Postprint: Sensitive Image Detection Method Using Sparse Semantics and Two-Layer Deep Convolutional Neural Network

Authors: Ruxangül Abdurashit, Yasin Eziz, Sun Guozi

Date: 2019-04-01T00:00:00+00:00

Abstract

The rapid development of Internet technology has transformed sensitive content images from originally largely covert content exchange to massive-scale data sharing, rendering traditional sensitive content detection methods based on image feature extraction no longer applicable. To address these challenges, this paper proposes a sensitive content detection method that combines sparse semantics with a dual-layer deep convolutional neural network. The upper-layer network first performs preprocessing of training samples and constructs sparse semantic representations of images as input to the neural network, while the lower-layer network further considers third-party regulatory mechanisms (such as government proxies, etc.) and proposes a sensitive content image detection method for specific groups. Compared with existing commonly used sensitive content image detection methods, the proposed detection method can effectively reduce the number of training samples, and achieves an improvement in detection accuracy of over 7% compared with traditional image detection methods (such as the bag-of-visual-words approach, etc.).

Full Text

Preamble

Vol. 37 No. 5
Application Research of Computers

Sensitive Image Detection Method Based on Sparse Semantics and Double-Layer Deep Convolutional Neural Network

Ruxianguli · Abudurexiti[1], Yasen · Aizezi[1]†, Sun Guozi[2]
(1. Dept. of Information Security Engineering, Xinjiang Police College, Urumqi 830013, China;

2. Institute of Computer Technology, Nanjing University of Posts & Telecommunications, Nanjing 210003, China)

Abstract: With the rapid development of Internet technology, sensitive content images have transitioned from basic concealed content exchange to massive data sharing, rendering traditional sensitive content detection methods based on image feature extraction no longer applicable. To address these challenges, this paper proposes a sensitive content detection method that combines sparse semantic representation with a double-layer deep convolutional neural network. The upper network first performs preprocessing of training samples and constructs a sparse semantic representation of images as input to the neural network, while the lower network further considers third-party control mechanisms (such as government proxies) and proposes a detection method for sensitive content images targeted at specific groups. Compared with existing common sensitive content image detection methods, the proposed method can effectively reduce the number of training samples while improving detection accuracy by more than 7% compared to traditional image detection methods (such as the bag-of-visual-words approach).

Keywords: sensitive image detection; double-layer artificial neural network; deep learning algorithm; sparse semantic representation; visual word bag; skin detector

---

0 Introduction

In recent years, the rise and development of social networking platforms have greatly facilitated the global dissemination of media data through advances in communication and computer technologies. However, these technological developments have also made it easier for individuals to access image information containing sensitive content. For specific groups, particularly, such images can cause serious and irreversible negative influences on their development.

Since the mid-1990s, sensitive content images have undergone tremendous changes in generation, dissemination, and storage, evolving from basically concealed content exchange to large-scale massive data sharing. Therefore, accurate identification and detection of sensitive image content on the Internet is of significant research importance. In recent years, numerous researchers have proposed algorithms for sensitive image content detection. For example, Reference [1] utilized depthwise separable convolutional neural networks and the MobileNet model, combined with a GPU parallel computing architecture, to establish an identification model with high accuracy for sensitive images, but the proposed method struggled to effectively identify seemingly harmless and highly concealed sensitive images. Reference [2] proposed a data-oriented sensitive image detection method based on deep convolutional neural networks (CNN) to extract image features, using batch stochastic gradient descent to train the CNN. However, the trade-off between training accuracy and speed

in batch gradient descent algorithms relies on manual empirical settings and cannot guarantee obtaining a global optimal solution. Reference [3] proposed a fast filtering algorithm for sensitive images based on Support Vector Machine (SVM), employing a hybrid skin color model for rapid detection of exposed skin regions to extract facial position, shape, and image background features. However, SVM algorithms are not suitable for scenarios with large sample sizes, as they impose high requirements on computer memory and computation time. Similar related work on sensitive content image detection can be found in References [4-7]. Furthermore, the aforementioned literature does not consider sensitive content detection technology under government regulation, meaning they cannot perform manual intervention detection on sensitive content images.

To address these problems, this paper proposes a sensitive content image detection method based on sparse semantics and double-layer deep convolutional neural networks. The upper layer employs general training samples to train the artificial neural network, achieving preliminary classification of images into three categories: non-sensitive, specific-group, and seemingly harmless. The lower layer introduces an appropriate government proxy mechanism to achieve specialized adjustments for detecting sensitive content images targeted at specific groups. Practical examples demonstrate that the proposed sensitive image detection method can more effectively reduce training sample quantities compared to existing solutions while lowering detection errors. The proposed manual intervention mechanism enables flexible control by government agents (such as law enforcement agencies and professional departments).

Deep convolutional neural networks for image classification achieve high detection accuracy, but the requirement for massive amounts of raw training data in the initial training phase constrains their application in sensitive content image detection. For specific sensitive groups, identification and detection of sensitive content images require not only focusing on image classification but also implementing stricter control and screening of image content, which represents the current challenge in deep learning-based sensitive content image detection for specific groups.

1 Sparse Semantic Representation of Images

To facilitate subsequent neural network training and eventual classification of images into sensitive and non-sensitive content categories, this paper first performs sparse representation of image features. For a set containing k image features $\{f_1, f_2, ..., f_k\}$, assuming the codebook generated from the i-th (i=1,2) class of training images is $A_i = [w_{i1}, w_{i2}, ..., w_{iC_i}] \in \mathbb{R}^{128 \times C_i}$, where C_i is the number of words in the codebook for the i-th image feature class. For a sample image $y_0 \in \mathbb{R}^{128}$, it can be represented as a linear approximation of the codebook in A_i, i.e., $y_0 \approx \sum_{j=1}^{C_i} a_{ij} w_{ij} = A_i a_i$. More generally, for K classes of images, we have $A = [A_1, A_2, ..., A_K]$, and y_0 is linearly represented as $y_0 = Aa$, where $a = [a_1, a_2, ..., a_K]$, and a is a C_1-dimensional sparse solution vector that

satisfies the following optimal solution:

$$\hat{a} = \arg\min \|y_0 - Aa\|_2 \quad \text{s.t.} \quad \|a\|_0 \leq k$$

where the L_0 norm $\|\cdot\|_0$ measures the number of non-zero elements in the solution vector a. When the solution vector a is sufficiently sparse, the optimal linear representation using feature codebooks becomes solving for the minimum l_1 norm of a. Therefore, the problem can be transformed into the following optimization problem:

$$\hat{a} = \arg\min \|y_0 - Aa\|_2^2 + \lambda\|a\|_1$$

According to Lagrange' s theorem, the optimization problem in Equation (6) can be further solved by optimizing the following objective function:

$$E(a, \lambda) = \|y_0 - Aa\|_2^2 + \lambda\|a\|_1$$

where parameter $\lambda > 0$ is a balancing factor between the approximation term and the sparsity term.

The physical meaning of sparse semantic representation is that the non-zero coefficient values in vector a reflect the correlation between the image sample and the constructed image feature codebook. The stronger the correlation, the better the effect of describing image features using that class of codebook, meaning the reconstruction error in Equation (8) is smaller; conversely, the reconstruction error is larger.

2.1 Problem Analysis

Traditional image recognition algorithms distinguish subtle differences between images by extracting color, shape, texture, and other manually described image features. However, applying such methods to screen harmless and non-sensitive content is extremely difficult. A convenient approach is to detect the data volume size of Internet downloads and related services for labeling. To address these practical problems, this paper proposes the following solution: based on the fundamental deep learning algorithm architecture, but abandoning the traditional approach of training data samples from scratch, the weight coefficients of each neuron node in the neural network are initialized using side data from source problems. Subsequently, the trained samples (i.e., after weight coefficient adjustment) are fed back to the target problem, thereby improving detection accuracy for subtle differences between sensitive content images. Following this approach, this paper proposes a hierarchical image detection scheme based on deep convolutional neural networks. The upper layer is the weight coefficient determination stage, which uses deep convolutional neural networks to classify

and detect general images, preliminarily determining the weights of each neuron node in the neural network. The lower layer introduces a government agent control mechanism to manually fine-tune the weight coefficients for detecting sensitive content images targeted at specific groups and retrains the neural network. An obvious advantage of the proposed method is that since the upper neural network has already specialized in network weight initialization, the lower layer requires a more streamlined set of training samples.

2.2 Double-Layer Deep Convolutional Neural Network Architecture

The basic architecture of the proposed double-layer deep CNN is shown in Figure 1 [Figure 1: see original paper]. The first layer network is the source problem solving network, whose basic objective (i.e., source task) is to perform reasonable classification of images. Generally, the source task of image detection requires classifying images into 1000 daily categories covering pets, household items, food, transportation, etc. The proposed algorithm structure includes an initialization module, nine processing modules, two auxiliary classifiers, and a final classifier. Figure 2 [Figure 2: see original paper] shows the technical details of each module. The main functions of each module are explained as follows:

The initialization module primarily processes the input raw images through operations such as convolution, pooling, and local response normalization. This paper employs a 7×7 convolution kernel to parse the raw images, followed by gradual data quantization processing using convolution, normalization, and data pooling operations. After image initialization, the obtained data is learned and trained through seven core feature learning modules, each of which can be regarded as a small neural network capable of detecting specific image features and attributes in the image classification source task.

In addition to the seven feature learning modules, this paper also sets up two auxiliary classifiers after the third and sixth feature learning modules to amplify gradient signals through the network and achieve highlighting of early feature data. Similar to the feature learning modules, these two auxiliary classifiers are also small convolutional networks. Furthermore, the auxiliary classifiers in this paper include a softmax classification decision layer and a data reduction layer through average pooling with a 5×5 kernel. It should be noted that auxiliary classifiers are only used during training and are removed during testing.

The basic function of the final classifier is classification. In general source problems, the final classifier includes a data reduction layer through average pooling with a kernel, fully connected layers, and a linear layer with softmax loss to achieve identification of 1000 image types. However, considering the two target problems in this paper (i.e., non-sensitive content and specific-group sensitive content image detection), this paper designs the final classifier using a bidirectional nonlinear SVM with a Radial Basis Function (RBF) kernel, because this classifier achieves higher detection and identification rates when performing

sensitive content image detection for specific groups.

2.3 Upper-Layer Source Problem Network Training

The neural network training samples for the source problem require approximately 1.2 million images covering 1000 different categories. This process employs asynchronous stochastic gradient descent optimization with momentum (set to 0.9) and a fixed learning rate (decreasing the learning rate by 4% every 8 epochs). During the source problem network training phase, the observation information loss of each image from the two auxiliary classifiers, multiplied by a weight coefficient of 0.3, is added to the observation information loss of the final classifier. To collect more training samples, the initial pool is dynamically expanded using operations such as rotation, mirroring, and cropping, but excluding scaling and photometric distortion. Additionally, this paper employs a polynomial learning rate decay strategy instead of stepwise learning rate decay to achieve four times or even faster training speed.

The training steps for the upper neural network are as follows: for a given set of training input samples, all images are first resized while maintaining aspect ratio, with all input sample images larger than the set input threshold (224×224 pixels); subsequently, images are further cropped to the shape required by the specific convolutional network. It is worth mentioning that the preprocessing of input sample images is identical for both the source problem' s basic image classification and the specific-group image detection classification.

2.4 Transfer Learning and Adjustment Strategy for Target Problems

After the trained neural network can accurately classify daily images among the 1000 target categories, this paper further proposes fine-tuning the learning weights for the target detection problem, with the goal of using the same network architecture for sensitive image detection for specific groups. In this paper, the neuron weights of the lower-layer target problem solving network are not trained from scratch but are initialized using a two-way softmax classifier (i.e., outputting non-sensitive images or specific-group sensitive content), with fine-tuning of initial weights and backpropagation using a series of images labeled as non-sensitive or specific-group sensitive. Additionally, the two auxiliary classifiers also adopt two-way softmax classification functions.

When the network with readjusted weights can accurately detect non-sensitive content images (i.e., when the network converges), the final classification function is replaced with an SVM classifier with an RBF kernel. This operation enables the deep learning neural network to have weights specialized for non-sensitive content image detection and allows the neural network to work as a feature extractor for non-sensitive content image classification, while the SVM classifier learns the specific problem-solving model on these features. Since the

final goal of this paper is to detect sensitive images for specific groups, non-sensitive content image detection can serve as a test case for the neural network' s reliability. By inputting instances of specific-group sensitive images and non-specific-group sensitive images, image feature extraction and SVM classifier training for specific-group sensitive image detection are performed. The specific training steps are as follows:

a) Given an image set containing specific-group sensitive content and non-specific-group sensitive content, the trained neural network is used to extract image features, and these example images are used to train a nonlinear SVM classifier. To train the SVM classifier, this paper employs a 5-fold cross-validation strategy on the available training set and uses a grid search approach to find optimal classification parameters $C \in \{2^c : c \in [-5, -3, -1, ..., 15]\}$ and $\gamma \in \{2^i : i \in [15, 13, ..., 3]\}$. The final obtained first set of SVM classifier parameters is $\gamma = 0.0078125$ and $C = 8.0$. This detection network can be called a primary detector, capable of detecting non-sensitive content images and, to some extent, specific-group sensitive content images, but the neural network at this stage still does not possess final detection capability for sensitive content images for specific groups.

b) After adjusting the initial image classification neural network to a sensitive content classification detection neural network, this paper focuses on further fine-tuning the neural network to enable detection of sensitive content images for specific groups, proposing two solution approaches. For the first approach, starting from the network trained for object classification, its learned weights are fine-tuned to receive a series of images labeled as sensitive and non-sensitive content, with backpropagation using the images as input. Consistent with the aforementioned content, when the weights obtained from converged network training can achieve detection for the target problem (i.e., specific-group sensitive content image detection), the final layer classifier is replaced with an SVM classifier with an RBF kernel. This second-layer neural network is called the first-type specific-group sensitive content image detector (referred to as Scheme 1), because the target-specific group sensitive content image detection is directly adjusted from the source problem (image category classification → specific-group sensitive content image detection).

For the second solution approach, this paper starts from the network fine-tuned for non-sensitive content detection and further fine-tunes its weights to receive a series of images labeled as SEIC and non-SEIC content, again using backpropagation. Consistent with the aforementioned content, when the weights obtained from converged network training can achieve detection for the target problem (i.e., specific-group sensitive content image detection), the final layer classifier is replaced with an SVM classifier with an RBF kernel. This second-layer neural network is called the second-type specific-group sensitive content image detector (referred to as Scheme 2), because this approach uses a two-step weight adjustment process (image category classification → non-sensitive content detection

→ specific-group sensitive content image detection).

For a given set of specific-group sensitive content images or non-specific-group sensitive content images, this paper uses the trained neural network from Scheme 1 or Scheme 2 to extract image features and employs these image instances to train a nonlinear SVM classifier. For the training process, this paper also uses the available dataset for 5-fold cross-validation and employs a grid search approach to find optimal classification parameters $C \in \{2^c : c \in [-5, -3, -1, \dots, 15]\}$ and $\gamma \in \{2^i : i \in [15, 13, \dots, 3]\}$. The final calculated first set of SVM classifier parameters is $\gamma = 0.0078125$ and $C = 0.5$.

2.5 Sample Data Augmentation

Even when fine-tuning neural network weights rather than training from scratch, a large amount of sample data is still required. Therefore, for the proposed double-layer artificial neural network, this paper proposes the following solution to achieve sample data augmentation during weight coefficient readjustment. Since the dataset used for fine-tuning rarely contains actual negative examples, the initial training further enhances non-sensitive specific-group content, specifically non-sensitive images. Therefore, instead of generating additional images from original images (such as rotation, mirroring, or contrast adjustment), the dataset used in this paper adopts 20,000 images from the Microsoft Common Objects Dataset, containing 16,000 images across 91 common categories. For the first-layer source problem solving, this paper does not perform data augmentation; for the second-layer target problem solving, this paper conducts experiments with and without sample data augmentation.

3.1 Dataset

To verify the effectiveness of the proposed image classification and detection algorithm, this paper selects the Pornography-2k dataset [8] for experimental analysis. The dataset includes nearly 140 hours of videos, with 1000 containing sensitive content and 1000 not containing sensitive content, with video lengths ranging from 6 seconds to 33 minutes. This paper further decomposes the original videos into image frames, obtaining a total of 58,971 images, of which 33,723 images contain sensitive content, resulting in proportions of 57% sensitive images and 43% non-sensitive images. Figure 3 [Figure 3: see original paper] shows some example frames from the selected dataset. According to the algorithm steps described in the previous section, this paper divides the dataset into three types: training set (containing 33,646 images), validation set (containing 5,938 images), and test set (containing 19,387 images).

For sensitive content image classification and detection, the training set is first used to pre-train the convolutional neural network to obtain its key parameters. The validation dataset is then used to optimize and adjust these parameters. Finally, the test set is used for algorithm effectiveness verification and analysis. This paper employs a standard random 5×2 cross-validation mechanism,

evaluating algorithm effectiveness through random dataset splitting.

3.2 Classification Metrics

Using the default image classification metrics from the Pornography-2k dataset, namely the normalized metric ACC and the F2 metric. ACC represents the percentage of correctly classified images, while the F2 metric refers to the weighted harmonic mean between precision and recall, as shown in Equation (9), where $\beta = 2$.

$$F_2 = (1 + \beta^2) \times \frac{\text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}}$$

3.3 Forensic Tools

To evaluate the classification performance of the proposed scheme, forensic tools relying on visual content are selected, including NuDetective [9], MediaDetective [10], and Snitch Plus [11]. NuDetective is a readily available tool for research purposes that can detect sensitive image content, while MediaDetective and Snitch Plus are commercial solutions that enable relatively concentrated detection of sensitive image content. All these tools are based on skin detectors to identify sensitive content images.

For MediaDetective and Snitch Plus, image files are rated according to their degree of suspicion for sensitive image content (i.e., probability). If the returned probability for an image is equal to or greater than 50%, the image is marked as sensitive. NuDetective assigns binary labels to images: positive (i.e., the image is sensitive) or negative (i.e., the image is non-sensitive).

MediaDetective and Snitch Plus have four predefined execution modes, primarily depending on the strictness level of the skin detector. In these experiments, the strictest execution mode is selected, while default settings are used for NuDetective.

3.4 Skin Detector

Currently, various methods have been proposed to detect sensitive image content, but human skin detection technology [12], especially color-based detection, remains challenging in terms of recognition simplicity and accuracy improvement. Existing methods primarily define skin regions in RGB color space to propose skin classification. Skin classification models can be broadly divided into three criteria:

Criterion 1:
$R > 95 \wedge G > 40 \wedge B > 20 \wedge \max(R, G, B) - \min(R, G, B) > 15 \wedge |R - G| > 15 \wedge R > G \wedge R > B$

Criterion 2:
$R > 220 \wedge G > 210 \wedge B > 170 \wedge |R - G| \leq 15 \wedge R > B \wedge G > B$

Criterion 3:
$\max(R, G, B) - \min(R, G, B) > 15$

where R, G, and B represent pixel values in RGB color space, ranging from 0 to 255. For convenient comparison, this paper selects a widely used and easily implementable skin detector that classifies images as sensitive if the percentage of skin pixels in the image is equal to or greater than 13%, performing cross-validation on the training set to select the optimal threshold.

3.5 Bag of Visual Words

Before the introduction of deep learning technology, the Bag of Visual Words (BoVW) [13] modeling method was the most commonly used approach for describing image content, representing images as a space of quantized local descriptors.

This paper evaluates two classic BoVW methods: the hard coding method and the average pooling method. First, image samples are preprocessed to obtain a dataset with a maximum of 100,000 pixels. Then, local feature quantities (using Speeded-Up Robust Features (SURF) descriptors) are extracted using grids of 24×24, 32×32, 48×48, 68×68, and 96×96 pixels generated with strides of 4, 6, 8, 11, and 16 pixels, respectively. Principal Component Analysis (PCA) is used to reduce the dimensionality of SURF to 32 dimensions.

The visual vocabulary is learned from over 1 million randomly sampled PCA descriptors using a K-means clustering tool based on Euclidean distance, extracting 2,048 visual words, while the classifier employs an SVM classifier. By default, this paper uses a nonlinear SVM classifier with an RBF kernel and employs a grid search approach to find optimal classification parameters $C \in \{2^c : c \in [-5, -3, -1, \ldots, 15]\}$ and $\gamma \in \{2^i : i \in [15, 13, \ldots, 3]\}$. The final calculated SVM classifier parameters are $\gamma = 0.0078125$ and $C = 32$.

4 Experiments and Results Discussion

To verify the superiority and effectiveness of the proposed method in sensitive image detection, experiments are designed to compare with other currently popular detection methods, forensic tools, and commercial software, analyzing experimental results based on sensitive image content detection. This paper implements the proposed algorithms using Matlab 2014a, with a computer configuration of: Windows 10 operating system, 2 GB memory, 256 GB hard disk, CPU model I5-3230M (2.6 GHz, dual-core), and GPU model Nvidia GeForce GT 740M (2 GB).

4.1 Non-Sensitive Content Detection Results

First, the Pornography-2k dataset (totaling 257,522 frames) is used to analyze different non-sensitive content image detection performances. The detection method proposed in Reference [14] is selected for comparison with the proposed method. Table 1 shows the ACC metric and F2 metric for both methods, demonstrating that the proposed method achieves better image detection accuracy than the method in Reference [14]. In terms of computational resources, the proposed method is also more economical: Reference [14] uses a cluster of 16 GPUs for neural network training, while this paper uses only a single GPU. Additionally, although both networks achieve similar detection performance for actual non-sensitive content detection, the proposed method' s detection accuracy for intercepting sensitive content images is 3.5% higher than that of Reference [14].

Table 1 Sensitive Content Detection Results

Method	TPR	TNR	ACC	F2
Reference [14] Method	-	-	-	-
Proposed Method	-	-	-	-

TPR denotes True Positive Rate; TNR denotes True Negative Rate; ACC denotes Accuracy; F2 denotes F2 Measure

4.2 Specific-Group Sensitive Image Detection

Further considering the final goal of this paper—sensitive content image detection—to demonstrate the superiority of the proposed method, commonly used methods shown in Table 2 are selected for comparison. As shown in Table 2, the proposed skin detector-based system achieves considerable performance improvement compared to other systems, with significantly enhanced detection accuracy for specific-group sensitive content.

For non-sensitive content image classification performance, compared with BoVW and the algorithm developed by Yahoo in Reference [14], the ACC and F2 metrics are improved by 7.7% and 6.5%, respectively. For specific-group sensitive content image classification performance, the proposed classification algorithm improves the ACC and F2 metrics by 8.6% and 7.0%, respectively, compared to Reference [14].

Figure 4 [Figure 4: see original paper] shows histograms of image detection error rates for different classifiers. Compared with the skin detection method, the error rate of the proposed specific-group sensitive content detection method is reduced from 46.9% to 3.9%, representing a 70.4% error reduction. Compared with the BoVW-based method, the error rate is reduced by 56.2%, demonstrating that image detection accuracy based on BoVW technology is higher than that of skin classifier detection, while the detection accuracy based on deep

convolutional neural networks can further reduce image false detection rates. Additionally, from Table 2 and Figure 4, it can be observed that the proposed detection method for specific-group sensitive content can further improve the detection accuracy of the neural network for non-sensitive content detection.

Table 2 Results of Sensitive Image Content Detection

Method	Non-Sensitive Content Classifier	Specific-Group Sensitive Content Classifier
BoVW [15]	-	-
Reference-[14]		-

5 Conclusion

Although deep convolutional neural networks for image classification achieve high detection accuracy, the requirement for massive amounts of raw training samples in the initial training phase constrains their application in sensitive content image detection. To address this problem, this paper proposes a solution combining sparse semantics with deep convolutional neural networks and introduces a double-layer neural network algorithm for optimizing neuron weight coefficients when solving target problems. Simulation results demonstrate that the proposed method achieves significant improvements in detection accuracy compared to conventional algorithms, with detection performance for specific-group sensitive content images superior to conventional detection methods based on bag-of-visual-words and skin classification.

References

[1] Xing Yanfeng, Zhuo Wenxin, Duan Hongxiu. Design of sensitive image recognition system based on MobileNet [J]. Television Technology, 2018, 42(7): 53-56.

[2] Yu Mingyang, Yang Peng, Wang Yijun. Pornographic image detection based on convolution neural network [J]. Computer Applications and Software, 2018, 35(1): 232-236.

[3] Zhou Jianzheng, Chen Faye, Yao Jinliang. A network bad image filtering method based on SVM [J]. Computer application and software, 2012, 29(5): 251-253.

[4] Wang Jingzhong, Zhou Jing. Research on network bad image filtering algorithm based on proportional feature [J]. Computer Engineering and Science, 2016, 38(3): 514-519.

[5] Bissias G, Levine B, Liberatore M, et al. Characterization of contact offenders and child exploitation material trafficking on five peer-to-peer networks [J]. Child Abuse & Neglect, 2015, 52: 185-199.

[6] Liu Xingwang, Wang Jiangqing, Xu Ke. A hybrid algorithm combining AutoEncoder and CNN for image feature extraction [J]. Application Research of Computers, 2017, 34(12): 3839-3843.

[7] Hu Changsheng, Zhan Shu, Wu Zhongzhong. Image super-resolution reconstruction based on depth feature learning [J]. Journal of Automation, 2017, 43(5): 814-821.

[8] Jia Xu, Sun Fuming, Li Haojie, et al. Universal improved non-negative matrix decomposition image feature extraction method [J]. Computer applications, 2018, 38(1): 233-237.

[9] Ren Xiaofang, Qin Jianyong, Yang Jie, et al. Application of LS-TSVM based on energy model in human motion recognition [J]. Application Research of Computers, 2016, 33(2): 598-601.

[10] Perez M, Avila S, Moreira D, et al. Video pornography detection through deep learning techniques and motion information [J]. Neurocomputing, 2016, 230(C): 279-293.

[11] Kanthan S, Graham J A, Azarchi L. Media detectives: bridging the relationship among empathy, laugh tracks, and gender in childhood [J]. Journal of Media Literacy Education, 2016, 8(2): 35-53.

[12] Niezgoda S, Way T P. SNITCH: a software tool for detecting cut and paste plagiarism [J]. ACM SIGCSE Bulletin, 2006, 38(1): 51-55.

[13] Chen Xiaoxue, Wei Yongqing, Ren Min, et al. Weighted K-means algorithm based on firefly optimization [J]. Application Research of Computers, 2018, 35(2): 466-470.

[14] Mahadeokar J, Pesavento G. Open sourcing a deep learning solution for detecting images [EB/OL]. (2016-09-30). https://noise.getoto.net/2016/09/30/open-sourcing-a-deep-learning-solution-for-detecting-nsfw-images/.

[15] Wang Qi, Chen Yimin, Ding Youdong. Vehicle recognition algorithms based on visual word bag model in complex environment [J]. Computer Applications, 2018, 38(5): 1299-1303.

Note: Figure translations are in progress. See original paper for figures.