 CMC, vol.55, no.1, pp.1-16, 2018

A Fusion Steganographic Algorithm Based on Faster R-CNN

Ruohan Meng[1, 2], Steven G. Rice[3], Jin Wang[4] and Xingming Sun[1, 2, *]

Abstract: The aim of information hiding is to embed the secret message in a normal cover media such as image, video, voice or text, and then the secret message is transmitted through the transmission of the cover media. The secret message should not be damaged on the process of the cover media. In order to ensure the invisibility of secret message, complex texture objects should be chosen for embedding information. In this paper, an approach which corresponds multiple steganographic algorithms to complex texture objects was presented for hiding secret message. Firstly, complex texture regions are selected based on a kind of objects detection algorithm. Secondly, three different steganographic methods were used to hide secret message into the selected block region. Experimental results show that the approach enhances the security and robustness.

Keywords: Faster R-CNN, fusion steganography, object detection, CNNs, information hiding.

1 Introduction

With the rapid development of the internet and information technology, how to protect the transmission of important information has gained a lot of attentions. Information hiding is an approach of putting the secret information into a carrier (such as digital images) from sender to receiver. The receiver extracts and restores the original embedded secret information through a specific method. The carrier that carrying secret information is called cover, which has certainly significance in itself. For example, it can be an image, a document, etc. The carrier adds secret information called stego. In the ideal situation, stego does not arouse suspicion by attacker in the dissemination process. According to the different use of information hiding, it is divided into steganography and digital watermarking, the former is mainly used for the transmission of secret information, and the latter is mainly used for the protection of intellectual property. According to the technology of information hiding, it can be divided into a variety of steganographic modes in spatial domain, transform domain and compressed domain. In the early days, the

[1] School of Computer and Software, Nanjing University of Information Science and Technology, Ning Liu Road, No. 219, Nanjing, 210044, China.

[2] Jiangsu Engineering Centre of Network Monitoring, Ning Liu Road, No. 219, Nanjing, 210044, China.

[3] Department of Mathematics and Computer Science, Northeastern State University Tahlequah, OK 74464, USA.

[4] School of Computer & Communication Engineering, Changsha University of Science & Technology, Changsha, 410114, China.

[*] Corresponding author: Xingming Sun. Email: sunnudt@163.com.

simplest and most representative method of information hiding in space domain is to hide the information by using the least significant bit (LSB) of the image or all bit algorithms of multiple bit planes. However, these algorithms are not robust enough to keep statistical characteristics, so that attackers can accurately estimate the embedding length according to the statistical detection methods. The typical detection algorithms are RS (regular and singular) groups' method [Fridrich and Goljan (2002)], SPA (sample pair analysis) method [Dumitrescu, Wu and Wang (2003)] and JPEG compatibility analysis [Fridrich, Goljan and Du (2001)] and so on. With the continuous development of steganography, the newly proposed steganography algorithm can maintain more complex image statistical features. For example, HUGO [Pevný, Filler and Bas (2010)], WOW [Holub and Fridrich (2012)], SUNIWARD [Holub, Fridrich and Denemark (2014)] and other content adaptive steganographic algorithms proposed in recent years can automatically embed the secret information into the rich noisy texture area on the cover image to maintain high-level statistics.

Deep learning is well known as a revolution in machine learning, especially in the field of computer vision. In traditional approaches of image feature extraction, the features of SIFT and BOW are usually used as representation of images. In recent years, the applications of these features extraction include image retrieval, image forensics and other privacy protection fields. Xia et al. [Xia, Zhu, Sun et al. (2018); Xia, Xiong, Vasilakos et al. (2017)] apply image retrieval to cloud computing. Zhou et al. [Zhou, Yang, Chen et al. (2016); Zhou, Wu, Huang et al. (2017); Zhou, Wu, Yang et al. (2017); Zhou, Wang, Wu et al. (2017); Cao, Zhou, Sun et al. (2018); Zhou, Mu and Wu (2018)] proposed to apply the traditional features of the image to coverless information hiding. Yuan et al. [Yuan, Li, Wu et al. (2017)] use CNN to detect fingerprint liveness. However, the appearance of image depth learning features makes the feature extraction more rapid and accurate. In the field of image classification, deep convolutional neural networks (DCNN) such as VGGNet [Simonyan and Zisserman (2014)], GoogLeNet [Szegedy, Liu, Jia et al. (2015)] and AlexNet [Krizhevsky, Sutskever and Hinton (2012)] are performance excellent. Gurusamy et al. [Gurusamy and Subramaniam (2017)] used a machine learning approach for MRI Brain Tumor Classification. Based on the achievements above, object detection has been rapidly developed that detects semantic objects of a class (such as a dog, vehicle, or person) in digital images and video. For texture-rich target images, the appearance features of object instances are mainly identified and detected by extracting stable and abundant feature points and corresponding feature descriptors. In deep learning files, Faster R-CNN [Ren, Girshick, Girshick et al. (2017)], R-FCN [Dai, Li, He et al. (2016)] and SSD [Liu, Anguelov, Erhan et al. (2016)] three object detection models that most widely used currently.

Aiming at information hiding, steganalysis approaches based on deep learning has appeared one after another. These methods have a very good test for the current steganography algorithm. For example, Ye et al. [Ye, Ni and Yi (2017)] proposed to improve the CNN model to detect hidden information has reached 99 % of the detection rate. At the same time, some researchers turned their attention to how to hide information based on deep learning, which makes it safer and more robust. Tang et al. [Tang, Tan, Li et al. (2017)] use generative adversarial network (GAN) to achieve end-to-end information hiding. Baluja [Baluja (2017)] use neural network to determine the embedding secret

information in the location of the image, train an encoder to embed information. Uchida et al. [Uchida, Nagai, Sakazawa et al. (2017)] embed watermark into the depth neural network model. Therefore, the combination of deep learning and information hiding has become the focus of this paper.

The existing methods usually adopt a hidden mode to hide the secret information or watermarking in the entire image [Xia, Wang, Zhang et al. (2016); Wang, Lian and Shi (2017); Chen, Zhou, Jeon et al. (2017)]. In order to enhance the security and complexity of information, it is straightforward to design a fusion hiding strategy that employs different steganography algorithms to hide information on different areas. In addition, the complexity of the image is closely related to human visual effects. The more complex the image, the more information it carries, but the visual information of people does not increase with the increase of image complexity. Using the effect of human visual redundancy, the information is hidden in the more complicated area of the image texture, so as to increase the robustness and anti-detectability of hiding process. Therefore, this work proposes an approach to detect the more complex area to hide the information by the method of object detection. We adopt the fusion method of multiple steganographic algorithms as multiple steganography Algorithms based on ROI (MSA_ROI) to hide the information.

2 Related works

2.1 Object detection

The aim of object detection is to find the location of all the targets and specify each target category on a given image or video. It is mainly divided into two tasks, target positioning and target category detection. In traditional methods, the object detection is mainly use sliding window framework. The common algorithm is DPM (Deformable Part Model) [Felzenszwalb, Girshick, Mcallester et al. (2010)]. In face detection, pedestrian detection and other tasks, DPM have achieved good results. But DPM is relatively complex that the detection speed is relatively slow. With the development of deep learning, object detection has entered a new era. Object detection methods related to deep learning can be divided into two categories, one is based on the regional nomination, such as R-CNN (Region-based Convolutional Neural Networks) [Girshick, Donahue, Darrell et al. (2014); He, Zhang, Ren et al. (2015)], SPP-net [He, Zhang, Ren et al. (2015)], Fast R-CNN [Girshick (2015)], Faster R-CNN, etc. The other is end-to-end approach, Such as YOLO [Redmon, Divvala, Girshick et al. (2016)] and SSD. At present, Faster R-CNN model is a typical object detection model based on deep learning. From R-CNN, Fast R-CNN to Faster R-CNN which used in this paper, the four basic steps of object detection (region proposal, feature extraction, classification and rectangles refine regression) are finally unified into a deep network framework to achieve end-to-end object detection.

2.2 Spatial domain steganography

The current steganography methods are mainly divided into two types: spatial domain and transform domain. In transform domain, main application is JPEG. JPEG is a type of image compression method that divides an image into several matrices, performs a discrete cosine variation on each matrix and transforms the specific pixel values on the image completely

into the frequency domain. Since human eye is sensitive to low frequency but insensitive to high frequency, all the high frequency information is eliminated so as to achieve the purpose of image compression. Hidden information will be embedded in the intermediate frequency area (high-frequency area of anti-attack, low-frequency changes in the region is easy to be perceived by the human eye), which will be embedded information dispersed throughout the image. In spatial domain, embedding information is changing pixel values directly. Adaptive steganography is to automatically select the carrier image which is not easily found by the attacker based on the content of the cover image features on the region of interest and embed secret information. Modifying pixels of the image causes less distortion to the image on the complexity of the rich texture area and the edge area. At the same time, it is difficult for the attacker to detect secret information.

The steganography that based on the principle of minimizing the embedded distortion cannot only ensure the minimum distortion rate but also achieves secret communication. Existing steganography methods that based on the principle of minimizing embedded distortion include: HUGO, WOW, S-UNIWARD, MVGG [Li, Wang, Huang et al. (2015)] and so on. The ultimate goal is the same: minimize the distortion function and embed it in noisy or complex textures region of the cover image.

2.2.1 S-UNIWARD algorithm

S-UNIWARD is a content adaptive steganography based on wavelet transform, which proposes a general distortion function independent of the embedded domain. The embedding distortion function of S-UNIWARD as a whole is:

$$D(X,Y) \triangleq \sum_{k=1}^{3}\sum_{u=1}^{n_1}\sum_{v=1}^{n_2}\frac{\left|W_{uv}^{(k)}(X)-W_{uv}^{(k)}(Y)\right|}{\sigma+\left|W_{uv}^{(k)}(X)\right|} \tag{1}$$

Here, $K^{(1)},K^{(2)},K^{(3)}$ represent horizontal, vertical, diagonal direction of the filter in three directions, calculated by $K^{(1)}=h\cdot g^T, K^{(2)}=g\cdot h^T, K^{(3)}=g\cdot g^T$, where h represents a one-dimensional wavelet decomposition low (high) pass filter. Parameter X represents the carrier image, the image size is $n_1\times n_2$, and parameter Y represents the image after embedding the message. Parameter $W_{uv}^{(k)}(X), W_{uv}^{(k)}(Y)$ represent the image of the carrier image and the encrypted image after wavelet transform. Formula (2) is used to calculate the wavelet coefficients of pixels in three directions of the original image.

$$W_{uv}^{(k)}=K^{(k)}*X_{uv}, (1\le u\le n_1, 1\le v\le n_2, 1\le k\le 3) \tag{2}$$

When the wavelet coefficients of pixels (such as texture regions) with complex content areas are changed, the distortion calculated by formula (1) will be small, indicating that the region is suitable for hiding information. However, when the pixel wavelet coefficients of the texture smoothing region are changed, the distortion will be very large, indicating that secret information should be avoided when embedding these pixels.

2.2.2 HUGO algorithm

HUGO is considered to be one of the most secure steganographic techniques. It defines a distortion function domain by assigning costs to pixels based on the effect of embedding some information within a pixel, the space of pixels is condensed into a feature space using

a weighted norm function. HUGO algorithm is the use of SPAM steganalysis feature design distortion cost function. It is considered to be one of the most secure steganographic techniques. According to the additive distortion function:

$$D(X,Y)=\sum_{i=1}^{n}\rho_i\left|x_i-y_i\right| \tag{3}$$

Here, the constant $0\le\rho_i\le\infty$ is a fixed parameter that represents the amount of distortion that results from a pixel change. When $\rho_i=\infty$, the pixel is the so-called wet pixel, and the wet pixel does not allow modification during embedding. The minimum expected distortion function is:

$$D_{\min}(m,n,p)=\sum_{i=1}^{n}p_i\rho_i \tag{4}$$

Where $p_i=\frac{e^{-\lambda\rho_i}}{1+e^{-\lambda\rho_i}}$ is the probability that the i-th pixel changes. Parameter $\rho=(\rho)_{i=1}^{n},(0\le\rho_i\le\infty)$ is the set of additive distortion metrics formula (3), where $i\in\{1,\ldots,n\}$. Parameter $m(0\le m\le n)$ is the number of bits to be passed when using binary embedding operations.

2.2.3 WOW algorithm

WOW (weight acquisition wavelet) is another method of steganography, which depending on the complexity of the region. It will be covered embedding information into an image. If one area of the image is more complex than the other, the pixel values in that area will be modified. WOW steganography algorithm mainly from the perspective design of the distortion function. The additive distortion function is:

$$D(X,Y)=\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\rho_{ij}(X,Y_{ij})\left|X_{ij}-Y_{ij}\right| \tag{5}$$

Where ρ_{ij} are the costs of changing pixel X_{ij} to Y_{ij}.

2.3 Quality assessment

The experimental results in this paper used the following indexes to evaluate the quality of stego images.

2.3.1 Mean Square Error (MSE)

The following expression is the Mean Square Error (MSE) between the images AI and BI. Supposing that the pixel value of the bearer image is $\{AI(i,j),0\le j\le N-1\}$ and the pixel value of the corresponding stego image is $\{BI(i,j),0\le i\le M-1,0\le j\le N-1\}$, the error image is

$\{e(i,j)=AI(i,j)-BI(i,j),0\le i\le M-1,0\le j\le N-1\}$, then the mean square error is expressed as:

$$MSE = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} e^2(i,j) = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left[AI(i,j) - BI(i,j)\right]^2 \tag{6}$$

Lower is the MSE, higher is the similarity between the images.

2.3.2 Peak Signal to Noise Ratio (PSNR)

The following expression is the Peak Signal to Noise Ratio (PSNR) between the images AI and BI . Setting $AI_{\max} = 2^k - 1$, where *K* represents a number of bits for all pixels, the PSNR is defined as:

$$PSNR = 101g\frac{AI_{\max}^2}{MSE} = 101g\left[\frac{AI_{\max}^2 MN}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left[AI(i,j) - BI(i,j)\right]^2}\right](dB) \tag{7}$$

In many video sequences and commercial image acquisition applications, often take k = 8. So, for an 8-bit binary image, $AI_{\max} = 255$, substituting formula (7) into:

$$PSNR = 101g\frac{255^2}{MSE} = 101g\left[\frac{255^2 MN}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left[AI(i,j) - BI(i,j)\right]^2}\right](dB) \tag{8}$$

Higher is the PSNR, higher is the similarity between the images.

2.3.3 Structural Similarity Index Measure (SSIM)

The Structure Similarity Index Measure (SSIM) between the images OI and WI which are of size M×N is given by the following expression. Given two images AI and BI , the structural similarity of the two images can be found in accordance with the following formula:

$$SSIM(AI, BI) = \frac{(2\mu_{AI}\mu_{BI} + c_1)(2\sigma_{AIBI} + c_2)}{(\mu_{AI}^2 + \mu_{BI}^2 + c_1)(\sigma_{AI}^2 + \sigma_{BI}^2 + c_2)} \tag{9}$$

Here, μ_{AI} a is the average of AI , μ_{BI} is the average of BI , σ_{AI}^2 is the variance of AI , σ_{BI}^2 is the variance of BI , and σ_{AIBI} is the covariance of AI and BI . $c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$ are constants used to maintain stability. L is the dynamic range of pixel values. $k_1 = 0.01, k_2 = 0.03$. Structural similarities range from -1 to 1. When two images are identical, the value of SSIM equals one.

3 Proposed method using multiple algorithms based on Faster R-CNN

This section discusses our steganographic scheme, the models we use and the information each party wishes to conceal or reveal. After laying this theoretical groundwork, we present experiments supporting our claims. The overall framework for this article is shown in Fig. 1. Firstly, the whole image is input into the CNN model, and then the feature is extracted. Then, the Proposal is generated through the RPN network, the Proposal is mapped to the last layer of convolution of CNN, and each proposal is made into a fixed-size feature maps

through Rol pooling layer. Finally, use Softmax classification and bounding box regression to get the part of the carrier we need. Next, the steganographic algorithms are matched for each of the obtained carrier parts to finally obtain stego images.

3.1 Extract object by Faster R-CNN

In the field of computer vision, object detection mainly solves two problems: the location of multiple objects on the image and the categories of each object. Faster R-CNN introduced the region proposal network (RPN) based on Fast R-CNN, replacing the slow search selective search algorithm. Region proposal uses information such as the texture, edge, and color in the image to find out the position where the target on the way may appear beforehand, and can guarantee a higher recall rate with fewer windows selected (a few hundred or even a few thousand). This greatly reduces the time complexity of follow-up operations, and obtains the candidate window than the sliding window of higher quality. In a sense, Faster R-CNN = RPN + Fast R-CNN. Taking into account the target detection is based on the image texture, edge and determine the target. This paper argues that objects selected from the Faster R-CNN are more conducive to hiding information than the background. Therefore, Faster R-CNN is used in the method proposed here. Faster R-CNN's network model is shown in Fig. 2. Faster R-CNN is mainly divided into four contents:

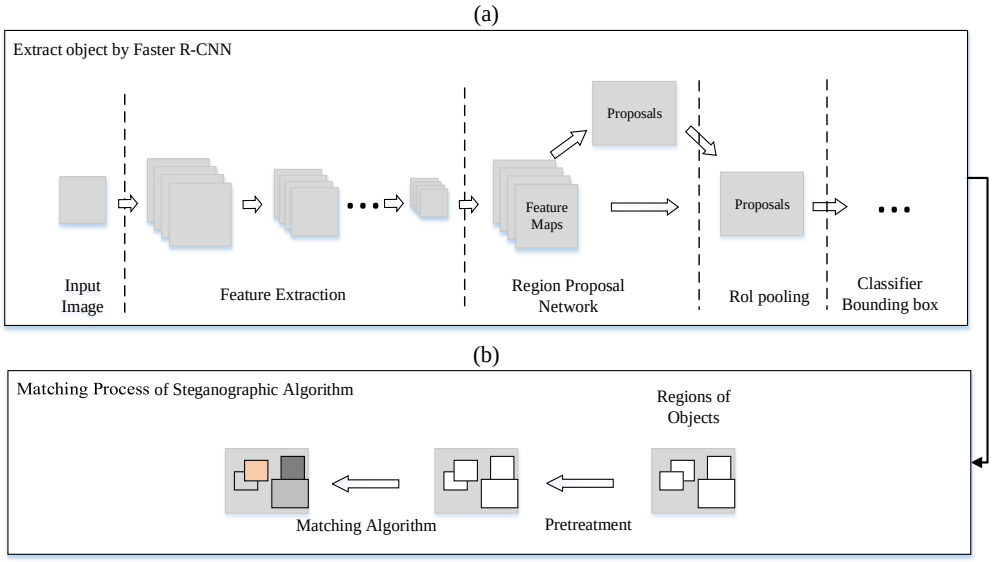


Figure 1: Proposed Steganographic architecture. (a) Target detection structure based on Faster R-CNN. (b) Steganography algorithm structure for the local area matching

Conv layers: Including the 13 *conv* layers +13 *relu* layers +4 *pooling* layers, used to extract image features maps. The feature maps are shared for subsequent RPN layers and full connectivity layers.

Region Proposal Networks: used to generate region proposals. The RPN network first passes a 3×3 convolutional layer, generating foreground anchors and bounding box

regression offsets, respectively, and then calculates proposals. Anchors belong to foreground or background via softmax function.

Roi Pooling: This layer according to feature maps and proposals to extract proposal feature maps, into the subsequent full connection layer.

Classification: Use proposal feature maps to calculate the type of proposal, and then use bounding box regression to get the final exact position of the test box.

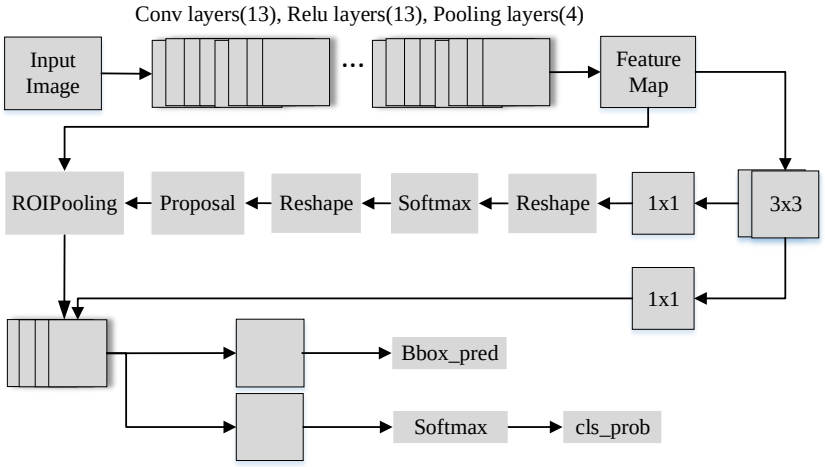


Figure 2: Network model of Faster R-CNN, including convolutional layers, region proposal network, RoI pooling and classification

The loss function of Faster R-CNN is:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{10}$$

p_i is the predicted probability of anchor i being an object. The ground-truth label:

$$p_i^* = \begin{Bmatrix} 0 & negative\,lable \\ 1 & positive\,lable \end{Bmatrix} \tag{11}$$

$t_i = \{t_x, t_y, t_w, t_h\}$ is a vector representing the 4 parameterized coordinates of the bounding box of the prediction. t_i^* is the coordinate vector of the ground truth bounding box corresponding to the positive anchor. $L_{cls}(t_i, t_i^*)$ is the logarithmic loss target and non-target:

$$L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1-p_i^*)(1-p_i)] \tag{12}$$

$L_{reg}(t_i, t_i^*)$ is the regression loss, calculated using $L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)$, where R is the smooth L1 function. $p_i^* L_{reg}$ means that only the foreground $anchor(p_i^* = 1)$ has a regression loss, and in other cases there is no $(p_i^* = 0)$. The outputs of *cls* and *reg* are composed of $\{p_i\}$ and $\{u_i\}$, respectively, then normalized by N_{cls} and N_{reg}.

3.2 Matching process of steganographic algorithm

From the first part, we get multiple texture complex regions. Next, we need to hide the information by matching the regions with different steganographic algorithms as shown in Fig. 3. First of all, the proposed method needs to judge whether there is overlap in the target area. If there are overlapping parts, the overlap part should be processed. Then the image should be grayscale. In the order stage, the hash algorithm is used to sort the target area in each cover image. Finally, the sorted target area and the steganography algorithm is sequentially matched to complete the concealment of the secret information.

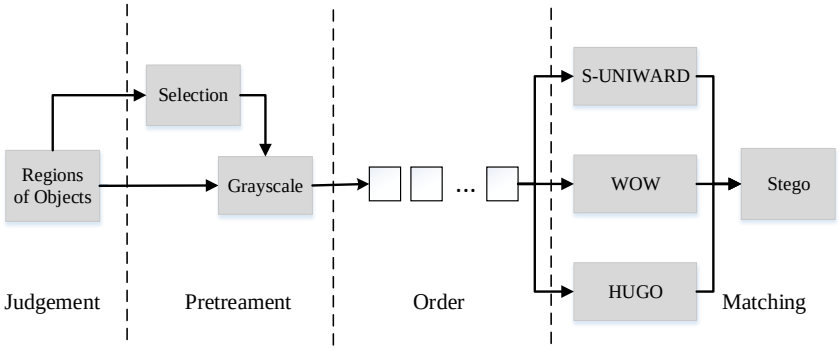
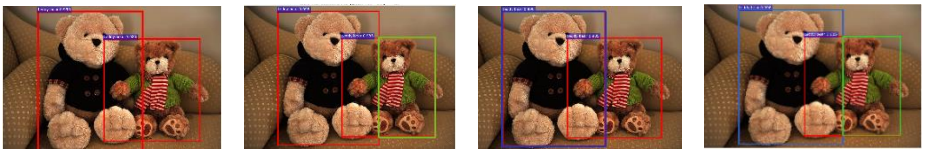


Figure 3: Flow chart of steganographic algorithm matching process

3.2.1 Selection of overlapped box

First of all, many of the target frames are overlapped after a target is detected by Faster R-CNN. Secondly, only one steganographic method can be used to hide the information in each target frame. Therefore, it needs to preprocess the overlapped box.

As shown in Fig. 4, we take the principle of maximum area, for overlapping target box, first calculate the area of overlapped box, with a large target box shall prevail, small target box by removing the remaining large area as a carrier.



(a) (b) (c) (d)

Figure 4: Preprocess for overlapping target box. (a) Target area obtained by Faster R-CNN. From the graph, we can see that there are overlapping areas in the red border. If the corresponding steganography algorithm is based on the box, it will cause the overlapping area to hide the information repeatedly. (b) Removing the overlapped area to get the green area. Due to the principle of choosing maximum area. (c), the result is shown in the purple area on the choice of maximum area. The blue and green areas in (d) are the areas of the final choice

3.2.2 Order of probability scores

After obtaining the grayscale regional target maps, we apply different steganographic algorithms to different regions for embedding information. Each target graph obtained by Faster R-CNN has a probability value of Softmax function. We sort different regions based on this value. As shown in Fig. 5, there are three boxes in the figure, from left to right named box 1, box 2, box 3. The probability scores of each box are 0.995, 0.968, and 0.994. According to descending order, the probability scores are 0.995, 0.994, and 0.0968. The sorted areas correspond to the boxes 1, 3, and 2, respectively numbered as 1, 2, and 3.

Figure 5: Target areas of the cover image. For different regions have different probability values, from left to right are 0.995, 0. 0968, 0.994

3.2.3 Match steganographic algorithms

After getting the order of probability scores, we use hash algorithm to match steganographic algorithms. The method is using division hash algorithm on the number, taking the remainder of 3. When the remainder is 0, S-UNIWARD algorithm is used. When the remainder is 1, WOW algorithm is used and when the remainder is 2, we use the HUGO algorithm. As shown in Fig. 6, the stego image is used by different matching steganographic algorithms to embed the information in different regions.

Figure 6: The stego image

4 Experiments

All the experiments GPU environment is NVIDIA GTX1080. The experiment training data set is COCO2014 dataset [Lin, Maire, Belongie et al. (2014)], and the target in the image is calibrated by exact segmentation. The image includes 91 categories of targets, 328,000 images and 2,500,000 labels. All experiments used the depth learning framework Caffe [Jia, Shelhamer, Jeff et al. (2014)]. The network used by the image feature extraction section when training Faster R-CNN is VGG16 [Simonyan and Zisserman (2014)] network.

4.1 Object extraction

When we extract the target area, we use the Faster R-CNN model on COCO2014 data set. When training, the learning rate (base_lr) is set to 0.001, gamma is set to 0.1, momentum is set to 0.9 and weight decay is set to 0.0005. After 200000 iterations, the network model file is obtained. When testing, we use this network model file to get the object detection area box of the image. Fig. 7 is the image obtained after training the test image through the Faster R-CNN model.

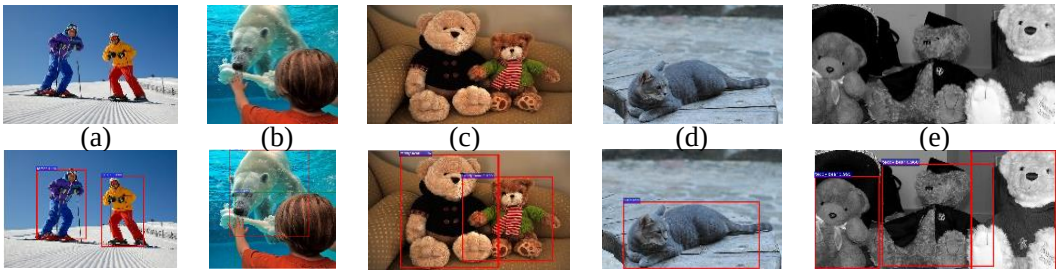


Figure 7: The first row is the five images from COCO2014 dataset used in testing process. The second row is the target area after the selection by Faster R-CNN

4.2 Steganographic process

After obtaining the coordinates of the object detection region of the image, three different steganographic algorithms are used to embed the secret information in the region to obtain the stego image. The three spatial steganographic algorithms are respectively S-UNIWARD, HUGO and WOW. According to the division hash, different regions correspond to different steganography algorithms. Fig. 8 shows the test results after steganography.

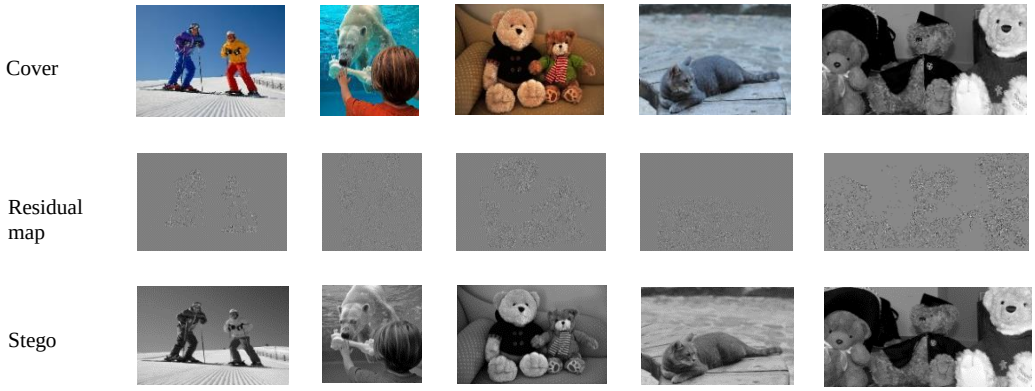


Figure 8: The first row shows the test images, the second row represents the residual images steganography by the proposed method, and the third row shows the stego images

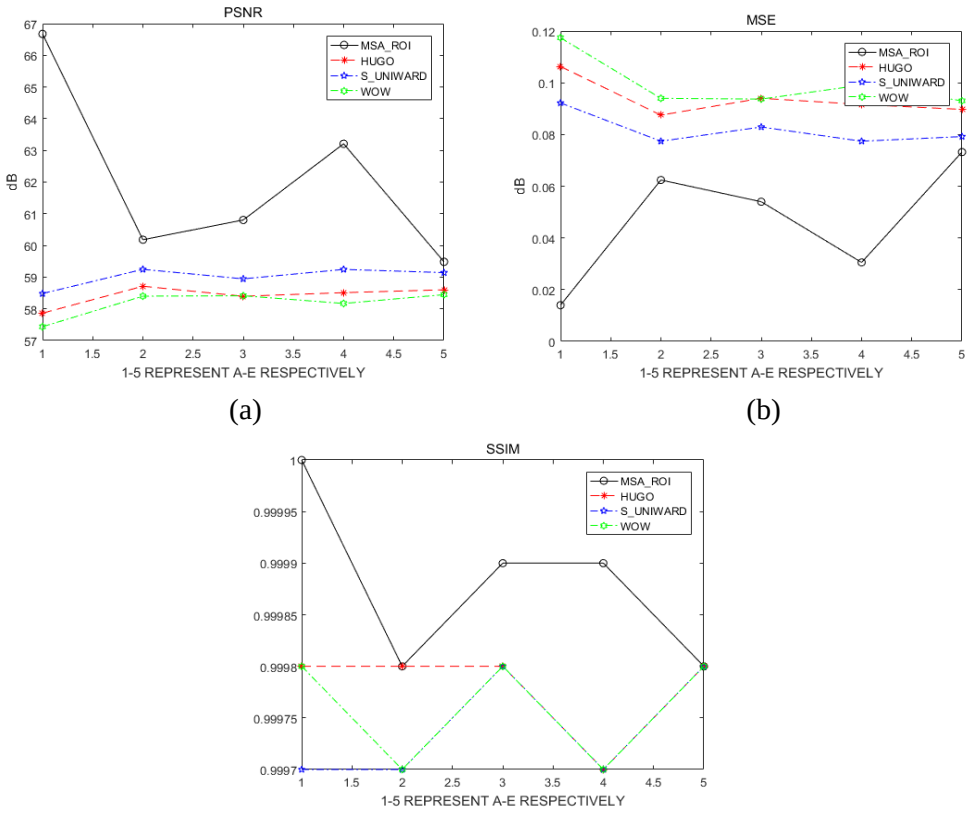


Figure 9: (a), (b) and (c) are line charts of the PSNR values, MSE values and SSIM values of the proposed method and the three traditional spatial algorithms, respectively

4.3 Analysis the quality of stego

In this section, three different image quality indexes are used to evaluate the quality of the stego images, which are MSE, PSRN, and SSIM. Through experiments, it is found that the proposed method is superior to HUGO, S-UNIWARD and WOW steganography in three indexes of PSNR, MSE and SSIM, indicating the least distortion of the method. As shown in Tab. 1, the PSNR value of SMSA_ROI is higher than that of HUGO, S-UNIWARD and WOW. The highest value is 66.6751db, which is higher than 9.24 db in WOW algorithms, indicating that the distortion of MSA_ROI method is the smallest. The maximum MSE value of the proposed method is 0.0732, which is lower than the other three methods, and the minimum can be as low as 0.0140. The last evaluation index SSIM, the four methods of SSIM value is not much difference, but MSA_ROI is still better than the other three algorithms. As shown in Fig. 9, the PSNR value of the proposed method is obviously higher than the three traditional spatial algorithms. The MSE value of the proposed method is obviously lower than the other three methods, and the lowest is 0.0140. The proposed method is higher than the three traditional spatial algorithms on the SSIM value, and the maximum is up to 1.0000.

Table 1: Comparison of the proposed algorithm and the single spatial algorithm in image quality

	PSNR	MSE
MSA_ROI_A	66.6751	0.0140
HUGO_A	57.8636	0.1063
S-UNIWARD _A	58.4827	0.0922
WOW_A	57.4344	0.1174
MSA_ROI_B	60.1815	0.0624
HUGO_B	58.7124	0.0875
S-UNIWARD _B	59.2457	0.0774
WOW_B	58.3996	0.0940
MSA_ROI_C	60.8032	0.0540
HUGO_C	58.3973	0.0940
S-UNIWARD _C	58.9462	0.0829
WOW_C	58.4119	0.0937
MSA_ROI_HUGO_D	63.0540	0.0322
MSA_ROI_ S-UNIWARD _D	63.5727	0.0268
MSA_ROI_WOW_D	62.9983	0.0326
HUGO_D	58.5103	0.0916
S-UNIWARD _D	59.2435	0.0774
WOW_D	58.1695	0.0991
MSA_ROI_E	59.4835	0.0732
HUGO_E	58.6009	0.0897
S-UNIWARD _E	59.1424	0.0792
WOW_E	58.4468	0.0930

5 Conclusion and future work

This paper has two main contributions. The first one is to combine the object detection method to select a complex texture region, which is suitable for hiding information. The second one integrates the existing multiple spatial steganography algorithms into a cover image. Experiments show that the proposed method is superior to the traditional spatial steganography algorithm. Future works to further move this research includes the following aspects. 1. Hide secret message in the foreground completely. 2. Switch to different object detection methods. 3. Adjust the steganography algorithm adaptively.

Acknowledgement: This work is supported, in part, by the National Natural Science Foundation of China under grant numbers U1536206, U1405254, 61772283, 61602253, 61672294, 61502242; in part, by the Jiangsu Basic Research Programs-Natural Science Foundation under grant numbers BK20150925 and BK20151530; in part, by the Priority

Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund; in part, by the Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET) fund, China.

References

Baluja, S. (2017): Hiding images in plain sight: Deep steganography. *Advances in Neural Information Processing Systems*, pp. 2066-2076.

Cao, Y.; Zhou, Z.; Sun, X.; Gao, C. (2018): Coverless information hiding based on the molecular structure images of material. *Computers, Materials & Continua*, vol. 54, no. 2, pp. 197-207.

Chen, B.; Zhou, C.; Jeon, B.; Zheng, Y.; Wang, J. (2017): Quaternion discrete fractional random transform for color image adaptive watermarking. *Multimedia Tools and Application*.

Dai, J.; Li, Y.; He, K.; Sun, J. (2016): R-FCN: Object detection via region-based fully convolutional networks. *Advances in Neural Information Processing Systems*, pp. 379-387.

Dumitrescu, S.; Wu, X.; Wang, Z. (2003): Detection of LSB steganography via sample pair analysis. *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1995-2007.

Felzenszwalb, P. F.; Girshick, R. B.; Mcallester, D.; Ramanan, D. (2014): Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 47, no. 2, pp. 6-7.

Fridrich, J.; Goljan, M. (2012): Practical steganalysis of digital images: state of the art. *Security and Watermarking of Multimedia Contents IV*, vol. 4675, no. 1, pp. 1-13.

Fridrich, J.; Goljan, M.; Du, R. (2001): Steganalysis based on JPEG compatibility. *Proc Spie*, vol. 4518, pp. 275-280.

Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. (2014): Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580-587.

Girshick, R. (2015): Fast R-CNN. *IEEE International Conference on Computer Vision*, pp. 1440-1448.

Gurusamy, R.; Subramaniam, V. (2017): A machine learning approach for MRI brain tumor classification. *Computers, Materials & Continua*, vol. 53, no. 2, pp. 91-108.

He, K.; Zhang, X.; Ren, S.; Sun, J. (2015): Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 9, pp. 1904-1916.

Holub, V.; Fridrich, J. (2012): Designing steganographic distortion using directional filters. *IEEE International Workshop on Information Forensics and Security*, vol. 2, no. 4, pp. 234-239.

Holub, V.; Fridrich, J.; Denemark, T. (2014): Universal distortion function for steganography in an arbitrary domain. *Eurasip Journal on Information Security*, vol. 2014, no. 1, pp. 1.

Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J. et al. (2014): Caffe: Convolutional architecture for fast feature embedding. *ACM International Conference on Multimedia*, pp. 675-678.

Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2012): Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, vol. 60, no. 2, pp. 1097-1105.

Li, B.; Wang, M.; Huang, J.; Li, X. (2015): A new cost function for spatial image steganography. *IEEE International Conference on Image Processing*, pp. 4206-4210.

Lin, T. Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P. et al. (2014): Microsoft COCO: Common objects in context. *European Conference on Computer Vision*, vol. 8693, pp. 740-755.

Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S. et al. (2016): SSD: Single shot multiBox detector. *European Conference on Computer Vision*, pp. 21-37.

Pevný, T.; Filler, T.; Bas, P. (2010): Using high-dimensional image models to perform highly undetectable steganography. *Lecture Notes in Computer Science*, vol. 6387, pp. 161-177.

Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. (2016): You only look once: unified, real-time object detection. *IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society*, pp. 779-788.

Ren, S.; Girshick, R.; Girshick, R.; Sun, J. (2017): Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149.

Simonyan, K.; Zisserman, A. (2014): Very deep convolutional networks for large-scale image recognition. *Computer Science*.

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S. et al. (2015): Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9.

Tang, W.; Tan, S.; Li, B.; Huang, J. (2017): Automatic steganographic distortion learning using a generative adversarial network. *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1547-1551.

Uchida, Y.; Nagai, Y.; Sakazawa, S.; Satoh, S. I. (2017): Embedding watermarks into deep neural networks. *ACM on International Conference on Multimedia Retrieval*, pp. 269-277.

Wang, J.; Lian, S.; Shi, Y. (2017): Hybrid multiplicative multi-watermarking in DWT domain. *Multidimensional Systems and Signal Processing*, vol. 28, no. 2, pp. 617-636.

Xia, Z.; Wang, X. H.; Zhang, L. G.; Qin, Z.; Sun, X. et al. (2016): A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2594-2608.

Ye, J.; Ni, J.; Yi, Y. (2017): Deep learning hierarchical representations for image steganalysis. *IEEE Transactions on Information Forensics & Security*, vol. 12, no. 11, pp. 2545-2557.

Yuan, C.; Li, X.; Wu, Q.; Li, J.; Sun, X. (2017): Fingerprint liveness detection from different fingerprint materials using convolutional neural network and principal component analysis. *Computers, Materials & Continua*, vol. 53, no. 4, pp. 357-371.

Zhou, Z.; Mu, Y.; Wu, Q. (2018): Coverless image steganography using partial-duplicate image retrieval. *Soft Computing*.

Zhou, Z.; Wang, Y.; Wu, Q.; Yang, C.; Sun, X. (2017): Effective and efficient global context verification for image copy detection. *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 48-63.

Zhou, Z.; Wu, Q.; Huang, F.; Sun, X. (2017): Fast and accurate near-duplicate image elimination for visual sensor networks. *International Journal of Distributed Sensor Networks*, vol. 13, no. 2, pp. 155014771769417.

Zhou, Z.; Wu, Q.; Yang, C.; Sun, X.; Pan, Z. (2017): Coverless image steganography based on histograms of oriented gradients-based hashing algorithm. *Journal of Internet Technology*, vol. 18, no. 5, pp. 1177-1184.

Zhou, Z.; Yang, C.; Chen, B.; Sun, X.; Liu, Q. et al. (2016): Effective and efficient image copy detection with resistance to arbitrary rotation. *IEICE Transactions on Information and Systems*, vol. E99-D, no. 6, pp. 1531-1540.