RESEARCH   **Open Access**

The study of security application of LOGO recognition technology in sports video

Zhi Li[1,2]

Abstract

With the rapid development of information technology and network technology and the rapid popularization of computers and intelligent devices, network video media have gradually replaced traditional media and became the main carrier of information production, storage, dissemination, sharing, and use. However, digital information can be easily copied, modified, and disseminated, which leads to the problem of network video copyright theft which is very prominent. Aiming at the copyright problem of sports video, this paper proposes an idea of using LOGO to identify the copyright of sports video, which can automatically and effectively detect whether the video is genuine or not. Firstly, encrypted LOGO is added to the video by encryption technology, and then, LOGO recognition technology is used to distinguish whether the LOGO is missing in the video, so as to distinguish the legitimacy of the video. In addition, aiming at the shortcomings of traditional LOGO recognition technology, this paper proposes a LOGO detection and recognition technology based on convolution neural network. This method improves the traditional histogram algorithm, proposes an analysis algorithm based on non-uniform block HSV histogram, and introduces region weighting coefficients to extract key frames from the video. Then, the convolution neural network is used to extract the features of the key frames of the video. Finally, the features of the sample LOGO are matched to complete the LOGO detection and recognition. Simulation results show that the effect of feature extraction using convolution neural network is better. Compared with the traditional LOGO recognition technology, the LOGO detection and recognition technology based on convolution neural network proposed in this paper has better detection and recognition performance. It effectively detects the LOGO information in the video and realizes the recognition of whether the video is genuine or not.

Keywords: Sports video, Copyright, LOGO recognition technology, Convolution neural network, Feature extraction

1 Introduction

Today, copyright protection will play a vital role in the increasingly prominent value of the copyright industry. As one of the important legal systems that encourage and protect intellectual innovation; promote economic, technological, cultural and social development; and enhance the country's core competitiveness, copyright protection is increasingly important and its role is becoming more and more prominent. At present, the development of copyright protection in China has entered a new historical stage, that is, the creation and use of intellectual achievements to promote economic development and

social progress. But now, our country is in the transition period, so we have encountered many problems. With the emergence of a large number of network video, the copyright problem of massive network video content is becoming more and more prominent. A technology that can automatically detect pirated videos is urgently needed to further enhance the efficiency, quality, and intensity of copyright management.

Our common logo for all kinds of goods is Logo, also known as the trademark name, which includes text, graphics, letters, numbers, three-dimensional logo, and color combinations, as well as the combination of the above elements. The Logo of the video is the symbol of the content of the video, which to some extent reflects the information of the content of the video.

Generally speaking, the following four principles should be followed in the design of Logo: [1–3]:

Correspondence: lizhi@nit.zju.edu.cn
[1]School of Media and Design, Ningbo Institute of Technology, Zhejiang University, No. 1 Qianhu South Rd. Yinzhou District, Ningbo, Zhejiang Province, China
[2]College of Media and International Culture, Zhejiang University, Hangzhou, Zhejiang Province, China

1. Simplicity: Logo, as a visual language, requires it to produce instantaneous effects and give a clear first impression when it comes to the audience. Therefore, the typical corporate logo is characterized by simplicity, clarity, and eye-catching features, while avoiding patterns that are too complex and too concealed. Our survey of the top 500 companies shows that logos in large companies are designed to be clear and eye-catching, which makes logos suitable for a variety of uses and seems to have excellent recognition from all directions and angles. At the same time, Logo designers must also consider the effect of Logo based on different media, in order to achieve communication convenience and consistency.

2. Uniqueness: The commercial purpose of Logo is to express the unique character of an enterprise, a product, or a brand. Logo, as a unique mark, needs to let consumers recognize the unique quality of the brand; therefore, Logo design must be novel and unique, showing the unique personality of enterprises, products, or brands. For these reasons, one of the most important things in Logo design is to avoid being similar to other Logos. The fundamental principle of logo design is creativity. In order to design a unique visual image with high visibility and strong expressive ability, designers should be good at using various techniques and eventually design a logo which is easy to recognize and memorize.

3. Artistry: Logo is not only a commercial logo, but also carries and conveys brand values. In order to achieve the above purposes, Logo design must conform to esthetic principles, so that people watch not only a Logo process, but also an esthetic process. The esthetic meaning of Logo is often expressed through the modeling of Logo. Modeling beauty is an important artistic feature of Logo. The modeling elements of Logo include four categories: point, line, surface, and body. With the help of these four elements and by mastering the relevant rules of different forms of modeling, the logo pattern has the esthetic characteristics independent of other things, expresses the connotation of the brand and embodies the spirit of the brand.

4. Development: With the change of the times or the development of the brand, the content and style reflected by Logo may not accord with the rhythm of the times and the change of the brand. Therefore, the content of Logo should be innovated so as to keep pace with the times. At present, many enterprises, including Huawei, have abandoned outdated logo designs in order to better conform to the spirit of the times and brand development, replacing them with logos that are more visually expressive and user-friendly in order to further enhance brand competitiveness.

The key point of Logo detection and recognition algorithm is the invariant feature extraction scheme, so some researchers have proposed SIFT (scale-invariant feature transform) feature descriptor [4], SURF (speeded up robust features) descriptor [5], ORB (oriented FAST and rotated BRIEF) feature descriptor [6], HOG (histogram of oriented gradient) feature descriptor [7], LBP (local binary pattern) feature description operator [8], and Haar feature description operator [9] for invariant feature extraction.

SIFT features are invariant in rotation, scale, translation, viewing angle, and brightness and have good robustness to parameter adjustment. The algorithm can effectively express the target feature information. The extraction of local feature points by SIFT algorithm mainly includes four steps: suspected feature points detection, removal of pseudo-feature points, matching of feature points' gradient and direction, and feature description vector generation. SURF description operator and ORB feature description operator are the improvements of SITF description operator, especially the ORB feature description operator, which makes the running time of target detection and recognition algorithm take a qualitative leap. The HOG feature description operator implements the object feature description by calculating and counting the gradient histogram of the local region of the image. Haar features can be divided into four categories: edge features, linear features, central features, and diagonal features. These features can be combined into feature templates to extract invariant features, which were first used in face detection by Papageorgiou et al. [10].

The effect of traditional Logo detection and recognition method is strongly dependent on the characteristics of artificial selection. The depth learning model proposed by Hinton in 2006 breaks this difficulty. He proposed a depth learning model using multiple hidden layers to learn high-level representations [11] and mentioned that deep neural networks could be used to automatically learn high-level features from large amounts of data. Compared with the characteristics of traditional methods, the deep learning model has richer self-learning features, higher abstraction level, and stronger expressive ability. Common deep learning models include deep learning models based on Restricted Boltzmann Machine (RBM) [12, 13], deep learning models based on autoencoder [14–16], and deep learning models based on the convolution neural network (CNN) [17–19].

With the continuous development of deep learning technology, researchers have found that using the deep learning model based on convolution neural network for target detection and recognition, the accuracy can be greatly improved. In particular, the R-CNN (regions with

CNN features) algorithm proposed by Girshick R in 2014 [20] has made breakthroughs in the application of deep learning in the field of target detection and recognition. In the R-CNN algorithm, the selective search algorithm is used to generate candidate regions instead of the traditional "pyramid" sliding window mechanism. In the whole network architecture of the R-CNN algorithm, the author uses Alex Net and VGG16 deep neural networks to build the target detection and recognition framework, and at the end of the network, BBox regression is used to improve the accuracy of the detection and recognition algorithm on the data set.

Aiming at the current problem of copyright piracy of network video, this paper applies logo recognition technology to the copyright identification of sports video and proposes an idea of video copyright identification using LOGO, which can automatically and effectively detect whether the video is genuine or not. The specific contributions of this paper are as follows:

1. Encryption technology is used to add encrypted LOGO to the video, and then, LOGO recognition technology is used to distinguish whether the LOGO is missing in the video, so as to distinguish the legitimacy of the video.
2. The key frames are extracted from the video, the traditional histogram algorithm is improved, and the analysis algorithm based on uneven block HSV histogram is proposed, and the region weighting coefficient is introduced, which highlights the importance of the feature of the center region of the image. The algorithm can achieve fast and good key frame extraction, greatly reduce the number of frames, and improve the processing speed.
3. To overcome the shortcomings of traditional LOGO recognition technology, this paper proposes a LOGO detection and recognition technology based on convolution neural network. This method uses convolution neural network to extract features of key frames of video and then matches the features of sample LOGO to complete the LOGO detection and recognition.

2 Proposed method
2.1 Video key frame extraction
2.1.1 Basic idea of algorithm
The basic idea of non-uniform block HSV histogram algorithm is to divide the image into several blocks and then extract its color histogram for each block, so that the extracted color histogram can better represent the color distribution in the image. Because pixels in different regions have different contributions to the image, the block histogram algorithm here does not simply divide the image into several uniform blocks, but divide the image into blocks of different sizes. In an image, the pixels that reflect the content of the image are basically concentrated in the middle of the image, and the contribution of the pixels in the surrounding area to the content of the image is slightly smaller.

The reason why the uneven block HSV histogram algorithm can be applied here mainly considers the advantages of the improved histogram algorithm:

1. The statistical process considers the change of local features and is more sensitive to the change of local features, so it is more scientific when it is used in the process of key frame extraction.
2. Because the image is divided into blocks for processing, multi-threading method can be introduced to speed up the processing.
3. Introducing weighting and discarding the edge and corner area of the image, highlighting the importance of the central area of the image, which is conducive to the video focus on the feature transformation of the central area of the image, so as to draw a conclusion.

2.1.2 Color space change
HSV color space is a kind of color representation in computer vision. The hue (H) is very similar to human visual angle, and the sensitive program of the precise human eye to color is much weaker than that of the human eye to brightness. Therefore, in order to better understand and analyze color processing, computer vision is usually using HSV color space. In HSV color space, each component is processed separately and is independent of each other. It is not like RGB color space. Although there are three channels, the three channels are worth getting from the color of one pixel. That is to say, the three channels are not independent of each other. The three channels must exist at the same time to form the effect. HSV color space can be described by a conical space model, which is very complex and difficult to describe. However, this model has its unique advantages, that is, it can really clearly show the changes in hue, brightness, and saturation. Therefore, in the HSV color space, the workload of image analysis and processing can be greatly simplified. The combination of HSV hues and saturation is often called chromaticity, which is used to indicate the type and depth of the color. They can be conveniently used in HSV color space. They can be processed separately and are independent of each other. Space and RGB color space are only different representations of the same physical quantity, so there is a transformation relationship between them. The transfer formula is shown in formula (1):

$$h = \begin{cases} \text{undefined}, & if \ \max = \ \min \\ 60^\circ \times \dfrac{G-B}{\max- \ \min} + 0^\circ, & if \ \max = R \ and \ G \geq B \\ 60^\circ \times \dfrac{G-B}{\max- \ \min} + 360^\circ, & if \ \max = R \ and \ G < B \\ 60^\circ \times \dfrac{G-B}{\max- \ \min} + 120^\circ, & if \ \max = R \\ 60^\circ \times \dfrac{G-B}{\max- \ \min} + 240^\circ, & if \ \max = B \end{cases} \tag{1}$$

$$s = \begin{cases} 0, & if \ \max = 0 \\ \dfrac{\max- \ \min}{\max}, & if \ \max \neq 0 \end{cases} \tag{2}$$

$$v = \ \max \tag{3}$$

In formula (1), max and min are the maximum and minimum values in R, G, and B channels. It is easy to convert RGB color space to HSV color space by the above formula.

In experiments, Euclidean distance is often used. Euclidean distance is mainly used to calculate the distance between two points. For the data of multi-dimensional spatial characteristics, it can also be well used. The calculation formula of Euclidean distance is as follows:

$$\text{dist} = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \tag{4}$$

Among them, n is the data dimension, and x_{1i} is the i data of the first point, and x_{2i} is the i data of the second point. But occasionally, dist/n is used instead of dist, because the results calculated from multi-dimensional data may be larger, so the Euclidean distance divided by its dimension is chosen to display the final result.

2.1.3 Algorithm implementation

In order to obtain a more coherent histogram, we divide the component values of HSV color space into larger intervals. In general, the value of the H component is 0 – 360° and the value of S is 0–1. But here, in order to facilitate the calculation of data, we reduce the error and use the following criteria: The H value range is 0~180, and the S value range is 0~255. In the experiment, we divide the H component into 16 intervals, divide the S component into 8 intervals, so that an image can be represented by 128-dimension vectors. Then, the Euclidean distance of 128-dimension vector of color feature corresponding to two frames can be calculated when the similarity of the two frames is compared. The similarity of the two frames can be judged by the threshold value of the distance. If the threshold value is exceeded, the shot switching will occur; otherwise, the shot switching will not occur.

Histogram extraction algorithm is as follows:

1. Read the image file
2. Get image size
3. w = pic.width; 1 = pic.width
4. The cvCvtColor function is called to convert the image to the HSV color space
5. The cvCvtpixplane function is invoked, and the HSV color is output to the specified array according to the three color channels of H, S, and V
6. Create histograms, set them into two dimensions, and divide each dimension equally
7. Histogram statistics for four surrounding areas
8. Calculate the histogram according to the two plane data of H and S, plus the weight of 0.8
9. The histogram of the statistical center region, with a weight of 1.2
10. The histogram is merged together according to the contribution value of the two parts to the image
11. The maximum value of histogram statistics is obtained, which is used to display histogram dynamically
12. Setting histogram and displaying histogram image
13. Output histogram and save histogram data for subsequent processing.

As long as the appropriate threshold of Euclidean distance is selected, the key frame extraction based on non-uniform block HSV histogram method can basically meet the requirements. Therefore, some experiments are needed to obtain the most appropriate distance threshold before the key frame extraction.

2.2 Image preprocessing
2.2.1 Image enhancement
Generally, spatial or frequency domain methods are used in image enhancement. The principle of spatial method is to enhance the image effect by converting the pixels in the image [21–23]. The formula is described as follows:

$$g(x, y) = f(x, y) \times h(x, y) \tag{5}$$

where $f(x, y)$ is the original image, $h(x, y)$ is a space conversion function, and $g(x, y)$ represents the processed image.

Compared with the spatial domain method, the frequency domain method is an indirect image enhancement method. The basic principle of the frequency domain method is to use some algorithms in the image frequency domain to perform some operations on the image transformation value and then change it back to the spatial domain. For example, the Fourier transform is used to transform the image into its frequency domain, and then, some filtering correction is made to the spectrum of the image. Finally, the filtered image is inversely changed to its spatial domain to enhance the image.

Fig. 1 Grayscale of original video frame. **a** Original image. **b** Grayscale processing

In this paper, we use image grayscale transformation, histogram equalization, and image 2 values to enhance image processing.

2.2.2 Image denoising

Because noise pollution seriously affects the quality of images, it is necessary to take appropriate methods to remove the image drying. It is very important to suppress the various interference signals which degrade the image, enhance the useful signals in the image, and correct the observed different images under the same constraint conditions. There are many algorithms for image denoising currently, and there are mainly three classical algorithms.

1. Gauss filtering method [24]: Gauss filter is a kind of linear smoothing filter, which is suitable for filtering Gaussian white noise. Gauss filter method has been widely used in the preprocessing stage of image processing. The pixel value of each pixel in the image is calculated by Gaussian filtering operation, the result is obtained by weighting the pixel gray value and other pixel gray values in the neighborhood of the pixel itself, and the weighted average weighting coefficient is obtained by sampling and normalizing the two-dimensional discrete Gaussian function.

2. Mean filtering method [25]: Mean filtering is also called linear filtering, and its main method of calculation is the neighborhood average method. The basic principle of mean filter is to replace each pixel value of the original image with the mean value, that

is, to treat the current pixel point (x, y), select a template, which is composed of several neighboring pixels, calculate the mean value of all pixels in the template, and then give the mean value to the current pixel point (x, y), which is the gray value $g(x, y)$ of the processed image at this point. That is, $g(x, y) = 1/m \sum f(x, y)$, where m is the total number of pixels in the template containing the current pixels. This method can smooth the image and is fast, and the algorithm is simple.

3. Median filtering method [26]: Median filtering is a nonlinear filtering technique, which sets the gray value of each pixel to the median value of all pixels in a neighborhood window of the point. The main principle is to sort the odd number of data from a sampling window in the image and then take the intermediate value after sorting as the gray value of the current pixel.

2.3 Feature extraction and recognition

The feature extraction stage mainly refers to the process of extracting the eigenvalue of the image input into convolution neural network. In the feature extraction stage, each candidate region will undergo several convolution operations and pooling operations in convolution neural network, so as to compute and extract the target features. Convolution can effectively extract the invariant features from the image, while pooling can reduce the dimension of the input image without changing the features. Finally, the network will use these invariant features to classify the target.
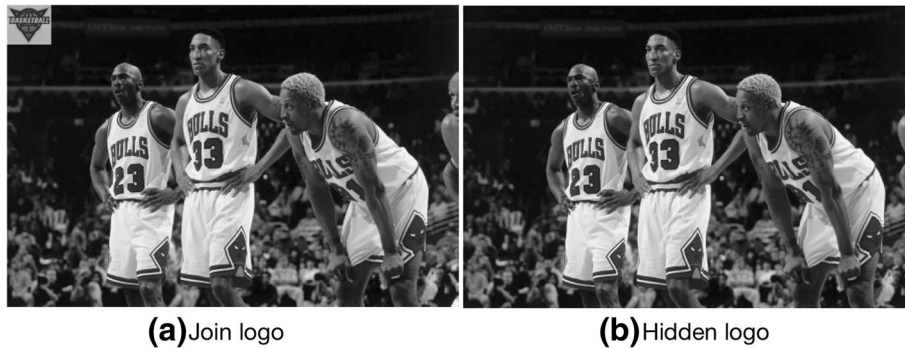

(a) Original image   **(b)**Grayscale processing

Fig. 2 Logo image grayscale. **a** Original image. **b** Grayscale processing

Fig. 3 Video frames are added to the logo and hidden logo. **a** Join logo. **b** Hidden logo

There are three main reasons for using an abstract neural network to abstract feature computation and extraction: Firstly, because of the inherent characteristics of convolution layer and pool layer of convolution neural network, image translation and other operations will not cause changes in convolution characteristics. Secondly, the use of convolution neural network to extract the independent learning features can avoid the use of artificial selection of fixed features, thus avoiding feature extraction as the bottleneck and ceiling of the later enhancement algorithm. Third, the size of eigenvectors of convolution layer, pool layer, and output layer can be controlled freely. The dimension of eigenvectors can be reduced when overfitting occurs in training, and the output of convolution layer can be improved when underfitting. Compared with other networks, convolution neural network has higher flexibility.

The basic network structure of convolution neural network can be divided into four parts: input layer, convolution layer, full connection layer, and output layer. The following sections describe the network parts in detail.

2.3.1 Input layer

Convolution input layer can act directly on the original input data; for the input layer image, the input data is the pixel value of the image.

2.3.2 Convolution layer

Convolution layer of convolution neural network, also known as feature extraction layer, consists of two parts. The first part is the real convolution layer in which the main role is to extract the input data features. The characteristics of each of the different convolution kernels extracting input data are different. The more the convolution layer's convolution kernels, the more features of the input data can be extracted. The second part is the pooling layer, also known as the downsampling layer, in which the main purpose is to retain useful information on the basis of reducing the amount of data processing and speed up the training network. Generally, convolutional neural networks contain at least four layers of convolution (in this case, the real convolution layer and the lower sampling layer are collectively called convolution layer), namely convolution layer, pooling layer, convolution layer, and pooling layer. The more the convolution level, the more abstract features can be extracted on the basis of the previous layer.

2.3.3 Fully connected layer

It can contain several full connection layers, which is actually the hidden layer of multilayer perceptron. Normally, the ganglion points in the posterior layer are connected to each ganglion point in the anterior layer, and there is no connection between the neuron nodes in
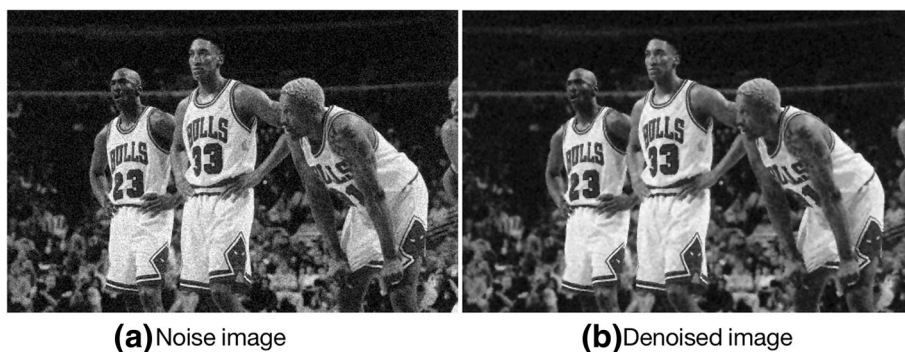
Fig. 4 Image denoising processing. **a** Noise image. **b** Denoised image
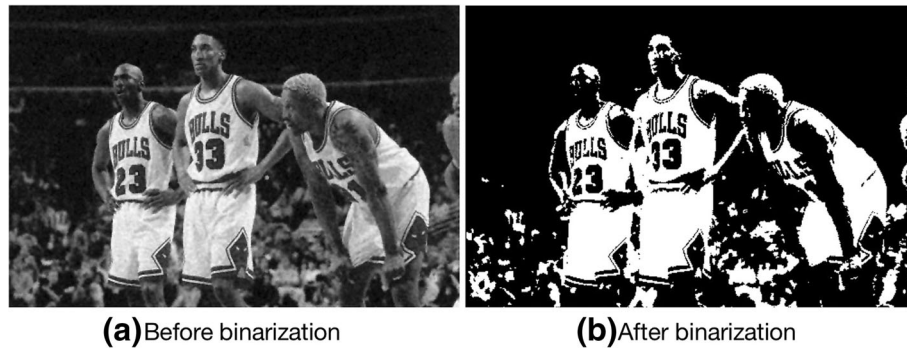
Fig. 5 Image binarization processing. **a** Before binarization. **b** After binarization

the same layer. Each layer of neuron node propagates forward through the weight of the connection line, and the weighted combination obtains the input of the next layer of neuron node.

2.3.4 Output layer

The number of neural nodes in the output layer is set according to specific application tasks. If it is a classification task, the output layer of convolution neural network is usually a classifier.

In order to match the features of the sample logo and realize the logo recognition, the SVM classifier is used for the binary classification operation.

3 Results and discussions

The sports video used in this simulation comes from the network, such as Tencent video, Youku video, and Aichi video. It adds an encrypted logo, and these pictures are downloaded from common media, not a special picture library, so it has a strong universality. The key pixels of the video keyframe are 500×334, and the pixel of logo is 50×45.

Before the simulation experiment, we need to add the logo to the video and encrypt it. First, the original picture of the video frame is grayed out, as shown in Fig. 1. Figure 1a is the original image, and Fig. 1b is the result of the graying process; it can be seen that the image information is basically retained, but the image is reduced from three dimensions to two dimensions. Figure 2 is a grayscale process of the logo designed herein, wherein Fig. 2a is an original image, and Fig. 2b is a result image after grayscale processing.

After graying the video frame image and the logo image, this paper adds the logo image to the video frame image and obtains Fig. 3a, then hides the logo, which is easy to keep a secret and does not affect the viewing effect. The hidden result is shown in Fig. 3b. Combining with Fig 1b and Fig. 3b, it can be seen that after adding logo, after security processing, the gray image of the original image is not different from the naked eye, so it is

feasible to add logo and secret operation, which shows the feasibility of the operation of this paper and lays a foundation for the following experimental simulation.

After the key frame is selected for the video, the image is obtained. In order to highlight the logo feature in the key frame image, the image is strengthened, the unimportant information is weakened or removed, the logo information is strengthened, and the key frame image is transformed into a more useful image, which is convenient for extracting the feature information of the image by using the convolutional neural network.

In the image processing process, the existence of noise is inevitable. In this paper, denoising processing is also needed in the preprocessing process. The result is shown in Fig. 4. It can be seen from the figure that after adding Gaussian noise, the picture information is blurred, which is unfavorable for the subsequent processing of the image, and after the denoising process, the gray image information is effectively restored, thereby avoiding the noise influence during the processing.

After the denoising process, the image is binarized, and the result is shown in Fig. 5, wherein Fig. 5a is an image before binarization processing, and Fig. 5b is a binarized image. It can be seen from Fig. 5 that after the grayscale processing of the image, there are 256 Gy levels, and the appropriate threshold is selected to divide the gray level of the grayscale image into two parts, thereby obtaining the binarization of the image. It keeps the region of interest in the image to the greatest extent and shields all the irrelevant information. In addition, it can be seen that after binarization, the data volume of the algorithm is less, and the sensitive region of the image can be highlighted, and the gray level of the whole image is reduced to the binary

Table 1 Effect of different logo recognition methods

Method	Accuracy (%)
Feature extraction based on convolution neural network	91.7
Feature extraction based on SITF algorithm	88.4
Feature extraction based on template matching	82.1

Fig. 6 Comparison of three kinds of feature extraction methods
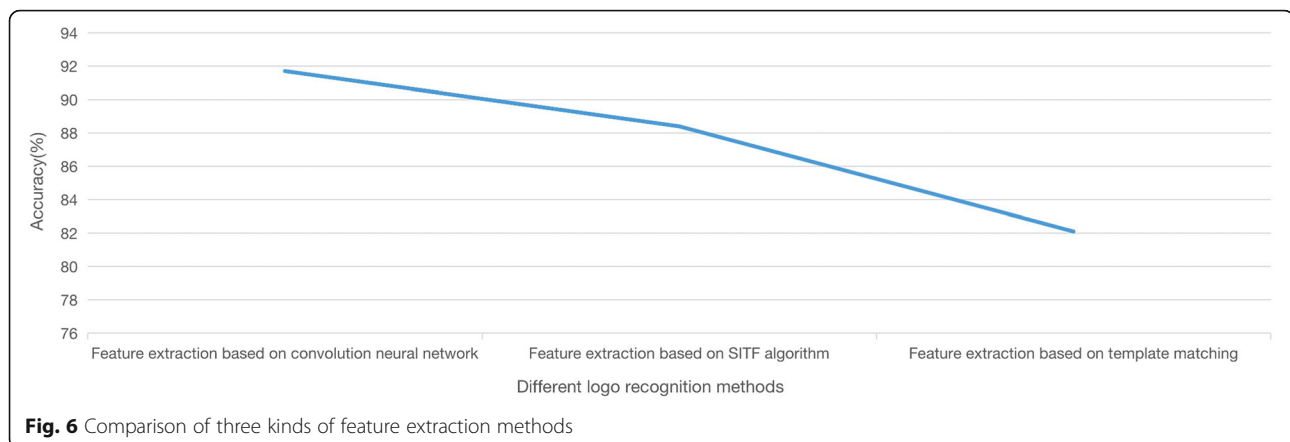
dimension, which makes the subsequent feature extraction algorithm work greatly simplified.

In order to show the superiority of using convolution neural network to extract feature information, this paper compares it with the logo recognition method based on template matching and the logo recognition method based on SIFT algorithm. The data is processed by the average value of 100 times. The comparison result is shown in Table 1. The effect comparison chart is shown in Fig. 6. Combining Table 1 and Fig. 6, we can see that the feature extraction method based on convolution neural network has the highest accuracy, the feature extraction method based on SIFT algorithm takes the second place, and the feature extraction method based on template matching is the worst. SIFT features have scale invariance, many matching points, good performance, more stable, and other excellent performance. However, the accuracy of logo recognition based on convolution neural network feature extraction method is 3.3% higher than that based on template matching method, and the accuracy of logo recognition is 9.6%. It can be seen that the feature extraction using the convolutional neural network has good performance because of the inherent characteristics of the convolutional neural network convolutional layer and the pooled layer, so that the translation of the image does not cause changes in the convolution characteristics. And using the autonomous learning feature extraction of the convolutional neural network can avoid the use of artificially selected fixed features, thereby avoiding the feature extraction becoming the bottleneck and ceiling of the effect of the later lifting algorithm. In addition, it can reduce the dimension of feature vectors when overfitting occurs in training and improve the output of convolution layer when underfitting, which makes convolution neural network more flexible.

A video with a logo is selected to test the encrypted video. The logo is intact with positive samples, and the logo is destroyed with negative samples. Video logo detection
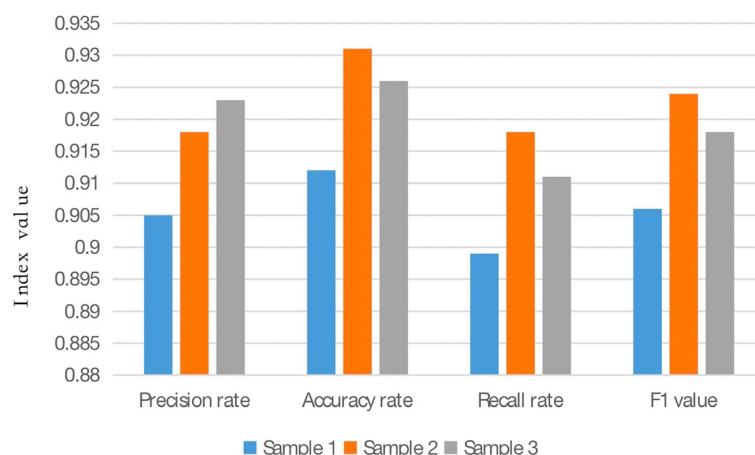
experiment is carried out through three video sample libraries. The positive and negative samples in sample 1 are 100, the positive and negative samples in sample 2 are 500, and the positive and negative samples in sample 3 are 1000. In addition, this paper uses the precision, accuracy, recall, and comprehensive evaluation index F1 to evaluate the performance of the algorithm. The precision, accuracy, and recall rate are all information retrieval. The accuracy rate is the ratio of the relevant samples detected by the system to the total number of samples detected by the system, while the recall rate is the ratio of the relevant samples detected by the system to the total number of samples related to the system. The accuracy rate and recall rate are interrelated. In general, the accuracy rate is how much the retrieved sample is accurate, and the recall rate is how many accurate samples are retrieved. The $F1$ value is a comprehensive index that combines these two indicators to reflect the overall index. Of course, we hope that the higher the accuracy and recall, the better, so that its comprehensive evaluation index $F1$ will be higher. The accuracy rate is expressed in P, and the recall rate is expressed in R. The definition of $F1$ is as follows:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{6}$$

where P denotes the accuracy and R denotes the recall rate.

Table 2 Simulation test results

Sample grouping	The number of positive samples detected correctly	The number of negative samples detected correctly	Precision rate	Accuracy rate	Recall rate	F1 value
1	89	92	0.905	0.912	0.899	0.906
2	457	461	0.918	0.931	0.918	0.924
3	919	927	0.923	0.926	0.911	0.918

Fig. 7 Histogram of test results of performance index

The results of the simulation test are shown in Table 2, and the histogram of four performance indicators is shown in Fig. 7.

Combining Table 2 and Fig. 7, we can see that the performance of three groups of samples is not very different, among which the precision rate of three groups of samples is 91.5%, the accuracy rate is 92.3%, and the recall rate is 90.9%, which shows that the logo recognition copyright detection method based on convolutional neural network has not only good stability, but also good recognition performance. In addition, from the value of the comprehensive evaluation index $F1$, the $F1$ values of the three groups of samples are more than 90%, with an average of 91.6%. This shows that the method of copyright piracy detection designed in this paper is effective. In summary, the analysis shows that the logo recognition copyright detection method based on convolutional neural network designed in this paper has good stability and recognition performance.

4 Conclusions

Nowadays, network video has spread all over the country, but because it is easy to be copied, modified, and disseminated, the problem of copyright theft of network video is very prominent. Aiming at the copyright problem of sports video, this paper proposes a logo recognition technology based on convolution neural network and uses the method of adding encrypted logo to the video to realize the video copyright detection function and automatically and effectively detect whether the video is genuine. In this paper, the analysis algorithm based on non-uniform block HSV histogram is adopted, the region weighting coefficient is introduced to extract the key frame from the video, and the key frame of the video is simulated. The simulation results show that the image preprocessing method proposed in this paper can effectively enhance the image feature information and

highlight the key areas, and by comparing with several commonly used methods for logo recognition and detection, we can see that the logo recognition technology based on convolution neural network proposed in this paper has better performance. Finally, the effectiveness and superiority of this method are illustrated by the simulation of video with encrypted logo.

About the author
Zhi Li was born in Shishou,Hubei, P.R. China, in 1988. He received the Ph.D from Macau University of Science and Technology, Macau China. Now, he studies in Ningbo Institute of Technology, Zhejiang University, and College of Media and International Culture, Zhejiang University. His research interest include Computer network security, International communication, and Healthy Communication.

Authors' contributions
The author LZ wrote the first version of the paper, did part of the experiments of the paper, and revised the paper in different versions. The author read and approved the final manuscript.

Competing interests
The author declares that he has no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. J.R. Jiang, G.L. Amp, The survival and development of anti conventional logo design[J]. Art & Design, 33–35 (2017)
2. C. Yang, Value embodiment of logo design in the clothing brand design[J]. Tianjin Textile Science & Technology, 43–45 (2017)
3. M. Wang, F.H. Chen, S.O. Design, et al., Regeneration of the regional tradition cultural symbol and logo design of Meishan Ecological Culture Park[J]. Packaging Engineering (2016)
4. Lin T, Huang G, Hao S, et al. Application of Scale-Invariant Feature Transform Algorithm in Image Feature Extraction[J]. Journal of Computer Applications, 2016
5. Boulkenafet Z, Komulainen J, Hadid A. Face anti-spoofing using speeded-up robust features and Fisher vector encoding[J]. IEEE Signal Processing Letters, 2017, PP(99):1–1
6. B.A.I. Xuebing, C.H.E. Jin, M.U. Xiaokai, Z.H.A.N.G. Ying, Improved feature points matching algorithm based on speed-up robust feature and oriented fast and rotated brief[J]. Journal of Computer Applications (2016)
7. S. Tian, U. Bhattacharya, S. Lu, et al., Multilingual scene character recognition with co-occurrence of histogram of oriented gradients[J]. Pattern Recogn. **51**(C), 125–134 (2016)
8. Z.G. Liu, Y. Yang, X.H. Ji, Flame detection algorithm based on a saliency detection technique and the uniform local binary pattern in the YCbCr color space[J]. Signal Image & Video Processing **10**(2), 1–8 (2016)
9. A. Huckleberry, A. Püttmann, M.R. Zirnbauer, Haar expectations of ratios of random characteristic polynomials[J]. Complex Analysis & Its Synergies **2**(1), 1–73 (2016)
10. J. Salmen, L. Caup, C. Igel, Real-time estimation of optical flow based on optimized Haar wavelet features[J]. Lect. Notes Comput. Sci **6576**, 448–461 (2011)
11. Y. Bai, Z. Chen, J. Xie, et al., Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models[J]. J. Hydrol. **532**, 193–206 (2016)
12. Li N, Shi J, Gong M. Change detection in synthetic aperture tadar images based on fuzzy restricted Boltzmann machine[J]. 2016:438–444
13. M.-A. Côté, H. Larochelle, An infinite restricted Boltzmann machine[J]. Neural Comput. **28**(7), 1 (2016)
14. B. Zhao, S. Lin, X. Qi, R. Wang, X. Luo, A novel approach to automatic detection of presentation slides in educational videos. Neural Comput. & Applic. **29**(5), 1369–1382 (2018)
15. Mona M. Soliman, Aboul Ella Hassanien, Hoda M. Onsi. An adaptive watermarking approach based on weighted quantum particle swarm optimization. Neural Computing and Applications 27(2): 469–481 (2016)
16. V. Patraucean, A. Handa, R. Cipolla, Spatio-temporal video autoencoder with differentiable memory[J]. Computer Science **58**(11), 2415–2422 (2016)
17. Varior R R, Haloi M, Wang G. Gated Siamese convolutional neural network architecture for human re-identification[J]. 2016:791–808
18. S. Zhang, Z. Wei, Y. Wang, T. Liao, Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary. Future Generation Comp. Syst. **81**(395–403) (2018)
19. P. Moeskops, M.A. Viergever, A.M. Mendrik, et al., Automatic segmentation of MR brain images with a convolutional neural network[J]. IEEE Trans. Med. Imaging **35**(5), 1252–1261 (2016)
20. Li J, Liang X, Shen S M, et al. Scale-aware fast R-CNN for pedestrian detection[J]. IEEE Transactions on Multimedia, 2015, PP(99):1–1
21. A. Dutle, A. Narkawicz, J. Upchurch, Unmanned aircraft systems in the national airspace system: a formal methods perspective[J]. Acm Siglog News **3**(3), 67–76 (2016)
22. L.I. Min-Le, B.I. Da-Ping, D.U. Hao, An optimized airspace search method in radar EW reconnaissance based on MDL-ADT[J]. Electronics Optics & Control (2017)
23. Y. Mao, Y. Gu, C. Xu, Validation of frequency-domain method to compute noise radiated from rotating source and scattered by surface[J]. AIAA J. **54**(4), 1–10 (2016)
24. Shao D, Deng Y, Xiang Y, et al. Speckle reduction based on adaptive Gauss filtering[J]. Journal of Data Acquisition & Processing, 2017
25. Q. Song, M.A. Li, J. Cao, et al., Image denoising based on wavelet transform and mean filtering[J]. Journal of Natural Science of Heilongjiang University (2016)
26. L.I. Yu-Qian, S.U. Guang-Da, Fast implementation of adaptive median filter based on neighborhood processor[J]. Instrument Technique & Sensor (2016)