [32] AND ZOU ET AL.'S [36] METHODS USING OUR FRAMEWORK IN PASCAL VOC2012 DATASET

Attack	Parameter	Liu's [29]	Liu's [29] with our framework(Δ)	Meng's [32]	Meng's [32] with our framework(Δ)	Zou's [36]	Zou's [36] with our framework(Δ)
CenterCropping	5%	83.8%	**88.9%**(+5.1%)	70.4%	**83.0%**(+12.6%)	99.2%	**100.0%**(+0.8%)
	10%	76.0%	**82.0%**(+6.0%)	61.4%	**76.4%**(+15.0%)	97.7%	**100.0%**(+2.3%)
	20%	73.9%	**75.5%**(+1.6%)	56.4%	**67.4%**(+11.0%)	89.8%	**98.4%**(+8.6%)
EdgeCropping	5%	66.9%	**84.9%**(+18.0%)	74.8%	**90.6%**(+15.8%)	100.0%	**100.0%**(+0.0%)
	10%	59.3%	**76.6%**(+17.3%)	75.2%	**88.6%**(+13.4%)	100.0%	**100.0%**(+0.0%)
	20%	54.1%	**62.1%**(+8.0%)	75.8%	**82.4%**(+6.6%)	100.0%	**100.0%**(+0.0%)
Rotation	10°	12.0%	**84.7%**(+72.7%)	82.4%	**92.2%**(+9.8%)	99.2%	**100.0%**(+0.8%)
	30°	4.3%	**73.9%**(+69.6%)	65.2%	**90.4%**(+25.2%)	85.9%	**100.0%**(+14.1%)
	50°	3.5%	**69.2%**(+65.7%)	60.0%	**90.0%**(+30.0%)	68.8%	**100.0%**(+31.2%)
Translation	(36,20)	7.3%	**64.6%**(+57.3%)	83.0%	**86.8%**(+3.8%)	100.0%	**100.0%**(+0.0%)
	(40,25)	5.1%	**66.3%**(+61.2%)	81.2%	**87.0%**(+5.8%)	100.0%	**100.0%**(+0.0%)
	(80,50)	4.0%	**56.9%**(+52.9%)	77.4%	**82.0%**(+4.6%)	100.0%	**100.0%**(+0.0%)
JPEGcompression	Q(90)	**97.8%**	97.7%(-0.1%)	**96.8%**	96.8%(-0.0%)	100.0%	**100.0%**(+0.0%)
Gauss noise	σ(0.001)	92.1%	**92.4%**(+0.3%)	93.8%	**94.4%**(+0.6%)	100.0%	**100.0%**(+0.0%)
S&P noise	σ(0.001)	94.6%	**97.0%**(+2.4%)	93.6%	**99.3%**(+5.7%)	100.0%	**100.0%**(+0.0%)
Speckle noise	σ(0.01)	90.0%	**90.9%**(+0.9%)	91.2%	**93.4%**(+2.2%)	100.0%	**100.0%**(+0.0%)
Median filter	3×3	93.2%	**94.7%**(+1.5%)	85.8%	**93.4%**(+7.6%)	100.0%	**100.0%**(+0.0%)
Mean filter	3×3	96.4%	**97.0%**(+0.6%)	83.0%	**96.9%**(+13.9%)	100.0%	**100.0%**(+0.0%)
Gauss filter	3×3	97.0%	**97.5%**(+0.5%)	85.0%	**97.9%**(+12.9%)	100.0%	**100.0%**(+0.0%)
Scaling	3.0	**99.2%**	99.1%(-0.1%)	**99.0%**	98.9%(-0.1%)	100.0%	**100.0%**(+0.0%)
C-H-E	-	**77.0%**	77.0%(-0.0%)	**84.6%**	84.6%(-0.0%)	**99.2%**	99.2%(-0.0%)
Gamma correction	0.8	**92.4%**	92.3%(-0.1%)	**94.0%**	93.9%(-0.1%)	100.0%	**100.0%**(+0.0%)
Average	-	62.7%	**82.8%**(+20.1%)	80.5%	**89.4%**(+8.9%)	97.3%	**99.9%**(+2.6%)

improve these results significantly. Zhang et al.'s [28] and Liu et al.'s [29] methods are also vulnerable to most geometric attacks in this dataset. When our framework is used for these methods, the improvement is significant. For example, the extraction accuracy for Zhang et al.'s [28] method can be improved by more than 50 percent in translation attacks with all parameters. As for Meng et al.'s [32] method, the original result is slightly lower than the previous two datasets. After using our framework, the increase is larger than that of the other two datasets. The reason is that only two categories of the ImageNet dataset are used in the experiments, which causes the simple content of the dataset.

In short, after using our framework on six typical coverless image steganography methods, the robustness of resisting image attacks is improved efficiently. On the one hand, the robustness of most existing methods against non-geometric attacks performs well, i.e., most of them can achieve 90% extraction accuracy, and the result can be further improved after using the proposed framework. On the other hand, most of the existing methods have poor resistance to geometric

TABLE VII

THE ROBUSTNESS COMPARISON OF ZHOU ET AL.'S [18], ZHANG ET AL.'S [28] AND KARIM ET AL.'S [31] METHODS USING OUR FRAMEWORK IN COCO2014 DATASET

Attack	Parameter	Zhou's [18]	Zhou's [18] with our framework(Δ)	Zhang's [28]	Zhang's [28] with our framework(Δ)	Karim's [31]	Karim's [31] with our framework(Δ)
CenterCropping	5%	34.0%	**86.0%**(+52.0%)	81.9%	**86.8%**(+4.9%)	33.2%	**86.8%**(+53.6%)
	10%	27.8%	**75.6%**(+47.8%)	74.2%	**79.5%**(+5.3%)	27.6%	**75.8%**(+48.2%)
	20%	7.6%	**54.4%**(+46.8%)	72.0%	**73.3%**(+1.3%)	7.6%	**55.4%**(+47.8%)
EdgeCropping	5%	69.0%	**95.2%**(+26.2%)	64.8%	**83.9%**(+19.1%)	78.8%	**96.2%**(+17.4%)
	10%	44.2%	**88.8%**(+44.6%)	57.5%	**73.4%**(+15.9%)	55.0%	**90.4%**(+35.4%)
	20%	21.6%	**76.2%**(+54.6%)	52.5%	**60.6%**(+8.1%)	25.8%	**75.6%**(+49.8%)
Rotation	10°	20.0%	**69.4%**(+49.4%)	9.7%	**54.1%**(+44.4%)	21.0%	**71.2%**(+50.2%)
	30°	3.2%	**41.2%**(+38.0%)	4.3%	**33.4%**(+29.1%)	3.2%	**43.8%**(+40.6%)
	50°	1.6%	**34.8%**(+33.2%)	3.2%	**26.9%**(+23.7%)	1.6%	**34.0%**(+32.4%)
Translation	(36,20)	20.0%	**36.2%**(+16.2%)	7.1%	**64.3%**(+57.2%)	20.4%	**43.4%**(+23.0%)
	(40,25)	16.0%	**33.8%**(+17.8%)	6.5%	**66.7%**(+60.2%)	17.2%	**38.8%**(+21.6%)
	(80,50)	4.8%	**18.0%**(+13.2%)	3.4%	**60.8%**(+57.4%)	5.2%	**20.0%**(+14.8%)
JPEGcompression	Q(90)	**99.6%**	99.6%(-0.0%)	**97.4%**	97.3%(-0.1%)	100.0%	**100.0%**(+0.0%)
Gauss noise	σ(0.001)	99.0%	**99.4%**(+0.4%)	93.2%	**93.4%**(+0.2%)	99.2%	**100.0%**(+0.8%)
S&P noise	σ(0.001)	**99.6%**	99.6%(-0.0%)	95.1%	**97.0%**(+1.9%)	100.0%	**100.0%**(+0.0%)
Speckle noise	σ(0.01)	98.2%	**99.6%**(+1.4%)	90.5%	**91.3%**(+0.8%)	99.0%	**99.6%**(+0.6%)
Median filter	3×3	97.2%	**98.4%**(+1.2%)	92.7%	**94.2%**(+1.5%)	97.6%	**99.2%**(+1.6%)
Mean filter	3×3	99.4%	**99.6%**(+0.2%)	96.5%	**97.2%**(+0.7%)	**100.0%**	99.8%(-0.2%)
Gauss filter	3×3	**99.2%**	99.2%(-0.0%)	97.0%	**97.4%**(+0.4%)	99.8%	**100.0%**(+0.2%)
Scaling	3.0	**99.6%**	99.6%(-0.0%)	**98.3%**	98.2%(-0.1%)	100.0%	**100.0%**(+0.0%)
C-H-E	-	**66.4%**	66.4%(-0.0%)	**76.1%**	76.1%(-0.0%)	**67.0%**	67.0%(-0.0%)
Gamma correction	0.8	**90.8%**	90.8%(-0.0%)	**92.3%**	92.2%(-0.1%)	**90.0%**	90.0%(-0.0%)
Average	-	55.4%	**75.5%**(+20.1%)	62.1%	**77.2%**(+15.1%)	56.8%	**76.7%**(+19.9%)

TABLE VIII

THE ROBUSTNESS COMPARISON OF LIU ET AL.'S [29], MENG ET AL.'S [32] AND ZOU ET AL.'S [36] METHODS USING OUR FRAMEWORK IN COCO2014 DATASET

Attack	Parameter	Liu's [29]	Liu's [29] with our framework(Δ)	Meng's [32]	Meng's [32] with our framework(Δ)	Zou's [36]	Zou's [36] with our framework(Δ)
CenterCropping	5%	81.9%	**86.8%**(+4.9%)	93.2%	**96.6%**(+3.4%)	98.4%	**100.0%**(+1.6%)
	10%	74.2%	**79.5%**(+5.3%)	88.2%	**94.4%**(+6.2%)	97.7%	**99.2%**(+1.5%)
	20%	72.0%	**73.3%**(+1.3%)	82.0%	**90.8%**(+8.8%)	93.8%	**99.2%**(+5.4%)
EdgeCropping	5%	64.8%	**83.9%**(+19.1%)	87.0%	**98.6%**(+11.6%)	100.0%	**100.0%**(+0.0%)
	10%	57.5%	**73.4%**(+15.9%)	88.6%	**95.6%**(+7.0%)	100.0%	**100.0%**(+0.0%)
	20%	52.5%	**60.6%**(+8.1%)	87.8%	**95.4%**(+7.6%)	100.0%	**100.0%**(+0.0%)
Rotation	10°	9.7%	**54.1%**(+44.4%)	91.0%	**96.4%**(+5.4%)	99.2%	**100.0%**(+0.8%)
	30°	4.3%	**33.4%**(+29.1%)	76.2%	**93.6%**(+17.4%)	84.4%	**93.8%**(+9.4%)
	50°	3.2%	**26.8%**(+23.6%)	63.6%	**89.6%**(+26.0%)	61.7%	**93.8%**(+32.1%)
Translation	(36,20)	7.1%	**64.4%**(+57.3%)	**96.8%**	96.0%(-0.8%)	100.0%	**100.0%**(+0.0%)
	(40,25)	6.5%	**66.7%**(+60.2%)	**96.4%**	96.2%(-0.2%)	100.0%	**100.0%**(+0.0%)
	(80,50)	3.4%	**60.7%**(+57.3%)	92.4%	**93.6%**(+1.2%)	100.0%	**100.0%**(+0.0%)
JPEGcompression	Q(90)	**97.4%**	97.3%(-0.1%)	99.8%	**99.8%**(+0.0%)	100.0%	**100.0%**(+0.0%)
Gauss noise	σ(0.001)	93.2%	**93.5%**(+0.3%)	99.0%	**99.6%**(+0.6%)	100.0%	**100.0%**(+0.0%)
S&P noise	σ(0.001)	95.1%	**97.0%**(+1.9%)	99.0%	**99.6%**(+0.6%)	100.0%	**100.0%**(+0.0%)
Speckle noise	σ(0.01)	90.5%	**91.3%**(+0.8%)	98.8%	**98.8%**(+0.0%)	100.0%	**100.0%**(+0.0%)
Median filter	3×3	92.7%	**94.2%**(+1.5%)	98.4%	**99.4%**(+1.0%)	100.0%	**100.0%**(+0.0%)
Mean filter	3×3	96.6%	**97.2%**(+0.6%)	98.4%	**99.6%**(+1.2%)	100.0%	**100.0%**(+0.0%)
Gauss filter	3×3	97.0%	**97.3%**(+0.3%)	98.6%	**99.6%**(+1.0%)	100.0%	**100.0%**(+0.0%)
Scaling	3.0	**98.3%**	98.2%(-0.1%)	99.8%	**99.8%**(+0.0%)	100.0%	**100.0%**(+0.0%)
C-H-E	-	76.1%	**76.1%**(+0.0%)	95.8%	**95.8%**(+0.0%)	97.7%	**97.8%**(+0.1%)
Gamma correction	0.8	**92.2%**	92.1%(-0.1%)	99.0%	**99.0%**(+0.0%)	100.0%	**100.0%**(+0.0%)
Average	-	62.1%	**77.2%**(+15.1%)	92.3%	**96.7%**(+4.4%)	96.9%	**99.3%**(+2.4%)

attacks. After using the proposed framework, the performance is improved significantly, i.e., the maximum increase could be close to 70 percent, which verifies the effectiveness of the proposed framework.

2) The Performance of Resisting More Complex Attacks: In order to showcase the superiority of the proposed framework, we have added more noise settings including the attacks with a higher intensity and the mix of existing attacks. The specific attack types are described in Table XI. The performance of the proposed framework in resisting these attacks is shown in Table XII and XIII. From these two tables, we can find that the proposed framework can improve the robustness of existing coverless image steganography efficiently with a higher attack intensity. Besides, for the mixed attacks, i.e., mixing between non-geometric attacks and mixing between geometric and non-geometric attacks, the proposed framework can still improve the extraction accuracy of different coverless image steganography methods. Especially for Zhou et al.'s [18] and Karim et al.'s [31] methods, the accuracy is improved by almost 20 percent for different mixed attacks.

TABLE IX
THE ROBUSTNESS COMPARISON OF ZHOU ET AL.'S [18], ZHANG ET AL.'S [28] AND KARIM ET AL.'S [31] METHODS USING OUR FRAMEWORK IN IMAGENET DATASET

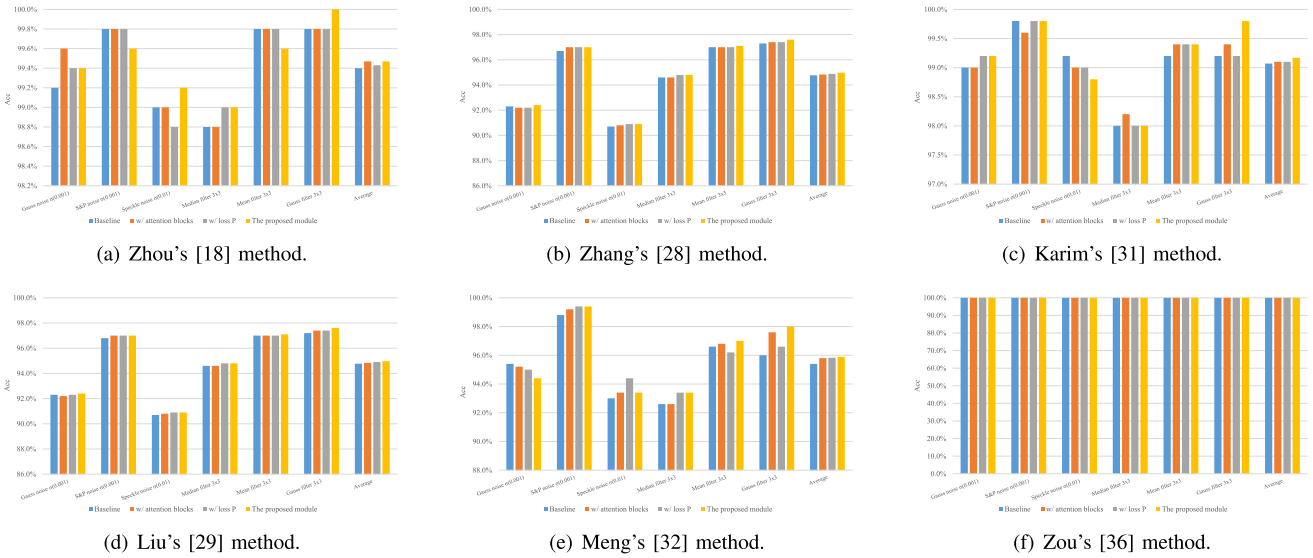
Attack	Parameter	Zhou's [18]	Zhou's [18] with our framework(Δ)	Zhang's [28]	Zhang's [28] with our framework(Δ)	Karim's [31]	Karim's [31] with our framework(Δ)
CenterCropping	5%	40.9%	**89.7%**(+48.8%)	82.5%	**87.5%**(+5.0%)	40.0%	**89.1%**(+49.1%)
	10%	33.7%	**81.4%**(+47.7%)	74.8%	**80.7%**(+5.9%)	33.1%	**80.6%**(+47.5%)
	20%	11.9%	**60.9%**(+49.0%)	72.5%	**73.9%**(+1.4%)	11.5%	**59.7%**(+48.2%)
EdgeCropping	5%	74.2%	**94.5%**(+20.3%)	63.1%	**83.6%**(+20.5%)	83.9%	**95.8%**(+11.9%)
	10%	51.8%	**89.7%**(+37.9%)	55.8%	**75.3%**(+19.5%)	60.6%	**91.7%**(+31.1%)
	20%	26.0%	**78.2%**(+52.2%)	50.9%	**59.1%**(+8.2%)	30.2%	**80.0%**(+49.8%)
Rotation	10°	25.5%	**76.3%**(+50.8%)	9.2%	**57.8%**(+48.6%)	27.2%	**78.0%**(+50.8%)
	30°	4.7%	**48.4%**(+43.7%)	3.4%	**38.7%**(+35.3%)	4.5%	**49.2%**(+44.7%)
	50°	2.2%	**38.5%**(+36.3%)	2.7%	**31.8%**(+29.1%)	2.2%	**39.3%**(+37.1%)
Translation	(36,20)	16.4%	**31.1%**(+14.7%)	4.8%	**55.2%**(+50.4%)	15.7%	**36.1%**(+20.4%)
	(40,25)	13.6%	**28.9%**(+15.3%)	4.1%	**62.0%**(+57.9%)	13.2%	**32.6%**(+19.4%)
	(80,50)	4.0%	**15.6%**(+11.6%)	2.7%	**53.2%**(+50.5%)	3.8%	**16.5%**(+12.7%)
JPEGcompression	Q(90)	**99.5%**	99.2%(-0.3%)	**97.6%**	97.1%(-0.5%)	**99.4%**	99.1%(-0.5%)
Gauss noise	σ(0.001)	98.9%	**99.4%**(+0.5%)	**92.5%**	92.5%(-0.0%)	98.9%	**99.2%**(+0.3%)
S&P noise	σ(0.001)	99.6%	**99.7%**(+0.1%)	94.7%	**96.3%**(+1.6%)	99.4%	**99.5%**(+0.1%)
Speckle noise	σ(0.01)	98.0%	**98.9%**(+0.9%)	91.0%	**91.5%**(+0.5%)	97.9%	**98.9%**(+1.0%)
Median filter	3×3	96.8%	**98.6%**(+1.8%)	91.8%	**93.3%**(+1.5%)	96.8%	**98.3%**(+1.5%)
Mean filter	3×3	99.3%	**99.5%**(+0.2%)	96.1%	**96.8%**(+0.7%)	99.1%	**99.3%**(+0.2%)
Gauss filter	3×3	99.4%	**99.5%**(+0.1%)	96.7%	**97.1%**(+0.4%)	**99.3%**	99.3%(-0.0%)
Scaling	3.0	**99.7%**	99.4%(-0.3%)	**98.3%**	97.8%(-0.5%)	**99.6%**	99.3%(-0.3%)
C-H-E	-	**72.6%**	72.5%(-0.1%)	**77.7%**	77.4%(-0.3%)	**72.2%**	72.1%(-0.1%)
Gamma correction	0.8	**91.4%**	91.2%(-0.2%)	**92.5%**	92.1%(-0.4%)	**90.3%**	90.1%(-0.2%)
Average	-	57.3%	**76.9%**(+19.6%)	61.6%	**76.9%**(+15.3%)	58.2%	**77.4%**(+19.2%)

TABLE X
THE ROBUSTNESS COMPARISON OF LIU ET AL.'S [29], MENG ET AL.'S [32] AND ZOU ET AL.'S [36] METHODS USING OUR FRAMEWORK IN IMAGENET DATASET

Attack	Parameter	Liu's [29]	Liu's [29] with our framework(Δ)	Meng's [32]	Meng's [32] with our framework(Δ)	Zou's [36]	Zou's [36] with our framework(Δ)
CenterCropping	5%	82.5%	**87.5%**(+5.0%)	47.0%	**68.2%**(+21.2%)	96.9%	**100.0%**(+3.1%)
	10%	74.8%	**80.7%**(+5.9%)	40.3%	**59.6%**(+19.3%)	93.8%	**96.9%**(+3.1%)
	20%	72.5%	**73.9%**(+1.4%)	30.1%	**50.4%**(+20.3%)	80.5%	**93.0%**(+12.5%)
EdgeCropping	5%	63.2%	**83.6%**(+20.4%)	62.8%	**81.8%**(+19.0%)	100.0%	**100.0%**(+0.0%)
	10%	55.8%	**75.3%**(+19.5%)	63.3%	**77.1%**(+13.8%)	100.0%	**100.0%**(+0.0%)
	20%	50.9%	**59.1%**(+8.2%)	63.4%	**71.3%**(+7.9%)	97.7%	**100.0%**(+2.3%)
Rotation	10°	9.2%	**57.8%**(+48.6%)	66.8%	**75.8%**(+9.0%)	99.2%	**100.0%**(+0.8%)
	30°	3.4%	**38.7%**(+35.3%)	57.6%	**68.1%**(+10.5%)	93.8%	**96.9%**(+3.1%)
	50°	2.7%	**31.8%**(+29.1%)	51.2%	**68.7%**(+17.5%)	74.2%	**91.4%**(+17.2%)
Translation	(36,20)	4.8%	**55.2%**(+50.4%)	68.0%	**75.1%**(+7.1%)	**99.2%**	99.2%(-0.0%)
	(40,25)	4.1%	**62.0%**(+57.9%)	66.7%	**74.5%**(+7.8%)	97.7%	**99.2%**(+1.5%)
	(80,50)	2.7%	**53.2%**(+50.5%)	59.8%	**66.5%**(+6.7%)	94.5%	**97.7%**(+3.2%)
JPEGcompression	Q(90)	**97.6%**	97.1%(-0.5%)	**92.1%**	91.9%(-0.2%)	100.0%	**100.0%**(+0.0%)
Gauss noise	σ(0.001)	**92.5%**	92.5%(-0.0%)	87.9%	**89.2%**(+1.3%)	100.0%	**100.0%**(+0.0%)
S&P pepper noise	σ(0.001)	94.7%	**96.2%**(+1.5%)	88.3%	**94.6%**(+6.3%)	100.0%	**100.0%**(+0.0%)
Speckle noise	σ(0.01)	91.0%	**91.5%**(+0.5%)	83.7%	**87.1%**(+3.4%)	100.0%	**100.0%**(+0.0%)
Median filter	3×3	91.8%	**93.3%**(+1.5%)	73.7%	**85.3%**(+11.6%)	100.0%	**100.0%**(+0.0%)
Mean filter	3×3	96.1%	**96.9%**(+0.8%)	70.0%	**90.0%**(+20.0%)	100.0%	**100.0%**(+0.0%)
Gauss filter	3×3	96.7%	**97.1%**(+0.4%)	73.7%	**92.0%**(+18.3%)	100.0%	**100.0%**(+0.0%)
Scaling	3.0	**98.3%**	97.8%(-0.5%)	**95.9%**	95.6%(-0.3%)	100.0%	**100.0%**(+0.0%)
C-H-E	-	**77.7%**	77.4%(-0.3%)	**72.5%**	72.3%(-0.2%)	**97.7%**	97.7%(-0.0%)
Gamma correction	0.8	**92.5%**	92.1%(-0.4%)	**85.1%**	84.9%(-0.2%)	100.0%	**100.0%**(+0.0%)
Average	-	61.6%	**76.9%**(+15.3%)	68.2%	**78.2%**(+10.0%)	96.6%	**98.7%**(+2.1%)

3) The Effect of Resolutions on the Proposed Framework: As the proposed framework resizes the attacked stego-images into the same size firstly, the effect of different image size is discussed. The experimental results are shown in Table XVI. To illustrate more intuitively, the average accuracy of resisting different attacks on six existing coverless image steganography methods is presented. It can be found that the performance is almost unchanged with the increase of resolution, which verifies the effectiveness of the proposed framework.

4) Ablation Study: In the proposed framework, the ablation study on the noise and filter restoration sub-module and the rotation restoration sub-module is discussed. For the noise and filter restoration sub-module, the operations of adding attention blocks and adding loss P are utilized to improve the restoration performance against noise and filter attacks. The specific results are shown in Fig. 7. The subfigures from Fig. 7(a) to Fig. 7(f) demonstrate the experimental results of resisting noise and filter attacks for six different coverless image steganography methods, respectively. "Baseline" in these subfigures represents the result of the proposed restoration sub-module without both attention blocks and loss P. It can be found that the average accuracy of the proposed module performs



(a) Zhou's [18] method. (b) Zhang's [28] method. (c) Karim's [31] method.

(d) Liu's [29] method. (e) Meng's [32] method. (f) Zou's [36] method.

Fig. 7. Ablation study of adding loss P and attention blocks for the noise and filter restoration sub-module.

TABLE XI
THE SPECIFIC PARAMETERS OF MORE ATTACK SETTINGS

Attack	Parameters
Gauss noise	The mean μ:0;the variance σ:0.01
S&P noise	The mean μ:0;the variance σ:0.01
Speckle noise	The mean μ:0;the variance σ:0.05
Gauss+S&P noise	The mean μ:0;the variance σ:0.001,0.001
Gauss+S&P+Speckle noise	The mean μ:0; the variance σ:0.001,0.001,0.01
Gauss+Median filter	The filter size: 3×3,3×3
Gauss+Median+Mean filter	The filter size: 3×3,3×3,3×3
S&P noise+Translation	The mean μ:0,the variance σ:0.001; Coordinate:(36,20)
S&P noise+Rotation	The mean μ:0, the variance σ:0.001; Rotation angle:10°
S&P noise+Center cropping	The mean μ:0,the variance σ:0.001; Ratio:5%
S&P noise+Edge cropping	The mean μ:0,the variance σ:0.001; Ratio:5%

best with these two operations simultaneously. The specific value of the average accuracy for different coverless image steganography methods can be achieved almost 95.0%. In most cases, the result of the proposed module is the best, and in others it is the second best. Especially for three kinds of filter attacks, i.e., Median filter, Mean filter, and Gauss filter, the improvement is more obvious compared with the baseline. For example, the extraction accuracy of Meng et al.'s [32] method for Gauss filter 3×3 increases 2.0% after using two operations. The second best performance of the proposed method focuses on repairing Zhou et al.'s [18], Karim et al.'s [31] and Meng et al.'s [32] methods. For Zhou et al.'s [18] method, Karim et al.'s [31] and Meng et al.'s [32] methods, they do not divide the image into several blocks. That is to say, the entire image is fed into a neural network or hash generation algorithm to generate hash sequences. When we add the attention mechanism during the noise and filter restoration, it is more difficult to capture the key information in the image compared with smaller blocks, e.g., Zhang's [28] and Liu's [29] methods. Therefore, the extraction accuracy slightly decreases after adding partial noise and filter, e.g., Gauss noise and median filter. Besides, adding these two operations also separately improves the performance generally. Among the six methods, the most significant improvement is for Meng et al.'s [32] method and both operations can improve the average accuracy by more than 0.4 percent. The reason for the improvement of adding attention blocks is that the attention mechanism can help the network focus on the key region of the image, which makes the restoration sub-module more accurately capture the difference between the original image and the attacked images. As for loss P, it enhances the constraint of minimizing the distance between the restored image and the original image.

As for the rotation restoration sub-module, the result of whether adding loss R is shown in Fig. 8. Similar to Fig. 7, six subfigures show the experimental results on six different coverless image steganography methods. '5w iters' and '10w iters' represent the number of training iterations is 5×10^4 and 10×10^4, respectively. It can be found that the results of '5w iters' for six coverless image steganography in resisting rotation attacks are below 40.0%. After adding the loss R, the proposed module can achieve a better restoration performance in 5w iterations, i.e., more than 40 percent improvement for all the rotation angles and different methods, which means that the convergence speed of the network is improved. In addition, the overall result at 10w iterations is better than the module without loss R, especially for a large rotation angle. For example, for Zhang et al.'s [28] method with rotation 50°, the improvement is 13.8%, which verifies the effectiveness of our operation. Therefore, the noise and filter restoration sub-module with attention blocks and loss P, and the rotation restoration sub-module with loss R are utilized in the comparison experiments.

C. Capacity

In coverless image steganography, the capacity is evaluated by the number of images for transmitting a fixed-length of the secret information L. Assume that the length of the hash sequences generated by different kinds of hash generation methods is ℓ, and the capacity can be defined as

$$N_h = \frac{L}{\ell}. \tag{20}$$

TABLE XII

THE PERFORMANCE OF THE PROPOSED FRAMEWORK IN RESISTING MORE COMPLEX ATTACKS, INCLUDING ATTACKS WITH A HIGHER INTENSITY AND THE MIX OF ATTACKS FOR ZHOU ET AL.'S [18], ZHANG ET AL.'S [28] AND KARIM ET AL.'S [31] METHODS IN PASCAL VOC2012 DATASET

Attack	Zhou's [18]	Zhou's [18] with our framework(Δ)	Zhang's [28]	Zhang's [28] with our framework(Δ)	Karim's [31]	Karim's [31] with our framework(Δ)
Gauss noise σ(0.01)	96.2%	**96.6%**(+0.4%)	82.0%	**82.8%**(+0.8%)	96.2%	**96.6%**(+0.4%)
S&P noise σ(0.01)	99.0%	**99.2%**(+0.2%)	87.5%	**93.4%**(+5.9%)	99.0%	**99.4%**(+0.4%)
Speckle noise σ(0.05)	94.4%	**96.0%**(+1.6%)	82.7%	**83.2%**(+0.5%)	94.8%	**96.0%**(+1.2%)
Gauss+S&P noise	99.2%	**99.4%**(+0.2%)	91.2%	**92.2%**(+1.0%)	98.2%	**98.8%**(+0.6%)
Gauss+S&P+Speckle noise	98.0%	**98.2%**(+0.2%)	87.2%	**87.7%**(+0.5%)	97.6%	**98.2%**(+0.6%)
Gauss+Median filter	99.2%	**99.8%**(+0.6%)	95.5%	**96.3%**(+0.8%)	98.4%	**99.2%**(+0.8%)
Gauss+Median+Mean filter	99.2%	**99.4%**(+0.2%)	94.3%	**94.9%**(+0.6%)	98.4%	**98.6%**(+0.2%)
S&P noise+Translation	13.8%	**33.4%**(+19.6%)	7.1%	**62.3%**(+55.2%)	15.0%	**36.4%**(+21.4%)
S&P noise+Rotation	23.4%	**73.6%**(+50.2%)	11.8%	**59.8%**(+48.0%)	24.2%	**74.4%**(+50.2%)
S&P noise+Center cropping	34.6%	**90.0%**(+55.4%)	81.1%	**84.2%**(+3.1%)	35.0%	**89.8%**(+54.8%)
S&P noise+Edge cropping	68.2%	**95.2%**(+27.0%)	65.8%	**82.9%**(+17.1%)	79.6%	**96.0%**(+16.4%)

TABLE XIII

THE PERFORMANCE OF THE PROPOSED FRAMEWORK IN RESISTING MORE COMPLEX ATTACKS, INCLUDING ATTACKS WITH A HIGHER INTENSITY AND THE MIX OF ATTACKS FOR LIU ET AL.'S [29], MENG ET AL.'S [32] AND ZOU ET AL.'S [36] METHODS IN PASCAL VOC2012 DATASET

Attack	Liu's [29]	Liu's [29] with our framework(Δ)	Meng's [32]	Meng's [32] with our framework(Δ)	Zou's [36]	Zou's [36] with our framework(Δ)
Gauss noise σ(0.01)	82.2%	**82.9%**(+0.7%)	82.2%	**86.0%**(+3.8%)	100.0%	**100.0%**(+0.0%)
S&P noise σ(0.01)	87.6%	**93.5%**(+5.9%)	88.4%	**95.0%**(+6.6%)	100.0%	**100.0%**(+0.0%)
Speckle noise σ(0.05)	82.7%	**83.2%**(+0.5%)	84.4%	**85.0%**(+0.6%)	100.0%	**100.0%**(+0.0%)
Gauss+S&P noise	91.2%	**92.2%**(+1.0%)	90.8%	**92.8%**(+2.0%)	100.0%	**100.0%**(+0.0%)
Gauss+S&P+Speckle noise	87.3%	**87.9%**(+0.6%)	88.8%	**89.2%**(+0.4%)	100.0%	**100.0%**(+0.0%)
Gauss+Median filter	95.7%	**96.5%**(+0.8%)	83.0%	**85.4%**(+2.4%)	100.0%	**100.0%**(+0.0%)
Gauss+Median+Mean filter	94.3%	**94.9%**(+0.6%)	80.0%	**82.8%**(+2.8%)	100.0%	**100.0%**(+0.0%)
S&P noise+Translation	7.1%	**62.3%**(+55.2%)	82.6%	**85.6%**(+3.0%)	100.0%	**100.0%**(+0.0%)
S&P noise+Rotation	11.8%	**59.8%**(+48.0%)	81.0%	**85.0%**(+4.0%)	99.2%	**100.0%**(+0.8%)
S&P noise+Center cropping	81.1%	**84.2%**(+3.1%)	71.4%	**81.4%**(+10.0%)	99.2%	**100.0%**(+0.8%)
S&P noise+Edge cropping	65.8%	**82.9%**(+17.1%)	85.2%	**89.2%**(+4.0%)	100.0%	**100.0%**(+0.0%)

TABLE XIV

THE PERFORMANCE OF THE PROPOSED FRAMEWORK IN RESISTING SOCIAL PLATFORM LOSSY FOR ZHOU ET AL.'S [18], ZHANG ET AL.'S [28] AND KARIM ET AL.'S [31] METHODS IN THE PASCAL VOC2012 DATASET

Attack	Zhou's [18]	Zhou's [18] with our framework(Δ)	Zhang's [28]	Zhang's [28] with our framework(Δ)	Karim's [31]	Karim's [31] with our framework(Δ)
Weibo	99.8%	99.8%(+0.0%)	99.1%	99.0%(-0.1%)	100.0%	100.0%(+0.0%)
WeChat	99.6%	99.6%(+0.0%)	93.0%	92.9%(-0.1%)	99.2%	99.2%(+0.0%)

TABLE XV

THE PERFORMANCE OF THE PROPOSED FRAMEWORK IN RESISTING SOCIAL PLATFORM LOSSY FOR LIU ET AL.'S [29], MENG ET AL.'S [32] AND ZOU ET AL.'S [36] METHODS IN THE PASCAL VOC2012 DATASET

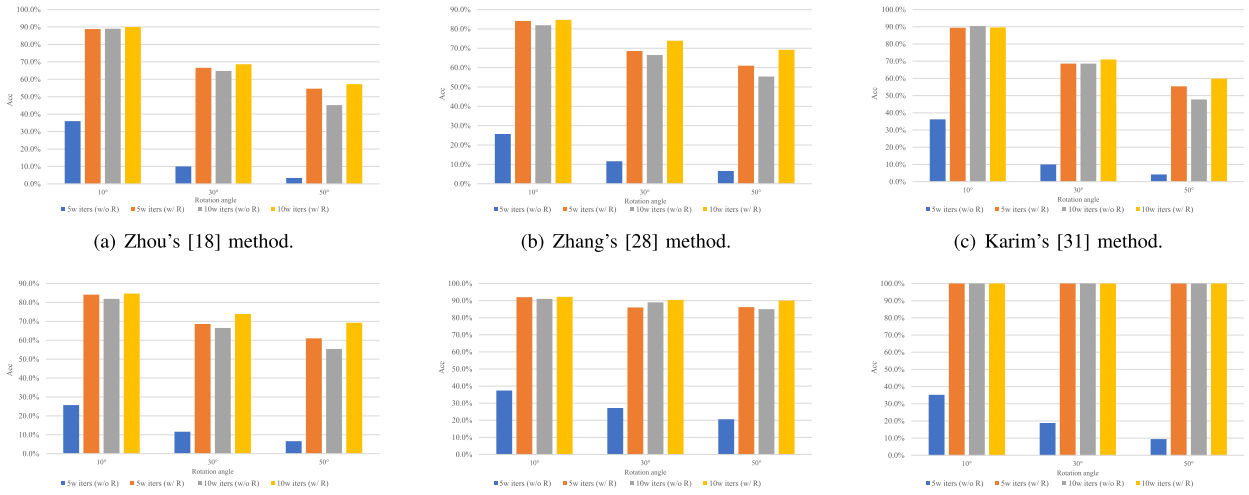
Attack	Liu's [29]	Liu's [29] with our framework(Δ)	Meng's [32]	Meng's [32] with our framework(Δ)	Zou's [36]	Zou's [36] with our framework(Δ)
Weibo	99.1%	99.0%(-0.1%)	99.0%	99.0%(+0.0%)	100.0%	100.0%(+0.0%)
WeChat	93.0%	92.9%(-0.1%)	94.0%	94.0%(+0.0%)	100.0%	100.0%(+0.0%)

TABLE XVI

THE AVERAGE ACCURACY OF THE PROPOSED FRAMEWORK WITH DIFFERENT RESOLUTIONS

Image size	Zhou's [18]	Zhang's [28]	Karim's [31]	Liu's [29]	Meng's [32]	Zou's [36]
128×128	79.2%	83.0%	79.6%	83.0%	89.5%	100.0%
256×256	79.0%	82.8%	79.4%	82.8%	89.4%	99.9%
512×512	78.9%	82.7%	79.2%	82.7%	89.2%	99.9%

From the above equation, it can be found that the capacity of different kinds of coverless image steganography is related to ℓ, which depends on the setup of the algorithm itself. For the proposed framework, there is no extra operations that affect the setting length of the hash sequence ℓ. Therefore, the capacity of different kinds of coverless image steganography using our framework remains unchanged. The specific capacity of the existing coverless



(a) Zhou's [18] method. (b) Zhang's [28] method. (c) Karim's [31] method.

(d) Liu's [29] method. (e) Meng's [32] method. (f) Zou's [36] method.

Fig. 8. Ablation study of adding loss R for the rotation restoration sub-module.

TABLE XVII

THE CAPACITY OF THE EXISTING COVERLESS IMAGE STEGANOGRAPHY WITH OUR FRAMEWORK

Algorithm	Length of hidden data				l
	1B	10B	100B	1KB	
Zhou's [18]	1	10	100	1024	8
Zhang's [28]	2~9	7~81	55~801	548~8193	1~15
Liu's [29]	2~9	7~81	55~801	548~8193	1~15
Karim's [31]	1	10	100	1024	8
Meng's [32]	2~9	7~81	55~801	548~8193	1~15
Zou's [36]	1	5	50	512	16

image steganography with our framework is shown in Table XVII.

D. Security

Compared with the modification image steganography, coverless image steganography does not modify the cover image and can resist steganalysis thoroughly. In the proposed framework, the restoration operation is used to repair the received stego-images before extracting the secret information at the receiving end, which is regarded as postprocessing. However, the steganalyzer usually accomplishes the detection process during transmission. Therefore, the ability of resisting steganalysis is not changed for all the existing coverless image steganography, i.e., all the steganalysis algorithms are expired. As for the transmission security, no more auxiliary information needs to be transmitted from the sender to the receiver. The transmission security of the existing coverless image steganography is not affected.

V. DISCUSSION

A. Discussion of Repairing Images With Attack Signs

In coverless image steganography, the original images are defined as the images obtained directly from the public dataset, e.g., in Zhou et al.'s [18], Zhang et al.'s [28], Liu et al.'s [29], Meng et al.'s [32], and Zou et al.'s [36] work. In our experiments, we randomly select images from three public datasets, i.e., the PASCAL VOC2012, MS COCO2014, and ImageNet. As the coverless image steganography needs to transmit a series of images, the sender will not actively tend to transmit a large number of images with attack signs, which will introduce suspicion to the monitor. However, it is inevitable that some images that show signs of being attacked will be transmitted. In fact, our datasets contain some images like being attacked, as shown in Fig. 9. We can see that Image 1 contains some noise, Image 2 contains black blocks and Image 3 is rotated. In our framework, the classification module is utilized to determine the attack type, and the experimental results show that our framework can classify these images correctly, i.e., they are all classified into "others". Moreover, the attacked version of these images and the restored images using our framework are shown in the middle and right columns of this figure, i.e., "Image 1 with speckle noise/Restored Image 1", "Image 2 with edge cropping/Restored Image 2", and "Image 3 with rotation/Restored Image 3", respectively. We can find that our framework does not over-correct these images, and the repaired results maintain a high similarity to the original images. Actually, the original images, which look like being attacked, often come from some natural scenes. For example, for Image 1 in Fig. 9, the noise-like snowflakes make the original images look like being attacked. For the black block in Image 2, it is actually a part of the natural scene, i.e., a part of the sunshade. As for Image 3, this rotation actually comes from the natural rotation of the camera when shooting. However, the attacks in the attacked stego-images often come from manual tampering, including adding speckle noise to Image 1, adding edge cropping to Image 2, and adding rotation to Image 3. It will inevitably leave tampering traces to images. In other words, the proposed framework can distinguish whether the image attack is added during transmission by learning learning the characteristics of tampering traces.

B. Discussion of Repairing "others" Attacks

In the proposed framework, JPEG compression, scaling, C-H-E and gamma correction are classified as "others". For these attacks, no more restoration sub-module is designed to repair, and the original stego-images are preserved. Due to the limitation of the classification module, the extraction accuracy of resisting these four attacks is slightly reduced, i.e., less than 0.1 percent. In fact, these four attacks are quite common in practical scenarios. Besides, there is still space to improve the robustness against these attacks for the

Fig. 9. The example of the original images, attacked images and restoration images.

existing coverless image steganography. Therefore, how to further improve the performance of resisting "others" attacks, i.e., JPEG compression, scaling, C-H-E and gamma correction is a subject worthy of further research.

C. Discussion of Repairing Social Platform Lossy

In practical application, social platforms such as Weibo, WeChat are common channels for image transmission. Therefore, we further test the performance of existing coverless image steganography methods in resisting social platform lossy. The experimental results are shown in Table XIV and XV. It can be found that the extraction accuracy of the existing coverless image steganography can reach over 92.0%. In fact, the transmission process of these social platforms uses JPEG compression with different parameters. As the proposed framework classifies the JPEG compression into "others", we do not use a specific restoration sub-module. Therefore, the performance of the proposed framework in repairing these attacks is similar to that of JPEG compression, i.e., with a slight reduction.

VI. Conclusion

In this paper, a universal framework for improving the robustness of coverless image steganography is proposed. The received attacked image is first sent to the classification module to obtain the attack type. Then, the corresponding restoration sub-module is used to repair the attacked stego-images to improve the extraction accuracy. It is worth noting that the restoration module adopts the blind image restoration, which means the process can be accomplished without parameters of attacks. As the restoration module is trained at the receiving end, no more auxiliary information needs to be transmitted by introducing the proposed framework to the existing coverless image steganography methods. Besides, the proposed framework can be used for almost all the existing coverless image steganography methods, which is the first universal framework in improving the robustness of coverless image steganography methods. The experimental results on three widely used image datasets show the efficiency. Further work will be focused on how to avoid compromising the extraction accuracy of the "others" attacks. Moreover, how to simplify the restoration module and develop a framework in coverless video steganography are also research directions in the future.

References

[1] L. Xiang, C. Ou, and D. Zeng, "Linguistic steganography: Hiding information in syntax space," *IEEE Signal Process. Lett.*, vol. 31, pp. 261–265, 2024.

[2] W. Peng, T. Wang, Z. Qian, S. Li, and X. Zhang, "Cross-modal text steganography against synonym substitution-based text attack," *IEEE Signal Process. Lett.*, vol. 30, pp. 299–303, 2023.

[3] W. Li, H. Wang, Y. Chen, S. M. Abdullahi, and J. Luo, "Constructing immunized stego-image for secure steganography via artificial immune system," *IEEE Trans. Multimedia*, vol. 25, pp. 8320–8333, 2023.

[4] K. Zeng, K. Chen, W. Zhang, Y. Wang, and N. Yu, "Robust steganography for high quality images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 9, pp. 4893–4906, May 2023.

[5] W. Tang, Z. Zhou, B. Li, K.-K.-R. Choo, and J. Huang, "Joint cost learning and payload allocation with image-wise attention for batch steganography," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 2826–2839, 2024.

[6] S. Li, Z. Wang, X. Zhang, and X. Zhang, "Robust image steganography against general downsampling operations with lossless secret recovery," *IEEE Trans. Dependable Secure Comput.*, vol. 21, no. 1, pp. 340–352, Jan./Feb. 2024.

[7] Z. Yang, K. Chen, K. Zeng, W. Zhang, and N. Yu, "Provably secure robust image steganography," *IEEE Trans. Multimedia*, vol. 26, pp. 5040–5053, 2024.

[8] J. Liu, Z. Li, X. Jiang, and Z. Zhang, "A high-performance CNN-applied HEVC steganography based on diamond-coded PU partition modes," *IEEE Trans. Multimedia*, vol. 24, pp. 2084–2097, 2022.

[9] Z. Li, X. Jiang, Y. Dong, L. Meng, and T. Sun, "An anti-steganalysis HEVC video steganography with high performance based on CNN and PU partition modes," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 1, pp. 606–619, Jan. 2023.

[10] Y. Dong, X. Jiang, Z. Li, T. Sun, and Z. Zhang, "Multi-channel HEVC steganography by minimizing IPM steganographic distortions," *IEEE Trans. Multimedia*, vol. 25, pp. 2698–2709, 2023.

[11] L. Meng, X. Jiang, T. Sun, Z. Zhao, and Q. Xu, "A robust coverless video steganography based on the similarity of inter-frames," *IEEE Trans. Multimedia*, vol. 26, pp. 5996–6011, 2024.

[12] S.-C. Liu and W.-H. Tsai, "Line-based cubism-like image—A new type of art image and its application to lossless data hiding," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1448–1458, Oct. 2012.

[13] L. Zhai, L. Wang, and Y. Ren, "Universal detection of video steganography in multiple domains based on the consistency of motion vectors," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1762–1777, 2020.

[14] W. Lu, R. Li, L. Zeng, J. Chen, J. Huang, and Y.-Q. Shi, "Binary image steganalysis based on histogram of structuring elements," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 9, pp. 3081–3094, Sep. 2020.

[15] W. You, H. Zhang, and X. Zhao, "A Siamese CNN for image steganalysis," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 291–306, 2021.

[16] K. Wei, W. Luo, S. Tan, and J. Huang, "Universal deep network for steganalysis of color image based on channel representation," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 3022–3036, 2022.

[17] Z. Zhang, H. Shi, X. Jiang, Z. Li, and J. Liu, "A CNN-based hevc video steganalysis against DCT/DST-based steganography," in *Proc. Int. Conf. Digit. Forensics Cyber Crime*, Singapore. Cham, Switzerland: Springer, Dec. 2022, pp. 265–276.

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

[19] D. Hu, L. Wang, W. Jiang, S. Zheng, and B. Li, "A novel image steganography method via deep convolutional generative adversarial networks," *IEEE Access*, vol. 6, pp. 38303–38314, 2018.

[20] J. Li et al., "A generative steganography method based on WGAN-GP," in *Proc. Int. Conf. Artif. Intell. Secur.*, Hohhot, China. Cham, Switzerland: Springer, 2020, pp. 386–397.

[21] X. Liu, Z. Ma, J. Ma, J. Zhang, G. Schaefer, and H. Fang, "Image disentanglement autoencoder for steganography without embedding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 2303–2312.

[22] X. Chen, Z. Zhang, A. Qiu, Z. Xia, and N. N. Xiong, "Novel coverless steganography method based on image selection and StarGAN," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 1, pp. 219–230, Jan. 2022.

[23] F. Peng, G. Chen, and M. Long, "A robust coverless steganography based on generative adversarial networks and gradient descent approximation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5817–5829, Sep. 2022.

[24] P. Wei, G. Luo, Q. Song, X. Zhang, Z. Qian, and S. Li, "Generative steganographic flow," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2022, pp. 1–6.

[25] Z. Zhou et al., "Secret-to-image reversible transformation for generative steganography," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 5, pp. 4118–4134, Sep./Oct. 2023.

[26] Z. Zhou et al., "Latent vector optimization-based generative image steganography for consumer electronic applications," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 4357–4366, Feb. 2024.

[27] S. Zheng, L. Wang, B. Ling, and D. Hu, "Coverless information hiding based on robust image hashing," in *Proc. Int. Conf. Intell. Comput.*, Liverpool, U.K. Cham, Switzerland: Springer, 2017, pp. 536–547.

[28] X. Zhang, F. Peng, and M. Long, "Robust coverless image steganography based on DCT and LDA topic classification," *IEEE Trans. Multimedia*, vol. 20, no. 12, pp. 3223–3238, Dec. 2018.

[29] Q. Liu, X. Xiang, J. Qin, Y. Tan, J. Tan, and Y. Luo, "Coverless steganography based on image retrieval of DenseNet features and DWT sequence mapping," *Knowl.-Based Syst.*, vol. 192, Mar. 2020, Art. no. 105375.

[30] Q. Liu, X. Xiang, J. Qin, Y. Tan, and Q. Zhang, "A robust coverless steganography scheme using camouflage image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 6, pp. 4038–4051, Jun. 2022.

[31] N. A. Karim, S. A. Ali, and M. J. Jawad, "A coverless image steganography based on robust image wavelet hashing," *TELKOMNIKA, Telecommun. Comput. Electron. Control*, vol. 20, no. 6, p. 1317, Dec. 2022.

[32] L. Meng, X. Jiang, Z. Zhang, Z. Li, and T. Sun, "A robust coverless image steganography based on an end-to-end hash generation model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 7, pp. 3542–3558, Jul. 2023.

[33] L. Zou, J. Sun, M. Gao, W. Wan, and B. B. Gupta, "A novel coverless information hiding method based on the average pixel value of the sub-images," *Multimedia Tools Appl.*, vol. 78, no. 7, pp. 7965–7980, Apr. 2019.

[34] Z. Zhou, Y. Cao, M. Wang, E. Fan, and Q. M. J. Wu, "Faster-RCNN based robust coverless information hiding system in cloud environment," *IEEE Access*, vol. 7, pp. 179891–179897, 2019.

[35] Y. Luo, J. Qin, X. Xiang, and Y. Tan, "Coverless image steganography based on multi-object recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 7, pp. 2779–2791, Jul. 2021.

[36] L. Zou, J. Li, W. Wan, Q. M. J. Wu, and J. Sun, "Robust coverless image steganography based on neglected coverless image dataset construction," *IEEE Trans. Multimedia*, vol. 25, pp. 5552–5564, 2023.

[37] G. Chen, F. Zhu, and P. A. Heng, "An efficient statistical method for image noise level estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 477–485.

[38] S. Nam, Y. Hwang, Y. Matsushita, and S. J. Kim, "A holistic approach to cross-channel image noise modeling and its application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1683–1691.

[39] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1712–1722.

[40] J. Xu, L. Zhang, D. Zhang, and X. Feng, "Multi-channel weighted nuclear norm minimization for real color image denoising," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1096–1104.

[41] N. Cai, Z. Su, Z. Lin, H. Wang, Z. Yang, and B. W.-K. Ling, "Blind inpainting using the fully convolutional neural network," *Vis. Comput.*, vol. 33, no. 2, pp. 249–261, Feb. 2017.

[42] S. Zhang, R. He, Z. Sun, and T. Tan, "DeMeshNet: Blind face inpainting for deep MeshFace verification," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 3, pp. 637–647, Mar. 2018.

[43] Y. Wang, Y.-C. Chen, X. Tao, and J. Jia, "VCNet: A robust approach to blind image inpainting," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 752–768.

[44] H. Zhao, Z. Gu, B. Zheng, and H. Zheng, "TransCNN-HAE: Transformer-CNN hybrid AutoEncoder for blind image inpainting," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 6813–6821.

[45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[46] 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. Cham, Switzerland: Springer, 2015, pp. 234–241.

[47] Y. Zeng, J. Fu, H. Chao, and B. Guo, "Aggregated contextual transformations for high-resolution image inpainting," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 7, pp. 3266–3280, Jul. 2023.

[48] L. Nie, C. Lin, K. Liao, S. Liu, and Y. Zhao, "Deep rotation correction without angle prior," *IEEE Trans. Image Process.*, vol. 32, pp. 2879–2888, 2023.

[49] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, Jan. 2015.

[50] T. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 740–755.

[51] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

Laijin Meng received the B.S. degree in communication engineering from Beijing Jiaotong University, Beijing, China, in 2020. He is currently pursuing the Ph.D. degree with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include steganography and steganalysis.

Fan Li is currently pursuing the master's degree with the School of Information and Communication Engineering, Hainan University, Haikou, China. His research interests include steganography.

Xinghao Jiang (Senior Member, IEEE) received the Ph.D. degree in electronic science and technology from Zhejiang University, Hangzhou, China, in 2003.

He is currently a Professor with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. He was a Visiting Scholar with New Jersey Institute of Technology, Newark, NJ, USA, from 2011 to 2012. His research interests include multimedia security, intelligent information processing, cyber information security, information hiding, and watermarking.

Qiang Xu (Member, IEEE) received the Ph.D. degree in cybersecurity from Shanghai Jiao Tong University, Shanghai, China, in 2021. After that, he was a Post-Doctoral Fellow with the City University of Hong Kong. He was also a Visiting Ph.D. Student with the Rapid Rich Object Search Laboratory, Nanyang Technological University, Singapore, from 2019 to 2020. He is currently an Assistant Professor with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University. His research interests include multimedia forensics and security.