Hindawi
Security and Communication Networks
Volume 2021, Article ID 3676720, 11 pages
https://doi.org/10.1155/2021/3676720

Research Article

SFRNet: Feature Extraction-Fusion Steganalysis Network Based on Squeeze-and-Excitation Block and RepVgg Block

Guiyong Xu, Yang Xu, Sicong Zhang, and Xiaoyao Xie

Key Laboratory of Information and Computing Science of Guizhou Province, Guizhou Normal University, Guiyang 550001, China

Correspondence should be addressed to Yang Xu; xy@gznu.edu.cn

Received 1 May 2021; Accepted 16 July 2021; Published 26 July 2021

Academic Editor: Zhaoqing Pan

In the era of big data, convolutional neural network (CNN) has been widely used in the field of image classification and has achieved excellent performance. More and more researchers are beginning to combine deep neural networks with steganalysis to improve performance in recent years. However, most of the steganalysis algorithm based on the convolutional neural network has only run test against the WOW and S-UNIWARD algorithms; meanwhile, their versatility is insufficient due to long training time and the limit of image size. This paper proposes a new network architecture, called SFRNet, to solve these problems. The feature extraction and fusion layer can extract more features from the digital image. The RepVgg block is used to accelerate the inference and increase memory utilization. The SE block improves the detection accuracy rate because it can learn feature weights to make effective feature maps with significant weights and invalid or ineffective feature maps with small weights. Experimental results show that the SFRNet has achieved excellent performance in the detection accuracy rate against four state-of-the-art steganography algorithms in the spatial domain, e.g., HUGO, WOW, S-UNIWARD, and MiPOD, under different payloads. The SFRNet detection accuracy rate achieves 89.6% against S-UNIWARD algorithm with the payload of 0.4bpp and 72.5% at 0.2bpp. As the same time, the training time of our network is greatly reduced by 35% compared with Yedroudj-Net.

1. Introduction

The rapid development of social networks provides convenience for users to exchange data. A large number of digital images are uploaded to the Internet every day. The proliferation of digital images provides a good medium for criminals to commit crimes using steganographic algorithms. Digital image steganography is a hiding technology that takes into account data security and communication, which uses the redundancy of the cover image to embed the secret information into the public carrier and transmits it through the public channel to ensure that the secret information is not discovered and intercepted by a third party. Image steganalysis is the opposite of image steganography, which can determine whether the image contains secret information by capturing minor disturbances in the stego image that are not easily perceivable by the human visual system. They provide a basis for extracting the secret information hidden in the image. In recent years, image steganalysis has played an increasingly important role in many information security systems and has attracted many researchers [1]. At the same time, the fast-developing adaptive steganography algorithm uses syndrome-trellis code (STC) [2] to minimize distortion and retains more complex image statistical properties. The current typical spatial adaptive steganography algorithms include HUGO [3], S-UNWIWARD [4], WOW [5], and MiPOD [6]. They make the secret information more cleverly hidden in the area where it is difficult to establish a steganalysis model, which improved the security of steganography algorithms and brought significant challenges to steganalysis.

Traditional steganalysis models include subtractive pixel adjacency matrix (SPAM) [7], spatial rich model (SRM) [8], max spatial rich model (maxSRM) [9], and its variant maxSRMd2, which are all feature extraction methods based on manual design. In recent years, convolutional neural network (CNN) has been widely used in image and video processing and has achieved excellent performance [10–15].

Steganalysis can be considered as a two-class problem of images. Since convolutional neural networks can extract features in the spatial and frequency domains of images, more and more researchers are beginning to combine deep neural networks with steganalysis to improve performance. The signal noise processed by steganalysis is a weak signal which will be affected by image content so that it will be ignored by the traditional classification network. The network needs to be specially modified before it can be used in steganalysis, such as suppressing the image content and enhancing the steganographic noise signal.

Qian et al. [16] proposed a steganalysis network Qian-Net, based on a convolutional neural network, using a Gaussian activation function to replace the rectifying linear unit (ReLU) [17] activation function. Xu et al. [18] proposed Xu-Net based on the Qian-Net framework. The high-pass filter is used to extract noise residuals in the preprocessing layer. Simultaneously, the network adds the absolute value (ABS) layer and TanH-ReLU hybrid activation function. Jian et al. [19] proposed Ye-Net with a deeper network structure, using high-pass filters as the preprocessing layer and the truncated linear unit (TLU) as the activation function, introducing the selection channel. Boroumand et al. [20] proposed a 48-layer deep learning steganalysis framework SRNet, which obtains filters through learning to improve the detection accuracy rate of the network against steganography algorithms. Yedroudj et al. [21] proposed a network architecture based on the concept of Alex-Net [22], called Yedroudj-Net. In addition to using ABS layer and TLU activation function, three fully connected layers are also used. Zhang et al. [23] proposed Zhu-Net, which optimizes the filter kernel of the preprocessing layer and uses pyramid pooling [24] to obtain excellent detection accuracy rate on the S-UNIWARD and WOW algorithms.

The problem of steganographic analysis tool with neural networks is that it is impossible to analyze larger-sized images due to limitations in computer resources. And, the versatility of such steganographic analysis tools is not good. Most of the steganalysis algorithm has only run test against the WOW and S-UNIWARD algorithms. At the same time, the training time of the neural network is too long. To enhance the practicality and universality of the steganalysis network framework, we propose a feature fusion steganalysis framework based on the network structure of the RepVgg [25] and squeeze-and-excitation [26] in this paper, which is called the SFRNet. Experimental results show that the SFRNet has achieved excellent performance in the detection accuracy rate of four different steganography algorithms under some different payloads. The SFRNet detection accuracy rate achieves 89.6% against S-UNIWARD algorithm with the payload of 0.4bpp and 72.5% at 0.2bpp. In summary, we make the following contributions in this paper:

(i) Instead of using the image as an input, we extract and merge the feature of images into a feature matrix through the rich model and use the generated feature matrix as the actual input of the work, which solves the dependence of the deep neural network on the size of the input image.

(ii) We propose SFRNet, a simple architecture with favorable speed-accuracy trade-off compared to the state of the arts, which uses the RepVgg block as the convolution layer of the network and uses the squeeze-and-excitation (SE) block to improve the detection accuracy rate.

(iii) We show the effectiveness of the SFRNet in steganalysis and the efficiency and ease of implementation.

The rest of the paper is organized as follows. Section 2 introduces the prior knowledge including SRM and its several variants and the deep learning methods. In Section 3, the SFRNet is proposed. This section describes feature extraction-fusion and the detailed structure of SFRNet. In Section 4, the dataset partition, training details, and specific parameters of the SFRNet steganalysis framework are introduced. In Section 5, we validate the effective proposed model on several states-of-the-art steganographic algorithms and compare the performance of the SFRNet with several advanced steganalysis algorithms. The study is ends with the conclusion in Section 6.

2. Preliminaries

2.1. The Feature Extraction Method. Friedrich and Kodovsky [8] proposed the spatial rich model (SRM) based on the subtractive pixel adjacency matrix (SPAM) model, which designed various linear and nonlinear high-pass filters (HPF) in spatial domains. It uses these filters to filter the image to obtain a wide variety of residual images and then separately counts the frequency of occurrence of each adjacent residual sample pattern to get the co-occurrence matrix. Finally, the elements of the co-occurrence matrix are arranged into vectors as steganographic analysis features, as shown in Figure 1. The steganographic analysis features can comprehensively perceive the change of image adjacent pixel correlation caused by steganography algorithm. The SRM improves the detection accuracy rate of steganalysis algorithm, which has been used and improved by researchers of general steganalysis.

Denemark et al. [9] proposed the steganalysis method maxSRM combined with the channel selection strategy, which is a variant of the so-called SRM. The maxSRM and maxSRMd2 are built in the same manner as the SRM, but the process of forming the co-occurrence matrices is modified to consider the embedding change probabilities estimated from the analyzed image. The version of the maxSRM with all co-occurrence scan directions replaced with the oblique direction "d2," as shown in Figure 2, is called maxSRMd2. Compared with SRM, maxSRM and maxSRMd2 have significant performance improvement.

2.2. The RepVgg Block. A classic convolutional neural network (ConvNet), VGG [27], achieved massive success in image recognition with a simple architecture composed of a stack of Conv, ReLU, and pooling. With Inception, ResNet [28], and DenseNet, many research interests were shifted to

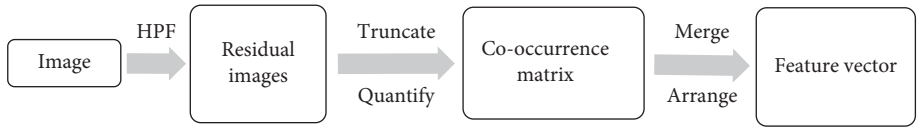

Figure 1: Feature extraction process.

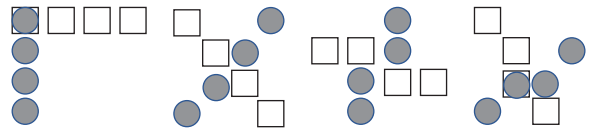

Figure 2: Four types of co-occurrence scan direction.

well-designed architectures, making the models more and more complicated. The complicated multibranch designs make the model difficult to implement, customize, slow down the inference, and reduce memory utilization. Ding et al. [25] presented RepVgg, a VGG-like inference-time body composed of nothing but a stack of 3×3 convolution and ReLU, while the training-time model has a multibranch topology.

In the SFRNet, we used the RepVgg block instead of the conventional convolution to accelerate the inference and increase memory utilization. The RepVgg block use ResNet-like identity and 1×1 branches so that the training-time information flow of a building block is $y = x + g(x) + f(x)$. It uses $W^{(3)} \in \mathbb{R}^{C_2 \times C_1 \times 3 \times 3}$ to denote the kernel of a 3×3 conv layer with C_1 input channels and C_2 output channels and $W^{(1)} \in \mathbb{R}^{C_2 \times C_2}$ for the kernel of 1×1 branch. It uses $\mu^{(3)}, \sigma^{(3)}, \gamma^{(3)}, \beta^{(3)}$ as the accumulated mean, standard deviation, learned scaling factor, and bias of the BN layer following 3×3 conv layer, $\mu^{(1)}, \sigma^{(1)}, \gamma^{(1)}, \beta^{(1)}$ for the BN layer following 1×1 conv layer, and $\mu^{(0)}, \sigma^{(0)}, \gamma^{(0)}, \beta^{(0)}$ for the identity branch. The identity branch can be viewed as a 1×1 conv layer with an identity matrix as the kernel. Let $M^{(1)} \in \mathbb{R}^{N \times C_1 \times H_1 \times W_1}$ and $M^{(2)} \in \mathbb{R}^{N \times C_2 \times H_2 \times W_2}$ be the input and output and $*$ be the convolution operator:

$$\begin{aligned} M^{(2)} = {} & bn\left(M^{(1)} * W^{(3)}, \mu^{(3)}, \sigma^{(3)}, \gamma^{(3)}, \beta^{(3)}\right) \\ & + bn\left(M^{(1)} * W^{(1)}, \mu^{(1)}, \sigma^{(1)}, \gamma^{(1)}, \beta^{(1)}\right) \\ & + bn\left(M^{(1)}, \mu^{(0)}, \sigma^{(0)}, \gamma^{(0)}, \beta^{(0)}\right). \end{aligned} \tag{1}$$

Then, it obtains the final bias by adding up the three bias vectors and the final 3×3 kernel by adding the 1×1 kernels onto the central point of 3×3 kernel, which can be easily implemented by first zero-padding the two 1×1 kernels to 3×3 and adding the three kernels up [25], as shown in Figure 3.

2.3. The Squeeze-and-Excitation Block. He et al. [26] focus on the channel relationship and propose the "Squeeze-and-Excitation" block, which can learn to use global information to emphasize informative features and suppress less useful ones selectively. Liu et al. [29] construct a new effective network with diverse filter modules (DFMs) and squeeze-and-excitation modules (SEMs), called DFSE-Net, which can better capture the embedding artifacts. The experiments presented that networks can pay more attention to critical channels by SEMs.

The squeeze-and-excitation block is not a complete network structure, which can construct a squeeze-and-excitation network by simply stacking a collection of SE blocks. The SE block can learn feature weights to make effective feature maps with significant weights and invalid or ineffective feature maps with small weights, as shown in Figure 4.

3. SFRNet

The proposed network architecture is called SFRNet: feature extraction-fusion steganalysis network via squeeze-and-excitation block and RepVgg block. Firstly, we explain the method of preprocessing, i.e., how to get the feature matrix. Then, the architecture of network is demonstrated. At the same time, we explored the values of key parameters through experiments.

3.1. Feature Extraction and Fusion Layer. The steganography algorithm modifies the original image content as little as possible when embedding secret information in the cover image to avoid detection. In other words, the steganography algorithm introduces noise in the image, which usually cannot be perceived by the human perceptual system. And, the noise is also easily ignored by those image classification networks which focus on the content of the image. At the same time, it modifies the correlation between adjacent pixels of the original image while also modifying the correlation between adjacent pixels of the residual image. The SRM and its variants are used to process the image, mainly to suppress the relevance of image content. We propose a feature information fusion block inspired by [30].

We use the following steps to extract and fuse the feature to obtain the feature matrix as input of the model. First, the residual image of the stego image and the cover image is

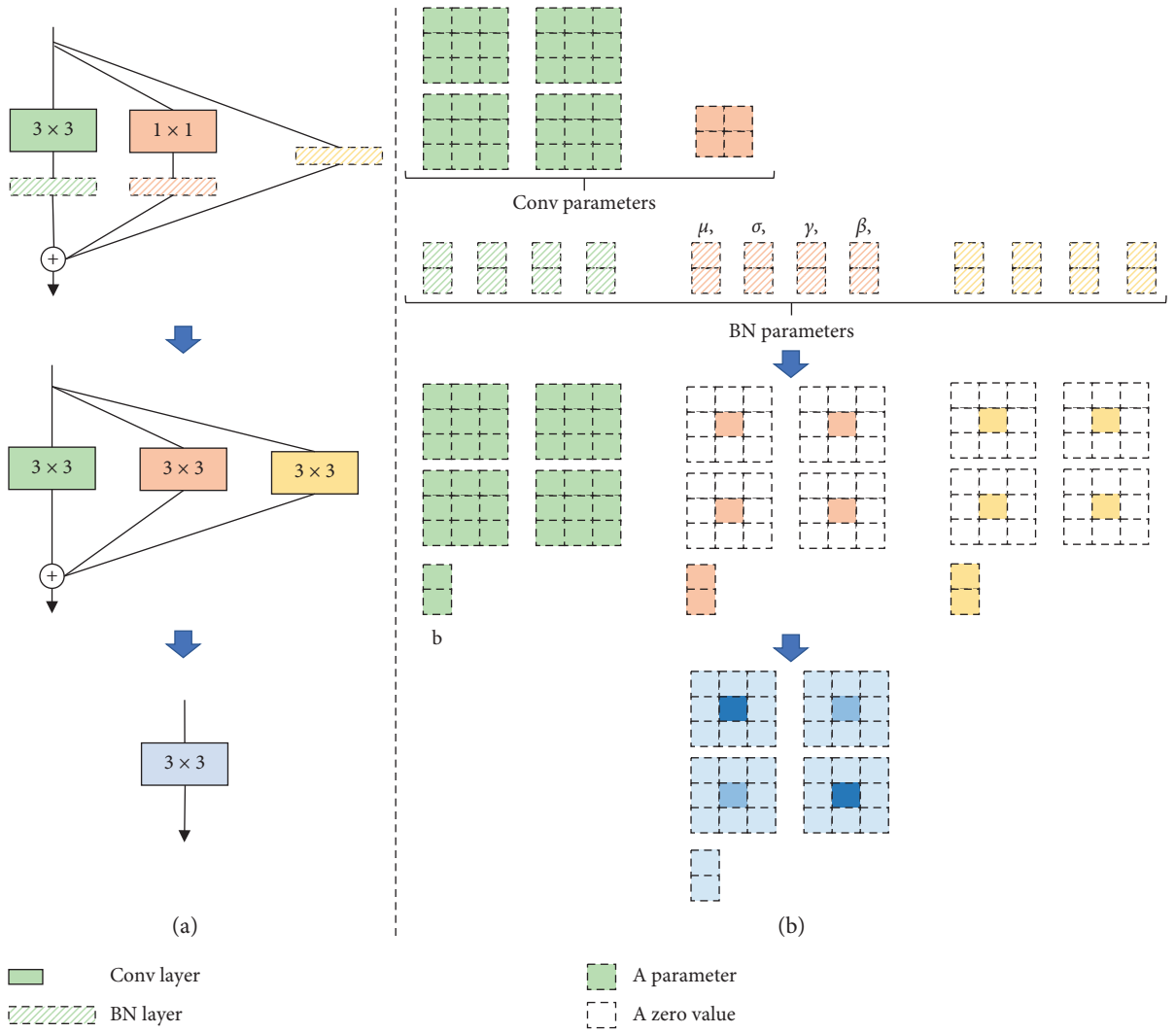

Figure 3: The RepVgg block.

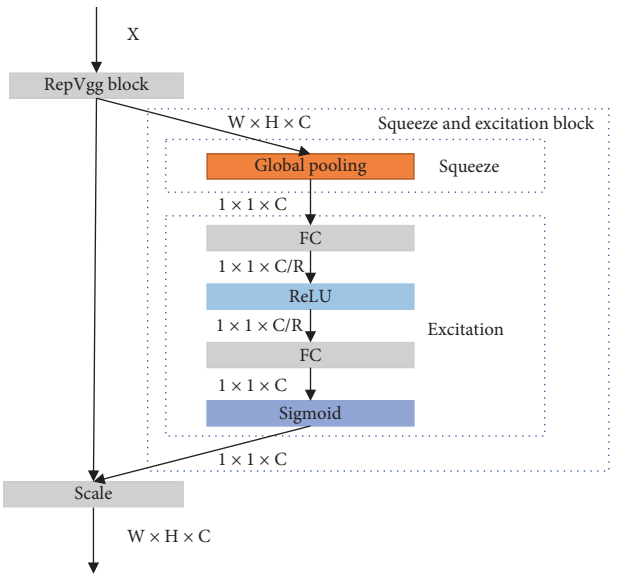

Figure 4: The squeeze-and-excitation block.

filtered by the high-pass filters to obtain submodels. Then quantize, round, and truncate each submodel and extract the co-occurrence matrix. Finally, the feature vectors are obtained by using the merging rules in SRM to process the co-occurrence matrix. The high-pass filters are shown in Figure 5.

The feature vector extraction process is defined as the follow equations:

$$R_{ij}^{k} = \widehat{X}_{ij}\left(N_{ij}\right) - cX_{ij} \Leftrightarrow R^{K} = X * K^{k}, \tag{2}$$

where X_{ij} represents the i, j pixel of the cover image, N_{ij} is the adjacent pixels of X_{ij}, $X_{ij} \neq N_{ij}$, $c \in \mathbb{N}$ is residual order, $\widehat{X}_{ij}(\cdot)$ is a predictor of cX_{ij} defined on N_{ij}, K^{k} is kth high-pass filters, and R^{K} is the residual filtered by the kth high-pass filter:

$$R_{ij}^{k} \leftarrow \mathrm{Trunc}_{T}\left(\mathrm{Round}\left(\frac{R_{ij}^{k}}{q}\right)\right), \tag{3}$$

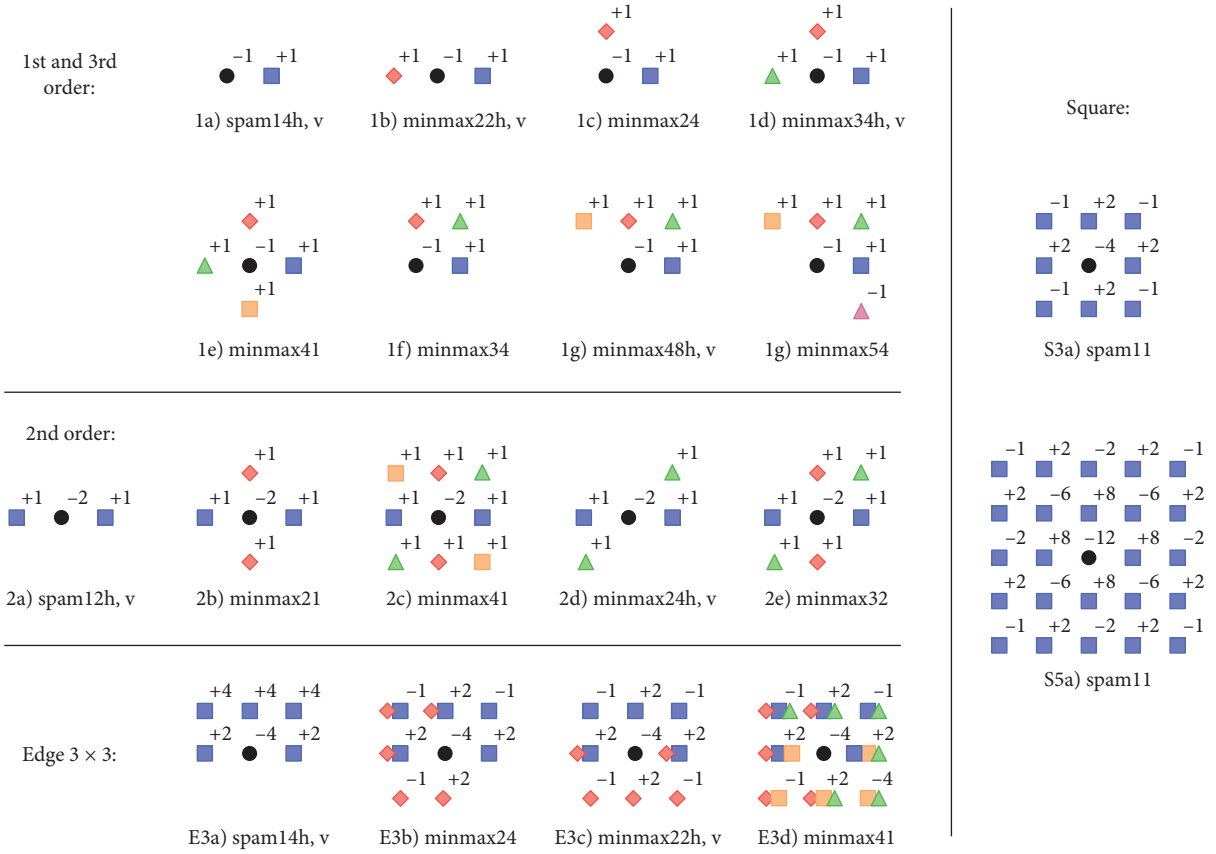

Figure 5: The high-pass filters.

where Round$(\cdot)$ means rounding up by element and Trunc$(\cdot)$ means a truncation operation by element. The purpose of truncation is to curb the dynamic range of residual to all description using co-occurrence matrices with a small T.

The SRM model extracts the co-occurrence matrix in the horizontal, defined as equation (4). The vertical co-occurrence, $C_d^{(v)}$, is defined analogically:

$$C_d^{(h)} = \sum_{i,j=1}^{n_1,n_2-3} \left[R_{i,j+n}^k = d_n, \forall n = 0,1,2,3\right]. \tag{4}$$

The maxSRM model extracts the co-occurrence matrix in the horizontal, defined by equation (5), where $\widehat{\beta}_{ij}^k$ is the embedding change probabilities in the kth high-pass filters; refer to [9], for details. The vertical co-occurrence, $C_d^{(v)}$, is defined analogically:

$$C_d^{(h)} = \sum_{i,j=1}^{n_1,n_2-3} \max_{n=0,1,2,3} \widehat{\beta}_{i,(j+n)}^k \left[\widehat{\beta}_{ij}^k = d_n, \forall n = 0,1,2,3\right]. \tag{5}$$

The scan direction of the maxSRMd2 model is different from SRM and maxSRM, which are replaced by "d2," as shown in Figure 2, so the co-occurrence matrix is defined as equations (6) and (7):

$$C_d^{(+)} = \sum_{i,j=1}^{n_{1-3},n_2} \overline{b}_{i,j} \times \left[R_{i,j}^k = d_0, R_{i,(j+1)}^k = d_1, R_{(i+1),(j+2)}^k = d_2, R_{(i+1),(j+3)}^k = d_3\right], \tag{6}$$

$$\begin{aligned} C_d^{(-)} &= \sum_{i,j=1}^{n_{1-3},n_2} \widetilde{b}_{i,j} \times \left[R_{i,j}^k = d_0, R_{i,(j+1)}^k = d_1, R_{(i+1),(j+2)}^k = d_2, R_{(i+1),(j+3)}^k = d_3\right], \\ \overline{b}_{i,j} &= \max\left\{\beta_{i,j}^k, \beta_{i,(j+1)}^k, \beta_{(i+1),(j+2)}^k, \beta_{(i+1),(j+3)}^k\right\}, \\ \widetilde{b}_{i,j} &= \max\left\{\beta_{(i-1),j}^k, \beta_{(i-1),(j+1)}^k, \beta_{i,j+2}^k, \beta_{i,(j+3)}^k\right\}. \end{aligned} \tag{7}$$

We choose $q \in [0.5, 1, 2], T = 2$, and $d = 4$ in all extraction methods to extract feature vector, getting 106 feature vectors. Among them, 17 are 338-dimensional feature vectors and 89 are 325-dimensional features vectors, and the feature vector is defined as

$$\vec{F}_k \leftarrow \text{Range}\left(\text{Merge}\left(C_d^{(h)}, C_d^{(v)}\right)\right), \tag{8}$$

where $\vec{F}_k$ is the feature vector calculated by using the kth submodel and Merge$(\cdot)$ merges two matrices into one by combining elements with the same or similar statistical laws in the horizontal co-occurrence matrices $C_d^{(h)}$ and vertical one $C_d^{(v)}$.We use the zero vector [0, 0] as the segmentation between each feature vector to fill it into a feature vector of 34,969 dimensions, and null values after the last feature in the vector are filled with the zero vector [0,0,0, . . ., 0,0]. It is defined as equation (9), where $*$ can denote SRM, maxSRM, and maxSRMd2:

$$F_* \leftarrow \left[\vec{F}_1, 0, 0, \vec{F}_2, 0, 0, \vec{F}_3, 0, 0 \ldots . \vec{F}_{106}, 0, 0, \ldots, 0, 0, 0\right]. \tag{9}$$

Then, we obtain the finally feature matrix fused by the three feature vectors, which are defined as

$$\text{MF}_{\text{cover}} = \begin{bmatrix} \text{Reshape}\left(F_{\text{SRM}}\right) \\ \text{Reshape}\left(F_{\max \text{SRM}}\right) \\ \text{Reshape}\left(F_{\max \text{SRM}d2}\right) \end{bmatrix}, \tag{10}$$

where MF_{cover} is the feature matrix of the cover image, MF_{stego} is the feature matrix of the stego image, and Reshape$(\cdot)$ converts a feature vector of 34,969 dimensions to a feature matrix of 187×187. Finally, our goal is to use SEFNet to train a mapping Map$(\cdot)$ based on the difference between them so that the mapping satisfies equations (11) and (12):

$$\text{Map}(.) \leftarrow \text{SEFNet}\left(\text{MF}_{\text{cover}}, \text{MF}_{\text{stego}}\right), \tag{11}$$

$$\begin{cases} \text{Map}\left(\text{MF}_{\text{stego}}\right) = 1, \\ \text{Map}\left(\text{MF}_{\text{cover}}\right) = 0. \end{cases} \tag{12}$$

3.2. The SFRNet Architecture. The overall structure of the SFRNet is presented in Figure 6. The SFRNet accepts an input image of size 256×256 and outputs two-class labels (stego and cover), composed of several number of layers, including one feature extraction-fusion block, five convolution blocks with different amounts of the RepVgg block, three SE blocks, and three fully connected layers.

The layer types and parameters are displayed inside boxes in Figure 6. N×(C×W×H) means that the number of batch size is N, the number of channels is C, and the height and width of the feature matrix is W and H. RepVgg denotes the RepVgg block. The details of the RepVgg block and SE block are described below. The full name of AVG is Average Pooling. Similarly, GAP is global average pooling.

3.2.1. The SE Blocks. Squeeze is achieved by using global average polling to generate channel-wise statistics. The statistic Z is generated by squeezing the input U through its spatial dimensions H×W, and the cth element of z is calculated by

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j). \tag{13}$$

Excitation is used to fully capture channel-wise dependencies. First, it must be capable of learning a nonlinear interaction between channels, and second, it must learn a nonmutually exclusive relationship. The operations of excitation can be defined by

$$\begin{aligned} s &= F_{ex}(z, W) \\ &= \sigma(g(z, W)) \\ &= \sigma\left(W_2 \delta\left(W_1 z\right)\right), \end{aligned} \tag{14}$$

where δ refers to the ReLU function, W_1 and W_2 are the fully connected operation, and σ refers to the sigmoid function. The final output of the SE block is obtained by rescaling U with the activations s:

$$\widetilde{x}_c = F_{\text{scale}}\left(\mathbf{u}_c, s_c\right) = \mathbf{u}_c s_c, \tag{15}$$

where $F_{\text{scale}}(\mathbf{u}_c, s_c)$ refers to channel-wise multiplication between the scalar s_c and the feature map $\mathbf{u}_c$ and $\widetilde{\mathbf{X}} = [\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_C]$ is the final output of the SE block.

In our architecture, the SE block is followed by the first three stages, as shown in Figure 6. To show the performance of the SE block against steganalysis algorithm, we conducted a comparative experiment based on the SFRNet with the SE block and without the SE block. The result in Figures 7 and 8 show that the SE block accelerates the convergence and shows better performance against WOW algorithm at 0.4 bpp.

3.2.2. Nonlinear Activation Layer. Two different activation functions, TLU and ReLU, are used in the SFRNet. The classical ReLU can prevent gradient vanishing/exploding and accelerate network convergence. The ReLU is used in "stage3," which selectively responds to embedded signals among the input feature map and get more efficient feature. Note that the remaining layers do not use the activation function.

Compared with cover image content, the signal introduced by the embedded message is usually of low amplitude. The high-frequency stego noise adds to the cover as a weak signal, significantly impacted by the image content. Therefore, the TLU is used to reduce the dynamic range of input feature maps in "stage1" and "stage2," suppressing image content and extract embedding signals more effectively. It can be defined as

$$\text{TLU}(x) = \begin{cases} T, & x > T, \\ x, & T \geq x \geq -T, \\ -T, & x < -T, \end{cases} \tag{16}$$

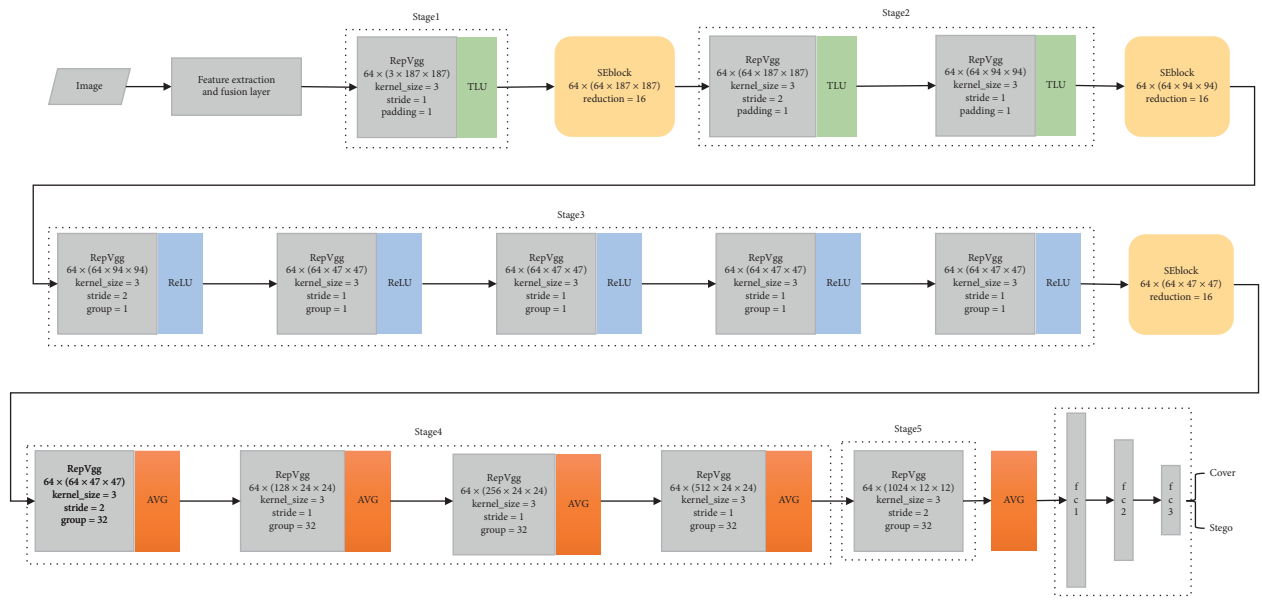

Figure 6: The SFRNet architecture.

where $T > 0$ is the threshold determined by experiments. To investigate the impact of parameter T in our network, we conduct several experiments with the SEFNet for a range of different T values. The results are shown in Table 1 and Figures 9 and 10. When the value of T is 1, the model achieves better performance and faster convergence.

4. Experiments

Python 3.8.3 was used for architecture construction, and the model was designed mainly with PyTorch 1.4.0. The operating system of the machine is Ubuntu 20.04 LTS, and the CUDA version is 11.0. The hardware of the machine has a GeForce RTX2080 SUPER with 8 GB and 250W, an Intel I7-9700k processor, and RAM with 32 GB (2 modules of 16 GB with 2666Mhz).

4.1. Dataset and the Steganographic Schemes. All experiments in this paper were evaluated and contrasted on the standard dataset BOSSBase ver. 1.01. This source contains 10,000 images acquired by seven digital cameras in the RAW format and subsequently processed by converting them to 8-bit grayscale, resizing, and central-cropping to 512×512 pixels. The image and camera information is shown in Table 2. The image source is widely used in research fields, such as information hiding, forensics, and steganalysis, which can be found at http://dde.binghamton.edu/download/.

Because other steganalysis algorithms use 256×256-size image as input, we decided to evaluate the effectiveness of all models on the images with a size of 256×256. To this end, we resized all the images into the size of 256×256 pixels using "imresize ()" in MATLAB with the default setting to generate the final datasets.

In our experiments, several state-of-the-art steganographic methods in the spatial domain, such as WOW, S-UNWARD, MiPOD, and HUGO, were employed to produce standard datasets. And, the embedding algorithms WOW and S-UNIWARD are implemented with STC simulator based on the publicly available codes, which can be found at the same URL of the BOSSBase original images. We use the MATLAB version rather than C++ implementation to avoid the problem as [31] that all images are embedded with the same key for all the steganographic algorithms. All methods were used to process the original images with two payloads: 0.2 bpp and 0.4 bpp. We use bit-per-pixel (bpp) to represent the size of secret data embedded into cover images in all experiments. For each steganography algorithm, we randomly select 5000 image pairs for training, 1000 image pairs for validating, and 4000 image pairs for testing, and the testing set was untouched during all of the training phases.

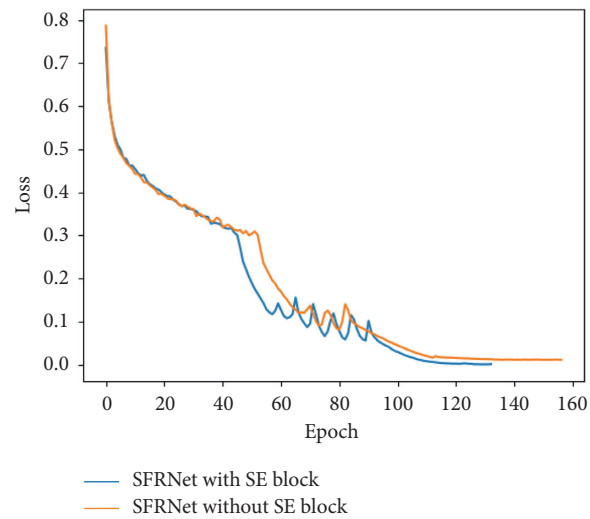

Figure 7: Comparing convergence performance of SFRNet with the SE block and SFRNet without the SE block against the WOW algorithm at 0.4 bpp.

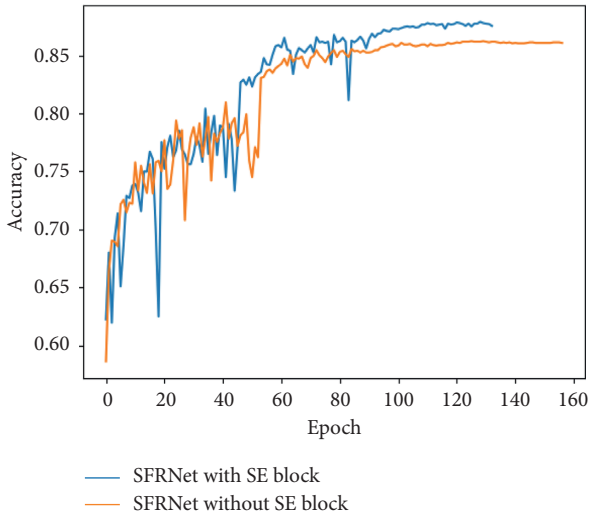

Figure 8: Comparing detection accuracy rate of SFRNet with the SE block and SFRNet without the SE block against the WOW algorithm at 0.4 bpp.

Table 1: The detection accuracy rate comparison of SFRNet against S-UNIWARD algorithm at 0.4bpp at different values of T.

T	1	2	3
Accuracy	0.8961	0.8869	0.8854

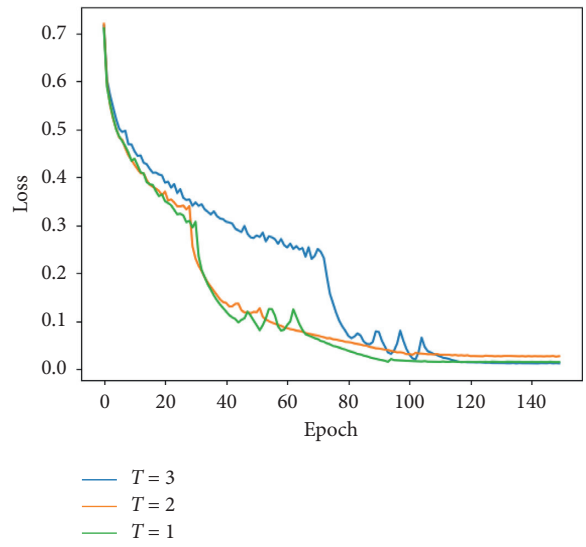

Figure 9: Comparing convergence performances of SFRNet with the different values of T against S-UNIWARD algorithm at 0.4 bpp.

Then, the feature extractor is employed to extract the feature for image steganalysis.

4.2. Hyperparameters. In SFRNet architecture, the Adam [32] optimizer is used to update the parameters of model in the learning phase since Adam can reach convergence faster than stochastic gradient descent (SGD). Due to GPU memory limitation, the minibatch size in training is set 64, containing

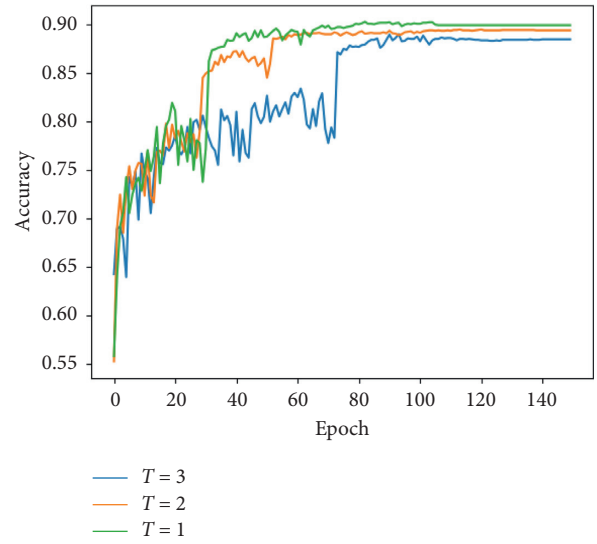

Figure 10: Comparing detection accuracy rate of SFRNet with the different values of T against S-UNIWARD algorithm at 0.4 bpp.

Table 2: Image and camera information.

Image number	Camera information
1-1354	Canon EOS 400D
1355-1415	Canon EOS 40D
1416-2769	Canon EOS 7D
2770-4811	Canon EOS DIGITAL REBEL XSi
4812-6209	PENTAX K2D
6210-7242	NIKON D70
7243-10000	M9 digital camera

32 cover images and their 32 corresponding stego images. The training dataset was shuffled after each epoch. Dropout is used, which followed every fully connected layer. Based on the above settings, the networks are then trained to minimize the cross-entropy loss. The SFRNet training is up to 150 epochs. We often stop training before 150 epochs to prevent overfitting. When the cross-entropy loss on the training set keeps decreasing, detection accuracy rate on validation begins declining, and we stop the training. The performance was evaluated by the testing accuracy rate, where the best validation model obtained during training was selected.

5. Experimental Results and Analysis

5.1. Comparison with the State-of-the-Art Steganalysis. We report the detection accuracy rate obtained when detecting S-UNIWARD and WOW embedding algorithms at 0.2 bpp and 0.4 bpp, as shown in Table 3. The steganalysis methods are Yedroudj-Net, SRNet, DFSE-Net, and Zhu-Net. The detection accuracy of the Zhu-Net is 0.3% higher than the SFRNet when applied to the WOW algorithm with 0.4 bpp. In addition to this case, the SFRNet generally has better performance than the other four steganographic analysis networks against WOW and S-UNIWARD algorithm at 0.2 bpp and 0.4 bpp, as shown in Table 3 and Figure 11.

Table 3: Performance comparisons between proposed network and several state-of-the-art models on S-UNIWARD and WOW at two different payloads.

Steganography	CNN model	0.2 bbp	0.4 bpp
S-UNIWARD	Yedroudj-Net	63.0	78.1
	SRNet	67.4	81.6
	Zhu-Net	71.5	84.7
	DFSE-Net	65.9	78.5
	SFRNet	**72.5**	**89.6**
WOW	Yedroudj-Net	72.3	83.1
	SRNet	75.4	86.9
	Zhu-Net	76.7	**88.2**
	DFSE-Net	75.3	85.1
	SFRNet	**76.8**	87.9

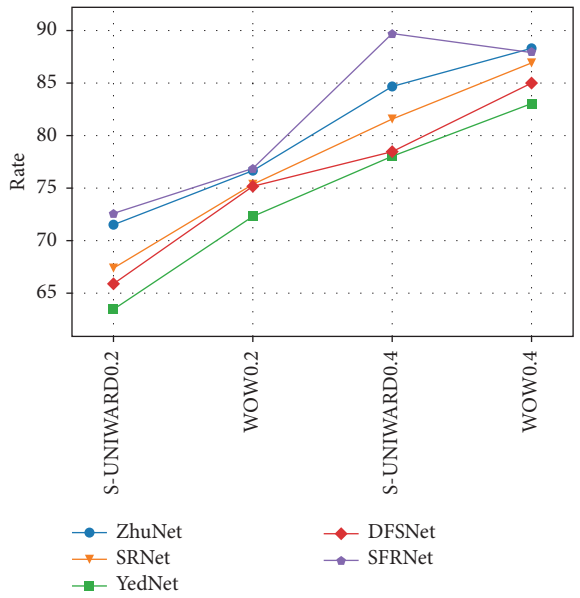

Figure 11: Comparison of the accuracy percentage of steganalysis algorithm among six steganalysis methods against two algorithms: S-UNIWARD and WOW at 0.2 bpp and 0.4 bpp.

The result shows that the proposed SFRNet outperforms Zhu-Net, SRNet against MiPOD, and HUGO embedding algorithms at 0.2 bpp and 0.4 bpp, as shown in Table 4 and Figure 12. The detection accuracy is increased by 8%–10% compared with the latest method Zhu-Net against MiPOD algorithm at 0.4 bpp and 0.2 bpp. Compared with Zhu-Net, the detection accuracy is increased by 4%–7% against HUGO algorithm at 0.2 bpp and 0.4 bpp. The good performance demonstrates the effectiveness of the network structure of the SFRNet in Figure 12.

5.2. The Time Consumption and Computational Complexity of SFRNet. We compare the number of parameters and times spent on network training and testing of the six types of steganalysis networks, as shown in Table 5. The SFRNet also reduces time consumption while improving accuracy. Although the SFRNet designed in this paper is a deeper network structure, the application of the RepVgg block

Table 4: Performance comparisons between proposed network and several state-of-the-art models on MiPOD and HUGO at two different payloads.

Steganography	CNN model	0.2 bbp	0.4 bpp
MiPOD	SRNet	64.3	75.1
	Zhu-Net	65.2	76.1
	SFRNet	**75.2**	**84.1**
HUGO	SRNet	67.1	78.7
	Zhu-Net	68.1	79.3
	SFRNet	**75.4**	**83.6**

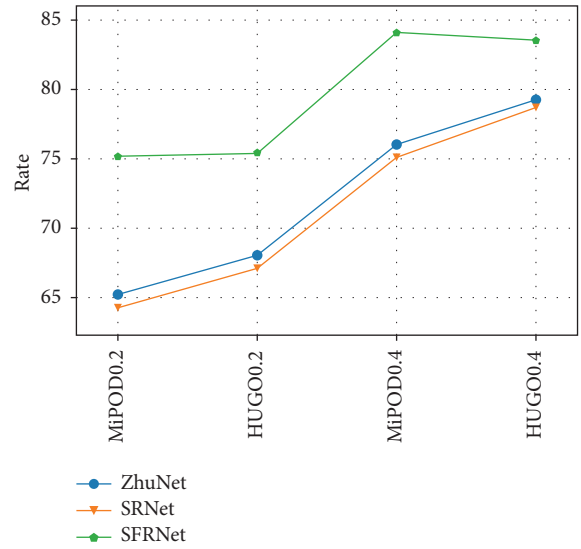

Figure 12: Comparison of the accuracy rate percentage of steganalysis algorithm among six steganalysis methods against two algorithms: MiPOD and HUGO at 0.2 bpp and 0.4 bpp.

Table 5: The computational complexity of the four networks.

Model	Parameter(10^4)	Train time(h)	Test time(s)
SFRNet	**150.5**	**2.91**	**9**
Zhu-net	287.1	8.65	33
SRNet	477.6	17.38	41
Yedroudj-Net	44.5	4.8	25

reduces the computational complexity and time consumption than the Zhu-net and SRNet while still ensuring considerable accuracy. Compared to Yedroudj-Net, training time is reduced by about 35%.

6. Conclusions

In this paper, a deep neural network with high accuracy and low time consumption is proposed for steganalysis. The feature extraction-fusion layers are used to extract features from original images and combine them into a feature matrix, which provides versatility for the steganographic analysis method. Furthermore, we use the SE block and the RepVgg block to construct the SFRNet, which significantly reduces the computational complexity while ensuring

accuracy. At the same time, the SE block is used to extract channel correlation of the feature matrix. The experimental result show that the SFRNet has excellent steganalysis performance in the spatial domain. Especially, compared with the latest algorithm under low payload, the detection accuracy has been improved by 10%. In the future, we would extend our methods to the frequency domain.

Data Availability

The steganography algorithm code and BOSSBase ver. 1.01. data used to support the findings of this study are available at http://dde.binghamton.edu/download/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Science Foundation of China, under Grant no. U1831131, the Special Funds of Central Government of China for Guiding Local Science and Technology Development, under Grant no. [2018]4008, and the Science and Technology Planned Project of Guizhou Province, China, under Grant no. [2020]2Y013.

References

[1] M. Chaumont, "Deep learning in steganography and steganalysis," *Digital Media Steganography*, Academic Press, Cambridge, MA, US, pp. 321–349, 2020.

[2] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 920–935, 2011.

[3] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Proceedings of the International Workshop on Information Hiding*, June 2010.

[4] V. Holub and J. Fridrich, "Digital image steganography using universal distortion," in *Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security*, Montpellier, France, June 2013.

[5] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS)*, December 2012.

[6] V. Sedighi, R Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, 2015.

[7] N. Jindal and B. Liu, *Review Spam Detection*, in *Proceedings of the 16th international conference on World Wide Web*, Alberta, Canada, 2007.

[8] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012.

[9] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," in *Proceedings of the 2014 IEEE International Workshop on Information Forensics and Security (WIFS)*, December 2014.

[10] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, June 2015.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[12] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, "Xnor-net: imagenet classification using binary convolutional neural networks," in *Proceedings of the European Conference on Computer Vision*, September 2016.

[13] K. Huang, X. Liu, S. Fu, D. Guo, and M. Xu, "A lightweight privacy-preserving CNN feature extraction framework for mobile sensing," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, 2019.

[14] Z. Pan, F. Yuan, J. Lei, W. Li, N. Ling, and S. Kwong, "MIEGAN: mobile image enhancement via A multi-module cascade neural network," *IEEE Transactions on Multimedia*, p. 1, 2021.

[15] Z. Pan, W. Yu, J. Lei, N. Ling, and S. Kwong, "TSAN: synthesized view quality enhancement via two-stream attention network for 3D-HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, p. 1, 2021.

[16] Y. Qian, J. Dong, W. Wang, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," in *Proceedings of SPIE–The International Society for Optical Engineering*, vol. 9409, San Francisco, CA, US, March 2015.

[17] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Madison, WI, US, June 2010.

[18] G. Xu, H.-Z. Wu, and Y.-Q. Shi, "Structural design of convolutional neural networks for steganalysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 708–712, 2016.

[19] Y. Jian, J. Ni, and Y. Yang, "Deep learning hierarchical representations for image steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2545–2557, 2017.

[20] M. Boroumand, M Chen, and J. Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 14, pp. 1181–1193, 2018.

[21] M. Yedroudj, F. Comby, and M. Chaumont, "Yedroudj-net: an efficient CNN for spatial steganalysis," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018.

[22] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," 2016, https://arxiv.org/abs/1602.07360.

[23] R. Zhang, F. Zhu, J. Liu, and G. Liu, "Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1138–1150, 2019.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[25] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: making VGG-style ConvNets great again," 2021, https://arxiv.org/abs/2101.03697.

[26] J. Hu, Li Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition, Salt Lake City, UT, USA, June 2018.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556, Article ID 1556.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 2016.

[29] F. Liu, X. Zhou, X. Yan, Y. Lu, and S. Wang, "Image steganalysis via diverse filters and squeeze-and-excitation convolutional neural network," *Mathematics*, vol. 9, no. 2, p. 189, 2021.

[30] Z. Pan, X. Yi, Y. Zhang, B. Jeon, and S. Kwong, "Efficient in-loop filtering based on enhanced deep convolutional neural networks for HEVC," *IEEE Transactions on Image Processing*, vol. 29, pp. 5352–5366, 2020.

[31] L. Pibre, J. Pasquet, D. Ienco, and M. Chaumont, "Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source-mismatch," *Electronic Imaging*, vol. 2016, no. 8, pp. 1–11, 2016.

[32] D. P. Kingma and B. Jimmy, "Adam: a method for stochastic optimization," 2014, https://arxiv.org/abs/1412.6980.