Research Article

Violation Detection of Live Video Based on Deep Learning

Chao Yuan^{1,2} and Jie Zhang^{1,3}

¹College of Economics and Management, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Avenue, Nanjing 211100, China

²School of Design, Jiangnan University, 1800 Lihu Avenue, Wuxi, Jiangsu 214122, China

³Nanjing University of the Arts, No. 74 Beijing West Road, Nanjing, Jiangsu 210013, China

Correspondence should be addressed to Jie Zhang; jiezhang@nuaa.edu.cn

Received 2 February 2020; Accepted 16 April 2020; Published 11 May 2020

Academic Editor: Chenxi Huang

Copyright © 2020 Chao Yuan and Jie Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of Internet technology, live broadcast industry has also flourished. However, in the public network live broadcast platform, live broadcast security issues have become increasingly prominent. The detection of suspected pornographic videos in live broadcast platforms is still in the manual detection stage, that is, through the supervision of administrators and user reports. At present, there are many online live broadcast platforms in China. In mainstream live streaming platforms, the number of live broadcasters at the same time can reach more than 100,000 people/times. Only through manual detection, there are a series of problems such as low efficiency, poor pertinence, and slow progress. This approach is obviously not up to the task requirements of real-time network supervision. For the identification of whether live broadcasts on the Internet contain pornographic content, a deep neural network model based on residual networks (ResNet-50) is proposed to detect pictures and videos in live broadcast platforms. The core idea of detection is to classify each image in the video into two categories: (1) pass and (2) violation. The experiments verify that the network proposed can heighten the efficiency of pornographic detection in webcasts. The detection method proposed in this article can improve the accuracy of detection on the one hand and can standardize the detection indicators in the detection process on the other. These detection indicators have a certain promotion effect on the classification of pornographic videos.

1. Introduction

At present, Internet technology and Internet of Things technology have achieved rapid development, and various products and consumption models based on these technologies have been deeply applied in various industries [1–12]. People's online entertainment has undergone tremendous changes. From simple video watching and single text reviews to video and commentary barrage videos, it has finally developed into today's barrage live broadcast. The interactive form is constantly approaching offline interaction, and the whole country has entered the "live +" era. According to network statistics, in 2015, China had nearly 200 online live broadcast platforms. The number of live broadcasts has reached 200 million, and the number of users in the online live broadcast industry has reached 300 million.

Large-scale live broadcast platforms have nearly 4 million simultaneous online users during peak hours every day, and the number of live broadcasts at the same time exceeds 3000. In 2016, the number of online live broadcast platforms in China increased to 300 and the number of users in the online live broadcast industry reached 300 million. By 2017, the number of live broadcast industry users has reached 500 million. With the vigorous development of live broadcast industry, the illegality of live broadcast content has become increasingly serious. In terms of live webcast video detection, it is still in the manual detection stage, mainly through the supervision of administrators and user reports. The main problems with manual supervision are (1) increased labor costs; (2) prone to missed inspections due to limited human energy; and (3) the huge number of live broadcast platforms and live video, resulting in a huge workload of manual

supervision. Relying solely on manual supervision cannot achieve the goal. In addition, there are a large number of online live broadcast platforms; the content is complex, and the workload of pure human supervision is too large to meet the regulatory requirements of live broadcast platforms. At present, there is an urgent need for pornographic detection methods for online live broadcast.

At present, various methods are used for pornographic image detection, and there are two classification methods in total. One is to divide the detection methods into the following three categories. The first is a rule-based approach [13]. This method first establishes a skin model that can filter nonskin areas and then obtains the skin area from the image to be identified. If it is larger than the threshold, it is considered to be an erotic picture. Although this method is intuitive and easy to implement, the skin area threshold is difficult to set accurately. Many nonerotic images also contain a large amount of skin area, so the accuracy of identifying pornographic images is low. The second method is based on image retrieval [14]. This method first establishes a database with pornographic images, then selects the appropriate image features, and then compares the image features to be identified with the image features in the database. If the similarity exceeds a certain threshold, it is determined as a pornographic image. Due to the diversity of pornographic images, in order to improve the accuracy, a huge database needs to be established. This will lead to excessive memory usage and long recognition time. The third is a learning-based approach [15, 16]. This method needs to design the visual features of pornographic images, and based on these visual features, use the method of machine learning [17–27] to obtain the learned model. In the end, the learned model is used for pornographic image recognition. It can be found through comparison that, although the learning-based method has higher accuracy and detection speed among the three methods, the image features obtained are generally artificially generalized. However, the generalization performance of artificially selected visual features is low and it is difficult to meet the actual requirements.

The second general classification divides pornographic recognition methods into the following three categories: pornographic recognition based on skin color, pornographic recognition based on manually extracted features, and pornographic recognition based on deep learning. The research based on skin color mainly focuses on the classification of pornographic pictures with a large number of skin pixels. Fprsyth and Fleck [28] identify skin tone region of image based on color and texture features to analyze the structure of the human body and the correlation between various parts of the human body, thereby implementing a naked recognition system. Yang and Ahuja [29] proposed using a Gaussian mixture model to learn the distribution of skin pixels of the training image under the CIELuv color space. The learned model is applied to skin pixel detection in the test image. Jones and Rehg [30] introduced the concept of histogram, extracting the skin color regions of erotic pictures and normal pictures and counting the color distribution histograms of their corresponding RGB color

spaces. According to the histogram, the distribution of human skin color in the RGB color space is obtained. According to this method, features such as skin color area, maximum skin color connected area, and average skin color probability are calculated. The neural network classifies the features to achieve pornographic detection of the image. Srisaan [31] and Basilio et al. [32] extract skin color regions in the YCbCr color space and HSV color space, respectively. In both studies, skin color detection is achieved by counting the distribution of skin color pixels. To sum up, the detection of pornographic images based on skin color detection relies on the judgment criteria of skin color pixels. Skin color is easily affected by light, and it is difficult to use a simple skin color detection for high-accuracy pornographic images with complex textures. Nonerotic images containing a large number of skin-colored pixels are prone to misjudgment during the detection process. Therefore, it is not reliable to use skin color as the criterion for identifying pornographic images. Pure skin color detection has certain limitations in pornographic image recognition. To enhance recognition rate, more features in image need to be extracted for multidimensional analysis. Therefore, pornographic recognition based on manual feature extraction has appeared. The framework mainly includes two links: feature extraction and classifier, as shown in Figure 1.

Pornographic recognition based on manually extracted features mainly uses different feature extraction algorithms to extract prominent features such as color, texture, and shape and uses different classifiers for the extracted features to realize pornographic image recognition. Karavarsamis et al. [33] obtained the skin color convex hull region of the image by ROI localization, calculated 15 features such as the expectation and variance of the convex hull region and the ratio of nonskinned pixels in the convex hull region in the RGB color space. Finally, the random tree forest was used for classification. Wang et al. [34] proposed a model for identifying nude images based on navel and torso features. Deselaers et al. [35] proposed to extract image information based on visual bag of words (BoW) model, feature vectors in image blocks were used as bag of words, and SVM was used to classify visual word histograms.

Pornographic image recognition based on deep learning mainly classifies pornographic images or videos through convolutional neural network (CNN) and LSTM network models. Moustaf [36] uses the NPDI [37] dataset for fine tuning on the AlexNet [38] and GoogleNet [39] models. The two models are mixed according to different thresholds, and finally the classification is achieved. The model is shown in Figure 2. Wehrmann et al. [40] proposed the Adult Content Recognition with deep neural network model, which is used to extract the features of key frames in the NPDI dataset. Long- and short-term memory networks are used to classify the extracted features. Perez et al. [41] divided video information into static information and dynamic information. Video frames are fed into a CNN model to extract static information, and optical flow and MPEG motion vectors are used to describe video motion information. The motion information is sent to a CNN to extract video motion features. Finally, support vector machines [42] are used to

FIGURE 1: Flowchart of pornographic recognition based on manually extracted features.

classify the extracted features. Pornographic image detection based on deep learning has higher accuracy than skin color and feature-based detection methods.

ResNet-50 learning model based on residual networks is introduced in this study and uses this learning model to detect and classify pictures and videos. The advantages of the model are as follows: (1) it has higher recognition accuracy and efficiency, and the accuracy reaches more than 95%; (2) due to the addition of a residual module in this model, there is relatively little information loss in the process of information transfer between layers, which can avoid the problem that it is difficult for deep networks to accept the information of the previous layers; (3) the model is no longer limited by the dimensions of the image; and (4) by improving the ordinary CNN, the generalization ability of the model after adding the residual network is improved. In this paper, the proposed network model is applied to pornographic content detection on live broadcast platforms, and experiments verify the effectiveness of the proposed scheme. The specific contributions of this article are as follows:

- (1) The detection algorithm based on deep residual network can process high-dimensional pictures, and the accuracy and efficiency of recognition are better than other similar algorithms.
- (2) Generally, as the number of layers of the learning network increases, the learning efficiency will become lower and lower. The advantage of the residual network is that it is very simple to perform identity transformation in the network structure, reducing the defined network structure by the increase of the number of network layers so that the information in the network structure can reduce the transmission loss as much as possible.
- (3) By improving the network model, the recognition accuracy and efficiency of the two types of pictures are further improved.

FIGURE 2: AGNet erotic image classification model.

2. Related Work

2.1. Picture Detection. Pictures are one of the main information carriers that people come into contact with daily, and it is necessary to detect pornographic pictures on the Internet. There are many types of erotic pictures, and the most common of which are pictures of naked bodies. Therefore, some scholars have proposed that the area of exposed skin and the area of all bodies in the picture can be calculated to find the proportion of the area of exposed skin. Given a certain threshold range to determine whether a picture violates the rules, this is a very simple method of judgment, but the disadvantages are also obvious. The detection of pictures is not targeted and the accuracy is difficult to guarantee.

Given a certain number of samples, mark whether these samples are illegal pictures. Before classification, the image must be preprocessed, and the features of the image are extracted and compared using a convolution layer to determine whether the image is illegal. Such an algorithm is more accurate than threshold classification. The specific process is shown in Figures 3 and 4.

2.2. Video Detection. Video is another representation of multimedia and one of the information carriers often seen in people's lives, but the detection of video is relatively tedious. The basic unit of a video is a frame, that is, a picture, which means that a video is composed of many pictures. The video detection process can be broken down into the following steps: first divide the video into frame-by-frame pictures, then preprocess each frame of pictures, and finally classify each frame of pictures. If an offending picture is detected, the administrator can be notified and the offending video can be processed in a timely manner. The flow chart of video detection is shown in Figure 5.

FIGURE 3: Image preprocessing flowchart.

3. Way of Solving the Problem

3.1. Network Structure. For the detection of illegal videos, the use of deep learning networks is feasible. However, as the depth of the network gradually increases, the efficiency of feature extraction will become lower and lower and the efficiency of learning will gradually decrease. In summary, a deep-learning network alone cannot complete the task well. To address these challenges, this article proposes a new solution; using ResNet-50 [43] based on residual networks, both deep-layer networks can be used to ensure the characteristics of the extracted pictures and as far as possible in the deep-layer network structure so that the defined network structure is not affected by the increase in the number of network layers, thereby improving learning efficiency and accuracy.

3.2. Image Preprocessing and Feature Extraction. Feature extraction the core factor for effective classification. Before feature extraction, a high-quality picture needs to be selected. Pick some pictures with sharp textures and sharp edges. Cut the picture to a suitable size and denoise the picture through a filter. In order to make the effect of feature extraction better, this article performs image enhancement on all pictures. Common methods are contrast raising, histogram equalization, etc. The pictures are divided into two types; the first type is the training set and the second type is the test set. ResNet-50 is used for feature extraction on the training set in deep-learning network. This process is to tell the detection system whether some marked pictures are illegal pictures or qualified pictures, thereby facilitating the detection of the test set and the accuracy of the detection. It directly affects the quality of the detection system. The image processing flow is shown in Figure 6.

3.3. Image Dimensionality Reduction. High-dimensional data is a challenge for detection. Assuming that the dimensions of the image are $512 * 512$, the corresponding one-dimensional vector calculation is 262144 dimensions.

FIGURE 4: Picture classification flowchart.

FIGURE 5: Video preprocessing flowchart.

Such dimensions make it difficult to extract features. Dimension reduction is an essential link. The reduction of dimensions is mainly done by the pooling layer (sampling layer). There are many commonly used feature extraction methods, and the classical ones are the CNN [44], the principal component analysis (PCA) method [45] and kernel principal component analysis [46] (KPCA) and so on. The pooling layer can be directly used to reduce the dimension, and no additional image-processing steps are required. The pooling layer (sampling layer) mainly depends on the size of the convolution kernel and the stride of the convolution kernel to reduce the dimensionality. The reduced dimensional image is easier to process.

FIGURE 6: Image processing flowchart.

3.4. Residual Network. For a common network structure, when the number of layers is appropriately increased within a certain range, the detection performance can be improved. However, it is not allowed to arbitrarily increase the number of layers of CNN. Continuously increasing the number of layers of the neural network will inevitably cause the accuracy on the training set to saturate or even decrease. When the number of neural network layers continues to increase, it will cause problems such as disappearance or degradation of gradients, gradient explosions, and network training costs, so the detection performance cannot be improved as the number of network layers increases. Based on this background, this paper introduces residual networks. It reduces the number of calculations and computational difficulty. For the residual network, the identity transformation of deep network is used to make the deep network degenerate into a relatively shallow network. The residual network mainly can perfectly solve the side effect caused by the deep network, namely, the degradation problem; then, we can improve the network performance by increasing the network depth.

3.5. Picture Classification. Classify the frames in the video, and the classification results are qualified and illegal. In the residual network structure, a fully connected layer is used to complete the classification. Fully connected layer processes and stores the data and uses it as input to the layer. Then, output a multidimensional vector, which is the number of classifications. Since this experiment is a binary classification problem, the multidimensional vector in the experiment is a two-dimensional vector. In the experiment, softmax [47] was used for discrimination. Since the result of the softmax output is a probability, the probability is used to represent the detection result in a two-dimensional vector. Figure 7 shows specific identification process.

FIGURE 7: Identification flowchart.

4. Introduction to Detection Algorithms

In order to verify the performance of the proposed detection method, this paper proposes five comparison models, which are (1) pornographic image recognition algorithms based on skin color features; (2) direction gradient histogram (HOG) + support vector machine (SVM); (3) CNN; (4) CNN + residual network algorithm; and (5) detection algorithm based on VGG-16 network. This paper proposes a network structure based on ResNet-50 and introduces five comparison algorithms first.

4.1. Contrast Algorithm 1: Pornographic Image Detection Algorithm Based on Skin Color Features. According to the author's research on pornographic images, erotic images are pictures with more naked body parts. Based on the survey, the author first set a certain threshold and then calculated the proportion of naked leakage in each picture. If the proportion of naked leakage is greater than the set threshold, the picture is considered to be a violation. This method will have certain results when identifying the offending pictures. However, this method has some drawbacks. If the clothes worn by the person in the picture are close to the skin, the recognition result will have a greater impact, and it is likely that the picture is illegal.

4.2. Contrast Algorithm 2: Image Detection Algorithm Based on Directional Gradient Histogram (HOG) and Support Vector Machine (SVM). HOG [48] is the histogram of direction gradient. This method is currently a method in the field of computer vision and pattern recognition. This method can be used to describe the local texture characteristics of the image. The specific description steps are as follows:

Step 1: segment the image.

According to the strategy of segmenting images, there are two kinds of segmentation strategies; the first is nonoverlap and the second is overlap, which means that features can be extracted as many times as possible after image segmentation. This experiment uses the overlap strategy.

Step 2: calculate the direction gradient histogram of the segmented block.

The specific calculation formula is as follows:

$$M(x, y) = \sqrt{I_x^2 + I_y^2},$$

$$\theta(x, y) = \tan^{-1} \frac{I_y}{I_x} \in [0, 360^\circ), \quad (1)$$

$$\text{or } \in [0, 180^\circ).$$

Among them, I_x and I_y represent the gradient values in the horizontal and vertical directions, $M(x, y)$ represents the magnitude of the gradient, and $\theta(x, y)$ represents the direction of the gradient.

Step 3: composition characteristics.

According to the block segmentation of the image, the results of the block segmentation are connected end-to-end to form the characteristics that describe the image; then, these characteristics are the basis for our judgment of image classification, which is the role of the image example mentioned earlier.

SVM is support vector machine. Take binary classification as an example. The data participating in binary classification are called positive samples and negative samples, respectively. If positive and negative samples are given in the experiment, the role of the support vector machine is to find a "hyperplane." Classify the positive and negative samples as well as possible so that the positive and negative samples are distributed as uniformly as possible on the hyperplane.

4.3. Contrast Algorithm 3: Image Detection Algorithm Based on Convolutional Neural Network. CNN is an excellent network structure among many neural networks. For CNN, the experiments in this paper mainly use convolutional layers, pooling layers (sampling layers), and fully connected layers. The CNN is specifically introduced below.

For the convolutional layer, the features of the original image are extracted by setting the size of the convolution kernel and the step size during the convolution. After the convolution, the dimension of the image will be slightly reduced according to the size of the convolution kernel.

The primary role of the pooling layer or sampling layer is to reduce the dimensions and reduce the image; secondly, simplify the complexity of the network, reduce the amount of calculation, and reduce memory consumption; at the same time, expand the receptive field to achieve translation invariance, rotation invariance, and scale invariance; implement nonlinearity and more.

The fully connected layer maps the learned "distributed feature representations" into the sample space and acts as a "classifier" in the entire CNN to find out whether the picture is specifically qualified or unqualified.

4.4. Contrast Algorithm 4: Image Detection Algorithm Based on Convolutional Neural Network and Residual Network. Through the improvement of the CNN, the residual network part is added to form a new algorithm. The network structure is introduced below. Since the CNN was introduced above, only the residual part is introduced in this section.

When we use a standard optimization algorithm to train a common network, such as using gradient descent (or other popular optimization algorithms), without using residuals, or without these shortcuts or jump connections, you will find that when the network depth deepens, mistakes first decrease and then increase. However, in theory, as the network depth deepens, the better the training effect is. However, in fact, if there is no residual network, for a normal network, the deeper the depth, the harder it is to train with an optimization algorithm. Moreover, as the depth of the network deepens, training errors will increase.

However, the situation is different by using ResNets. Even if the network is deeper, the training performance remains stable, and the training error is reduced, even if the network is deep to 100 layers. Some people have even performed experiments in 1000-layer neural networks, which allow us to guarantee good performance while training deeper networks. As the network gets deeper, the scale of the network connection will become huge, but ResNet is indeed very effective in training deep networks.

4.5. Contrast Algorithm 5: Image Detection Algorithm Based on VGG-16. This model has a total of 16 layers of network structure, so it is called VGG-16 [49]. It was proposed by Oxford University in 2014, and its network is simple and practical. Taking a $224 \times 224 \times 3$ picture as an example, the specific calculation process is as follows:

Step 1: input a $224 \times 224 \times 3$ picture, and after two convolutions with 64 convolution kernels, use one pooling. After the first convolution, c1 has $(3 \times 3 \times 3)$ trainable parameters.

Step 2: after two convolution kernel convolutions of 128, one pooling is used.

Step 3: after three convolutions of 256 convolution kernels, use pooling.

Step 4: repeat the three 512 convolution kernels twice and perform pooling again.

Step 5: the last entry is three full connections.

4.6. Algorithm in This Paper: Image Detection Algorithm Based on ResNet-50. The ResNet algorithm proposed in 2015 is relatively simple and very practical. Based on this, ResNet-50 or ResNet-101 is derived, and many methods are constructed based on this. Detection, segmentation, recognition,

TABLE 1: Key parameters of the experiment.

Experimental parameters	Number of initial convolution kernels	Training times	Learning efficiency	Image size	Classification number	Classification
Value	64	10,000	$1e-4$	208×208	2	Softmax

TABLE 2: Detection results of 6 algorithms.

Experiment name	Accuracy $p1$ (%)	Recall rate $p2$ (%)	Accuracy $p3$ (%)	Misclassification $p4$ (%)	F1 score (%)
CNN	85.39	88.24	86.26	14.61	86.79
CNN + ResNet	92.51	91.91	92.37	7.87	92.21
HOG + SVM	81.65	80.15	83.21	18.35	80.89
Skin color features	72.50	74.56	70.15	27.50	73.52
VGG-16	—	—	—	—	—
ResNet-50	97.08	97.79	96.18	2.92	97.43

and other application scenarios are beginning to use ResNet. All this shows that ResNet is really easy to use and has great potential for development. Theoretically, with the deepening of the network layers, the accuracy of image classification should gradually increase, but in practice, with the deepening of the network, the accuracy of the training set decreases. However, it is definitely not the case of the network due to overfitting because overfitting will cause the training set to have a high-accuracy rate. Traditional convolutional networks or fully connected networks will have more or less problems of information loss and loss when transmitting information and also cause “gradient disappearance” or “gradient explosion,” resulting in deep network training. ResNet solves the abovementioned problems to a certain extent. By using the idea of bypassing the information at the input to the output, the integrity of the information can be protected. The entire network only learns the difference between input and output, thereby streamlining learning goals and reducing complexity.

In view of the abovementioned situation, the residual network will have a relatively better result. The residual unit establishes a direct correlation channel at the data input and data output ends. This makes a powerful network structure focus on learning the residuals between input and output. For example, generally we use $F(X, W_i)$ to represent the residual mapping, then the output is $Y = F(X, W_i) + X$. When the number of input and output channels is the same, we can naturally use X to add directly. When the number of channels between them is different, we need to consider establishing an effective mapping function so that the number of processed input X and output Y channels is the same, that is, $Y = F(X, W_i) + W_s \times X$.

According to the abovementioned introduction of ResNet, it can solve the problem of the deepening of the recognition accuracy of the network. One of the reasons is that the residual network can directly transfer shallow data information to relatively deep levels through associated channels, unlike traditional networks. Shallow layers of information need to be transmitted backward layer by layer. When the information is transmitted to the deeper layer, the information in the ordinary network is bound to cause losses. This explains why the recognition accuracy of ordinary networks decreases as the network deepens.

ResNet-50 is a 50-layer network structure based on residual networks, including 50-layer network structures such as convolutional layers, pooling layers, and residual networks. It is a relatively mature network model. In this paper, the network model is used to detect illegal videos. According to the experimental results, ResNet-50 will have a relatively good effect.

5. Experiment and Analysis

5.1. Dataset. The dataset is used as the basis for data identification. A data-rich dataset has a decisive effect on the recognition result. The dataset in this article contains a training dataset and testing one. In the training dataset, two types of pictures are classified and labeled and used as training. Then, we use the test dataset to check the trained network. Since there is no official library for this kind of dataset, all the datasets are used to find images by themselves, filter the images, and perform a series of pre-processing processes on the images to make ordinary images a usable dataset.

The dataset contains a training set: 280 images, of which 140 are qualified pictures and 140 are unqualified pictures. The test set contains 267 images, of which 136 are qualified images and 131 are unqualified images. For data of the training set, image enhancement is required. All the images of the training set are enhanced trying to avoid the phenomenon of network “overfitting.” By transforming the training image to obtain a network structure with stronger generalization ability, it is also a means to enhance the network when the number of datasets is not large. There are several methods for data enhancement:

- (1) Image enhancement using standardization
- (2) Use geometric transformation (translation, flip, and rotation) to enhance the data of the image
- (3) Use random adjustment of brightness to enhance the image
- (4) Use random adjustment contrast to enhance the image

FIGURE 8: Example 1 of violation results detected.

5.2. Evaluation Criteria of Experimental Results. For the test results, the following evaluation criteria have been formulated.

Criterion 1. Accuracy $P(1)$ = number of correctly identified pictures/number of participating pictures $\times 100\%$.

However, the recognition accuracy $P(1)$ cannot be simply used as the evaluation criterion. First of all, a single evaluation standard will bias the evaluation results, and if for a wrong classifier, the detection results of all pictures are qualified, then when the test set of qualified pictures is detected, it will show excellent performance. Then, we cannot consider such a classifier to be a good classifier, and such a classifier is often not a qualified classifier. As a result, the classification accuracy rate of qualified pictures and violation pictures and the misclassification rate of the overall classification of the classifier appear, respectively. The specific definitions are as follows.

Criterion 2. Recall rate $P(2)$ = number of correctly identified qualified pictures/number of participating qualified pictures $\times 100\%$.

Criterion 3. Recognition accuracy $P(3)$ = number of correctly identified offending pictures/number of participating in identifying offending pictures $\times 100\%$.

Criterion 4. Misclassification rate $P(4)$ = number of mis-predicted pictures/total number $\times 100\%$.

Criterion 5. F_1 is the harmonic mean of precision and recall, and the specific formula is as follows:

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \quad (2)$$

where precision is the accuracy rate, that is, the proportion of the number of correctly identified violation pictures in the identification of violation pictures. Recall is the recall rate, that is, the proportion of correctly identifying qualified pictures among the proportion of participating in identifying qualified pictures.

5.3. Experimental Parameters. The key parameters used in the experiment are shown in Table 1. The number of initial convolution kernels determines the number of features extracted by the convolution layer, so as many features as possible can be extracted to facilitate the subsequent detection and classification steps, so this study uses 64 initial convolution kernels. Learning efficiency is the speed of network learning. Due to the number of trainings, the learning efficiency can be set smaller. If you set a large learning efficiency, as the number of learning times increases, it will be difficult to reduce the loss function, and it will always oscillate in an interval.

FIGURE 9: Example 2 of violation results detected.

FIGURE 10: Continued.

FIGURE 10: Example 3 of violation results detected.

5.4. Experimental Environment. The compiler used in this experiment is Matlab and Pycharm; operating system is Windows 10, 64 bit system, Linux 16.5; version of language is Matlab2018a, python3.7, and running environment is Intel Core i7 CPU, 32 GB RAM.

5.5. Experimental Results and Analysis. According to the evaluation criteria of the abovementioned experiments, the three models are now tested separately, and the test results of 6 models are shown in Table 2.

It can be seen from the multiple indicators in Table 2 that ResNet-50 has better recognition effect than other algorithms. In this paper, three illegal videos are selected and continuous frames are detected for the three videos. The specific video detection results are as follows.

The detection result of violation video 1 is shown in Figure 8.

The detection result of violation video 2 is shown in Figure 9.

The detection result of violation video 3 is shown in Figure 10.

6. Conclusions and Improvements

6.1. Conclusions. With the rapid development of online live broadcast, live broadcast has become an indispensable part of people's entertainment and enriches people's lives. In practical application scenarios, this research can be used to detect illegal video on live broadcast platforms and assist network police in intelligent detection of pornographic videos. Violation video detection can be added as a module to the network public security big data platform to capture real-time video from the network in real time and detect whether the video is illegal. If a video violation is detected, an alert will be issued immediately to assist the public security in maintaining the network environment. However, the quality and quality of live broadcast does not have a good detection platform. It currently stays at the stage of manual detection. Rely on manual testing. Then, the content and quality of the live broadcast will be uneven, especially the pornographic live broadcast, which will affect the user's

visual experience and will endanger the health of children. Therefore, this paper proposes a new detection strategy based on the residual network (ResNet-50), which detects the live broadcast content of the anchor and improves the efficiency of detecting pornographic live broadcasts. As can be seen from the results, the accuracy of the detection of illegal and qualified pictures has reached more than 95%. The detection strategy proposed in this paper is better than traditional methods, especially for deep networks. The addition of residual network modules will increase the network's generalization level and reduce the complex calculation process of deep networks. ResNet-50 is a relatively mature network structure. According to the experimental results, it can be seen that the network model performs relatively well for various indicators of video detection.

6.2. Outlook and Improvement. Judging from the detection results and experimental accuracy, the algorithm in this article, based on the ResNet-50 network, shows good performance for video detection, but this algorithm also has some problems. Because many datasets are not public and the datasets are not complete, the accuracy of detection needs to be further improved. This article focuses on some detections of the video screen during live broadcast, and the target area is relatively small. It may also be possible to detect and recognize voice and text simultaneously in the future. For these problems, we can make related improvements in the future to improve the detection system. In traditional convolutional neural networks, convolution kernel size must be set in each convolution. In other words, if the size of the convolution kernel is variable, it may be useful to detect, rather than when it is fixed.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported in part by the National Social Science Fund Project under Grant 17BGL102; Excellent Project of Jiangsu Province Social Science Union under Grant 15SYC-043; Soft Science Research of Wuxi Science and Technology Association under Grant KX15-B-01; and Fundamental Research Funds for the Central Universities under Grant 2015ZX18.

References

- [1] H. Gao, W. Huang, and X. Yang, "Applying probabilistic model checking to path planning in an intelligent transportation system using mobility trajectories and their statistical data," *Intelligent Automation and Soft Computing (Autosoft)*, vol. 25, no. 3, pp. 547–559, 2019.
- [2] H. Gao, W. Huang, Y. Duan, X. Yang, and Q. Zou, "Research on cost-driven services composition in an uncertain environment," *Journal of Internet Technology (JIT)*, vol. 20, no. 3, pp. 755–769, 2019.
- [3] X. Ma, H. Gao, H. Xu, and M. Bian, "An IoT-based task scheduling optimization scheme considering the deadline and cost-aware scientific workflow for cloud computing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, 2019.
- [4] H. Gao, Y. Duan, L. Shao, and X. Sun, "Transformation-based processing of typed resources for multimedia sources in the IoT environment," *Wireless Networks*, 2019.
- [5] H. Gao, Y. Xu, Y. Yin, W. Zhang, R. Li, and X. Wang, "Context-aware QoS prediction with neural collaborative filtering for internet-of-things services," *IEEE Internet of Things Journal*, 2019.
- [6] J. Yu, L. Jing, Y. Zhou, and Q. Huang, "Multimodal transformer with multi-view visual representation for image captioning," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.
- [7] J. Yu, M. Tan, H. Zhang, D. Tao, and Y. Rui, "Hierarchical deep click feature prediction for fine-grained image recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [8] J. Yu, C. Zhu, J. Zhang, Q. Huang, and D. Tao, "Spatial pyramid-enhanced NetVLAD with and weighted triplet loss for place recognition," *IEEE Transactions on Neural Networks and Learning Systems*, 2019.
- [9] Y. Yin, J. Xia, Y. Li, Y. Xu, W. Xu, and L. Yu, "Group-wise itinerary planning in temporary mobile social network," *IEEE Access*, vol. 7, pp. 83682–83693, 2019.
- [10] Y. Chen, S. Deng, H. Ma et al., "Deploying data-intensive applications with multiple services components on edge," *Mobile Networks and Applications*, vol. 25, no. 2, pp. 426–441, 2019.
- [11] C. Zhang, H. Zhao, and S. Deng, "A density-based offloading strategy for IoT devices in edge computing systems," *IEEE Access*, vol. 6, pp. 73520–73530, 2018.
- [12] L. Kuang, X. Yan, X. Tan, S. Li, and X. Yang, "Predicting taxi demand based on 3D convolutional neural network and multi-task learning," *Remote Sensing*, vol. 11, no. 11, p. 1265, 2019.
- [13] J. Pei, J. Han, B. Mortazavi-Asl et al., "PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth," in *Proceedings of the International Conference on Data Engineering*, p. 215, IEEE Computer Society, Washington, DC, USA, 2001.
- [14] J. Ayres, J. Flannick, J. Gehrke et al., "Sequential pattern mining using a bitmap representation," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, Edmonton, Canada, pp. 429–435, July 2002.
- [15] B. Le, M. T. Tran, and D. Tran, "A method for early pruning a branch of candidates in the process of mining sequential patterns, intelligent information and database systems, ACIIDS 2017," *Lecture Notes in Computer Science*, p. 10191, Springer, Berlin, Germany, 2017.
- [16] M. Zihayat, Y. Chen, and A. An, "Memory-adaptive high utility sequential pattern mining over data streams," *Machine Learning*, vol. 106, no. 6, pp. 799–836, 2017.
- [17] P. Qian, Y. Jiang, Z. Deng et al., "Cluster prototypes and fuzzy memberships jointly leveraged cross-domain maximum entropy clustering," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 181–193, 2015.
- [18] P. Qian, Y. Jiang, S. Wang et al., "Affinity and penalty jointly constrained spectral clustering with all-compatibility, flexibility, and robustness," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 5, pp. 1123–1138, 2016.
- [19] P. Qian, K. Zhao, Y. Jiang et al., "Knowledge-leveraged transfer fuzzy C-means for texture image segmentation with self-adaptive cluster prototype matching," *Knowledge-Based Systems*, vol. 130, pp. 33–50, 2017.
- [20] P. Qian, C. Xi, M. Xu et al., "SSC-EKE: semi-supervised classification with extensive knowledge exploitation," *Information Sciences, Information Sciences*, vol. 422, , pp. 51–76, 2018.
- [21] P. Qian, S. Sun, Y. Jiang et al., "Cross-domain, soft-partition clustering with diversity measure and knowledge reference," *Pattern Recognition*, vol. 50, pp. 155–177, 2016.
- [22] P. Qian, J. Zhou, Y. Jiang et al., "Multi-view maximum entropy clustering by jointly leveraging inter-view collaborations and intra-view-weighted attributes," *IEEE Access*, vol. 6, pp. 28594–28610, 2018.
- [23] P. Qian, F. L. Chung, S. Wang, and Z. Deng, "Fast graph-based relaxed clustering for large data sets using minimal enclosing ball," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 3, pp. 672–687, 2012.
- [24] Y. Jiang, D. Wu, Z. Deng et al., "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2270–2284, 2017.
- [25] Y. Jiang, Z. Deng, F.-L. Chung et al., "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 1, pp. 3–20, 2017.
- [26] Y. Jiang, F. L. Chung, S. Wang, Z. Deng, J. Wang, and P. Qian, "Collaborative fuzzy clustering from multiple weighted views," *IEEE Transactions on Cybernetics*, vol. 45, no. 4, pp. 688–701, 2014.
- [27] Y. Jiang, F. L. Chung, H. Ishibuchi, Z. Deng, and S. Wang, "Multitask TSK fuzzy system modeling by mining intertask common hidden structure," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 534–547, 2015.
- [28] D. A. Forsyth and M. Fleck, "Finding naked people," in *Proceedings of the 4th European Conference on Computer Vision*, no. 2, pp. 593–602, Cambridge, UK, April 1996.
- [29] M.-H. Yang and N. Ahuja, "Gaussian mixture model for human skin color and its application in image and video database," *SPIE Storage and Retrieval for Image and Video Database*, vol. 3656, pp. 45–466, 1999.
- [30] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *International Journal of Computer Vision*, vol. 1, no. 46, pp. 81–96, 2002.

- [31] C. Srisaan, "A classification of Internet pornographic images," *International Journal of Electronic Commerce Studies*, vol. 7, no. 1, pp. 95–104, 2016.
- [32] J. A. M. Basilio, G. A. Torres, G. S. Perez et al., "Explicit content image detection," *Signal & Image Processing*, vol. 1, no. 2, pp. 47–58, 2011.
- [33] S. Karavarsamis, N. Ntarmos, K. Blekas, and I. Pitas, "Detecting pornographic images by localizing skin ROIs," *International Journal of Digital Crime and Forensics*, vol. 5, no. 1, pp. 39–53, 2013.
- [34] X. Wang, C. Hu, S. Yao et al., "An adult image recognizing algorithm based on naked body detection," in *Proceedings of the ISECS International Colloquium on Computing, Communication, Control, and Management*, IEEE, Sanya, China, pp. 197–200, August 2009.
- [35] T. Deselaers, L. Pimenidis, and H. Ney, "Bag-of-visual-words models for adult image classification and filtering," in *Proceedings of the 19th International Conference on Pattern Recognition*, IEEE, Tampa, FL, USA, pp. 1–4, December 2008.
- [36] M. N. Moustafa, "Applying deep learning to classify pornographic images and videos," in *Proceedings of the Pacific-rim Symp on Image and Video Technology, PSIVT*, Auckland, New Zealand, November 2015.
- [37] *NPDI Pornography Database*, <https://sites.google.com/site/pornograohydatabase/>, 2013.
- [38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolution neural networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, NIPS, Lake Tahoe, NV, USA, 2012.
- [39] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Boston, MA, USA, pp. 1–9, June 2015.
- [40] J. Wehrmann, G. S. Simões, R. C. Barros, and V. F. Cavalcante, "Adult content detection in videos with convolutional and recurrent neural networks," *Neurocomputing*, vol. 272, no. 272, pp. 432–438, 2018.
- [41] M. Perez, S. Avila, D. Moreira et al., "Video pornography detection through deep learning techniques and motion information," *Neurocomputing*, vol. 230, no. 230, pp. 279–293, 2017.
- [42] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, Berlin, Germany, 1995.
- [43] L. V. Fulton, D. Dolezel, J. Harrop, Y. Yan, and C. P. Fulton, "Classification of alzheimer's disease with and without imagery using gradient boosted machines and ResNet-50," *Brain Sciences*, vol. 9, no. 9, 2019.
- [44] I. Amerini, C.-T. Li, and R. Caldelli, "Social network identification through image classification with CNN," *IEEE Access*, vol. 7, pp. 35264–35273, 2019.
- [45] A. S. Hess and J. R. Hess, "Principal component analysis," *Transfusion*, vol. 58, no. 7, pp. 1580–1582, 2018.
- [46] L. Xie, J. Z. Li, U. Zeng, and U. Kruger, "Block adaptive kernel principal component analysis for nonlinear process monitoring," *AIChE Journal*, vol. 62, no. 12, pp. 4334–4345, 2016.
- [47] Y. Zhou, X. Wang, M. Zhang, J. Zhu, R. Zheng, and Q. Wu, "MPCE: a maximum probability based cross entropy loss function for neural network classification," *IEEE Access*, vol. 7, pp. 146331–146341, 2019.
- [48] H. Nezamabadi-pour and E. Kabir, "Image retrieval using histograms of uni-color and bi-color blocks and directional changes in intensity gradient," *Pattern Recognition Letters*, vol. 25, no. 14, pp. 1547–1557, 2004.
- [49] Z. Kolar, H. Chen, and X. Luo, "Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images," *Automation in Construction*, vol. 89, pp. 58–70, 2018.