Journal of Cyber Security
DOI:10.32604/jcs.2020.012275
Article

High Visual Quality Image Steganography Based on Encoder-Decoder Model

Yan Wang*, Zhangjie Fu and Xingming Sun

School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China
*Corresponding Author: Yan Wang. Email: wyoookk@163.com
Received: 23 June 2020; Accepted: 21 July 2020

Abstract: Nowadays, with the popularization of network technology, more and more people are concerned about the problem of cyber security. Steganography, a technique dedicated to protecting peoples' private data, has become a hot topic in the research field. However, there are still some problems in the current research. For example, the visual quality of dense images generated by some steganographic algorithms is not good enough; the security of the steganographic algorithm is not high enough, which makes it easy to be attacked by others. In this paper, we propose a novel high visual quality image steganographic neural network based on encoder-decoder model to solve these problems mentioned above. Firstly, we design a novel encoder module by applying the structure of U-Net++, which aims to achieve higher visual quality. Then, the steganalyzer is heuristically added into the model in order to improve the security. Finally, the network model is used to generate the stego images via adversarial training. Experimental results demonstrate that our proposed scheme can achieve better performance in terms of visual quality and security.

Keywords: Steganaography; visual quality; cyber security

1 Introduction

Cyber security is proposed based on the development of the Internet and the arrival of the network society to face the new challenges of information security. Steganography is a technique that prevents anyone other than the intended receiver from knowing the event of transmission or the content of the secret information, which is a hot issue of cyber security. Specifically, due to the network environment full of multimedia (such as audio, video and images), steganography allows secret information to be embedded in a digital medium without compromising the quality of its cover.

LSB is the most famous traditional image steganography algorithm, which embed the secret information into the least significant bit of the pixel value of the cover image. Although this steganographic algorithm has the least influence on the visual quality of the cover image. However, this method is easy to change the statistical characteristics of the image, which cannot be avoided by the detection of the steganalyzer. In order to improve the performance of the steganographic algorithms, the goal of steganography is to shift from guaranteeing visual quality to reduce the impact on statistical analysis. It is found that changing the texture and noise area in the image will bring great challenges to the steganalyzer. The steganography based on distortion minimization framework was emerged, which heuristically embeds secret information into the texture and noise regions of the image and slightly changes the statistical characteristics of cover images. With the development of deep learning, instead of relying on hand-crafted embedding algorithms, we can automatically embed secret information into the cover through a well-trained neural network.

Generally, there are several important indexes to measure the performance of the steganographic algorithm, such as visual quality, security and hiding capacity. The visual quality is the most important

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

and basic requirement of steganography, which means that people's visual system cannot detect the changes of the cover image easily. Security is usually the capability of steganographic algorithm to resist steganalyzer based on statistical characteristic. Besides, hiding capacity refers to the maximum capacity of secret information that can be embedded in the cover image under the condition of ensuring visual quality and security. Both the traditional image steganography or the steganography based on deep learning has its own advantages and disadvantages. Although traditional image steganography algorithms have high security, their performance is easily affected by the original cover and hiding capacity. Although steganography based on deep learning can maintain the statistical properties of images better, the visual quality of stego images is not good enough.

To address the above limitations, we propose a new steganographic neural network based on encoder-decoder model. In our proposed scheme, the secret message can be concealed into a color cover image, and the stego image generated by our model has high visual quality and security. Comparing with previous works, our work has the following contributions:

1. In order to improve the stego image's visual quality, we improve the encoder module by introducing the structure of U-Net++.
2. In order to guarantee the security of our scheme, the steganalyzer based on deep learning is added into the encoder-decoder model.
3. The experimental results show that our scheme is effective and efficient.

2 Related Work

In this section, we introduce the latest research results from steganography and steganalysis respectively.

2.1 The Steganography Based on GAN

Since the generative adversarial network (GAN) [1] was proposed in 2014, many researchers have combined steganography with GAN and made excellent achievements. Volkhonskiy et al. [2] firstly proposed steganographic model, which called SGAN (Steganographic GAN), in 2017. a steganalyzer discriminator was added on the basis of GAN, which was used for steganalysis of stego images generated by the generator, making the generated stego images resistant to steganalyzer. Hayes et al. [3] proposed the HayesGAN model in 2017, which could directly generate stego images via adversarial learning. The same year, Tang et al. [4] combined the generative adversarial network with adaptive steganographic distortion and proposed ASDL-GAN (Automatic Steganographic Distortion Learning Framework with GAN) to learn the distortion cost. In 2018, Yang et al. [5] improved the ASDL-GAN and used Tanh simulator as an activation function to replace TES (Ternary Embedding Simulator), and considered the SCA (Selection-Channel-Aware) in the design of discriminator, so that the designed algorithm can resist detection of steganalysis based on the SCA.

2.2 The Steganography Based on Encoder-Decoder Model

Encoder-Decoder model [6] is a common framework of deep learning. This model can abandon the professional knowledge in the field of information hiding to a certain extent. In 2017, Google Research [7] proposed the steganographic network model based on encoder-decoder model for the first time, which could put a full-size color image into another image of the same size. Atique et al. [8] also proposed a steganographic model based on encoder-decoder model in the same period. They can embed the gray image into another color image and restore the gray image with higher accuracy. In 2018, Zhu et al. [9] put forward a model called HiDDeN, which adds a noise layer on the basis of encoder-decoder model, so that stego images can still extract binary secret information with high precision even after geometric attacks such as JPEG compression, Gaussian blurring and pixel-wise dropout. In 2019, Zhang et al. [10] added a critic into the encoder-decoder model and designed the loss function from multiple perspectives. In this scheme, the hiding capacity is increased to 4.4 BPP on the basis of ensuring the visual quality. Steganography

algorithm based on Encoder-Decoder has a high visual quality, but its security and the ability of resisting geometric attacks are far from enough.

2.3 The Steganalysis Based on Deep Learning

In 2014, Tan et al. [11] first proposed a steganalysis method based on deep learning, which is referred to as TanNet. TanNet consists of three convolutional layers and a full connection layer. Qian Net was proposed in 2005, which uses uniform pooling to reduce information loss. In 2016, Xu et al. [12] proposed XuNet based on convolutional neural network and added a fixed high-pass filtering layer (KV core) to the front end of the network. XuNet is a milestone work. Then, Xu et al. [13] also proposed an overlapped pooling to solve the problem of excessive information loss in the traditional pooling process, and utilized the integrated learning method of convolutional neural network to improve its detection capability. In the same year, Pibre et al. [14] proposed the use of large-size convolution kernel in convolutional neural networks. YNY Net [15], proposed in 2017, combines the selection-channel-aware with convolutional neural network to improve the detection accuracy, marking a major breakthrough in deep learning in the field of steganalysis. In 2018, Wu et al. [16] proposed the use of deep residual network to construct steganalysis model to avoid the diffusion of gradients that is prone to occur in deep network. Besides, Fridrich et al. [17] proposed a steganalysis model of deep residual network with the latest detection accuracy in 2019.

3 Architecture

The overall architecture of our proposed scheme is shown in Fig. 1. Our proposed architecture comprises the following three modules. Firstly, the encoder receives a cover image and a data tensor converted from secret message, and outputs a stego image of the same size as the cover image. Secondly, the steganalyzer that distinguish the difference between the cover and stego images. Finally, the stego image is taken as the input by the decoder, whose job is to recover the data tensor.

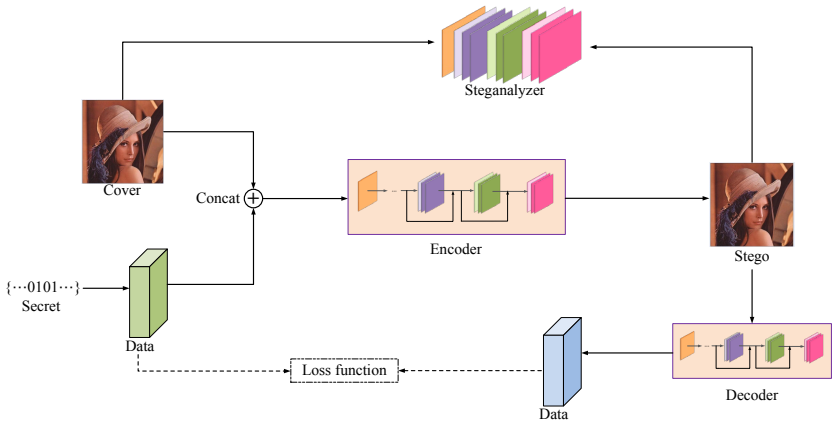

Figure 1: The overall framework of the proposed scheme

Encoder Based on the basic model, we have improved the encoder module by introducing U-Net++. Compared to the original U-Net network, it connects all the 1–4 layers of U-Net. The advantage of this structure is that it allows the network to learn features of different depths, regardless of which depth is effective. The second advantage is that it shares a feature extractor, trains only one encoder instead of a few blocks of U-Net, and its different levels of features are restored by different decoder paths. U-net ++ can capture features of different levels and integrate them through feature superposition. The edge information of large objects and small objects themselves are easily lost by the deep network sampling down and sampling up again and again, at this time, different size of receptive fields are needed to extract features. The architecture of our proposed encoder is similar with the U-Net++ network structure. Given k bits of secret message and an image of size (m, n), the secret information is converted into a data tensor of

shape (k, m, n). Then the encoding network adopts a structure similar to U-Net++ to improve the precision of shallow feature and deep feature extraction. At the same time, the structure of "Convolution-Activation Function-BN" is used to speed up network training.

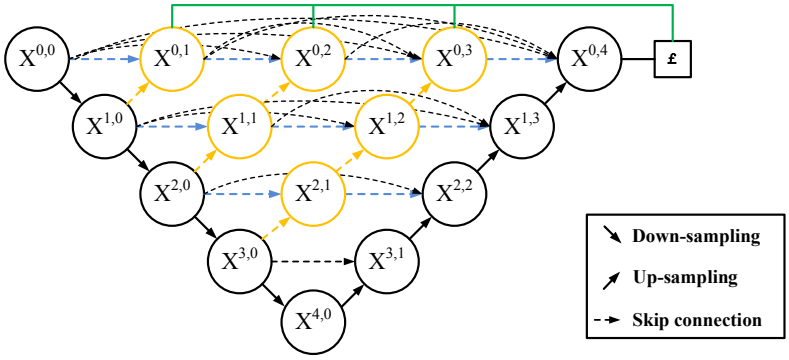

Figure 2: The structure of original U-Net++

Decoder The decoder takes the stego image as input and outputs the reconstructed data tensor. Several convolutional layers are applied to produce the feature maps. After the convolution operation, the global spatial average pool is applied to generate the data tensor with the same shape as the secret. Finally, the reconstructed message is produced by the single linear layer.

4 Evaluation Metrics

This section describes some of the metrics that evaluate the performance of the proposed model. These metrics will measure the performance of the model in terms of visual quality, security and hiding capacity.

SSIM (Structural Similarity Index) SSIM is a measure of the similarity of two images, which is based on three comparative measurements between samples x and y: luminance, contrast, and structure.

$$l(x,y)=\frac{2\mu_x\mu_y+c_1}{\mu_x^2+\mu_y^2+c_1} \tag{1}$$

$$c(x,y)=\frac{2\sigma_x\sigma_y+c_2}{\sigma_x^2+\sigma_y^2+c_2} \tag{2}$$

$$s(x,y)=\frac{2\sigma_{xy}+c_2}{2\sigma_x\sigma_y+c_2} \tag{3}$$

$$SSIM(x,y)=[l(x,y)\cdot c(x,y)\cdot s(x,y)] \tag{4}$$

Among them, the representations of the symbols are shown in Tab. 1.

Table 1: The representations of the symbols

Symbol	Representation
μ_x	The mean of the image x
μ_y	The mean of the image y
σ_x^2	The standard deviation of the image x
σ_y^2	The standard deviation of the image y
σ_{xy}	The covariance of the image x and y
c_1, c_2	$(k_1L)^2$, $(k_2L)^2$, in addition to avoid zero

PSNR (Peak Signal to Noise Ratio) PSNR is the most common and widely used objective measure of image quality., which is often used for the objective evaluation of image degradation before and after compression. Between the two images, the larger the PSNR value is, the more it tends to have no deterioration. In other words, the worse the image, the value is close to zero. Given two images of size (m, n), the PSNR is computed by the mean squared error (MSE). Formula is as follows:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-K(i,j)]^2 \tag{5}$$

$$PSNR = 10\cdot\log_{10}(\frac{MAX_I^2}{MSE}) \tag{6}$$

BPP (Bits Per Pixel) BPP is the number of bits of secret information embedded per pixel. In the field of information hiding, BPP is often used to measure the information capacity actually embedded into the cover.

$$bpp = \frac{bits}{m\times n} \tag{7}$$

5 Experiment and Analysis

5.1 Experimental Setups and Dataset

The experiments are conducted on the hardware environment consisted of a GeForce RTX TiTan GPU and an Intel i9 CPU. Our proposed model is trained on the COCO dataset, and the images are cut to experiment-specific size. There are 10,000 images for training and 1,000 images for testing. The batchsize of proposed model is set to 24 and the train epoch is 100. We use Adam optimizer with a learning rate of 10^{-3}, which is descended after 15 epochs. The loss curve of our proposed model during the training evolution is shown on Fig. 3. Besides, Figs. 4 and 5 show the experimental results of our proposed model.

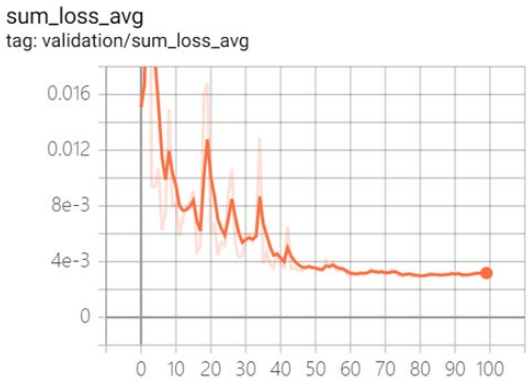

Figure 3: The loss curve of our proposed model during the training evolution

Figure 4: The results generated by our scheme in the middle of training. The first line represents the cover images, and the second line represents the stego images

Figure 5: The results generated by our scheme after the stable training. The first line represents the cover images, and the second line represents the stego images

5.2 Experimental Results and Analysis

In one hand, SSIM is a number between 0 and 1. The larger the SSIM, the smaller the difference between the two images. In the other hand, PSNR value range: 20–40. The higher the value, the closer the image quality is to the original image. We tested 200 pairs of images for SSIM and PSNR. The results of the structural similarity index (SSIM) and the peak signal to noise ratio (PSNR) are shown in Tab. 2. From the table we can see that the average SSIM of the cover images and our generated stego images is around 0.9 and the average PSNR is around 35. These results show that our images have higher visual quality. In addition, different embedding capacities have been applied in our scheme. We find that with the increase of embedding capacity, the image quality begins to decrease. However, due to the error of neural network, the secret information cannot be recovered completely, which does not affect the transmission of the message content.

Table 2: The average value of PSNR and SSIM

Scheme	PSNR (db)	SSIM
Our model	35.3	0.92

6 Conclusion and Future Work

The paper proposes a novel steganographic framework based on the encoder-decoder model. We improve the encoder module by introducing the structure of U-Net++ and introduce the steganalyzer based on deep learning into the encoder-decoder model. The experiment of SSIM and PSNR shows that the visual quality of the proposed achieves better performance.

However, the security has not been proven to be better than other algorithms and steganalysis experiments will be conducted in the future. At the same time, the security of our proposed steganographic framework still can be improved in some ways. For instance, there are many steganalyzers based on deep learning that we can use, such as Xu-Net, Ye-Net and SRNet and so on. In addition, the module of decoder can attempt to adopt different network structures. All in all, the encoder-decoder model holds great promise.

Funding Statement: This work is supported by the National Natural Science Foundation of China under Grant Nos. U1836110, U1836208.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley *et al.,* "Generative adversarial nets," in *Proc. the 27th Annual Conf. on Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, pp. 2672–2680, 2014.

[2] D. Volkhonskiy, I. Nazarov and E. Burnaev, "Steganographic generative adversarial networks," arXiv preprint arXiv: 1703.05502, 2017. https://arxiv.org/abs/1703.05502.

[3] J. Hayes and G. Danezis, "Generating steganographic images via adversarial training," in *Proc. the 29th Annual Conf. on Neural Information Processing Systems*, Long Beach, California, USA, pp. 1954–1963, 2017.

[4] W. Tang, S. Tan, B. Li and J. Huang, "Automatic steganographic distortion learning using a generative adversarial network," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1547–1551, 2017.

[5] J. Yang, K. Liu, X. Kang, E. Wong and Y. Shi, "Spatial image steganography based on generative adversarial network," arXiv preprint arXiv: 1804.07939, 2018. https://arxiv.org/abs/1804.07939.

[6] I. Sutskever, O. Vinyals and Q. Le, "Sequence to sequence learning with neural networks," in *Proc. the 27th Annual Conf. on Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, pp. 3104–3112, 2014.

[7] S. Baluja, "Hiding images in plain sight: Deep steganography," in *Proc. the 29th Annual Conf. on Neural Information Processing Systems*, Long Beach, California, USA, pp. 2069–2079, 2017.

[8] R. Rahim and S. Nadeem, "End-to-end trained CNN encoder-decoder networks for image steganography," in *Proc. the 15th European Conf. on Computer Vision*, Munich, Munich, Germany, pp. 723–729, 2018.

[9] J. Zhu, R. Kaplan, J. Johnson and F. Li, "Hidden: Hiding data with deep networks." in *Proc. the 15th European Conf. on Computer Vision*, Munich, Munich, Germany, pp. 657–672, 2018.

[10] K. Zhang, A. Cuesta-Infante, L. Xu and K. Veerama-chaneni, "SteganoGAN: high capacity image steganography with gans," arXiv: 1901.03892, 2019. https://arxiv.org/abs/1901.03892.

[11] S. Tan and B. Li, "Stacked convolutional auto-encoders for steganalysis of digital images," in *Proc. Signal and Information Processing Association Annual Summit and Conference*, Chiang Mai, Thailand, pp. 1–4, 2014.

[12] Y. Qian, J. Dong, W. Wang and T. Tan, "Learning representations for steganalysis from regularized cnn model with auxiliary tasks," in *Proc. the Int. Conf. on Communications, Signal Processing, and Systems*, Berlin, Heidelberg, pp. 629–637, 2016.

[13] G. Xu, H. Z. Wu and Y. Shi, "Ensemble of CNNs for steganalysis: An empirical study," in *Proc. the 4th ACM Workshop on Information Hiding and Multimedia Security*, Vigo, Galicia, Spain, pp. 103–107, 2016.

[14] M. Salomon, R. Couturier, C. Guyeux, J. Couchot and J. Bahi, "Steganalysis via a convolutional neural network using large convolution filters for embedding process with same stego key: A deep learning approach for telemedicine," *European Research in Telemedicine/La Recherche Européenne en Télémédecine*, vol. 6 no. 2, pp. 79–92, 2017.

[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] S. Wu, S. Zhong and Y. Liu, "Deep residual learning for image steganalysis," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 10437–10453, 2018.

[17] M. Boroumand, C. Mo and F. Jessica, "Deep residual network for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, 2018.