EURASIP Journal on Image and Video Processing

RESEARCH Open Access

A video coverless information hiding algorithm based on semantic segmentation

Nan Pan, Jiaohua Qin*, Yun Tan, Xuyu Xiang and Guimin Hou

*Correspondence:
qinjiaohua@163.com
College of Computer Science and Information Technology, Central South University of Forestry & Technology, South Shaoshan Road, Changsha, 410004, China

Abstract

Due to the fact that coverless information hiding can effectively resist the detection of steganalysis tools, it has attracted more attention in the field of information hiding. At present, most coverless information hiding schemes select text and image as transmission carriers, while there are few studies on emerging popular media such as video, which has more abundant contents. Taking the natural video as the carrier is more secure and can avoid the attention of attackers. In this paper, we propose a coverless video steganography algorithm based on semantic segmentation. Specifically, to establish the mapping relationship between secret information and video files effectively, this paper introduces the deep learning based on semantic segmentation network to calculate the statistical histogram of semantic information. To quickly index the sender's secret message to the corresponding video frame, we build a three-digit index structure. The receiver can extract the valid video frame from the three-digit index information and restore the secret information. On the one hand, the neural network is trained through the original image and the noisy image in this scheme; therefore, it can not only effectively resist the interference of noises, but also accurately extract the robust deep features of the image. The frames of video generate the robust mapping to the secret information after the semantic information statistics. On the other hand, semantic segmentation belongs to pixel-level segmentation, which has high requirements for network parameters, so it is difficult for attackers to decrypt and recover secret information. Since this scheme does not modify the primitiveness of video data, it can effectively resist steganalysis tools. The experimental results and analysis show that the video coverless information hiding scheme has a large capacity and a certain resistance to noise attack.

Keywords: Semantic segmentation, Coverless, Information hidden

1 Introduction

In recent years, with the proliferation of portable multimedia devices , the application of 4G network and a large number of videos with artistic expression have been spread at an amazing speed, involving various fields such as television circle, social production, personal life, and network broadcast. Meanwhile, the video generates a variety of security and privacy issues, which are as follows:

1. *Privacy security.* In the era of information and data flooding, personal privacy and industrial copyright are valued by the public, which is embodied in reducing the risk of

information disclosure and cracking to prevent the theft, abusement, and infringement of video files.

2. *Verify integrity.* In the process of long-distance transmission and storage of video, video files are vulnerable to compression noise or other attacks, resulting in the loss or deletion of video parts. Ensuring that video files are complete means that the integrity of the video can be judged.

Information hiding is an effective way to meet the above requirements. Early secure communication technologies are based on cryptography, such as DES (data encryption standard) and RSA (public-key encryption algorithm). The model is shown in Fig. 1.

The sender uses the encryption algorithm to encrypt the output confidential information, and the receiver decrypts the received ciphertext through a corresponding key to obtain the information. However, the secret information encryption method is "graffiti" encryption with unreadable cipher, which explicitly told the attacker what information is important information and aroused the suspicions of the attacker. Besides, with the continuous improvement of computer performance and the ability to process large amounts of data, ciphertext information is more likely to be cracked. Once the ciphertext message is deciphered, the information security will suffer a devastating attack.

For the problems discussed above, information hiding technology has been put on the agenda.

Compared with the traditional secret science technology, information hiding technology adopted, the encryption model is difficult for attackers to find the important information. Information hiding technology [1, 4] is using the redundancy of the carrier signal itself and human visual sensitivity to embed the secret information vehicle the information existing in numerous other media information, which makes it hard for the attacker to find the target to attack and decipher. Based on the redundancy of the carrier, the carrier is modified to some extent. However, under the background of big data and cloud computing, especially the general analysis and blind information hiding analysis method, the traditional information hiding method is also effective against some types of steganalysis. Once an attacker detects the presence of private information, even though the attacker cannot decrypt the hidden information, he can attack these carriers. The rea-

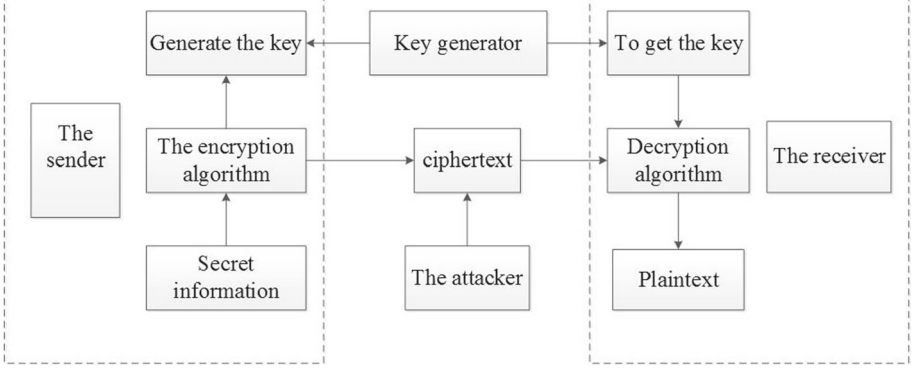

Fig. 1 Early information hiding algorithm model

son why the traditional hiding algorithm cannot resist the steganalysis is the modification caused in the process of hiding.

In order to fundamentally resist steganalysis, the coverless information hiding scheme proposed by Sun and other scholars does not mean that it does not require the carrier. It directly establishes the mapping relationship between secret information and hidden carrier according to the characteristics of the carrier rather than modifying the carrier. Since coverless information hiding technology does not modify carrier information, an attacker cannot obtain the information even if he may get the original carrier that includes the secret information. Therefore, coverless information hiding technology has a unique anti-steganalysis ability. Both theoretical research and the maturity of coverless still have a long way to go; it is still a relatively new field with potential value.

In the early stage, some scholars proposed a zero-watermark algorithm similar to the coverless hiding information. Huang et al. [5] proposed a VQ-Based robust multi-watermarking algorithm. In order to send secret information, the sender converted the secret information into a binary hash sequence and divided it into several equal-length segments. For each segment, search for it in the inverted index structure to find the image with the same hash sequence of that segment. A series of stego-images related to images is collected and transmitted. For the receiver, the hash sequences of these received images are generated by the same hash algorithm. The existing coverless information hiding schemes are mainly based on text and image, few relevant literature are based on video. Zhou et al. proposed steganography of coverless information hiding based on gray images [6]. Luo et al. [7] proposed a coverless real-time image information hiding algorithm based on image block matching and dense convolution network. Yi et al. [8] proposed a coverless information hiding algorithm based on Web text. Zhang et al. [9] proposed a robust coverless image steganography based on DCT and LDA topic classification scheme. Liu [10] proposed a coverless steganography based on image retrieval of DenseNet features and DWT sequence mapping. Ruan et al. [11] proposed a GIF-based method for information hiding without carrier. This method quantifies each GIF image in the existing carrier image library and extracts the attribute value of its extension to hide secret information. Zhou et al. [12] proposed a coverless information hiding method based on hog hashing, which is generated by using a hog-based hash algorithm. Zheng et al. proposed a coverless information hiding method based on robust image hash [13]. Duan et al. [14] proposed a coverless information hiding method based on the generation model.

At present, coverless information hiding algorithm about videos has been proposed, and some researchers have proposed some zero-watermark algorithms for video copyright protection. This technology constructs watermark by extracting video features without modifying any video data. Jiang et al. proposed an improved pseudo 3D-video zero DCT domain watermarking algorithm [15]. The Euclidean distance between frame method is adopted to select key frames, and the three-frame difference method is used to get the moving target. Bu et al. proposed a video zero-watermark algorithm [16] based on contourlet transformation. This algorithm made contourlet changes for each frame in the original video, then took its low-frequency coefficient, and calculated the coefficient to construct zero-watermark. It is worth mentioning that digital watermarking should focus on avoiding data modification and deletion, improving the robustness of various attacks. These schemes cannot meet the requirements of resisting steganalysis.

Video not only has rich semantic features such as image texture, shape, and color, but also has continuity in time and space. Semantic segmentation is an important method for the content analysis of video images. In this work, we study the semantic information of video image and map it with hidden information to realize the information hiding. The major contribution of this work is we build a framework based on semantics of video segmentation without covering information hiding, which introduced the MobilenetV2 [17] convolution neural network in video coding, feature extraction. Because MobilenetV2 neural network belongs to lightweight neural network, it is easy to handle the video file with its superior performance. At the same time, video is decoded and segmented by upsampling and 1×1 convolution module. One convolution module model can effectively obtain the classification scores of pixels in rough positions and obtain the final pixel-intensive output through upsampling. Meanwhile, this paper proposes an information mapping algorithm based on semantic information statistical histogram. Finally, the scheme analyzes the influence of different parameters on capacity and the robustness.

The structure of this paper is as follows: Section 2 introduces the basic content, Section 3 introduces the proposed methods, and Section 4 gives the experimental results and comparison. We conclude this article in Section 5.

2 Preliminaries

2.1 Semantic segmentation based on convolutional neural network

The application field of convolutional neural network is quite extensive, such as image recognition [18–20], image classification [21–24], target tracking [25–28], text analysis [29–32], target detection, and image retrieval [33, 34]. It is a powerful tool for image processing and research.

In this paper, neural network is used to understand and segment the image semantically, and the semantic features are extracted to hide secret information. Semantic segmentation aims to understand the image at the pixel level and expects all pixels in the image to be labeled with the target category. MobilenetV2 convolutional neural network classifies images at the pixel level, thus solving the problem of semantic segmentation at the semantic level. Due to the above characteristics of semantic segmentation, it has high security when applied to information hiding. From the perspective of attackers, it is difficult for them to crack secret information from the semantic level, and it is difficult to restore secret information with different neural networks and parameters.

Each layer of the convolutional network used for semantic segmentation is an $h \times w \times d$ 3d array, where h and w are spatial dimensions and d is the feature or channel dimension. The first layer is the image with $h \times w$ pixels and d color channels. Locations at high levels correspond to locations in the image where they are connected, called the receiving region.

The convolution net is based on translation invariance. By convolution, pooling and excitation function and other basic layers in the local input domain only depend on the relative space coordinates. In a certain layer x_{ij} for coordinates (i, j) of vector data, computation formula is as follows:

$$y_{ij} = f_{ks} \cdot (\{x_{si} + \delta_i, s_j + \delta_j\}, 0 \le \delta_i, \delta_j \le k) \tag{1}$$

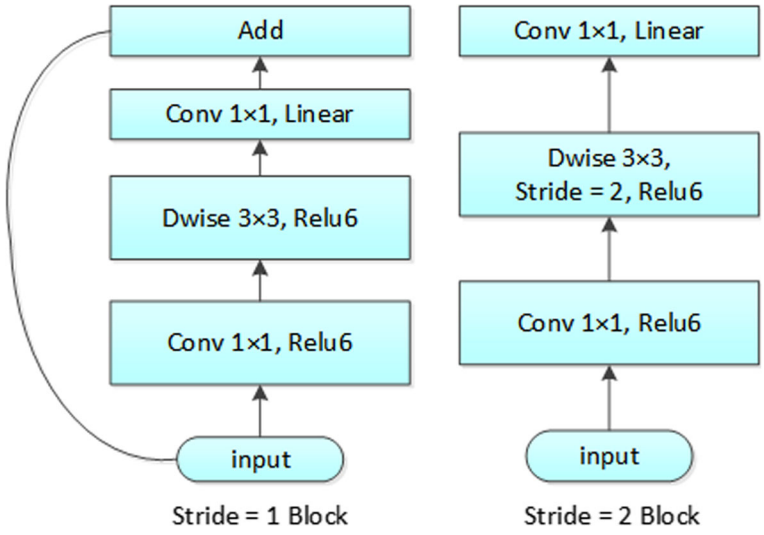

Fig. 2 Linear bottlenecks + inverted recurrent block

where k is the convolution size, s is the step size or downsampling factor, and f_{ks} determines the type of layer, such as matrix multiplication or average pooling, maximum pooling, or a nonlinear excitation function.

MobilenetV2 adopted in this paper is a lightweight network, which not only solves many parameter problems, but also solves the problem of low latitude data collapse caused by relu layer and how to use features for reuse by proposing linear bottlenecks + inverted recurrent block as the basic network structure. Data collapse is a problem encountered in deconvolution. At the same time, if the weight value of a convolution node changes to 0 in convolution network training, the output of the node is 0 for any input, the gradient value through the relu layer is 0, and the node will never recover, which can be effectively solved by feature multiplexing. Figure 2 shows the basic structure of MobilenetV2.

The complete parameters of MobilenetV2 are shown in the Table 1. The bottlenecks of MobilenetV2 are the most basic unit in the module. Bottleneck modules are respectively by the expansion, convolution, and compression, and MobilenetV2 uses bottleneck module to simplify the calculation of network.

Table 1 MobilenetV2 structural parameters

Input	Operator	Extension ratio	Number of output channels	Times of repetition	Stride
$224^2 \times 3$	Conv2d	–	32	1	2
$112^2 \times 32$	Bottleneck	1	16	1	1
$112^2 \times 16$	Bottleneck	6	24	2	2
$56^2 \times 24$	Bottleneck	6	32	3	2
$28^2 \times 32$	Bottleneck	6	64	4	2
$28^2 \times 64$	Bottleneck	6	96	3	1
$14^2 \times 96$	Bottleneck	6	160	3	2
$7^2 \times 160$	Bottleneck	6	320	1	1
$7^2 \times 320$	Conv2d 1 × 1	–	1280	1	1
$7^2 \times 1280$	Avgpool 7 × 7	–	–	1	–
$1 \times 1 \times k$	Convd2d 1 × 1	–	K	–	–

In the one convolution module adopted by the decoder, a convolution of 1×1 with 150 channel dimensions is added at the end to predict the score of each classification, followed by a deconvolution layer for bilinearly sampling rough output to pixel dense output.

As shown in the structure diagram, the neural network is to extensively train encoder and decoder, and then extract and screen out the most representative features layer by layer in frames. It is the training of neural network that makes the secret information obtained from the mapping of carrier feature also certainly robust.

2.2 Statistics of semantic information

As mentioned in Section 2.1, semantic segmentation refers to understand the images at the pixel level. The specific process is to comprehensively consider the weight of the relationship between pixels, than divide the pixels or even the whole picture according to the given threshold. Given that the shot is continuous, the semantic information of the front and rear frames, such as the scene category, and the subtle difference of target location, can be considered as the same semantic information. According to the above considerations, semantic information is calculated to improve algorithm robustness. Figure 3a represents the original picture frame. Figure 3b represents the segmentation graph, which have been divided into sub-blocks of $M \times M$ blocks, where the size of M is 3. Figure 3c represents the maximum semantic percentage value in each sub-block. The color represents the semantic type, and the value represents the ratio of that semantic type to the partitioned sub-block size. Figure 3d is the semantic histogram of the sub-block.

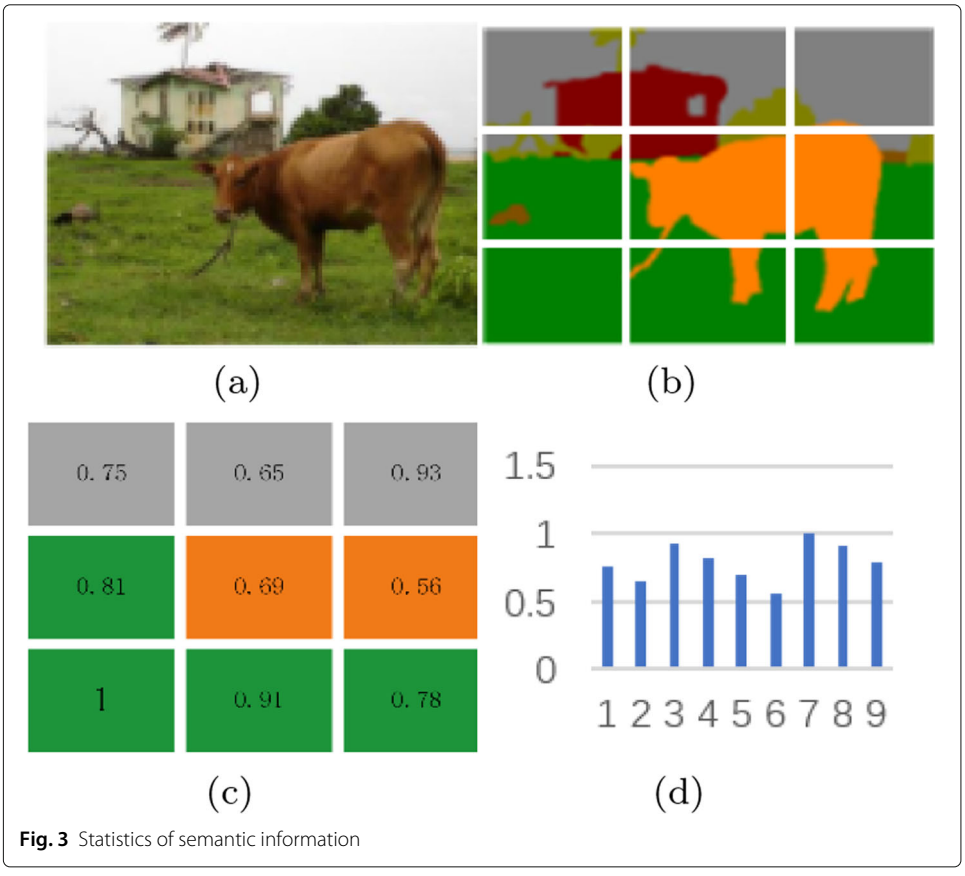

Fig. 3 Statistics of semantic information

Suppose that each frame image has $row \times col$ pixels, and each pixel belongs to a category. We will use MobilenetV2 for pixel-level semantic segmentation. For every frame of the image semantic segmentation image, the semantic image segmentation can be divided into 3×3 of the same size of the semantic chunk. Semantic information histogram can be obtained by counting of the highest semantic ratio.

The steps of semantic information statistics are as follows:

1. The size of the image is $(row \times col)/9$ after the semantic segmentation graph is segmented.
2. Semantic types of sub-blocks of segmentation graph after block segmentation.
3. The number of semantic types in a partitioned graph sub-block.
4. Statistics of the maximum semantic occupancy of all segmented sub-blocks.

3 Methods

In this section, we describe the process of hiding and extracting secret information. The video coverless information hiding scheme proposed is shown in Fig. 4. First, we set up a video database consisting of videos of various topics, which is shared by the sender and receiver and stored on the cloud platform, thus effectively saving storage space. Secondly, the image frame of each video file is segmented to obtain the statistical histogram of semantic information. The histogram feature is mapped to hash sequence, and the video index database is established. For the sender, the secret message is sliced into segments, and each bit group can be mapped to a histogram of the semantic information. By searching the index library, the appropriate video is selected as the carrier its mapping index records as auxiliary information to send to the receiver. The received auxiliary information will be used to determine the video frame, and the secret information can be restored by calculating the hash sequence from the video carrier. Throughout the process, the carrier video remains original without any modifications. Therefore, it can resist the detection of steganalysis.

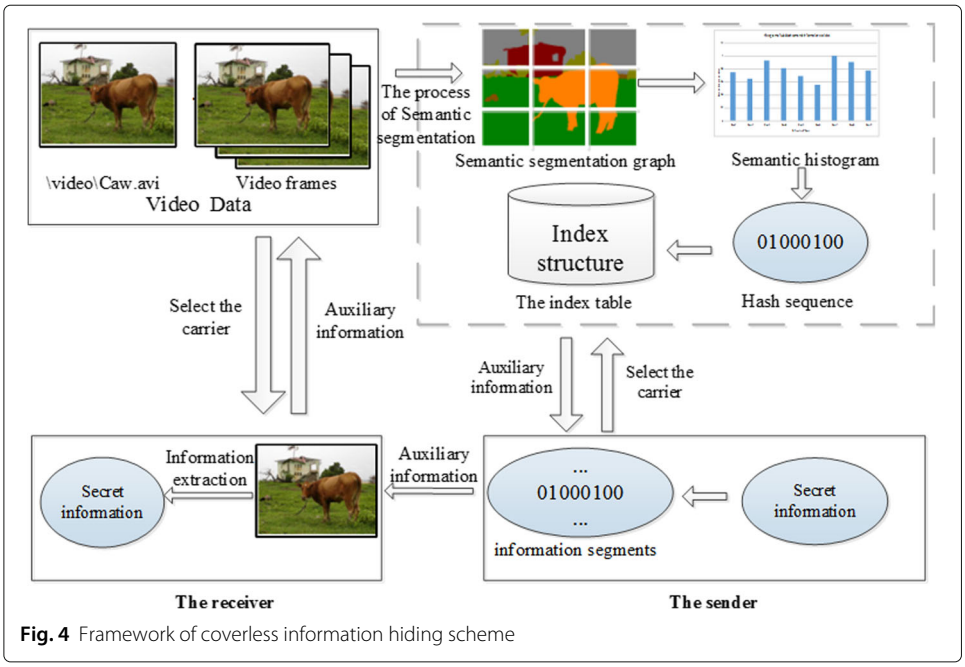

Fig. 4 Framework of coverless information hiding scheme

3.1 Hash sequence mapping based on semantic statistics

As described in Section 2.2, semantic information can reflect the content information of video frame, and the hash sequence of semantic information can be generated from Fig. 5.

First, each frame of the video image is extracted. Secondly, we carry out semantic segmentation and finally divide the semantic segmentation graph into sub-blocks of $M \times M$ blocks of the same size, and count the semantic information block by block.

Through the process shown in Section 2.2, the corresponding statistical histogram of semantic information of each block is obtained. The histogram is composed of statistical semantic types and semantic proportions: suppose we count one of the $M \times M$ blocks image segmentation graphs, we will get that there are n semantic types, and the semantic ratios corresponding to each semantic type in the segmentation block are $H = \{h_1, h_2, \ldots, h_n\}$. At the same time, we keep all the segmentation sub-block semantic ratio in the largest proportion, in regular scanning way for sorting, the corresponding semantic largest proportion also sorting; we will get the maximum $M \times M$ semantics of graph. Finally, we compare the size of the graph with the largest semantics before and after, and the biggest semantic of the size of the chart, and the resulting length of $M \times M - 1$ the hash sequence. In Fig. 5, we assume that M is equal to 3, so the length of the hash sequence is 8. The size of the vertical graph of semantic proportion is compared from left to right to generate a hash sequence of {01000100}.

1. For a given video file, the frame is expressed as $PV = \{p_1, p_2, \ldots, p_m\}$. Semantic segmentation network is used to obtain semantic information images. The performance of these semantic segmentation networks is different, which can be selected according to the actual situation. It is shown as follows:

$$PM_i = \text{Segmentation}(PV), 1 \leq i \leq m \tag{2}$$

where PM_i is the semantic image of frame i in the video file of PV, m is the number of frames of the video, and Segmentation() is the semantic segmentation function, which is selected according to the requirements.

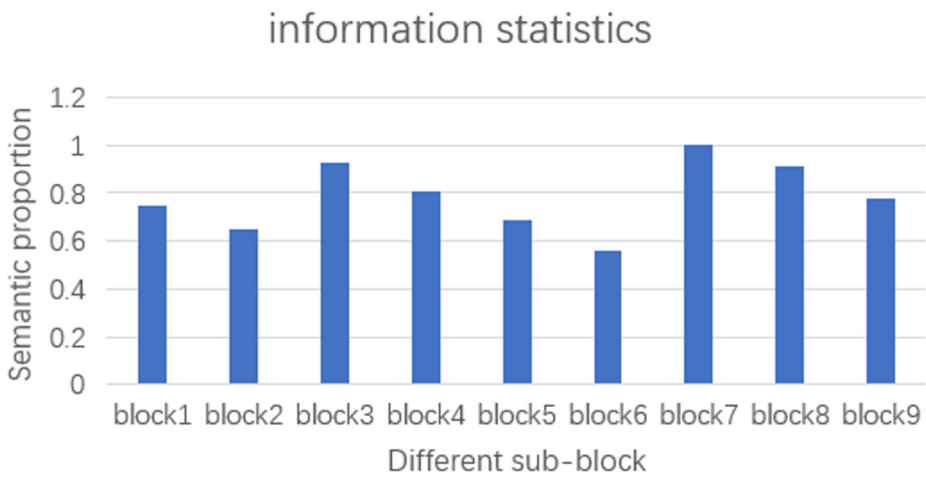

Fig. 5 The process of hash sequence generation

2. According to the obtained semantic segmentation image PM_i, assuming that the size of each frame is $row \times col$, the PM_i is uniformly divided into a number of semantic segmentation blocks of $M \times M$, with the size of each block being $\frac{row}{M} \times \frac{col}{M}$, and M can be any positive integer greater than 1. We customize a scan to arrange the semantic segmentation block B_i in the specified order from top to bottom and from left to right:

$$B_i = \{B_1, B_2, B_3, \ldots, B_{M\times M}\}, M \in (1, 2, \ldots) \tag{3}$$

3. According to the obtained semantic image sub-block B_i, we need to calculate all the semantic proportions h in the corresponding block:

$$h = \{h_1, h_2, h_3, \ldots, h_j\} \tag{4}$$

4. Maximum semantic proportion HM in statistical segmentation sub-blocks:

$$HM = max\{h_1, h_2, \ldots, h_j\} \tag{5}$$

where max is a function to calculate the maximum value, and the maximum semantic occupancy ratio in all segmented sub-blocks B_i is calculated to obtain the semantic statistical information H of PM_i:

$$H = \{HM_1, HM_2, \ldots, HM_i\}, i = M \times M \tag{6}$$

5. By comparing before and after the vertical graph of semantic information statistics, the K-bit hash sequence B_s is obtained:

$$B_S = \begin{cases} 1, \text{if} & HM_i \geq HM_{i-1} \\ 0, \text{if} & HM_i < HM_{i-1} \end{cases}, 1 \leq S \leq M \times M - 1 \tag{7}$$

3.2 Establishment of video index database

The key step of the method in this paper is to search the qualified video in video database and transmit it.

We mainly take index ID + video ID + number of frame ID, where the index ID refers to the ID number of the index, which is incremental and represented by w. The video ID represents the path to the folder and the file name, which is represented by p. Number of frame ID means that frame v hides secret information in video with path p, and we use v to represent frame ID. In order to ensure efficient and accurate search, we build the index structure as shown in Fig. 6. For example, if the hash sequence of a frame in Video Walking.avi after image segmentation is {00000001}, its corresponding index ID is 2, that is, w is 2. At the same time, the video corresponding to video ID information p is the path corresponding to the index ID of 2. If this frame appears in frame 25, then

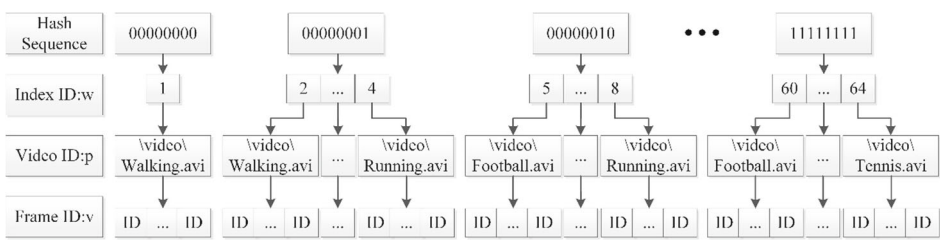

Fig. 6 Structure of video index database

V is 25. Then, the corresponding auxiliary information of this frame is (w:2,p:...*video*\Walking.avi,v:25), which also included in the ID of the index information table.

3.3 Coverless video steganography algorithm

In the process of hiding secret information, this paper effectively solves the mapping problem of secret information to video. The process is as follows:

(1) Establish a public video library, which is shared by the sender and receiver.

(2) For each video in the library, semantic segmentation is carried out according to Section 2.1.

(3) For each semantic segmentation image, statistical semantic information is shared.

(4) Establish the histogram according to the semantic information after each statistic, hash code according to the method is described in Section 3.1, and finally get the hash sequence.

(5) Establish video index database, as described in Section 3.2.

(6) The process of secret information matching carrier video is as follows:

First, the secret information S with length g is divided into large M binary information segments:

$$M = \begin{cases} \frac{g}{N}, & \text{if} \quad g\%N = 0 \\ \frac{g}{N} + 1, & \text{otherwise} \end{cases} \tag{8}$$

where N is the length of each binary information segment. When the total length g of secret information cannot be divisible, 0 is added in the last segment to form a sequence of length N, and the number of 0 is recorded.

(7) For each information fragment, we search the corresponding item in the index database, which is equal to the information fragment of secret information. It is possible to have multiple index entries mapped to the same hash sequence. To improve the efficiency of information extraction, we try to select index entries with the same video file. For the same video file, select the index with the smallest index ID and larger than the previous information segment. Algorithm 1 describes the sending process in detail.

3.4 Information extraction

For the receiver end, the hidden information can be successfully extracted by computing the hash sequence of the carrier video. The secret information extraction process is as follows:

(1) Corresponding frames can be found according to video ID and frame ID, and the semantic segmentation graph can be obtained according to the description in Section 2.1.

(2) Obtain the semantic information histogram according to Section 2.2.

(3) According to the hash method described in the histogram and Section 3.1, the hash sequence of the corresponding frame is obtained.

(4) Repeat the above steps until all hash sequences are obtained.

(5) After connecting the hash sequence and removing the populated 0 bit from the tail, the bitstream of the secret information flow was successfully restored.

Algorithm 2 describes the process of secret information extraction algorithm:

Algorithm 1 : Information hiding

Input:Video databases $V = \{V_1, V_2, \ldots, V_k\}$,The secret information bit sequence is $B = \{b_1, b_2, \ldots, b_d\}$.

Output:Mapping the index $I = \{Ind_1, \ldots, Ind_i\}$.

for $i = 1 : k$
  Decompose video to picture:$P = VideoToFrame(V_k)$
  for $j = 1 : FramreNum$
    Semantic segmentation processing $S_i = Semanticsegmentation(P_j)$
    Calculate the gradient histogram of the massage of semantic segmentation in S_i: $H_i = GradientHist(S_i)$
    $Hash_i = HashCalc(H_i)$
    Update index database:
    Index item $- >$ $\{indexID, VideoID, FrameID\}$
  end
end
padding secret information bits stream:$B' = pad(B)$
Divide B' to M segments
for $i = 1 : M$
  Search corresponding index item and update the index ID set as:$I = \{I, Ind_i\}$
end
Send Mapped Index ID set I to the receiver

Algorithm 2 : Information Extraction

Input:Video databases $V = \{V_1, V_2, \ldots, V_k\}$, Mapping the index $I = \{Ind_1, \ldots, Ind_m\}$.

Output:The secret information bit sequence is $B = \{b_1, b_2, \ldots, b_d\}$

for $i = 1 : m$
  get video ID,Frame ID from Index item i
  $P = VideoToFrame(V_{videoID})$
  Semantic segmentation processing $S_i = Semanticsegmentation(P_{Frame\_ID})$
  Calculate the gradient histogram of the massage of semantic segmentation in S_i: $H_i = GradientHist(S_i)$
  $Hash_i = HashCalc(H_i)$
end
Connect all the segments as: $Hash_1, Hash_2, \ldots, Hash_m$
Remove padding bits and the secret information bits stream are recovered as: $B = \{b_1, b_2, \ldots, b_m\}$

4 Experimental results and discussions

Experimental environment: Intel(R) Core(TM) i7-7800x CPU @3.50 ghz, 64.00 GB RAM, and two Nvidia GeForce GTX 1080 Ti GPUs were used in the experiment. Deep learning adopts Pytorch framework, which is a high-level neural network API.

We can use TensorFlow with Pytorch. All experiments were completed in MATLAB 2016a and Pycharm.

Training, validation data set: the data set used in this paper for training is the ADE20K data set published by MIT. ADE20K has more than 25,000 images that are densely annotated with open dictionary tags. The annotated images cover the various scene category, with a rich variety of scenes suitable for target segmentation and part segmentation. Among them, the training data set has 20,210 images, the verification set has 2000 images, and the test set has over 2000 images. Figure 7 shows part of the ADE20K training data set.

Given that neural network has a strong anti-interference ability to noise, it can be learned from a large amount of data at the same time. Therefore, in order to improve the segmentation robustness of image with noise. In this paper, a new training data set is developed, which includes the original data and the noise image. On the basis of ADE20K data set, different factors of Gaussian noise, salt and pepper, speckle, and compressed images were added to train together; Fig. 8 shows the original image and the noise image. Based on the pre-training model, 30 cycles of training were conducted on the new training data set, with each period iterating for 5000 times. Finally, the training accuracy reaches 80% and verification accuracy reaches 75%.

The mobilenet network was used in this paper to test the DAVIS-2017 data set, and the DAVIS-2017 is shown in Fig. 9. DAVIS is a high-quality and full high-definition data set, including 90 video sequences of different scenes and a total of more than 6000 pixels of fully matched annotated images. In this paper, we will analyze and test the scheme from three aspects of capacity, robustness, and security on the DAVIS-2017 data set.

4.1 Capacity

This section examines the capacity of the program. The capacity of information hiding depends on the length of hash sequence, which is related to the semantic segmentation

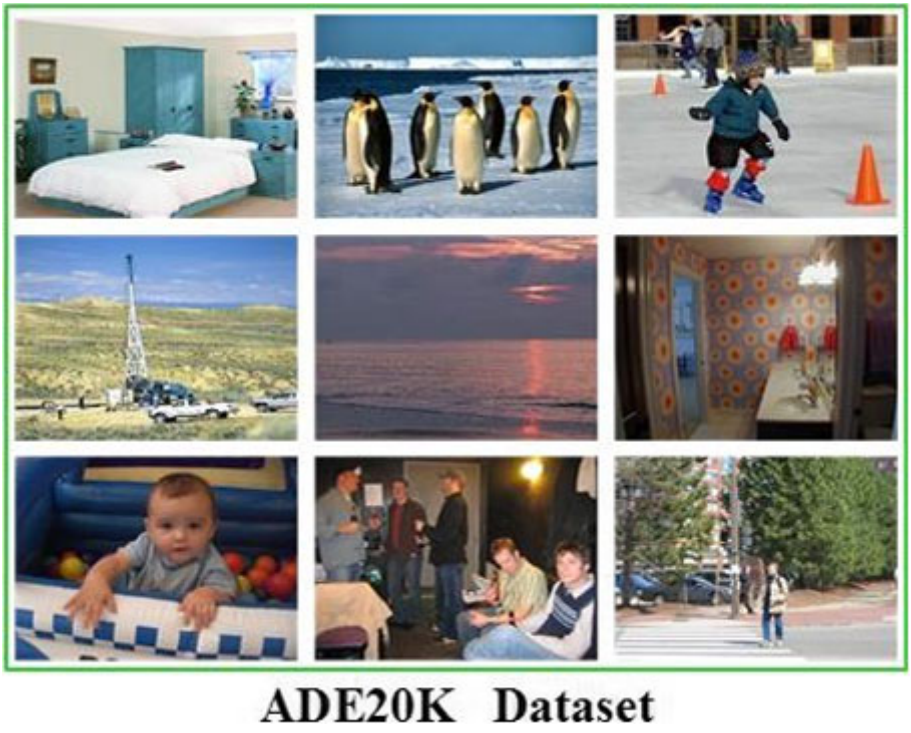

Fig. 7 ADE20K data set

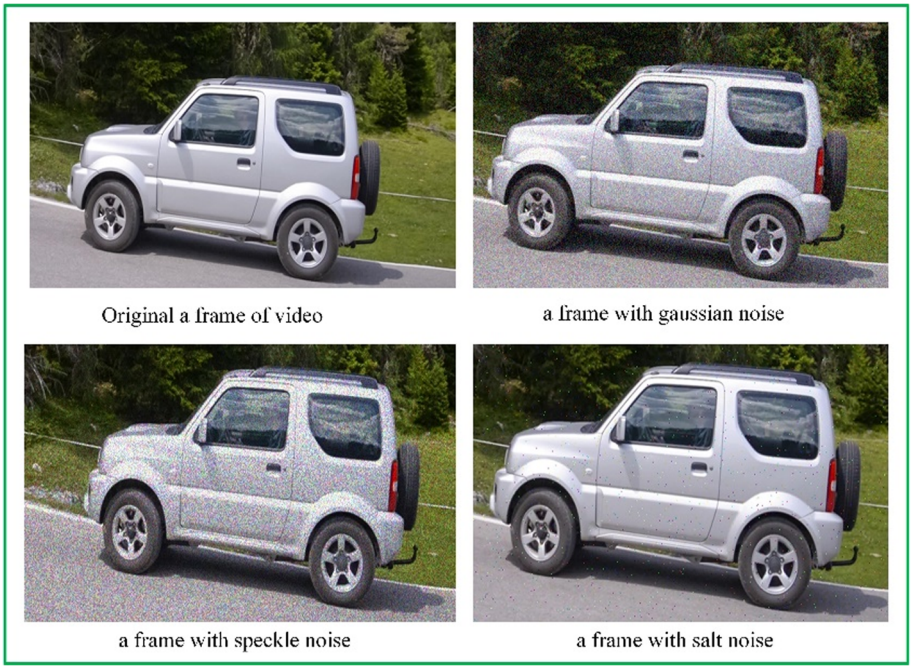

Fig. 8 Different noise frames

algorithm and the number of image blocks. Given that video frame is represented as $PV = \{p_1, p_2, \ldots, p_m\}$, assuming that the number of blocks of video frame segmentation graph in video is n, and we calculate the maximum semantic proportion of j block, then the mapping position of complete video can be expressed as:

$$C = p_1 \cdot (j-1) + p_2 \cdot (j-1) + \ldots + p_m \cdot (j-1), 1 \leq j \leq n \tag{9}$$

Capacity C refers to the total length of the bitstream that can be generated for a video file. It is easy to see that the length of the hash sequence is positively correlated with the number of blocks n, and the capacity increases with the increase of j. Since there is

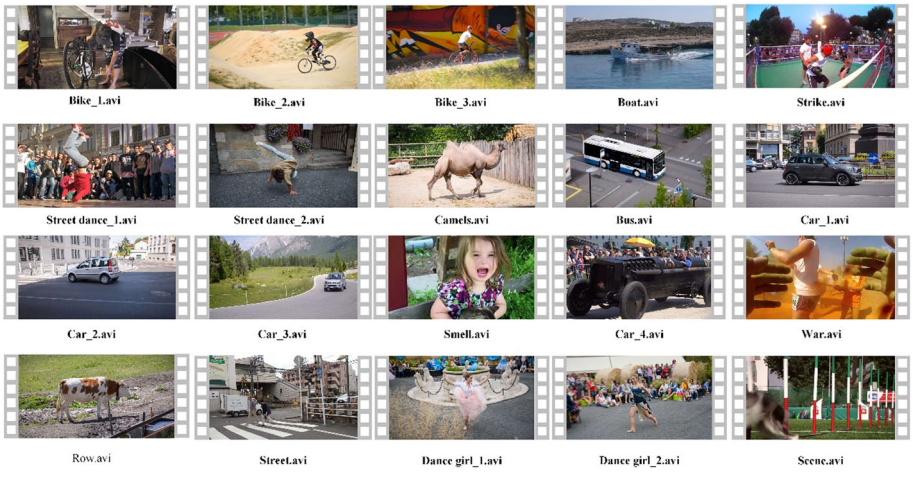

Fig. 9 DAVIS-2017 data set

Table 2 Hidden bit number comparison in one image

Methods	Pixel method	Image hashing method	Zheng's method	GIF method	Proposed method ($j = 9$)
Capacity	8 bit	18 bit	8 bit	7∼14 bit	8 bit

currently coverless steganographic scheme based on video, this scheme will be compared with the existing single image steganographic scheme. The results are shown in Table 2.

The capacity of the algorithm depends on the length of hash sequence. According to the hash algorithm in Section 3.1, the hash length of the image depends on the number of semantic segmentation blocks and the parameter j. Compared with other image steganography algorithms, our method is more extensive, and the specific results are shown in Table 3.

Ideally, the maximum available significant bits of capacity C are $m \times (J-1)$ bits. If the maximum value can be achieved, the $(J-1)$ bit information of each frame mapping is different, but the actual situation is that the semantic information between frames changes slowly, which is related to the number of frames of the camera. The possible situation is that the bitstream mapped by consecutive frames is the same. On the one hand, it is a waste of carrier resources. On the other hand, if the carrier lacks key frames due to transportation, we can extract the secret information we need from the adjacent frames to ensure the integrity of the secret information. Based on the above, we calculate the effective capacity C_E of the carrier. For a given video file, the frame is expressed as $PV = \{p_1, p_2, \ldots, p_m\}$. After the process of segmentation, we can get a bitstream $B_i = \{b_1, b_2, \ldots, b_8\}$ for every frame. The effective capacity can be expressed as follows:

$$D_{Bi} = \sum_{j=1}^{8} b_j \times 2^{8-j} \tag{10}$$

Here, D_{Bi} represents that the sequence information of the i frame is converted into decimal value. We can simply get the decimal number of the whole video $D = \{D_{B1}, D_{B2}, \ldots, D_{Bm}\}$. Then, the effective capacity is obtained:

$$C_E = \sum_{k=0}^{255} f(k), f(k) = \begin{cases} 1, & \text{if } k \text{ in } D \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

Effective capacity C_E refers to how many different mapping sequences can be generated in the entire video file. The effective capacity is related to block parameter M and parameter j, while the effective capacity is positively related to parameter j. The maximum hiding capacity of 8-bit secret information is 256, and the effective capacity of DAVIS data set is shown in Fig. 10. Here,we assume M is equal to 4.

Table 3 Number of images needed when the same information is hidden

Length of hidden data	1 B	10 B	100 B	1 KB
Pixel method [3]	1	10	100	1024
Zhou's method [6]	2	6	46	457
Zheng's method [13]	1	5	50	512
GIF method [11]	1∼2	6∼11	58∼115	586∼1171
Zhang's method [15]	2∼9	7∼81	55∼801	548∼8193
Proposed method	1	10	100	1024

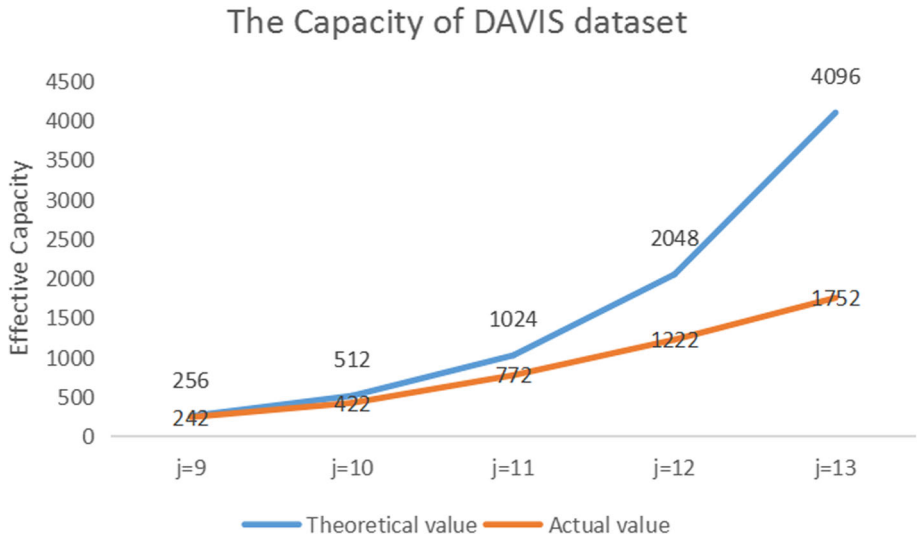

Fig. 10 The effective capacity of DAVIS data set

4.2 Robustness

This section describes the robustness to noise. The robustness of the algorithm largely depends on the ability of segmentation network to resist noise. In this experiment, we studied the robustness of different partition methods against pepper and salt noise, Gaussian noise, speckle noise, and JPEG compression attacks with different quality factors. The quality factor Q was 70% and 90%, respectively. Given that video frame is represented as $PV = \{p_1, p_2, \ldots, p_m\}$, the bitstream generated by each frame of image is $B_i = \{b_1, b_2, \ldots, b_8\}$; then, the calculation formula of accuracy rate is as follows:

$$ACC = \frac{\sum_{i=1}^{m} f(i)}{m}, f(i) = \begin{cases} 1, & \text{if } B_i = B_i' \\ 0, & \text{if } B_i \neq B_i \end{cases} \tag{12}$$

In this scheme, the results of different semantic segmentation networks against noise attacks are shown in Table 4. It can be seen that our scheme is robust. Different partitioning methods mean that semantic information is properly filtered and has different effects on robustness. If the video is irresistibly deleted, according to the hypothesis analysis before, we can extract the secret information we need from the consecutive frames before and after the index frame, which greatly improves the robustness.

Table 4 Robustness under different attack with different M

Attack	$M = 3$			$M = 4$	
	$j = 9$	$j = 9$	$j = 11$	$j = 13$	$j = 15$
Salt and pepper ($\sigma = 0.001$)	0.8332	0.8321	0.7917	0.7524	0.6856
Salt and pepper ($\sigma = 0.005$)	0.6169	0.6378	0.5660	0.4974	0.3930
Gauss ($\sigma = 0.001$)	0.3055	0.3593	0.2738	0.2018	0.1127
Gauss ($\sigma = 0.005$)	0.2975	0.3637	0.2776	0.2044	0.1226
Speckle ($\sigma = 0.01$)	0.4959	0.5386	0.4538	0.3803	0.2876
Speckle ($\sigma = 0.05$)	0.3002	0.3582	0.2745	0.2059	0.1316
JPEG ($\sigma = 70$)	0.7717	0.7593	0.7075	0.6573	0.5762
JPEG ($\sigma = 90$)	0.8833	0.8766	0.8497	0.8199	0.7713

At present, there is no research related to video coverless information hiding. This paper compares with the novel classical image coverless information hiding scheme. The specific results of the experiment are shown in Table 5. Under the attack of compressed noise, it has certain robustness. However this scheme is weak and sensitive to other noises. In Section 4.4, the reason of result is analyzed.

4.3 Security

The database adopted in this scheme is the video database uploaded on the cloud, which is shared by the sender and the receiver. Moreover, there is no modification in the carrier, so there is no steganalysis of video data. In these database, video can be uploaded in real-time, and the codebook can be updated.

In order to improve security, we can encrypt the auxiliary information. Meanwhile, we can use chaotic sequence encryption algorithm, RC4, SEAL, and other sequence encryption algorithms, which are determined according to the needs of users.

4.4 Discussion

In this section, we will further discuss some problems encountered in the experiment and give analysis and suggestions. On the one hand, deep learning and semantic segmentation are applied to the coverless information hiding scheme, which is a new coverless information hiding idea. This paper adopts lightweight semantic segmentation network, MobilenetV2, to train and segment all kinds of images with noise. The training of neural network is like a tool that can provide a lot of information materials. Firstly, it depends on the training data and, secondly, the depth of the network, and finally, it classifies the extracted information. This is like a heavy copper lock, with high security for information hiding. Semantic segmentation shows the rich semantic information of the video frame and capacity. It skillfully avoids steganalysis through the mapping rules only known by both sides. By applying coverless information hiding to video data, video data has large capacity and is not vulnerable to attack. This scheme has good robustness against compression attack of video data. On the other hand, the anti-interference ability of the scheme to some noises is not strong. One of the reasons may be the insufficient depth and breadth of the network segmentation. This paper uses Mobilenet, which belongs to portable network. The purpose of this paper is to apply it to video carriers. Second, the training data set is not comprehensive enough to cover all scenes of daily life. This will lead to low segmentation accuracy and robustness in the test data set.

Table 5 Robustness of different methods

DAVIS				
Attack	Ours	Zheng [13]	Zhou [6]	Zhang [9]
Salt and pepper ($\sigma = 0.001$)	0.8332	0.9436	0.9983	0.9982
Salt and pepper ($\sigma = 0.005$)	0.6169	0.9028	0.9967	0.9969
Gauss ($\sigma = 0.001$)	0.3055	0.7480	0.9693	0.9645
Gauss ($\sigma = 0.005$)	0.2975	0.7483	0.9703	0.9653
Speckle ($\sigma = 0.01$)	0.4959	0.8374	0.9780	0.9793
Speckle ($\sigma = 0.05$)	0.3002	0.6887	0.9294	0.9370
JPEG ($\sigma = 70$)	0.7717	0.9149	0.9912	0.9962
JPEG ($\sigma = 90$)	0.8833	0.9439	0.9935	0.9983

5 Conclusion

In this paper, a video coverless information hiding scheme based on semantic segmentation is proposed. In the direction of coverless information hiding, we make a preliminary attempt in the fields of video, neural network, and semantic segmentation. Using video as transmission carrier has the advantages of large capacity and not easy to be detected by attackers. We conduct semantic segmentation on video and obtained the statistical histogram of semantic information. The hidden bit sequence is mapped to the hash sequence through the histogram, which provides a new idea for coverless steganography. The receiver can extract the secret information via calculating the semantic information from the carrier videos. In the whole process of secret information transmission, the carrier videos have not been modified. Therefore, the scheme can effectively resist the steganalysis. The scheme applied deep learning to information security, which can resist the attack of noise to a certain extent, but the robustness of the scheme largely depends on the network. In the future, we will try to improve the capacity and robustness. How to give full play to the advantages of neural network to enhance the robustness is also the focus of our future work.

Abbreviations

DES: Data encryption standard; GIF: Graphics interchange format; DCT: Discrete cosine transformation; MIT: Massachusetts Institute of Technology

Acknowledgements

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

Authors' contributions

NP designed the algorithm. JHQ carried out the experiments and drafted the manuscript. XYX gave suggestions on the structure of the manuscript and participated in modifying the manuscript. YT built the data set and gave suggestions on the experimental analysis. GMH participated in the survey of large motion and gave suggestions on the experimental analysis. All authors read and approved the final manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China under grant 61772561, in part by the Key Research and Development Plan of Hunan Province under grants 2018NK2012 and 2019SK2022, in part by the Science Research Projects of Hunan Provincial Education Department under grants 18A174 and 19B584, in part by the Degree & Postgraduate Education Reform Project of Hunan Province under grant 2019JGYB154, in part by the Postgraduate Excellent teaching team Project of Hunan Province under grant [2019]370-133, and in part by the Postgraduate Education and Teaching Reform Project of Central South University of Forestry & Technology under grant 2019JG013.

Availability of data and materials

Please contact author for data requests.

Competing interests

The authors declare that they have no competing interests.

Received: 26 March 2020 Accepted: 24 May 2020
Published online: 16 June 2020

References

1. J. Qin, Y. Luo, X. Xiang, Y. Tan, H. Huang, Coverless image steganography: a survey. IEEE Access. **7**, 171372–171394 (2019). https://doi.org/10.1109/ACCESS.2019.2955452
2. Y. Tan, J. Qin, X. Xiang, W. Ma, W. Pan, N. N. Xiong, A robust watermarking scheme in ycbcr color space based on channel coding. IEEE Access. **7**(1), 25026–25036 (2019). https://doi.org/10.1109/ACCESS.2019.2896304
3. Z. Xia, X. Wang, X. Sun, Steganalysis of lsb matching using differences between nonadjacent pixels. Multimedia Tools Appl. **75**(4), 1947–1962 (2016). https://doi.org/10.1007/s11042-014-2381-8
4. B. Wang, W. Kong, W. Li, N. N. Xiong, A dual-chaining watermark scheme for data integrity protection in internet of things. Comput. Materi. Cotinua. **58**(3), 679–695 (2019). http://doi.org/10.32604/cmc.2019.06106
5. H. Huang, F. Wang, J. Pan, A vq-based robust multi-watermarking algorithm. IEICE Trans. Fundam. Electron Commun. Comput. Sci. **E85-A**(7), 1719–1726 (2002)
6. Z. Zhou, H. Sun, R. Harit, X. Chen, X. Sun, Coverless image steganography without embedding. Cloud Comput. Secur, 123–132 (2016). https://doi.org/10.1007/978-3-319-27051-7_11

7. Y. Luo, J. Qin, X. Xiang, Y. Tan, Q. Liu, L. Xiang, Coverless real-time image information hiding based on image block matching and dense convolutional network. J. Real-Time Image Process. **17**(1), 125–135 (2020). https://doi.org/10.1007/s11554-019-00917-3
8. Y. Long, Y. Liu, Y. Zhang, X. Ba, J. Qin, Coverless information hiding method based on web text. IEEE Access. **7**(1), 31926–31933 (2019). https://doi.org/10.1109/ACCESS.2019.2901260
9. X. Zhang, F. Peng, M. Long, Robust coverless image steganography based on dct and lda topic classification. IEEE Trans. Multimed. **20**(12), 3223–3238 (2018)
10. Q. Liu, X. Xiang, J. Qin, Y. Tan, J. Tan, Y. Luo, Coverless steganography based on image retrieval of densenet features and dwt sequence mapping. Knowl.-Based Syst. **192**(2020), 105375–105389 (2020). https://doi.org/10.1016/j.knosys.2019.105375
11. S. Ruan, Z. Qin, Coverless covert communication based on gif image. Commun. Technol. **50**(7), 1506–1510 (2017)
12. Z. Zhou, Q. M, J. Wu, C. Yang, X. Sun, Z. Pan, Coverless image steganography using histograms of oriented gradients-based hashing algorithm. J. Internet Technol. **18**(5), 1177–1184 (2017). https://doi.org/10.6138/JIT.2017.18.5.20160815b
13. S. Zheng, L. Wang, B. Ling, D. Hu, Coverless information hiding based on robust image hashing. Intell. Comput. Methodol, 536–547 (2017). https://doi.org/10.1007/978-3-319-63315-2_47
14. M. Liu, M. Zhang, J. Liu, Y. Zhang, Y. Ke, Coverless information hiding based on generative adversarial networks. Cryptogr. Secur, 371–382 (2018)
15. Y. Jiang, C. Song, An improved video zero watermarking algorithm for pseudo-3d-dct domain. Comput. Eng. Sci. **39**(9), 1721–1728 (2017). https://doi.org/10.3969/j.issn.1007-130X.2017.09.019
16. X. Chen, G. Bu, H. Li, A video zero-watermark algorithm based on the contourlet transform (2013). https://doi.org/10.2991/icmt-13.2013.27
17. M. Sandler, A. Howard, M. Zhua, A. Zhmoginov, L. Chen, Mobilenetv2:inverted residuals and linear bottlenecks. IEEE Conf. Comput. Vis. Pattern Recogn, 4510–4520 (2018)
18. W. Ma, J. Qin, X. Xiang, Y. Tan, Y. Luo, N. N. Xiong, Adaptive median filtering algorithm based on divide and conquer and its application in captcha recognition. Comput. Mater. Continua. **58**(3), 665–677 (2019). https://doi.org/10.32604/cmc.2019.05683
19. L. Pan, J. Qin, H. Chen, X. Xiang, C. Li, R. Chen, Image augmentation-based food recognition with convolutional neural networks. Comput. Mater. Continua. **59**(1), 297–313 (2019). https://doi.org/10.32604/cmc.2019.04097
20. J. Wang, J. Q. adn, X. Xiang, Y. Tan, N. Pan, Captcha recognition based on deep convolutional neural network. Math. Biosci. Eng. **16**(5), 5851–5861 (2019). https://doi.org/10.3934/mbe.2019292
21. J. zhang, W. Wang, C. Lu, J. Wang, A. Sangaiah, Lightweight deep network for traffic sign classification. Ann. Telecommun (2019). https://doi.org/10.1007/s12243-019-00731-9
22. W. Pan, J. Qin, X. Xiang, Y. Wu, Y. Tan, L. Xiang, A smart mobile diagnosis system for citrus diseases based on densely connected convolutional networks. IEEE Access. **7**(1), 87534–87542 (2019). https://doi.org/10.1109/ACCESS.2019.2924973
23. J. Qin, W. Pan, X. Xiang, Y. Tan, G. Hou, A biological image classification method based on improved cnn. Ecol. Inform (2020). https://doi.org/10.1016/j.ecoinf.2020.101093
24. Y. Chen, W. Xu, J. Zuo, K. Yang, The fire recognition algorithm using dynamic feature fusion and iv-svm classifier. Cluster Comput. **22**(S3), 7665–7675 (2019). https://doi.org/10.1007/s10586-018-2368-8
25. Y. Chen, J. Wang, R. Xia, Q. Zhang, Z. Cao, K. Yang, The visual object tracking algorithm research based on adaptive combination kernel. J. Ambient Intell. Human. Comput. **10**(12), 4855–4867 (2019)
26. J. Zhang, Y. Wu, W. Feng, J. Wang, Spatially attentive visual tracking using multi-model adaptive response fusion. IEEE Access. **7**(1), 83873–83887 (2019)
27. F. Liu, Y. Guo, Z. Cai, N. Xiao, Z. Zhao, Edge-enabled disaster rescue: a case study of searching for missing People. ACM Trans. Intell. Syst. Technol. **11**(4), 1–26 (2020)
28. J. Zhang, X. Jin, J. Sun, J. Wang, A. Sangaiah, Spatial and semantic convolutional features for robust visual object tracking. Multimed. Tools Appl (2018). https://doi.org/10.1007/s11042-018-6562-8
29. L. Xiang, G. Guo, J. Yu, V. Sheng, A convolutional neural network-based linguistic steganalysis for synonym substitution steganography. Math. Biosci. Eng. **17**(2), 1041–1058 (2020)
30. Y. Zhang, W. Lu, W. Ou, G. Zhang, X. Zhang, J. Cheng, W. Zhang, Chinese medical question answer selection via Hybrid models based on cnn and gru. Multimedia Tools Appl (2020). https://doi.org/10.1007/s11042-019-7240-1
31. Y. Tong, Y. Liu, J. Wang, G. Xin, Text steganography on rnn-generated lyrics. Math. Biosci. Eng. **16**(5), 5451–5463 (2019)
32. Z. Zhou, J. Qin, X. Xiang, Y. Tan, Q. Liu, N. N. Xiong, News text topic clustering optimized method based on tf-idf algorithm on spark. Comput. Mater. Continua. **62**(1), 217–231 (2020). https://doi.org/10.32604/cmc.2020.06431
33. H. Li, J. Qin, X. Xiang, L. Pan, W. Ma, N. N. Xiong, An efficient image matching algorithm based on adaptive threshold and ransac. IEEE Access. **6**(1), 66963–66971 (2018). https://doi.org/10.1109/ACCESS.2018.2878147
34. J. Qin, H. Li, X. Xiang, Y. Tan, W. Pan, W. Ma, N. N. Xiong, An encrypted image retrieval method based on harris corner optimization and lsh in cloud computing. IEEE Access. **7**(1), 24626–24633 (2019). https://doi.org/10.1109/ACCESS.2019.2894673

Publisher's Note