IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Multi-frequency Residual Convolutional Neural Network for Steganalysis of Color Images

JIALIN LIN, YU YANG

School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Yu Yang (e-mail: yangyu@bupt.edu.cn).

This work was supported in part by the National Key R&D Program of China under Grant 2017YFB0802803, Open Foundation of Guizhou Provincial Key Laboratory of Public Big Data under Grant 2018BDKFJJ014.

ABSTRACT Current steganalyzers based on deep learning mostly adopt wider or deeper designs to improve detection performance. However, an overly complex network increases the training cost and is not conducive to its expansion and optimization. Moreover, steganalysis pays more attention to high-frequency information corresponding to the image texture. However, the deeper the network, the more likely it is to learn low-frequency information corresponding to the image content, which is inconsistent with the goal of steganalysis. In response to these problems, a multi-frequency residual deep convolutional neural network for steganalysis of color images called MFRNet is proposed in this paper. We apply the idea of multi-frequency analysis to steganography detection for the first time, effectively controlling the network scale. By designing columns of different depths, it can learn different frequency components of steganographic noise at the same time. The detection performance is better than the existing networks that only learn a single frequency component of steganographic noise at the same depth. Therefore, it can achieve a good detection performance with a lighter architecture. In addition, by designing residual basic blocks with different residual shortcuts, different scales of steganographic noise residuals can be calculated at the same time, which can effectively suppress the interference of image content, so as to better reduce the impact of steganography algorithm mismatch and payload mismatch than the existing methods. The experimental results on PPG-LIRMM-COLOR showed that the proposed MFRNet outperformed the state-of-the-art model WISERNet.

INDEX TERMS Color image steganalysis, deep learning, convolutional neural network, multi-frequency residual analysis, lightweight.

I. INTRODUCTION

Steganography and steganography have always been a focus of research in the field of information hiding in recent years. Steganography is a process of hiding a secret message in digital carries such as images, videos, and texts, and tries not to change the visual and statistical characteristics of the carriers[1]. Since images are easy to obtain, the related steganography algorithm has the most studies. According to its embedding principle, it can be divided into steganography algorithm in the spatial domain and steganography algorithm in the frequency domain. The earliest steganography algorithm in the spatial domain is the Least Significant Bit (LSB) algorithm, which embeds secret information by modifying the least important bit in the image bit layer, and usually performs a ±1 operation on the least significant bit [2]. However, it is easy to be detected because of its large changes to the statistical characteristics of the image. At present, the safest steganography algorithm recognized is the adaptive steganography algorithm based on the distortion minimization framework and "STC" coding. By limiting the embedding changes to the complex texture area of the image, it can improve the security of the steganography algorithm [3], [4]. The current typical adaptive steganographic algorithms in the spatial domain include S-UNIWARD [5], WOW [6], HILL [7], and HUGO [8], etc.

As an effective detection method for steganography, steganalysis and steganography have been developing in confrontation. Traditional steganalysis algorithms are mainly divided into two steps: steganographic feature extraction and classification. For example, the classic Spatial Rich Model (SRM) [9] and its variants [10], [11]. They usually use hand-designed high-pass filters to effectively extract high-dimensional steganographic features. However, manual feature extraction cannot get rid of the dependence on expert experience. With the development of deep learning technology, its ability to process complex data has brought new possibilities for steganalysis[32][33]. The steganalysis algorithm based on deep learning can automatically learn steganographic features through network training, and simultaneously perform feature extraction and classifier optimization. It can avoid the dependence of traditional manual extraction of features on experience and become a new development trend of steganalysis [12]. The deep learning steganalysis algorithms that have been proposed with good performance include YeNet [13], YedroudjNet [14], XuNet [15], SRNet[16], WangNet[43], etc.

Unfortunately, the existing deep learning steganalysis algorithms are mostly designed for grayscale images. There is only a little research on steganalysis algorithms for color images. In addition, most of the existing steganalysis algorithms for color images are designed to detect LSB steganography, and do not consider the embedding characteristics of the adaptive steganography algorithm.WISERNet [17] is currently recognized as the best-performing color image steganalysis algorithm, which was proposed in 2019. The channel-wise convolution method proposed in that paper which is used to suppress the image content has very good detection performance when dealing with color image steganalysis.

Meanwhile, we found that the existing steganalysis network based on deep learning generally has the following problems: First, in order to improve the detection performance, the networks are usually designed too deep or too wide. For example, SRNet [16] uses up to 22 layers of convolutional layer, while WISERNet [17] outputs up to 1152 feature maps. This makes the network complexity, and excessive training costs are not conducive to network expansion and optimization. Moreover, steganalysis pays more attention to the high-frequency texture information of the image, which is the position where the secret information embedding probability is higher in the adaptive steganography algorithm. According to the frequency principle [20], the deeper the network, the more likely it is to learn low-frequency image content information, which is inconsistent with the goal of steganalysis. Therefore, a too deep network limits the improvement of steganographic detection performance. Second, the actual steganography detection task is faced with the problems of steganography algorithms mismatching and payload mismatching, and the existing steganalysis methods have not well resolved them.

In order to solve the above problems, we introduce the idea of multi-frequency residual analysis into steganalysis, and a residual steganalysis convolutional neural network called MFRNet based on multi-frequency residual analysis suitable for color images is proposed. The main contributions of the proposed method are:

1. For the first time as far as we know, the idea of multi-frequency residual analysis is applied to steganalysis， and a more lightweight network is proposed. Through the idea of multi-frequency residual analysis, the network scale is effectively controlled. Our network learns simultaneously through columns of different depths, which can synthesize the learning results of different frequencies of steganographic noise components. Therefore, our network can achieve slightly better detection performance with a more lightweight architecture than the existing best networks that learn single-frequency steganographic noise components.
2. Aiming at the payload mismatch problem in actual steganography detection tasks, we use multi-scale analysis technology to extract different scales of steganographic noise residuals by designing basic blocks with different residual shortcuts. Further suppressing the interference of image content, and comprehensively learning the steganographic noise residual components of different frequencies and different scales, and improving the detection performance on images with small payload. This effectively improves the detection ability of the network model in the case of small payloads. The accuracy is greatly improved compared with the existing model when the steganography algorithm mismatch and the payload mismatch.

The remaining sections have the following order: Section II introduces the existing related work and the overall architecture and key part design of the proposed MFRNet. Section III introduces the experimental design, including experimental data sets, parameter settings, the learning effect of the proposed model, and comparative experimental results. The network proposed in this paper is mainly compared with the state-of-the-art model WISERNet。 Section IV gives conclusions and future research directions.

II. THE PROPOSED NETWORK

A. RELATED WORK

The earliest deep learning model used for image steganalysis in the spatial was QianNet [42] in 2015. It uses a high-pass filter to reduce image content and enhances steganographic noise, and uses a Gaussian activation function for feature extraction. Its detection performance is better than SRM based on manual feature extraction [9]. Since then, steganalysis based on deep learning has received widespread attention.

In 2016, Xu et al. [15] proposed a new deep learning network XuNet for steganalysis. It uses an absolute value layer (ABS) after the second convolutional layer, which effectively improves the performance of model detection. Batch Normalization(BN) and TanH activation function are used in feature extraction, while BN and ReLU activation function are used in other layers, which improves the performance by nearly 4% compared with QianNet.

YeNet [13] proposed in 2017 has two points of innovation. One is to use the SRM filter to initialize the residual co-occurrence matrix instead of the traditional high-pass filter to extract the steganographic noise. The other is to introduce the knowledge of the selected channel into the feature extraction. The performance is improved by embedding probability maps to weight the image texture characteristics. The adaptive steganalysis models designed afterward mostly borrowed the idea of SRM filter initialization and introduction of channel selection technology in YeNet.

After that, YedroudjNet [14] proposed in 2019 combines the advantages of XuNet and YeNet, and once again improves the accuracy of steganographic detection compared to YeNet.

SRNet[16] proposed in the same year is a good performance network suitable for the steganalysis of gray-scale images. It uses residual shortcuts to strengthen the extraction of steganographic noise, so as to avoid the shortage of traditional steganalyzers that the design of steganographic noise residual extraction is too dependent on expert experience and hand-designed elements. It can be used for steganography detection in the space and frequency domain. But the depth of SRNet is too deep, that its convolutional layer reaches 22 layers. In fact, Xu [20] pointed out that according to the frequency principle, the deeper the network is more inclined to learn low-frequency signal components, corresponding to the image content part, which is not conducive to learning high-frequency information components. While steganalysis pays more attention to the high-frequency signal components that correspond to the image texture. Therefore, the excessively deep network structure of SRNet limits its ability to learn steganographic noise to a certain extent. In addition, an overly complex network increases the training cost, which is not conducive to the expansion, and optimization of the network.

Sign et al. [19] applied the idea of the self-similar from FractalNet[18] to steganalysis, designed and implemented a steganalyzer SFNet in 2020. SFNet studies the proportional relationship between network depth and width. It points out that a deep network designed with a fixed ratio of width and depth can get better detection results than a simple deeper network. However, the design of the basic module of SFNet is relatively simple, including only two types of basic blocks: CBR(Convolution, BN, ReLU) and CABR(Compared with CBR, there is one more ABS layer). And the whole network also lacks the design of residual extraction, so it can not reduce image content very well, which is not conducive to the detection of steganographic noise. Meanwhile, SFNet is difficult to converge when the width and depth are small.

In the same year, Wang et al. [43] proposed a CNN steganalysis model, WangNet, which uses a joint domain detection mechanism and a nonlinear detection mechanism. Its innovation lies in the ability to capture steganographic features in the spatial domain and the transform domain, thereby simultaneously detecting the spatial domain and the transform domain. And it improves the generalization performance of the network. Meanwhile, the detection performance of images with low steganography is improved. When the S-UNIWARD and WOW adaptive steganography algorithms of 0.2 bpp are detected, the detection accuracy is higher than that of YedroudjNet.

Different from steganalysis for gray-scale images, steganalysis for color images are more complex. Color images can be divided into three channels, R, G, and B (Red, Green, and Blue). Therefore, steganalysis for color images need to consider the three channels at the same time. However, gray-scale images have only one color channel. Thus, the conventional steganalyzers for gray-scale images are not suitable for steganalysis of color images.

There are lots of researches on deep learning steganalysis for grayscale images. In 2018, Aljarf et al. [34] proposed an image steganalysis system that combines the features of color gradient co-occurrence matrix (CGCM) and the number of histogram features. It extracts a variety of CGCM and histogram features by using the histogram of the difference image, and can well detect color images in multiple formats (BMP, JPG) under the LSB algorithm. In the same year, Rasool et al. [35] proposed a steganalysis model that uses an enhanced grayscale statistical feature set to detect uncompressed RGB color images. This model shows better performance when detecting BMP format color images under the LSB steganography algorithm. In 2019, Renad et al. [38] proposed a model that can well detect LSB steganography algorithms with small payloads. It uses statistical texture features and machine learning methods to detect the presence of hidden data in the RGB color image benchmark dataset. However, the above models have a common problem, that they are not designed for adaptive steganalysis[42].

WISERNet [17] proposed in 2019 is currently recognized as the best-performing color image steganalysis algorithm. It theoretically proves that the summation in ordinary convolution can be seen as a linear collusion attack, which actually reduces the signal-to-noise ratio and is not conducive to the extraction of steganographic noise. Therefore, in the underlying convolutional layer, a channel-wise convolution method is used to suppress the image content. In the higher-level convolutional layer, ordinary convolution is used to effectively improve the potential perception of steganographic features, so as to retain and train rich steganographic features. and achieve a good color image steganalysis effect. In addition, WISERNet improves network performance by widening the width of the upper convolutional layer to increase the output involved in the summation. However, a too wide network structure will increase the parameter scale, which is not conducive to network expansion and optimization.

Since then, a number of steganalysis models for color images designed for the transform domain have been proposed [36][37]. Or they realize a multi-domain joint detection model [40][41]. However, there is currently no steganalysis scheme for color images in the spatial domain with better performance than WISERNet.

Therefore, there is a lot of room for research on adaptive steganalysis of color images in the spatial domain. This paper introduces multi-frequency residual analysis and proposes a new solution suitable for the steganalysis of spatial color images.

B. ARCHITECTURE

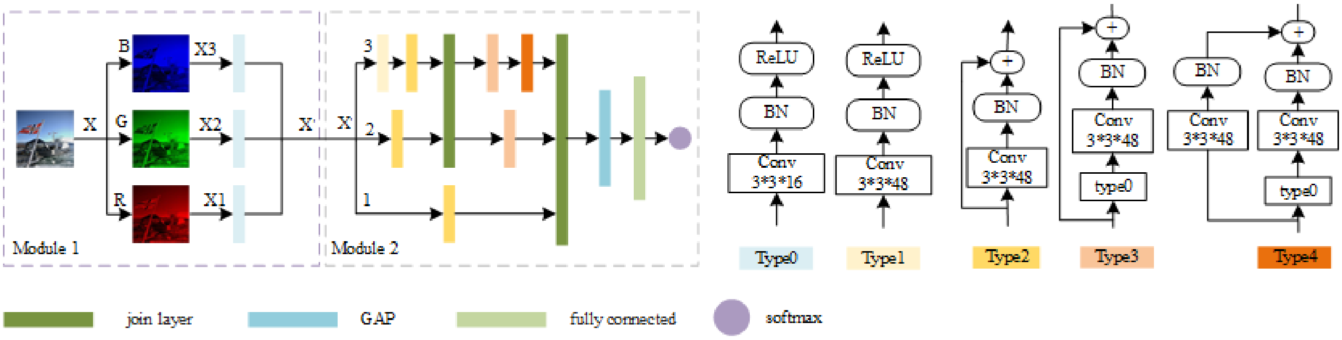

FIGURE 1. Overview design of the proposed MFRNet and four basic block

The core idea of the proposed MFRNet for color image steganalysis in this paper is multi-frequency residual analysis. The network consists of two modules, the color image preprocessing module, and the multi-frequency steganographic noise residual extraction and classification module. The overall design is shown in Fig. 1.

1) THE COLOR IMAGE PREPROCESSING MODULE

The color image preprocessing module is used to reduce the image content, enhance the steganographic noise, thereby improving the steganographic noise ratio. By analyzing the different frequency components of the input signal, we found that if the input signal contains strong interference from image content information, the learning effect will decrease. Therefore, preprocessing the input image is very helpful to improve the performance of detection. WISERNet [15] has proved that channel-wise convolution operations can reduce the image content and enhance steganographic signals more effectively than ordinary convolution operations when detecting color images. Therefore, we also select channel-wise convolution in the color image preprocessing module.

Specifically, we design a Type 0 basic block for the color images preprocessing part to reduce the image content and enhance steganographic signal. It contains a convolutional layer, a BN layer, and a ReLU activation. The size of the convolution kernel here is set to 3×3×16, the stride is set to 1.

The input of the network is a three-channel RGB color image X, the definition domain of which can be expressed as $(C \times I^m \times I^n)$. Among them, C represents the number of input channels, and C=3. I represents the value range of the input image, and $I \in [0,255]$. m, n respectively represent the width and height of the input image.

First, the input color image X is divided into RGB three channels, X_1 to X_3, which definition domain can be expressed as $(C \times I^m \times I^n)$, C=1. Then Type 0 is applied to the three channels for channel-wise convolution, and merge the output results of the three channels to finally get the output X' of color image preprocessing module, whose value range can be expressed as $(C \times I^m \times I^n)$, and C=3. X' is used as the input of the multi-frequency steganographic noise residual extraction and classification module.

2) THE MULTI-FREQUENCY STEGANOGRAPHIC NOISE RESIDUAL EXTRACTION AND CLASSIFICATION MODULE

This module is the core of MFRNet. It learns multi-frequency steganographic noise residuals at the same time through columns of different depths to ensure that the network learns more high-frequency steganographic information while eliminating the interference of low-frequency image content as much as possible. So as to achieve a better learning effect with a relatively lightweight structure. This part is mainly to stack columns of different depths by designing 4 types of basic blocks Type 1-Type 4 with residual shortcuts. The kernel size of the convolution is set to 3*3*48, the stride is set to1.

A key issue of this module is the calculation of residuals. In order to construct columns of different depths, the residual calculation method adopted by each type of basic block should be conducive to the expansion of the network. Therefore, the traditional method using the SRM filter [9] is not suitable for the proposed MFRNet. After comparison, a method of calculating residuals through residual shortcuts proposed in SRNet [14] is more conducive to the overall design of the network. Therefore, we use the residual shortcuts to calculate the residual.

The input of this part is denoted as X', the domain of definition is $(C \times I^m \times I^n)$, and C=3. We express a convolution operation as $f(X')$. For ease of understanding, other non-convolutional operations such as the BN layer and activation function are ignored here. Then Type 1 to Type 4 can be expressed as:

$$\begin{gathered} F_1 = f(X') \\ F_2 = f(X') + X' \\ F_3 = f^{(2)}(X') + X' \\ F_4 = f^{(2)}(X') + f^{(1)}(X') \end{gathered} \tag{1}$$

Here F_1 to F_4 respectively represent the output of Type1 to Type4, including the direct mapping part and the residual part, corresponding to the direct branch and bypass of each basic block in Figure 1. "+" means short-cut, that is, unit plus here. The superscript n in $f^{(n)}(X')$ represents the number of convolution operations, abbreviated as f^n. Because multiple convolutions are equivalent to re-extracting and refining the feature map extracted by the previous convolution, the more convolution operations experienced, the deeper the residual level obtained. As shown in (1), Type 2 to Type 4 can learn different levels and types of residuals, and Type 4 has the deepest level of residuals, which is conducive to MFRNet extracting rich multi-scale residual features.

For this network, the input of the next layer of convolution is the output of the previous layer of convolution. Therefore, we denote the output f^n of the nth convolution and the result of m times convolution again as f^{n+m}.

Then, the output J_1 of the first connection layer, that is, the synthesis (arithmetic average) of the output of Column 1-Column 2 can be expressed as (2).

$$J_1 = \left\{\frac{1}{2}[(f^1 + X') + ((f^1 + X')f^1)]\right\}$$
$$= \frac{1}{2}(f^1 + X' + f^2 + f^1) \tag{2}$$

And Column 1-Column 3 before the last join layer can be expressed as:

$$C_1 = f^1 + X'$$
$$C_2 = \frac{1}{2}(f^2 + X')J_1 = \frac{1}{2}f^4 + f^3 + f^2 + f^1 + \frac{1}{2}X'$$
$$C_3 = \frac{1}{2}(f^2 + f^1)[(f^2 + X')J_1] = (f^2 + f^1)C_2$$
$$= \frac{1}{2}f^6 + \frac{3}{2}f^5 + 2f^4 + 2f^3 + \frac{3}{2}f^2 + \frac{1}{2}f^1 \tag{3}$$

Among them, $C_1 - C_3$ represent the output of columns 1-3 before the last join layer. It can be seen from (3) that as the depth deepens, the level of convolution in each column of the network is also deepening. The selection of the basic blocks that constitute each column here has been carefully designed. When the level of the basic block extraction residuals used in each column deepens with the deepening of the network level, the level of the learned steganographic noise residuals is also continuously deepening.

Then, the learning results of each column are synthesized through the join layer, which retains the steganographic noise as much as possible through the arithmetic average operation and eliminates the interference of the image content. The final output of the network can be expressed as:

$$C_{\text{out}} = \frac{1}{3}(C_1 + C_2 + C_3)$$
$$= \frac{1}{6}f^6 + \frac{1}{2}f^5 + \frac{5}{6}f^4 + f^3 + \frac{5}{6}f^2 + \frac{5}{6}f^1 + \frac{1}{2}X' \tag{4}$$

It can be seen from (4) that the final output combines the learning results of $C_1 - C_3$, and contains the residual components of steganographic noise of different frequencies. The level of convolution ranges from 6 times to 0 times, and the extracted features are very rich. Most of the existing networks only have one very deep column, so they can only learn the residual components of the steganographic noise of one frequency (usually low frequency). Therefore, although the depth of the network is deeper, it is not as rich as the residual features learned by our network.

The join layer is followed by the global average pooling layer(GAP), which accepts the input feature map of *C*=48, and obtains the output feature map of *C*=1 after dimensionality reduction. The output of GAP is sent to the fully connected layer, and then the final classification label is obtained through the softmax layer for two classification.

C. MULTI-FREQUENCY RESIDUAL ANALYSIS

The main innovation of the proposed MFRNet is to introduce the idea of multi-frequency residual analysis to steganalysis for the first time, so as to improve the overall performance of the network by simultaneously learning the steganographic noise components of different frequencies. In the previous section, we demonstrated through mathematical analysis how the idea of multi-frequency analysis works. This section will further demonstrate the rationality and feasibility of introducing the idea of multi-frequency analysis into steganalysis through spectrogram and Grad-CAM.

1) APPLICATION EFFECT OF MULTI-FREQUENCY RESIDUAL ANALYSIS

According to the frequency principle [20], the deeper the network, the more likely it is to learn low-frequency information, corresponding to the part of the image content, which is meaningless to steganalysis. Steganalysis focuses on the texture area of the image, which corresponds to high frequency information [21], [22]. Therefore, too deep network limits the improvement of steganalysis performance. The proposed MFRNet uses the idea of multi-frequency analysis and uses columns of different depths to learn information of different frequencies at the same time, which can achieve good steganographic detection performance through a relatively lightweight network.

In order to verify the application of multi-frequency analysis in the network, a sine wave S(t) containing low-frequency, intermediate-frequency, and high-frequency signal components is designed as the input of the network, as shown in (5):

$$S(t) = sin(10\pi t) + sin(30\pi t) + sin(90\pi t) \tag{5}$$

t represents the time (t∈[0s,256s]) and S(t) represents the change of the constructed sine wave signal over time in (5). The signal waveform and spectrogram are shown in Fig. 2. It can be seen from Fig.2 that the main frequency components of S(t) are 5, 15, 45 (Hz), which are consistent with the signal constructed in (5). (The interference of the signal with frequency 0 is not considered.). The value of S(t) is normalized to 0-255 to generate the corresponding gray-scale image, and send it to the model for testing. The Grad CAM [24] of the detection result of MFRNet is converted into a 1D signal wave, and the Fourier Transform spectrogram [25], [26]is shown in Fig. 3. It can be seen that after the signal S(t) passes through MFRNet, the components of different frequencies can be maintained, and the high-frequency components are retained to the greatest extent (The peak of the frequency in the histogram of Fig. 3(a) is at the position corresponding to the high frequency). Among them, the first column learns the highest frequency component of the signal, and the second column learns the middle frequency component, the third column learns the low frequency component. It shows that the idea of multi-frequency analysis is well applied in MFRNet.

2) THE INFLUENCE OF MULTI-FREQUENCY RESIDUAL ANALYSIS ON THE EFFECT OF NETWORK LEARNING

From the above analysis, we can see that the idea of multi-frequency residual analysis has indeed been correctly applied in MFRNet. In order to better show the significance of this idea for improving the performance of steganographic detection, this section compares the MFRNet with the idea of

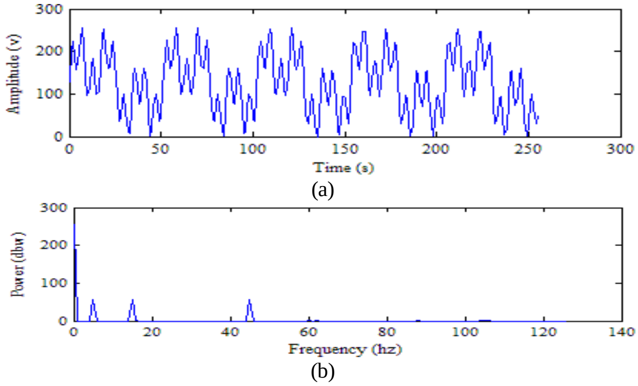

FIGURE 2..The signal waveform and spectrogram of the S(t) containing multi-frequency signals.(a) shows the signal waveform of S(t), and (b) shows the spectrogram of S(t).

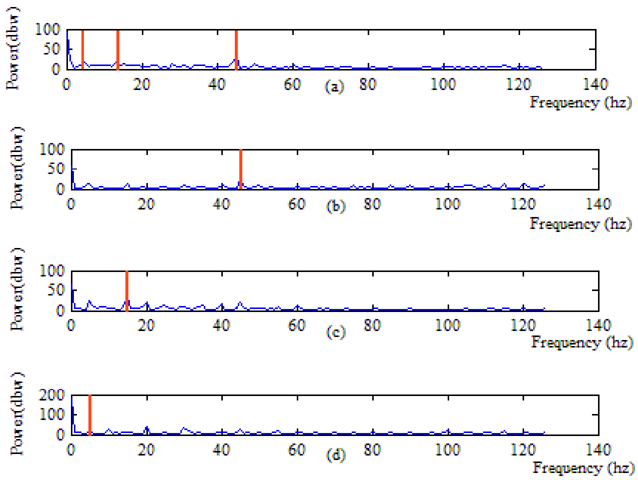

FIGURE 3. The spectrogram of the proposed MFRNet and each column. (a).MFRNet. (b).Column 1. (c).Column 2. (d).Column 3.

TABLE 1. The detection accuracy of each column of MFRNet under different payload (bpc, %).

Model/Payload	0.1	0.2	0.4	0.6	0.8
Column 1	52.6	56.7	67.3	75.0	80.7
Column 2	56.5	68.2	82.6	89.2	92.0
Column 3	63.5	78.1	90.1	95.8	97.4
MFRNet	**71.6**	**90.1**	**97.2**	**99.2**	**99.2**

multi-frequency analysis and the network that only learns the residual of the steganographic noise of a single frequency.

Specifically, extract the three columns of MFRNet for separate training, respectively corresponding to different depths of the network that only learns the residual component of the single-frequency steganographic noise. In order to ensure the normal convergence of the network, only the initial learning rate is reduced to 0.001, and the other experimental data sets and parameter settings are the same as III A. The trained model detects the images generated by the channel-by-channel S-UNIWARD algorithm under different payloads, and the results are shown in Table 1.

As is shown in Table 1, when the idea of multi-frequency noise residual analysis is not applied, the detection performance of a single column of MFRNet for images with low steganography is not good. As the depth increases, the

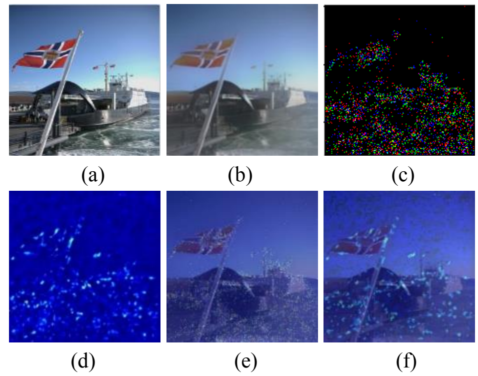

FIGURE 4. The learning effect shown by Grad-CAM of the proposed MFRNet. (a). Cover. (b). Cover(blurred). (c). Steganographic area. (d).Grad-CAM. (e) Steganographic area + cover. (f). Grad-CAM + cover.

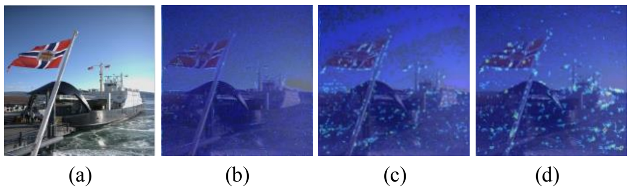

FIGURE 5. THE LEARNING EFFECT SHOWN BY GRAD-CAM OF COLUMNS 1-3 OF MFRNET. (A). COVER.(B). COLUMN 1. (C). COLUMN 2. (D). COLUMN 3.

learning effect of the column that only learns a single frequency steganographic noise component is getting better. And the deeper the network (Column 3), the better the image detection effect for small payloads, which is in line with existing research. However, the learning effect of the complete MFRNet is better than the learning effect of any single column, especially when detecting images with a small payload, the detection accuracy is greatly improved.

Though Accuracy can be used to evaluate the learning effect of the network, but it cannot visually show which areas the network pays attention to. Thus, we use CAM to display the learning effect.

The idea of CAM was first proposed by Selvaraj in [23]. By visualizing the learning results after the last convolutional layer, the content that you pay attention to during network learning can be effectively displayed. However, CAM requires modifying the original model structure: replace the fully connected layer with a global average pool. In order to save training costs, Grad-CAM [24] is used instead, which has been proved that is equivalent to CAM.

We first show the overall learning effect of the network. Take the 55th color image in the PPG-LIRMM-COLOR dataset with 0.4bpc channel-by-channel S-UNIWARD as an example for detecting, and use Grad-CAM to show the output feature map after the last connection layer of MFRNet. In order to better show the learning effect, the original image is blurred and converted to BGR format (displayed in blue). The Grad-CAM is superimposed on the processed original image for visualization; as a comparison, the steganographic position map is converted into a pseudo-color image and superimposed on the processed original image. The effect is shown in Fig. 4.

In Fig 4 (d), the brighter the color represents the area with higher attention. Comparing Fig 4 (e) and (f), it can be seen that the actual learning effect of the network is accurate and effective after introducing multi-frequency residual analysis.

Furthermore, the learning effect of column 1-3 is shown by Grad-CAM. As shown in Fig. 5, from left to right, the learning effect of the last convolutional layer in column 1-3. The first row is the Grad-CAM map, and the second row is the Grad-CAM superimposed to the cover image. It can be seen that the steganographic information learned by different columns is different, corresponding to different frequency components of the steganographic noise. And as the depth of the column deepens, it will learn more steganographic information while containing more interference from the image content, so the effect is not as good as MFRNet for multi-frequency analysis.

It proves that the proposed multi-frequency analysis method can indeed effectively improve the learning effect of the network under the lighter network scale. And effectively avoid the poor learning effect of a single column.

D. DESIGN DETAILS

This section will discuss other design details of the proposed MFRNet. The models used for testing here were trained under the same settings as III A, using detection accuracy as an evaluation indicator.

1) DESIGN OF MULTI-SCALE RESIDUALS

While learning multi-frequency residuals, MFRNet calculates residuals of different scales by designing residual basic blocks with different types of shortcut connections, which effectively improves the detection accuracy of the network. In order to evaluate the help of obtaining multi-scale residuals to improve the accuracy of network detection, we keep other parts of the network unchanged, and removes the Type 4, Type 3 and Type 2 residual basic blocks from the network respectively. The results are shown in Table 2. When only Type 1 is used, the network does not calculate residuals, the accuracy of detecting 0.4 bpc S-UNIWARD channel-by-channel algorithm is less than 90%, and the learning effect is not good. As the types of residual basic blocks used increase, the scales for calculating residuals are more abundant. When using Type 2-Type 4 at the same time, compared to using Type 2 only, the detection accuracy is improved by nearly 7%. It can be seen that compared with single-scale residual learning, multi-scale residual extraction can effectively improve the detection performance of the network.

2) ACTIVATIONS

The activations can add a non-linear factor to the network. By adding an activation after the convolutional layer, it is helpful for feature extraction and network convergence, and avoids problems such as overfitting and gradient disappearance.

In order to explore the influence of the activations on the network, we compare the influence of adding activations at different locations of the network performance. As shown in Table 3, when all the basic blocks use the activations, the network performance is reduced by 1% compared to without activations at all. When only Type 1 uses ReLU, the network performance is the best. It can be seen that the activations are beneficial to the ordinary convolution block, but not conducive to the calculation of the residual. Therefore, Type 2-Type 4 will no longer use ReLU activation after adding residual shortcuts.

TABLE 2. The impact of multi-scale residuals on network performance. Model1: Replace Type2-4 with Type 1. Model 2: Replace Type 3-4 with Type 2. Model 3: Replace Type 4 with Type 3. Model 4: Use Type 1-Type 4 at the same time.

Model	Accuracy(%)
Model 1	89.7
Model 2	90.3
Model 3	--
Model 4	**97.2**

TABLE 3. The influence of activation on network performance. Model 1: Without ReLU. Model2: With ReLU. Model 3: Only Type 1 with ReLU.

Model	Accuracy(%)
Model 1	92.6
Model 2	91.7
Model 3	**97.2**

In addition, in addition to the linear activation ReLU, we also compare the effects of the hyperbolic tangent activation tanh, Leaky-ReLU [30], and ELU [31]. However, they do not bring performance improvements to the network. Therefore, MFRNet finally only uses the ReLU activations in Type 1.

3) BN LAYERS

In order to explore the impact of the BN layers, we compare the detection accuracy of the network with and without the BN layers, which are 97.2% and 90.4%. Moreover, when the BN layer is not used, the network starts to converge when it is trained to 60k steps. And when the BN layer is used, the network has converged at 40k steps. It can be seen that the use of the BN layer can make the network converge faster and better, and improve the network learning effect. Therefore the BN layer is used after convolution layers in all types of basic blocks.

4) OPERATION FOR JOIN LAYERS

In order to explore the impact of different operations of the join layers, we compare the maximum, minimum, arithmetic average and addition operations. The detection accuracy are 89.0%, 75.0%, 97.2% and 86.6%. It is found that the max and min operations will overemphasize the dominant signal, and add operation retains all the learned information. However, the use of the arithmetic average can better average the learning results of each column. It retains as much steganographic information as possible while effectively eliminating the interference of image content. Therefore we adopt arithmetic average as the best selection operation in the join layers.

5) POOLING LAYER

In order to confirm the impact of the pooling layers on network performance, the paper tried different pooling operations. When we try to use average pooling layers to gradually reduce the dimensionality, but the network could not converge. It is because the average pool is equivalent to a low-pass filter, which enhances the image content and suppresses steganographic noise by averaging adjacent embedding changes. It is unfavorable for steganalysis. This view is also confirmed in SRNet [16]. MFRNet only uses a GAP to reduce the dimensionality of the feature map.

6) OPTIMIZER

We compare two optimizers, Adam and Adamax [27]. Finally, Adamax is chosen because the detection performance is higher and it can converge faster.

7) NUMBER OF COLUMNS

We try to expand the number of columns of the network, but found that it does not significantly help the improvement of performance. When the number of columns is set to 4, the detection performance of the network is only improved by less than 1%. Therefore, we finally set 3 columns to achieve a better balance between network scale and performance.

III. EXPERIMENTS

A. DATASETS AND PARAMETERS

The dataset used in our experiments is PPG-LIRMM-COLOR [28], which contains 10000 color images in ppm format of different types (people, landscapes, buildings, etc.) with a size of 512*512. Ppm is a simple image format that can retain stego signal to the utmost extent, and is often used for steganography and steganalysis of color images. In order to facilitate the learning of CNN network, the size of the images is all converted to 256*256 in the experiment.

Our main detection target in the experiment is the channel-by-channel adaptive steganography algorithm (S-UNIWARD, HILL, WOW, HUGO) with different payloads (0.1-0.8 bpc). The training set contained 6,000 cover images and 6,000 stego images. Both verification set and test set contained 2,000 cover images and 2,000 stego images.

The steganalyzers used for comparison are mainly WISERNet [17]. WISERNet was proposed in 2019 as the state-of-the-art adaptive steganalysis model for color images. The underlying channel-by-channel convolution strategy proposed by it has a significant effect on steganalysis of color images.

Meanwhile, we also select YedroudjNet [14]、SRNet [16], and WangNet [43] several spatial or joint domain gray image steganalysis models. YedroudjNet proposed in 2018 is a representative network with nice performance using SRM filters. And SFNet proposed in 2019 has better performance than SRNet when the network width and depth are small. WangNet is a cross-domain joint detection model proposed in 2020. It improves the detection accuracy of images with small steganography rates through a joint domain detection mechanism and a non-linear detection mechanism, which is better than other solutions. We apply the same channel-wise convolution to the lowest level convolution because the original SRNet and YedroudjNet are designed for images of gray-scale. They are expressed as color-YedroudjNet, color-SRNet and color-WangNet in the experiment.

The above-selected model covers the latest and most representative solutions in the image steganalysis field in recent years, and is reproduced strictly in accordance with the parameter settings in the original paper. Thus making our experimental data more convincing.

The proposed model is trained using images generated by the channel-by-channel S-UNIWARD steganography, 0.4 bpc. The max generations of epoch is200 and the batch-size was 20. The learning rate is 0.01. The learning rate decline method is exponential decline and the decay rate is 0.95. The parameters of color-YedroudjNet, color-SRNet, color-WangNet, and WISERNet are set according to [14], [16], [43], and [17]. The learning rate is set to 0.4, 0.01,0.4 and 0.001 respectively.

B. COMPARISION OF NETWORK SIZE

One of the main problems of the existing steganalysis algorithms based on deep learning is that the model is too complex, which makes the training cost very high, which is not conducive to the expansion and optimization of the model. Thus, it is necessary to reasonably control the complexity of the network. This part we evaluate the complexity of the model from three aspects: the depth and width of the network, the scale of the parameters, and the average training time cost.

1) DEPTH AND WIGTH

Table 4 displays the comparison of the width and depth of the network. The definition of the width here relates to WISERNet [17], which refers to the number of channels of the convolution kernels in each convolutional layer of the network. While the depth is counted by the number of convolutional layers used by the network. The greater the depth and width of the network, the more complex the network. Here color-WangNet uses 2D and 3D convolution at the same time, so it is not suitable for comparison.

As is shown in Table 4, the width of the proposed MFRNet is much lower than WISERNet and color-YedroudjNet, and the depth is much lower than color-SRNet. Thus it can be seen that both in terms of depth and width, MFRNet has more advantages.

2) PARAMETER SCALE

In addition, the parameter scale is used to evaluate the complexity of the model, which includes the number of model parameters and the required GPU storage space. It can be used to measure the resource occupation of different models and facilitate the training and optimization of the model on the GPU.

Considering that the color image preprocessing module is shared by all networks participating in the comparison, and is also the core part of the color picture steganalysis network, it is convenient for comparison and has representative significance. Therefore we only compared the parameter scale of the color image preprocessing layer here. The number of convolution kernel channels is all recorded as 3 in this experiment. Set the original cover image to be 55.ppm of PPG-LIRMM-COLOR. Set batch size to 4. Set the computing device to be a TITAN X model GPU server with 12G video memory.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3119664, IEEE Access

TABLE 4. Comparison of network depth and width

Model	Width	Depth (Shallowest/deepest)
color-YedroudjNet	3*30/32/64/128	6
color-SRNet	3*16/48	22
color-WangNet	--	--
WISERNet	72/288/1152	3
MFRNet	3*16/48	2/7

TABLE 5. Comparison of parameter scale of preprocessing layer

Model	Kernels Size	Parameter Scale	CPU Space/MB
color-YedroudjNet	(3,30,5,5)	23595210	94.38
color-SRNet	(3,16,3,3)	12583344	50.33
color-WangNet	(3,30,5,5）	23595210	94.38
WISERNet	(3,30,5,5）	23595210	94.38
MFRNet	(3,16,3,3)	12583344	50.33

TABLE 6. Comparison of training time(s)

Model	Training Time Scale
color-YedroudjNet	152.3
color-SRNet	346.028
color-WangNet	353
WISERNet	239
MFRNet	346

The parameter scale of the network preprocessing layer is given in Table 5.The parameter scale of the proposed MFRNet is reduced by about half compared with WISERNet and color-YedroudjNet._Combined with the smaller network depth and width of MFRNet, it can be concluded that the overall parameter scale of MFRNet is much smaller than other existing networks.

3) TRAINING TIME SCALE

Finally, the training time scale is compared. The size of dataset for training is 6000 pairs of cover/stego images, and the batch size is 20. The average time of 10 epochs after the model is running stably is used to measure the training time scale. As shown in Table 6, the proposed MFRNet is in the same order of magnitude as the existing network in terms of training time scale, and the gap is within an acceptable range.

Moreover, combined with subsequent analysis, the performance of MFRNet is slightly better than that of WISERNet, which means that MFRNet can efficiently detect steganographic images with a lighter structure.

C. PERFORMANCE OF MFRNET WHEN DETECTING ADAPTIVE STEGANOGRAPHY ALGORITHMS

As can be seen from the previous section, MFRNet is more lightweight than existing steganalysis networks in terms of model complexity. But for a steganalysis algorithm, the detection accuracy is also very important. This section we first test the performance of the MFRNet detection performance on adaptive steganography algorithm proposed in the most ideal situation. That is, when the parameters of the dataset during model training and detection are consistent.

Table 7 shows the performance of each model trained with 0.4 bpc S-UNIWARD to detect 0.4 bpc steganographic images. The index is accuracy. It can be seen that the detection performance of the proposed MFRNet is 1.4% better than the existing best model WISERNet, and it is better than the existing gray-scale image steganalysis scheme combined with the color image steganalysis strategy. It can be seen that MFRNet can achieve better performance with a more lightweight network structure. the detection accuracy of MFRNet is about 1.4% higher than WISERNet.

D. PERFORMANCE OF MFRNET IN ACTUAL STEGANOGRAPHY DETECTION TASKS

As can be seen from the previous experimental results, MFRNet can be implemented in an ideal environment with a more lightweight architecture due to the performance of existing solutions. However, in actual steganography detection tasks, the detection conditions are often more demanding than the experimental environment. For example, the steganography algorithm and payloads in the application are often unknown. Therefore, in order to ensure the practicability of the designed steganalysis algorithm, it is usually necessary to consider the influence of the mismatch of the steganalysis algorithm and the mismatch of the payloads. This section will give the performance of MFRNet when the payload is mismatched and when the steganography algorithm is mismatched.

1) PAYLOAD MISMATCH

The payload mismatch is a common situation in actual steganography detection tasks and a relatively low detection difficulty, so we will discuss it first. Generally speaking, the smaller the payload is, the less secret information is embedded and the more difficult it is to detect. Therefore, the existing models generally have poor detection performance on images with low steganography rates, and generally have better detection performance on images with larger steganography rates. How to improve the detection performance of images with low steganography rate. Adaptive steganalysis is a problem that needs to be solved urgently.

We use the trained models to detect S-UNIWARD adaptive steganography images of 0.1, 0.2, 0.6, and 0.8 bpc respectively, and the detection results are shown in Table 8.

It can be seen that when the payload is mismatched, MFRNet will have less impact. Especially for images with low payloads, its performance is much better than existing steganalysis networks. When the payload is 0.2 bpc, the detection accuracy of MFRNet is 90.1%, which is 25.7% higher than WISERNet, and 17.1% higher than color-WangNet._Even compared with WangNet, which has good performance in detecting small payloads, our MFRNet has better performance. Moreover, when detecting 0.6 and 0.8 bpc images with lower difficulty, the performance of the proposed MFRNet has also been improved and is higher than other existing solutions.

TABLE 7. Model performance when detecting 0.4 bpc S-UNIWARD adaptive steganography algorithm (%)

Model	Accuracy
color-YedroudjNet	90.8
color-SRNet	96.2
color-WangNet	87.0
WISERNet	95.8
MFRNet	**97.2**

TABLE 8. Detection accuracy of channel-by-channel S-UNIWARD algorithm under different payload when payload mismatch (bpc, %)

Model/bpc	0.1	0.2	0.6	0.8
color-YedroudjNet	53.0	64.6	97.3	98.0
color-SRNet	54.0	72.6	98.0	97.9
color-WangNet	54.5	68.2	91.1	91.9
WISERNet	52.2	64.4	97.2	96.9
MFRNet	**71.6**	**90.1**	**99.2**	**99.2**

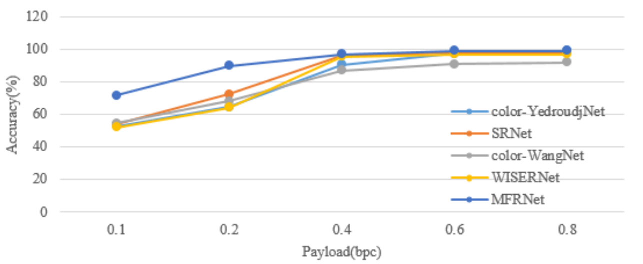

FIGURE 6. The detection accuracy of the channel-by-channel S-UNIWARD algorithm under different payloads

TABLE 9. Detection accuracy of different channel-by-channel steganography under 0.4 bpc (%)

Model	HILL	WOW	HUGO
color-YedroudjNet	90.6	91.0	89.2
color-SRNet	96.0	96.6	95.7
color-WangNet	85.9	86.9	87.8
WISERNet	94.8	95.3	94.8
MFRNet	**96.5**	**96.7**	**97.2**

TABLE 10. Comparison of network performance between MFRNet and steganalyzers with selection channel(%,bpc).Model 1 represents color-YedroudjNet with selection channel. Model 2 represents MFRNet

Algorithm	Model/Payload	0.1	0.2	0.3	0.4	0.5
S-UNIWARD	Model 1	56.3	69.4	84.5	91.2	93.2
	Model 2	**71.6**	**90.1**	**94.8**	**97.8**	**98.1**
HILL	Model 1	59.2	74.5	86.0	91.3	92.8
	Model 2	**69.7**	**86.5**	**93.8**	**96.5**	**97.5**
WOW	Model 1	60.7	76.8	86.4	89.8	90.6
	Model 2	**73.5**	**89.1**	**95.1**	**96.8**	**97.9**

Fig. 6 shows a more intuitive display. It can be seen that MFRNet is competitive when the payload is mismatched. As analyzed in II D, the proposed MFRNet can achieve such an excellent detection performance of steganography images when payload mismatch is related to the idea of multi-frequency analysis. The experimental results in this section also prove this point well.

2) STEGANOGRAPHY ALGORITHM MISMATCH

In the actual steganography detection task, a situation that is also common and more difficult to detect is the mismatch of the steganography algorithm. Therefore, the detection accuracy when detecting different steganography algorithms an important indicator for evaluating the practicability of the network. Table 9 shows the detection accuracy of the 0.4 bpc adaptive steganography algorithm (S-UNIWARD, HILL, WOW, HUGO).

As shown in Table 9, under 0.4bpc, when detecting all channel-by-channel steganography algorithms, MFRNet can achieve better performance than other existing networks. And the best color image steganalysis algorithm WISERNet is around 1.5%.

More importantly, the mismatch of the steganography algorithm has little impact on the performance of MFRNet, only about 1% drop). It can be seen that MFRNet can well deal with the problem of algorithm mismatch in actual steganographic detection tasks.

3) COMPARING WITH THE METHOD OF COMBINING THE SELECTION CHANNEL

In order to improve the detection performance of the steganalysis model detection adaptive steganography algorithm in the actual steganalysis detection task, a commonly used solution is to introduce the channel selection technology. Selection channel is a technology that uses prior knowledge such as embedded probability maps to highlight the contribution of steganographic noise in complex regions of the image texture, and can effectively improve the performance of the steganalysis method. Based on multi-frequency residual analysis, MFRNet can better learn the content of the steganographic noise suppression image, and can even be better than the steganalysis method that uses the selected channel technology to enhance image texture.

Thus, we select color-YedroudjNet with better performance, and use the selection channel to enhance its performance, called SCA-color-YedroudjNet as a comparison. The results are shown in Table 10. It can be seen that when the proposed MFRNet without selection channel detects multiple steganography algorithms with different payloads(0.1 bpc - 0.5 bpc), the performance is better than YedroudjNet combined with the selection channel. This proves once again that the network we proposed has a strong competitive advantage.

E. COMPARISON OF THE OUTPUT FEATURE MAP

In the previous experiment, we have described the performance of the proposed MFRNet in the ideal and actual environment through the detection accuracy index. However, there is still no more intuitive feeling for the actual learning effect of the network. We usually think that the learning ability of convolutional neural networks is usually unable to accurately describe, but the feature map can describe the feature extraction process very well.

In order to show the feature extraction capabilities of the proposed MFRNet more clearly, the paper visualizes the network feature map of each convolutional layer. We still choose color-YedroudjNet with better performance as a comparison. We visualized the output feature map of the last convolutional layer before the GAP (Global Average Pooling)

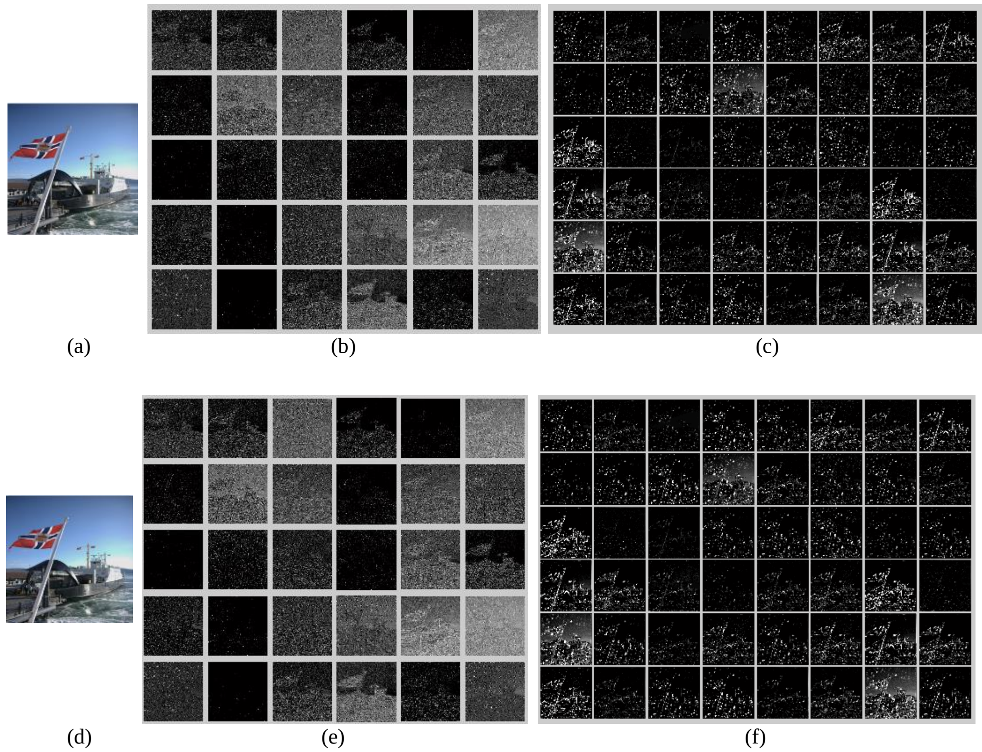

FIGURE 7. The comparison of feature maps between YedroudjNet and MFRNet. (a). Cover image. (b). The feature map of cover generated by color-YedroudjNet. (c). The feature map of cover generated by MFRNet. (d). Stego image. (e). The feature map of stego generated by color-YedroudjNet. (f). The feature map of stego generated by MFRNet.

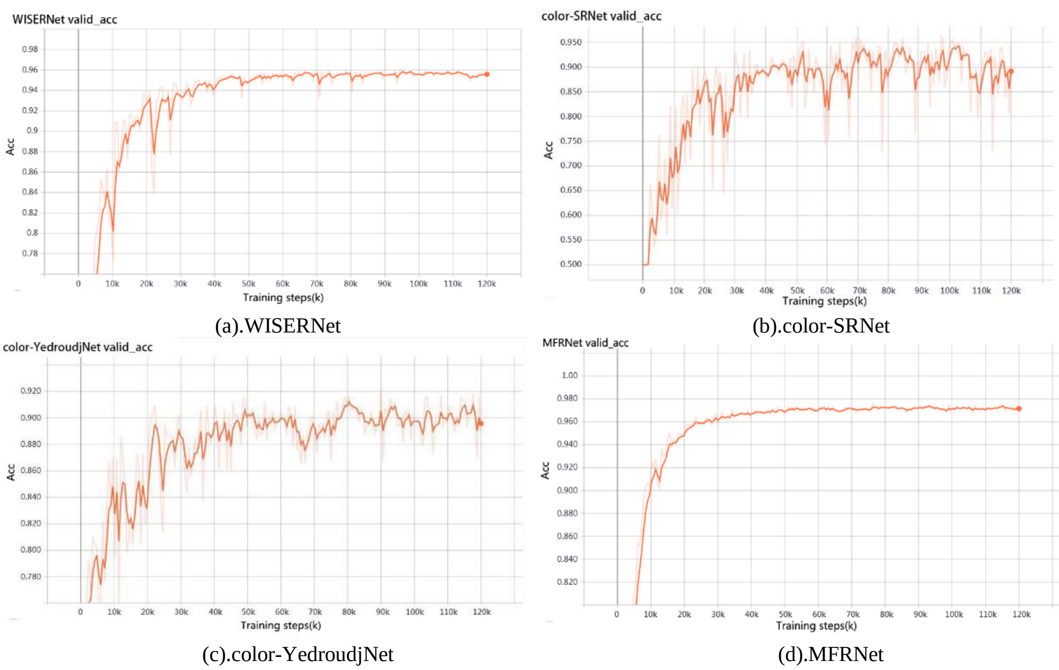

FIGURE 8. Comparison of network convergence

layer (layer 3 for color-YedroudjNet, and layer 6 for MFRNet.) Because convolutional layers before GAP are mainly used to extract steganographic features. When GAP is added, the feature map output by CNN is difficult to interpret and visualize. A feature map can describe the feature extraction process well. Both MFRNet and color-YedroudjNet are trained under S-UNIWARD 0.4 bpc. The comparison result of feature maps generated based on natural images and steganographic images is shown in Fig. 7.

The results show that the feature map generated by MFRNet retains less image content information, and the signal-to-noise ratio between the steganographic signal and the image signal continues to increase. Whether for steganographic images or natural images, this method can extract features with strong expressive ability. At the same time, the similarity between each feature map is relatively low, which is convenient for subsequent convolution and classification. In contrast, the feature map generated by color-YedroudjNet retains more image content information, and the difference between the feature maps is not obvious.

F. COMPARISION OF NETWORK CONVERGENCE

In the above experiment, we have demonstrated the excellent performance of the network detection adaptive steganography algorithm. This section further demonstrates the performance of MFRNet through the convergence of the network.

In this part, the accuracy of the verification set during the training process is visualized through TensorBoard, and the smoothness value is 0.6. The training set is a randomly selected 6000 PPG-LIRMM-COLOR cover images and the corresponding stego images generated by 0.4 bpc channel-by-channel S-UNIWARD. The batch size is set to 20, and the total model is 200 iterations (120K steps) to train WISERNet, color-SRNet, color-YedroudjNet and MFRNet.

As shown in Fig. 8, the three models all begin to converge at about 40k steps. And MFRNet achieves the same fast convergence speed when the depth is much lower than that of color-SRNet. The stability of MFRNet is also very well.

IV. CONCLUSION

In this paper, a multi-frequency residual convolutional neural network MFRNet for color image steganalysis is proposed. For the first time, the idea of multi-frequency residual analysis is introduced into steganalysis, and the performance of WISERNet is no less than the existing color image steganalysis models in spatial domain with a more lightweight network architecture. And compared with the existing network, it can better deal with the mismatch problem of steganography in actual steganography detection tasks.

In future work, we will aim to study the steganalysis in the frequency domain. In addition, we will study the detection of the more secure steganography algorithm for color images, named vector steganography algorithm. It better preserves the correlation between the pixels in the neighborhood of the color image, and the detection difficulty is also higher.

REFERENCES

[1] T. S. Reinel, R. P. Raul, and I. Gustavo, “Deep Learning Applied to Steganalysis of Digital Images: A Systematic Review,” *IEEE Access*, pp.99, 2019.

[2] A. D. Ker, “Steganalysis of LSB matching in grayscale images,” *IEEE Signal Process*. Lett., vol. 12, no. 6, pp. 441–444, Jun. 2005.

[3] L. M. Zhai, J. Jia, W. X. Ren, Y. B. Ren and L. N. Wang, “Recent advances in deep learning for image steganography and steganalysis,” *Journal of Cyber Security*, vol.3, no.6, 2018.

[4] Z. J. Fu, F. Wang, X. M. Sun and Y. Wang, “Research on Steganography of Digital Images based on Deep Learning,” *Chinese Journal of Computers*, vol.43, no.9, 2020.

[5] V. Holub, J. Fridrich and D. Tomá, “Universal distortion function for steganography in an arbitrary domain,” *Eurasip Journal on Information Security*, vol.2014, no.1, pp.1, 2014.

[6] V. Holub and J. Fridrich, “Designing Steganographic Distortion Using Directional Filters,” *IEEE Workshop on Information Forensic and Security*, 2012.

[7] B. Li, W. Ming, J. Huang and X. Li, “A new cost function for spatial image steganography,” *2014 IEEE International Conference on Image Processing (ICIP)*, 2015.

[8] T. Pevný, T. Filler and P. Bas, “Using High-Dimensional Image Models to Perform Highly Undetectable Steganography,” *Lecture Notes in Computer Science*, vol. 6387, pp.161-177, 2010.

[9] J. Fridrich and J. Kodovský, “Rich models for steganalysis of digital images,” *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 868–882, Jun. 2012.

[10] V. Holub and J. Fridrich, “Random projections of residuals for digital image steganalysis,” *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 12, pp. 1996–2006, Dec. 2013.

[11] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, “Selection-channel-aware rich model for steganalysis of digital images,” in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, pp. 48–53, Dec. 2014,

[12] J. F. Chen, Z. J. Fu, W. M. Zhang, X. Cheng and X. M. Sun, “Review of Image Steganalysis Based on Deep Learning,” *Journal of Software*, vol.32, no.2, pp.551−578, 2021.

[13] Y. Jian, J. Ni and Yang Y, “Deep Learning Hierarchical Representations for Image Steganalysis,” *IEEE Transactions on Information Forensics and Security*, vol.12, no.11, pp.2545-2557, 2017.

[14] M. Yedroudj, F. Comby, M. Haumont, “Yedroudj-Net: An efficient CNN for spatial steganalysis,” *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[15] Xu G, Wu H Z, Shi Y Q. Structural Design of Convolutional Neural Networks for Steganalysis. *IEEE Signal Processing Letters*, 2016:708-712.

[16] M. Boroumand, M. Chen and J. Fridrich, "Deep Residual Network for Steganalysis of Digital Images," *IEEE Transactions on Information Forensics and Security*, vol.14, no.5, pp.1181-1193, 2019.

[17] J. S. Zeng, S. Q. Tan, G.Q. Liu, B. Li and J.W. Huang, “WISERNet: Wider Separate-then-reunion Network for Steganalysis of Color Images,” *IEEE Transactions on Information Forensics and Security*, pp.99, 2018.

[18] G. Larsson, M. Maire and G. Shakhnarovich, “FractalNet: Ultra-Deep Neural Networks without Residuals,” *ICLR*, 2016.

[19] B. Singh, A. Sur and P. Mitra, "Steganalysis of Digital Images Using Deep Fractal Network," *IEEE Transactions on Computational Social Systems*, 2021.

[20] Z. Xu, Y. Zhang, T. Luo, Y. Xiao and Z. Ma, “Frequency Principle: Fourier Analysis Sheds Light on Deep Neural Networks,” *ICLR*, 2019.

[21] Z. Xu and H. Zhou, “Deep frequency principle towards understanding why deeper learning is faster,” 2020.

[22] Z. Q. Xu, Y. Zhang, Y. Xiao, “Training behavior of deep neural network in frequency domain,” 2018.

[23] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, “Learning Deep Features for Discriminative Localization,” *IEEE Computer Society*, 2016.

[24] U. Selvaraj, R. R. Selvaraju, M. Cogswel, A. Das, D. Parikh and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *International Journal of Computer Vision*. vol.128, no.2, pp.336-359, 2017.

[25] H. P. Ren, “MATLAB image processing based on Fourier transform,” *Science & Technology Information*, no.16, 2019.

[26] H. H. Chen, “Simple Analysis of Signal Spectrum by Fast Fourier Transform,” *Electronic Test*, 2020.

[27] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” CoRR, Dec. 2014. [Online]. Available: http://arxiv.org/abs/1412.6980.

[28] Hasan Abdulrahman, Marc Chaumont, Philippe Montesinos, and Baptiste Magnier, "Color Image Steganalysis Based on Steerable Gaussian Filters Bank," I*H&MMSec'2016, in Proceedings of the 4th ACM workshop on Information Hiding and Multimedia Security*, Vigo, Galicia, Spain, 6 pages, June 20-22, 2016.

[29] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *In NIPS'2012*, 2012.

[30] Andrew L. Maas, Awni Y. Hannum and Andrew Y. Ng, "Rectified NonlinearitiesImprove Neural Network Acoustic Models," *ICML*, 2013.

[31] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," CoRR, 2015.[Online]. Available: http://arxiv.org/abs/1511.07289.

[32] Z. Zhu, H. Wei, G. Hu, et al, "A Novel Fast Single Image Dehazing Algorithm Based on Artificial Multiexposure Image Fusion," *IEEE Transactions on Instrumentation and Measurement*, pp.99, 2020.

[33] M. Zheng, G. Qi, Z. Zhu, et al, "Image Dehazing by An Artificial Image Fusion Method based on Adaptive Structure Decomposition," *IEEE Sensors Journal*, pp.99, 2020.

[34] A. Aljarf, and S. Amin, "Steganalysis System for Colour Images Based on Merging the Colour Gradient Cooccurrence Matrix and Histogram of Difference Image," *2018 21st Saudi Computer Society National Computer Conference (NCC)*, pp. 1-7, 2018.

[35] Z. I. Rasool, M. M. Al-Jarrah and S. Amin, "Steganalysis of RGB Images Using Merged Statistical Features of Color Channels," *2018 11th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 46-51, 2018.

[36] Yassine Yousfi, Jan Butora, Jessica Fridrich, and Quentin Giboulot, "Breaking ALASKA: Color Separation for Steganalysis in JPEG Domain," *In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security. Association for Computing Machinery*, New York, NY, USA, pp. 138–149, 2019.

[37] N. Mohamed, T. Rabie, and I. Kamel, "A Review of Color Image Steganalysis in the Transform Domain," *2020 14th International Conference on Innovations in Information Technology (IIT)*, pp. 45-50, 2020.

[38] Renad M. Al-Manaseer, and Mudhafar M. Al-Jarrah, "Steganalysis of Color Images for Low Payload Detection. In Proceedings of the 2019 2nd International Conference on Information Hiding and Image Processing," *2019 Association for Computing Machinery*, New York, NY, USA, pp.35–38, 2019.

[39] T. S. Reinel, A. H. Brayan, B. M. Alejandro, et al., "GBRAS-Net: A Convolutional Neural Network Architecture for Spatial Image Steganalysis," in *IEEE Access*, vol. 9, pp. 14340-14350, 2021.

[40] K. Chubachi, "An Ensemble Model using CNNs on Different Domains for ALASKA2 Image Steganalysis," *2020 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1-6, 2020.

[41] R. Cogranne, Q. Giboulot and P. Bas, "ALASKA#2: Challenging Academic Research on Steganalysis with Realistic Images," *2020 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1-5, 2020.

[42] Y. Qian, J. Dong, W. Wang and T. Tan, "Deep learning for steganalysis via convolutional neural networks," *Proc. IS T Int. Symp. Electron. Imag. (EI)*, vol. 9409, 2015.

[43] X. Q. Deng, B. Chen, W. Q. Luo, and D. Luo, "Fast and Effective Global Covariance Pooling Network for Image Steganalysis," *In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security. Association for Computing Machinery*, New York, NY, USA, pp. 230–234, 2019.

[44] Z. Wang, M. Chen, Y. Yang, et al, "Joint multi-domain feature learning for image steganalysis based on CNN," *EURASIP Journal on Image and Video Processing*, vol,2020, no.1, 2020.