2018 International Conference on Computer, Communications and Mechatronics Engineering (CCME 2018)
ISBN: 978-1-60595-611-4

A LeNet Based Convolution Neural Network for Image Steganalysis on Multiclass Classification

Yen-Ting Chen*, Tung-Shou Chen and Jeanne Chen

Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taiwan, R.O.P China

*Corresponding author

Keywords: Data hiding, Deep learning, Convolutional neural network (CNN), LeNet, Steganalysis.

Abstract. It is increasingly difficult to identify data hiding techniques used in the stego-images in this era of great advancement in technology. The purpose of this paper is to identify the types of data hiding techniques of stego-images. This paper proposed the use deep learning and convolutional neural networks (CNN) to train and generate a binary classification model for prediction. This paper proposed three types of detection algorithms for stego-images. The algorithms are based on the LeNet technique of convolutional neural network (CNN) in deep learning. The first algorithm detects the DCT stego-images. The second detects the histogram stego-images. The third and final algorithm detects the LSB stego-images. Experimental results from LeNet deep learning show significantly high detection rate. The DCT stego-images showed a significant high detection accuracy of 99%, while the LSB and histograms stego-images showed 70% accuracy rate.

Introduction

To date, many data hidden techniques had been proposed to protect information. In order to pass confidential information, many departments hide important the information into images to avoid being discovered. Some use includes the military for passing secret action plan. Currently, data hiding is one of the most effective ways to deliver hidden messages. As computers evolve rapidly, data hiding techniques are becoming more diverse too.

Data hiding can be divided into reversible data hiding [1] and irreversible data hiding. Reversible data hiding hides confidential information into images, and after extracting confidential information hidden in the images, the original images would be restored without any damages. The irreversible data hiding is where the embedding and extraction processes will damage the original image. An image with hidden data is called a stego image.

With the development of image data hiding technology, the features of the newly proposed data hiding method are gradually becoming more complicated. At present, models such as SRM (Spatial Rich Model) [2] and PSRM (Projection Spatial Rich Model) [3] have achieved good detection results. However, in the traditional manual feature detection environment, the new type of data hiding can easily avoid the traditional detection method. As a result, it is impossible to adapt to the new data hiding techniques.

However, there are many hidden images today with no good detection mechanism. Therefore, this paper proposed deep learning to identify multiple stego images and to use convolutional neural networks for image training where important features of an image are extracted. The method identifies three different stego image, namely, DCT stego images [4-5], histogram stego images [6] and LSB stego images [7]. In training images by the convolutional neural network, three of binary classification models are generated for the predictive analysis. Various classification results are generated through the model to determine which kind of data hiding is detected in the stego image by deep learning.

Method

This paper uses image processing techniques and deep learning convolutional neural networks as the based architecture. The first step is image preprocessing. Image preprocessing will first process the pixel difference. The second step is to divide the total image set into training image set and test chart. The third step designs and parameterizes the convolutional neural network model. The final step is to complete the model and output the results and analysis results as shown in Figure 1.

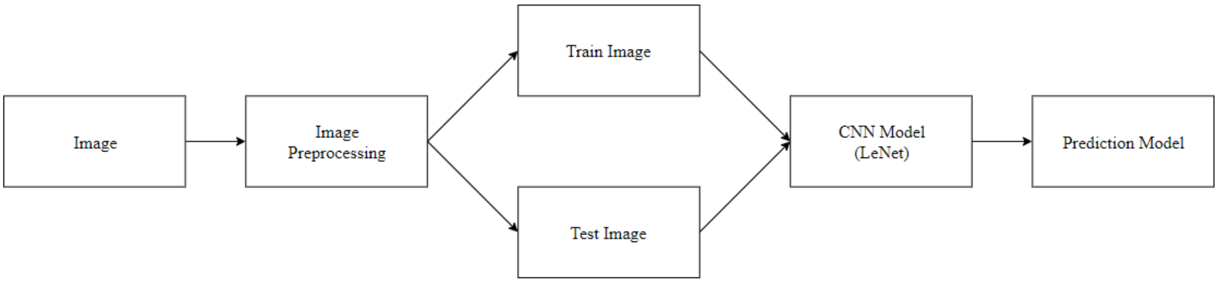
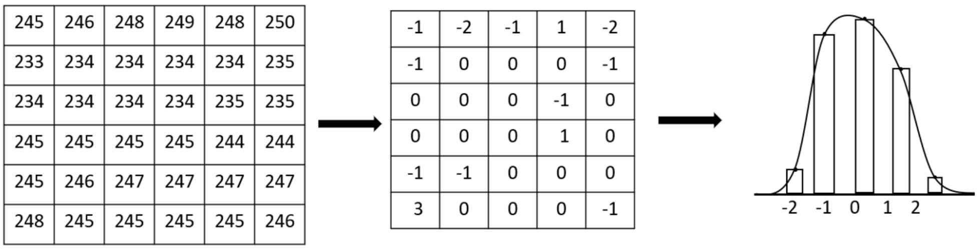


Figure 1. Overall architecture flow chart.

Image Preprocessing

The original image, DCT stego image, histogram stego image and LSB stego image each have 1000 pre-processing. The Subtractive Pixel Adjacency Matrix (SPAM) [8] is used to calculate four images respectively. The pixel difference is taken as its characteristic, and this image is regarded as the training image of the convolutional neural network in Figure 2.



245	246	248	249	248	250
233	234	234	234	234	235
234	234	234	234	235	235
245	245	245	245	244	244
245	246	247	247	247	247
248	245	245	245	245	246

-1	-2	-1	1	-2
-1	0	0	0	-1
0	0	0	-1	0
0	0	0	1	0
-1	-1	0	0	0
3	0	0	0	-1

Figure 2. Image Preprocessing for image.

Training Dataset and Testing Dataset

Through the image preprocessing method, each 1000 processed images are generated as the training image set of this paper, and the original image dataset and the DCT stego image dataset are the first model's dataset. The original image dataset and the histogram stego image are the datasets of the second model, and the original image dataset and the LSB stego image dataset are the third model's datasets in Figure 3.

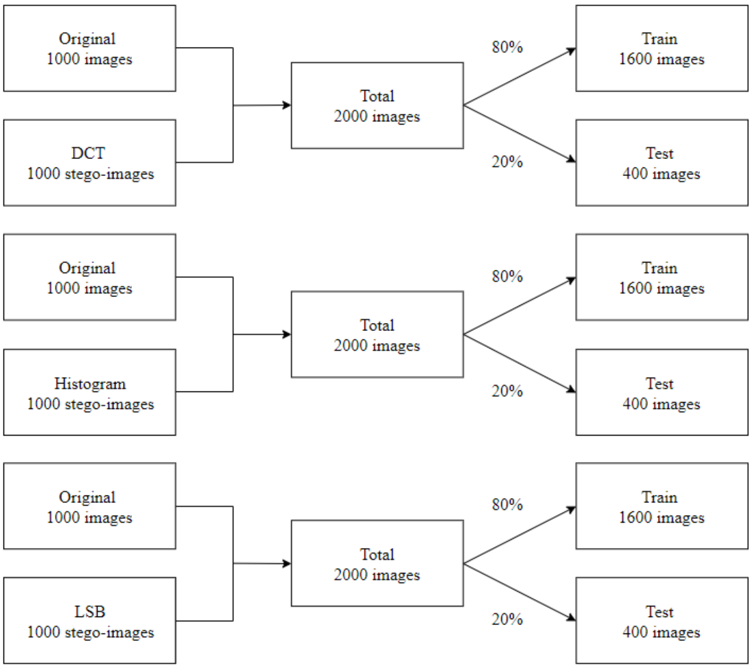


Figure 3. Training and testing dataset for model training.

Convolution Neural Network Model

This paper uses the LeNet [9] network architecture. The first and third layers are convolutional layers, the second and fourth layers are the largest pooling layer, and the last two layers are fully connected layers. The original image is transmitted through the model architecture. The output is a binary classification result in Figure 4.

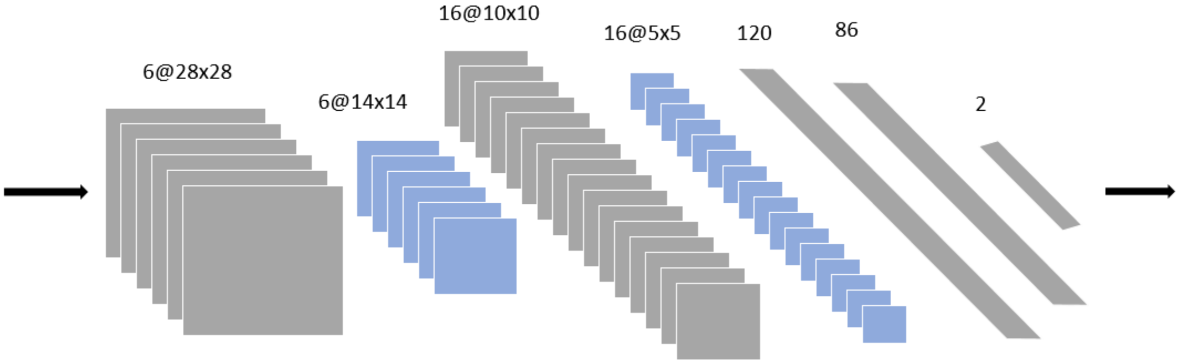


Figure 4. Overview of the LeNet architecture.

Convolutional Layer. Convolution is a feature extraction method, which requires training through a large number of pictures and repeated training and learning to achieve the best effect of the convolved features. However, the size of the image after convolution must also be set. Inputshape is the image size before input, Kernelshape is the size of the kernel, Padding is the number of zeros in the periphery of the image, and finally stride is the number of steps per step. The final Outputshape is the image size after the output, and the size of the image output through the convolution layer to Eq. 1.

$$Outputshape = \frac{Inputshape - Kernelshape + 2 \times Padding}{Stride} + 1 \qquad (1)$$

Max Pooling Layer. The difference between the max pooling layer and the convolution layer is that it does not do the feature extraction, but reduces the amount of image data and retains the most important information, and also reduces the size of the image output by the convolution layer. This paper used 2-layer max-pooling.

LeNet Network. This paper uses the LeNet network architecture, in which two layers of convolutional layers and two layers of the max pooling layer are finally two layers of fully connected layer as the LeNet parameters of this paper. Table 1 shows the descriptions of each layer.

Table 1. The architecture of the Convolution Neural Network.

Layer	Type	Maps & Neurons	Kernel
1	Convolution	6 Maps of 28x28 Neurons	5x5
2	Max-pooling	6 Maps of 14x14 Neurons	2x2
3	Convolution	16 Maps of 10x10 Neurons	5x5
4	Max-pooling	16 Maps of 5x5 Neurons	2x2
5	Fully-connected	120 Neurons	
6	Fully-connected	84 Neurons	

Prediction Model. The training image set uses the original images, the DCT stego images, the LSB stego images and the histogram stego images, and uses the binary classification method to train three models. The first model have the original images and the DCT stego images, the second model have the original images and the LSB stego images, and the third model have the original images and the histogram stego images, as the model's output and research analysis in Figure 5.

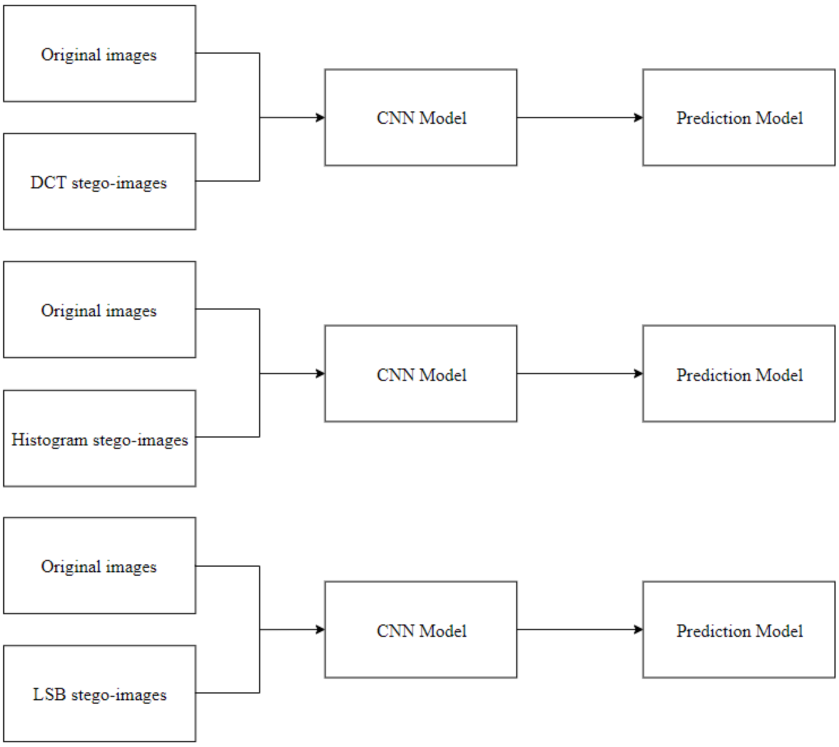


Figure 5. Prediction model.

Experimental Results

The data set is divided into two parts: training and testing, first using 80% of random samples for training models, 20% for testing models, and using another 2000 proofs to perform type testing, including original images, DCT stego image, histogram stego image and LSB stego image, each of which is 500 pictures. Using the model built by LeNet architecture, the accuracy of DCT stego images in 2000 test image sets is 99.00%, the histogram stego image is 68.00%, and the LSB stego image is 68.75%. It is a good effect in the discrimination of DCT stego images, and the discrimination of histograms and LSB stego images is about 70% accurate. Table 2 shows the accuracy of the model.

Table 2. Model testing accuracy.

Model	DCT Model	Histogram Model	LSB Model
Test images	2000	2000	2000
Accuracy(%)	99.0	68.0	68.75

In this experiment, 100x100 images were used for training. The Learning-rate was set to 0.001, and the experiment was performed with the epoch of 20.30.40.50. The training time and accuracy of the three models in each cycle were recorded separately. The records were relatively good accuracy, and finally take the best model test results from the experiment for verification. Table 3 shows the epoch for time and accuracy of the model.

Table 3. Model training time.

Model	DCT Model		Histogram Model		LSB Model	
Epoch	Time(s)	Accuracy(%)	Time(s)	Accuracy(%)	Time(s)	Accuracy(%)
20	312.62	97.75	303.03	49.50	287.25	78.00
30			499.28	80.00		
40			535.76	88.75	559.09	86.50
50					690.03	84.75

Conclusion

This paper proposed the use of deep learning and the convolutional neural networks training. DCT stego image has significant high recognition with high accuracy. However, the accuracy of the histogram embedding image and the LSB stego image is not as high as that of the DCT stego image. The pixel difference values proposed in this paper are very close to the histogram of stego image and the LSB of stego image, mutual interference during verification, resulting in reduced accuracy of both models. In the future, it is possible to analyze and improve the histogram and LSB collection and improve the overall accuracy.

Reference

[1] Z. Ni, Y. Q. Shi, N. Ansari, W. Su, Reversible data hiding, IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 3, (2006) 354-356.

[2] P. Wang, Z. Wei, L. Xiao, Pure spatial rich model features for digital image steganalysis, Multimedia Tools and Applications, vol. 75, no. 5 (2016) 2897-2912

[3] V. Holub, J. Fridrich, Random Projections of Residuals for Digital Image Steganalysis, IEEE Transactions on Information Forensics and Security, vol. 8, no. 12, (2013) 1996-2006

[4] C. C. Lin, P. F. Shiu, DCT-based Reversible Data Hiding Scheme, ICUIMC '09 Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication, vol. 5, no. 2, (2009) 327-335

[5] F. Huang, X. Qu, H. J. Kim, J. Huang, Reversible Data Hiding in JPEG Images, IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 9, (2016) 1610 – 1621

[6] Z. H. Wang, C. F. Lee, C. Y. Chang, Histogram-shifting-imitated reversible data hiding, Journal of Systems and Software, vol. 86, no. 2, (2013) 315-323

[7] C. K. Chan, L. M. Cheng, Hiding data in images by simple LSB substitution, Pattern Recognition, vol. 37, no. 3, (2004), 469-474

[8] T. Pevny, P. Bas, J. Fridrich, Steganalysis by Subtractive Pixel Adjacency Matrix, IEEE Transactions on Information Forensics and Security, vol. 5, no. 2, (2010) 215-224

[9] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE, vol. 86, no. 11, (1998) 2278 - 2324.