A Novel Spatial Image Steganalyzer with Adaptive Channel Attention

Fan Nie*, Chaoyang Zhu
Department of Information and Communication, China Electric Power Research Institute, Beijing, China
*Corresponding Author.

Abstract

For image steganalysis, many studies have showed that the superiority of the convolutional neural network over conventional methods based on artificially designed features. With the trend of the fusion of traditional steganalysis methods and some tricks used in classic computer vision tasks, such as SRNet equipped with residual modules and ZhuNet which used spatial pyramid pooling, more and more CNN architectures used for steganalysis are proposed. However, there still are some characteristics in most content-adaptive steganographic algorithms such as S-UNIWARD, HUGO, WOW, and tricks in designing network structure which can be used for steganalysis. Here, we propose a CNN network framework which can further improve the performance of spatial image steganographic algorithms. First, we utilize more SRM kernels to initialize the pre-processing layer than previous CNNs, and use an image padding method different from traditional modelsto preserve the integrity of image residuals as much as possible. Next, we use multiple channel attention layers which aim to discriminate the more informational features boosting the detection accuracy of network. Then, we deploy the spatial pyramid pooling layer before features are fed into the fully-connected layers, aiming to extract more features from the last feature mapsin several scales. Several experiments under different steganographic algorithms show that, the proposed CNN outperforms the other CNN-based steganalyzers such as YeNet, XuNet, YedroudjNet, SRNet and ZhuNet.

Keywords: Image steganalysis, Convolutional neural network, Channel attention layer

I. Introduction

Since the emergence of steganography and steganalysis, they have been developing in competition with each other. Image steganography is away of hiding secret information in different domains of an image by slightly making some unnoticeable modification (changing pixel values of image in the spatial domain, and DCT coefficients in the frequency domain). In the past several years, with the advent of multimedia, more and more types of information were used in communication, steganographic algorithms keep developing, from simplest LSB embedding[1,2] to content-adaptive algorithms. Nowadays, those steganographic approaches equipped with content-adaptive algorithms are more secure, which tend to conceal data in some highly textured regions, aiming to decrease the values of the specific pre-defined additional distortion function. Such as WOW[3], S-UNIWARD[4], HUGO[5], and the other methods[6] in spatial domain. In contrast, significant progress has also been made in the field of image steganalysis, whose propose is to detect the existence of hidden information in images. Before the advent of machine learning techniques, the most powerful analyzer of steganographic algorithms was pixel statistical methods, which took advantage of some defects in traditional steganographic algorithms, including RS analysis[7], chi-Square test. Later on, due to the development of machine learning, several novel steganalysis tools based on the content of images are proposed[8,9]. The most famous steganalyzer based on ML classifiers in spatial domain is the Spatial Rich Model[9] including its multiple improved variants[10-12], which are regarded as the milestone of the modern steganalysis. Most of these steganalyzers are formed by assembling several submodels constructed by different high-pass filters into a rich bigger model. However, the performance of the conventional steganalyzers with hand-crafted features rely strongly on the effect of feature engineering, in other words, getting more complete information of image, the model will get better performance. In contrast, due to limitation of memory and computational power, it is impossible to use traditional machine learning models directly with a huge amount of...
features which may lead to overfitting or the curse of the dimension.

Benefiting from the development of deep learning, there are several models of steganalysis using CNN[13-17]. In the traditional computer vision tasks, which include image segmentation[18-20], image classification[20,21] and object detection[22], CNN has been proved that it can efficiently extract the features of images and get better performance, compared with conventional methods. Unlike fixed hand-crafted filters used in traditional steganalysis detectors[9], these filters used in different layers of CNN can be optimized automatically by back propagation. Therefore, Diverse CNN architectures of steganalysis using different ways to improve the detection accuracy are proposed, such as high-pass KV filter[13,14] which is one of the filters in SRM [9], the absolute value layer(ABS layer)[14], more SRM kernels[15,17], integrating some modules of classical network architectures(such as ResNet[23], InceptionNet[24-26]).

Here, we propose a new network architecture implemented by a convolutional neural network (CNN) named NieNet for steganalysis to capture more comprehensive and discriminative features of images. There are multiple novel characteristics compared with other CNN-based image steganalyzers, which are detailed as follows:

1. In the pre-processing layer, we use 46 different filters of SRM(linear and non-linear kernels) to initialize the kernels to extract the residual of images and modify the shape of some convolution kernels which are 3rd spam kernels mentioned in [15,17]. Additionally, before convoluting the input images, the images will be mirror-padded to improve the robustness of our model and to achieve better accuracy.
2. We deploy a new network module that combines essences of two advanced CNN frameworks, ResNet and InceptionNet. This type of network architecture can be used for extracting image features from different perspectives, meanwhile avoid the problem of gradient vanishing.
3. The success of the attention mechanism in several computer vision tasks demonstrates that attention is important for neural network, we add the channel attention layer to each Res-Inception module similar to[27]. With the help of this layer, the weights of different types of SRM kernels will be assigned dynamically to increase the stability of our CNN.

On two datasets (BOSSBase[34] and BOWS2[35]), several experiments are conducted, in which we train multiple CNN-based steganalyzers including ours under the same settings and compare their detection performance. And experimental results show our proposed steganalyzer achieves state-of-the-art performance.

II. Related Works

Tan [28], uses a convolutional neural network equipped with four convolutional layers for image steganalysis, plus well-designed initialization using one of the hand-crafted filters. According to characteristics of Gaussian function, Qian[13] proposed that a CNN architecture of steganalysis with Gaussian activation function. But the accuracy of these models is still worse than traditional models which consist of SRM kernels and classic machine learning methods (Support Vector Machine, Fisher Linear Discriminator)[9]. Later, Xu[14] proved that the feasibility of batch normalization layer in CNN-based steganalyzer and introduced absolute value function into the network architecture.

Inspired by previous approaches of image residual extraction, Sedighi[29] proposed a CNN structure with initialized weights with SRM kernels and well-designed histogram layers, which projected feature maps into histogram maps. YeNet[15] proposed a CNN architecture with 30 high-pass pre-defined filters mentioned in SRM[9] for initializing weights of a pre-processing layer plus a truncated activation function (TLU). The Yedrouj-Net[30] achieves better performance than YeNet[15] by using multiple methods of data augmentation. A deep network for image steganalysis equipped with shortcut connections, SRNet, was proposed in [16] and significantly improve detection accuracy in both spatial domain plus JPEG domain. And it also proved that a model using randomly initialized kernels for the first pre-processing layer can also get good performance with the deeper network structure. Zhu[17] deployed more advanced layers including grouped convolution layers and spatial pyramid.
pooling layers to a CNN structure, it achieves better accuracy of detection and also regards images of arbitrary size as input. Wang[31] combined multiple domains of images and proposed WangNet whose the first layer is initialized with DCT coefficient and more SRM kernels, their experiments showed that with the increase of kernels initialized by SRM kernels, the model can achieve better accuracy.

2.1 Architecture

The architecture of the proposed CNN in demonstrated in Fig 1. It consists of multiple stacked layers including one image pre-processing layer with SRM initialization, several feature extraction layers containing two different types of Res-Inception layers with shortcut connection, a spatial pyramid pooling layer mentioned in ZhuNet, and there are two fully connected layers generating the probability of stego/cover.

There are two types of blocks named as ‘TypeA Inception’ and ‘TypeB Inception’ shown in Fig. 3 to extract spatial correlation with different kernels which have multiple shapes and finally feed these feature maps into fully connected layer. Each type of basic block consists of the following steps approximately:

2.1.1 Convolutional Layer

We utilize multiple small convolution kernels (e.g. 3*3) to achieve function of larger kernels instead of directly using large convolution kernel, preventing too many parameters to slow the time of convergence. By this way, this model can effectively extract different feature maps by using Inception layer, which contains several parallel convolution layers with kernels of different size. Specifically, the big difference between two types of blocks is that, the maxpooling operation is used in the TypeA block which aims to simulate the function of non-linear kernels mentioned in[9], and the TypeB block is equipped with depth-wise convolution layers similar to [17] aiming to extract features efficiently. And there is a hyperparameter named ‘reduction’ to be used for controlling the number of outputs’ channels, according to our setting, the hyper-parameter is set to 2.

2.1.2 Batch Normalization Layer

As Xu[14] mentions, CNN-based steganalyzer can take advantage of superiority of batch normalization layer, the use of batch normalization allows us to set a larger learning rate, and neural networks can converge quickly
benefiting from the uniform distribution of features.

2.1.3 Different Non-Linear Activation Function
Except the first preprocessing layer followed by a TLU function. For all other blocks in the proposed CNN architecture, we use ReLUs as the activation function after Batch Normalization operation, it has been proven in many computer vision tasks to have excellent characteristics, including avoiding gradient vanishing and accelerating model convergence.

2.1.4 Average Pooling Layer
In order to decrease the dimension of feature maps, a CNN usually is equipped with several pooling layers for down-sampling feature maps. For most circumstances, it would be max pooling layer, for in the field of CNN-based steganalyzers, average pooling layer is widely chosen for down-sampling features, because of the ability of preserving weakly steganographic signal. Therefore, we choose the average pooling layer for the proposed CNN instead of max pooling layer.

2.2 Diverse Kernels
The hiding operation of steganographic algorithms can be regarded as generating designed noises and adding them to the cover image. As mentioned in YeNet[15], ZhuNet[17], Yedoudj-Net[30], it is a great idea to use SRM kernels’ initialization for steganalysis based on neural network to extract residual features. However, the common of these former models is that all of them use 30 high-pass spam filters of SRM and their rotated counterparts, we deploy a set of high-pass filters to our CNN’s preprocessing layer (30 high-pass filters of SRM, similarly to YeNet[15], Yedoudj-Net[30] and ZhuNet[17], plus 14 non-linear minmax filters). Therefore, our proposed CNN is able to extract more noise residual maps from input images than other networks, which means more comprehensive information can be extracted by our CNN.

2.2.1 Improved Kernels
As some existing CNN-based steganalyzers demonstrate, the filter of size 3 × 3, such as “SQUARE 3 × 3”, “EDGE 3 × 3” and the remaining 13 filters of size 5 × 5 including “SQUARE5 × 5”, “EDGE5 × 5” and “SPAM 3rd” can efficiently extract discriminative features of image. Besides these SPAM kernels, we creatively add some non-linear kernels to the preprocessing layer to increase the number of noise residual maps. In practice, it can be easily implemented by using cross-channel max and min operations. As the additional kernels of SRM[9], we choose the limited numbers of non-linear kernels which include minmax2nd21, minmaxEDGE3 × 3 22v, minmaxEDGE5 × 5 22v, and minmax3rd 22v. As shown in Fig.2, there are two types of values of different colors performing the same operations just like linear kernels.

2.2.2 Mirror-Padding Operation
In many traditional computer vision models which take the pixel values of the image as input, the zero padding...
operation is most commonly used for the sake of simplicity. From the perspective of steganalysis, because of slight differences between cover images and stego images, modeling image residue will extract more discriminative features using some hand-crafted filters than directly modelling pixel values. In YedroudjNet[30] and XuNet[14], the weights of convolutional kernels in the preprocessing layer are initialized by some SRM kernels. In ZhuNet[17], there are more SRM filters used for initialization. However, there is a defect among these models, that almost CNN architectures use the zero padding operation before convoluting to keep the size of feature map, but this operation will damage the quality of our extracted pixels’ residues. For instance, if a zero value is padded to the image, the residue of this image will become large and cannot show the characteristics of this image. Considering the characteristic of noise residues extraction and the operation of some popular steganographic algorithms, such as HUGO[5], WOW[3], S-UNIWARD[4], which all use the mirror padding operation to calculate the embedding cost of the image, doing zero padding operation and optimizing kernels as usual will damage the quality of noise extraction. According to Table 1, by using mirror-paddingle operation, the performance of CNN will improve a lot.

| Different strategies of optimizing kernels | Our CNN | YeNet |
|-------------------------------------------|---------|-------|
| mirror-padding optimizing                  | 0.172   | 0.286 |
| zero-padding optimizing                    | 0.191   | 0.284 |
| mirror-padding fixed                       | 0.176   | 0.279 |

### 2.2.3 Inception Module with Separable Convolution

Inception module has recently been proved that it can improve the model accuracy significantly in traditional computer vision tasks, such as Inception[24-26], Xception[32]. Because of the wider parallel convolutional layers and shortcuts, this module can get more informational features from different scale than general convolution layers but also keep the whole model from getting deeper which will lead gradient vanishing. And in practice, we put these modules into deeper layers which can prevent from damage the residual information of cover/stego, for each Inception module, we just use small sized kernels which can efficiently reduce storage space and accelerate the training process. The another advantage of that is with use of multiple size of filters, the network will benefit from larger receptive field, and discriminate more informational features.

![Inception Module with Separable Convolution](image)

**Fig 3:** This caption has one line so it is centered. Two different Inception modules used to the proposed CNN. Left: Type A Inception module for shallow layers, Right: Type B Inception module for deeper layers.
As shown in Fig. 3, we apply two different types of Inception Module to our proposed network named "Incep-TypeA" and "Incep-TypeB", more details are shown in Fig. 3. In the more shallow layers which are quipped with 'Incep-TypeA' modules, The use of a combination of maxpooling operation and convolution operation will not only extract more features in the larger field, but also add more non-linear features by maxpooling, just like non-linear kernels do in the preprocessing layer. In the deeper layers which consist of several 'Incep-TypeB' modules, we apply more branches of convolution operations which include grouped convolution operations to these layers, aiming to extract more semantic features efficiently.

2.2.4 Channel Attention Layer
Due to the use of different attention modules which can make the model assign different weights to the content of different regions in the same input according to the task like a brain, the performance of the model is improved greatly, such as recurrent attention convolutional neural network [33], and Residual Attention Network [21]. At the point of steganalysis’s view, the actual input to our CNN is several residual maps which are calculated by the preprocessing layer’s multiple kernels. As [9] mentioned, the different residual map processed by different kernels has different performance of steganalysis. Inspired by those ideas, we proposed a new attention layer called channel attention layer (CAL) used for our CNN-based steganalyzer. The outputs of the CAL, the weight for each feature channel, will be used to output weighted feature maps. As shown in Fig. 4. Considering the number of additional trainable parameters brought by this layer, we use a simple strategy to construct this layer similar to [27] which is composed of global average pooling operation and several fully connected layers. But there are some differences between the CAL and SE block mentioned in [27]. In the CAL, using a depth-wise convolutional layer additionally will improve the robustness of this attention module rather than [27].

![Channel Attention Module Diagram](image)

Fig 4: A channel attention module used in the proposed CNN, similar to [27]

To explore the impact of using channel attention layers on CNN performance, we designed several CNN structures for steganalysis mentioned which are applied proposed attention layers. As shown in Table. 2, due to addition of the CAL, there is slight improvement among these three networks including our proposed CNN. For ZhuNet and Our NieNet under S-UNIWARD at 0.4 bpp, both of them equipped with CAL get better performances.

| Steganographic Algorithms | S-UNIWARD | WOW |
|---------------------------|-----------|-----|
| S-UNIWARD                 |           |     |
| WOW                       |           |     |

Table 2 Steganalysis error rate comparison of networks with channel attention layer against multiple steganographic algorithms at 0.4 bpp
2.2.5 Spatial Pyramid Pooling Module
Spatial pyramid pooling module (SPP) is proposed in Zhu-Net[17] and helps the model achieve better performance. By using the SPP module, this pooling operation can extract feature information in different scales such as $1 \times 1$, $2 \times 2$, $4 \times 4$, and better model the local features. Besides, Benefiting from the fact that spatial pyramid pooling operation can turn the output of feature maps into fixed sized tensor, our CNN can steganalyze arbitrary-size images.

2.2.6 The Difference between SRNet, ZhuNet and OurCNN
The comparison of three CNN-based steganalyzers is illustrated in Table 3. Compared to SRNet, the differences in Our CNN are using hand-crafted kernels initializing weights of the preprocessing layer proposed in SRM[9], mirror-padding operation, deploying two types of Inception Modules (with and without separable convolution) plus spatial pyramid pooling module. Compared to Zhu-Net, the uniqueness of our CNN is using improved kernels and additional non-linear kernels for initialization, using mirror-padding operation, and using shortcut connection for all left layers.

| Algorithm | Preprocessing module | Padding operation of pre-processing layer | Inception or shortcut | Pooling before fully connected layers |
|-----------|----------------------|------------------------------------------|-----------------------|---------------------------------------|
| SRNet     | Random initialization and update filters | zero padding | shortcuts | global average pooling |
| ZhuNet    | initialize with 30 specific kernels and update them | zero padding | depthwise separable convolutions and shortcuts | spatial pyramid pooling |
| NieNet    | initialize with 44 specific kernels and update them | mirror padding | inception separable module and shortcuts | spatial pyramid pooling |

### III. Experiments

Several experiments are conducted to show the effectiveness of our proposed CNN. We compare our model with several CNN-based steganalyzers: XuNet[14], YeNet[15], YedroudjNet[30], SRNet[16], ZhuNet[17]. All networks are trained and tested on the same datasets and same steganographic algorithms (WOW[3], S-UNIWARD[4]) for fair comparison.

3.1 The Environments

For steganographic algorithms, all of them are implemented on the publicly available codes. We choose the ones in Matlab implementation with random embedding key. As mentioned, if the embedding key is unchanged, the generalization of networks will decrease dramatically. And by using Nvidia Tesla V100 graphic card with 32GB memory for training, we can use larger batch size of training images to get more stable parameters of batch normalization layer.

3.2 Datasets

ISSN: 0010-8189
© CONVERTER 2020
www.converter-magazine.info
In this paper, we use the combination of two standard datasets which are commonly used for steganalysis to train these networks and test the performance of these networks. The first source of images is from the BOSSBase[34], it contains 10,000 512\times512 grayscale images taken by several devices and is commonly used in steganalysis. The other is BOWS2[35], it contains 10,000 512\times512 grayscale images whose distribution is similar to BOSSBase. Based on consideration of GPU memory and batch size of training process, we decide to conduct experiments on resized images of 256\times256 pixels created by Matlab with default settings. The setting of dataset of training and testing will be detailed later.

3.3 Experimental Parameters Setting

Stochastic gradient descent (SGD) approach is widely used in several field of deep learning as the optimizing algorithm. Therefore, at the training phase of our image steganalyzer, we choose SGD to optimize our model. According to previous experiences and memory limitation of our graphic processing unit, batch size of our training data is going to be set to 32 (16 cover/stego image pairs). The ultimate goal of our training phase is to minimize the cross-entropy loss of outputs of our CNN.

On our dataset, the model will be trained for 250 epochs with a learning rate of r= 0.0001. The learning rate will be modified (divided by 5) at epoch 100, 150, and 200 respectively. According to our practical experience, the CNN training will converge totally at about 200 epochs. We train and test all models used in this paper for several times using same dataset setting (uniform training/validation/testing sets). The final experimental results are concluded by averaging the testing results. The metric of these CNN models was defined by the steganalysis error rates \( P_e = 1 - P_{acc} \).

3.4 Experimental Results

3.4.1 BOSSBase Only

As shown in Table 4, the performance of several steganalyzers are reported in our uniform experiment settings. First, cover images of BOSSBase are randomly divided into three parts: 4000 cover/stego image pairs are used for training networks, 1000 cover/stego image pairs are used as validation set and 5000 pairs as test set. In this table, we use two steganographic algorithms S-UNIWARD and WOW to conduct experiments under different payloads.

| Algorithm     | Payload/bpp | WOW  | S-UNIWARD |
|---------------|-------------|------|-----------|
|               | 0.2         | 0.4  | 0.2       | 0.4       |
| XuNet         | 0.331       | 0.283| 0.354     | 0.312     |
| YeNet         | 0.324       | 0.277| 0.341     | 0.298     |
| YedroudjNet   | 0.317       | 0.252| 0.327     | 0.279     |
| SRNet         | 0.251       | 0.204| 0.294     | 0.243     |
| ZhuNet        | 0.215       | 0.176| 0.258     | 0.217     |
| NieNet        | 0.223       | 0.167| 0.247     | 0.213     |

In different experiment settings including different algorithms and different payloads, our CNN has gotten better detection performance over the other CNN-based steganalyzer. Specifically, under the circumstance of same payload and different algorithms (S-UNIWARD and WOW), the detection performance of our proposed CNN is ahead of other steganalyzers except ZhuNet, the proposed CNN and ZhuNet get comparable detection performance under same payload, (such as 21.5% versus 22.3% under 0.2 bpp with WOW, 17.6% versus 16.7% under 0.4 bpp with S-UNIWARD) Similar to other networks under same algorithm and different payloads, our proposed CNN can get better performance under larger payload, (for example 21.3% versus 24.7% under S_UNIWARD with 0.2 bpp and 0.4 bpp) We speculate that the reason for this phenomenon is that because of a small number of training datasets, these attention modules used are not well trained, when the number of training samples goes larger, this
phenomenon will not exist, and the next part of the experiment also verifies this idea.

3.4.2 BOSSBase and BOWS2
For well training networks with larger number of training samples, we add extra image sources BOWS2 (containing 10000 images which has the similar distribution with BOSSBase) into our training set. Therefore, the training set now is composed of 14000 pairs of cover/stego images.

Table 5 Steganalysis error rate comparison.

| Algorithm   | WOW Payload/bpp | S-UNIWARD Payload/bpp |
|-------------|-----------------|-----------------------|
| XuNet       | 0.318           | 0.243                 |
| YeNet       | 0.284           | 0.211                 |
| YedroudjNet | 0.278           | 0.177                 |
| SRNet       | 0.256           | 0.156                 |
| ZhuNet      | 0.241           | 0.127                 |
| NieNet      | 0.237           | 0.131                 |

As shown in Table 5, under the circumstance of larger training set, the performance of all networks has been improved simultaneously, in almost settings, our proposed CNN has outperformed other CNN-based steganalyzers. As we say above, once our attention modules can be well trained with larger datasets, our CNN will get better detection performance.

V. Conclusion

In this article, we apply more techniques used for traditional computer vision tasks to architecture of CNN-based steganalysis. The advantages of our proposed CNN are following: First, we deploy more pre-defined convolution kernels and use a novel padding operation in preprocessing layer, these operations make our CNN extract more discriminative features. More convolution kernels enhance the robustness of this model. Second, we continuously combine separable convolution mentioned in [17] and Inception Module to extract channel correlation. Finally, we use channel attention layer which will increase the computation slightly to make our model allocate different channel different weight. By using these tricks, the network performance is improved. Additionally, using SPP module, an arbitrary sized image can be steganalyzed by the CNN. In the future, the application of attention mechanism on image steganalysis should be exploited thoroughly, and we will design more powerful attention module for image steganalysis.

Acknowledgements

This research was partly supported by the National Key Research and Development Program of China (Grant No. 2020YFB0905900).

References

[1] H.-C. Wu, N.-I. Wu, C.-S. Tsai, and M.-S. Hwang, “Image steganographic scheme based on pixel-value differencing and lsb replacement methods,” IEE Proceedings-Vision, Image and Signal Processing, vol. 152, no. 5, pp. 611–615, 2005.
[2] Y.-K. Lee and L.-H. Chen, “High capacity image steganographic model,” IEE Proceedings-Vision, Image and Signal Processing, vol. 147, no. 3, pp. 288–294, 2000.
[3] V. Holub and J. Fridrich, “Designing steganographic distortion using directional filters,” in 2012 IEEE International workshop on information forensics and security (WIFS). IEEE, 2012, pp.234–239.
[4] V. Holub, J. Fridrich, and T. Denemark, “Universal distortion function for steganography in an arbitrary domain,” EURASIP Journal on Information Security, vol. 2014, no. 1, p. 1, 2014.
[5] T. Pevnỳ, T. Filler, and P. Bas, “Using high-dimensional image models to perform highly undetectable steganography,” in International Workshop on Information Hiding. Springer, 2010, pp. 161–177.

[6] B. Li, S. Tan, M. Wang, and J. Huang, “Investigation on cost assignment in spatial image steganography,” IEEE Transactions on Information Forensics and Security, vol. 9, no. 8, pp. 1264–1277, 2014.

[7] A. D. Ker, “Quantitative evaluation of pairs and rrssteganalysis,” in Security, Steganography, and Watermarking of Multimedia Contents VI, vol. 5306. International Society for Optics and Photonics, 2004, pp. 83–97.

[8] T. Pevny, P. Bas, and J. Fridrich, “Steganalysis by subtractive pixel adjacency matrix,” IEEE Transactions on Information Forensics and Security, vol. 5, no. 2, pp. 215–224, 2010.

[9] 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.

[10] V. Holub, J. Fridrich, and T. Denemark, “Random projections of residuals as an alternative to co-occurrences in steganalysis,” in Media Watermarking, Security, and Forensics 2013, vol. 8665. International Society for Optics and Photonics, 2013, p. 86650L.

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

[12] M. Goljan, J. Fridrich, and R. Cogranne, “Rich model for steganalysis of color images,” in 2014 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, 2014, pp. 185–190.

[13] Y. Qian, J. Dong, W. Wang, and T. Tan, “Deep learning for steganalysis via convolutional neural networks,” in Media Watermarking, Security, and Forensics 2015, vol. 9409. International Society for Optics and Photonics, 2015, p. 94090J.

[14] 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.

[15] J. Ye, J. Ni, and Y. Yi, “Deep learning hierarchical representations for image steganalysis,” IEEE Transactions on Information Forensics and Security, vol. 12, no. 11, pp. 2545–2557, 2017.

[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, 2018.

[17] 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.

[18] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[19] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 12, pp. 2481–2495, 2017.

[20] D. Ciregan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012, pp. 3642–3649.

[21] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, “Residual attention network for image classification,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3156–3164.

[22] P. Viola, M. Jones et al., “Robust real-time object detection,” International journal of computer vision, vol. 4, no. 34-47, p. 4, 2001.

[23] 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, 2016, pp. 770–778.

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.
[25] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” arXiv preprint arXiv:1502.03167, 2015.

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

[27] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.

[28] S. Tan and B. Li, “Stacked convolutional auto-encoders for steganalysis of digital images,” in Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific. IEEE, 2014, pp. 1–4.

[29] V. Sedighi and J. Fridrich, “Histogram layer, moving convolutional neural networks towards feature-based steganalysis,” Electronic Imaging, vol. 2017, no. 7, pp. 50–55, 2017.

[30] M. Yedroudj, F. Comby, and M. Chaumont, “Yedroudj-net: An efficient cnn for spatial steganalysis,” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 2092–2096.

[31] Z. Wang, M. Chen, Y. Yang, M. Lei, and Z. Dong, “Joint multi-domain feature learning for image steganalysis based on cnn,” EURASIP Journal on Image and Video Processing, vol. 2020, no. 1, pp. 1–12, 2020.

[32] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251–1258.

[33] J. Fu, H. Zheng, and T. Mei, “Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4438–4446.

[34] P. Bas, T. Filler, and T. Pevný, “’break our steganographic system’: the ins and outs of organizing boss,” in International workshop on information hiding. Springer, 2011, pp. 59–70.

[35] P. Bas and T. Furon, “Bows-2,” 2007.