Video Quality Diagnosis System Based on Convolutional Neural Network

Hu Yi(✉) and Xiaodong Zhan

Training Center Beijing Polytechnic, Beijing 100176, China
huyi@bpi.edu.cn

Abstract. With the rapid development of modern society, people demand higher and higher performance of various products, there are many quality problems in the process of practical application. Therefore, in order to improve user experience and improve this situation this paper proposes a video quality diagnosis system based on convolutional neural network. The design includes various construction methods, several main framework structures and related databases. This paper takes the video quality during video conferencing as the research object, hopes to build a video quality diagnosis system using the theory of convolutional neural network.

Keyword: Convolutional Neural Network · Video Quality Diagnosis System

1 Introduction

Convolutional neural networks have achieved unprecedented success in the field of computer vision, many traditional computer vision algorithms have been replaced by deep learning convolutional neural networks. Meanwhile, with the development of network technology, video conferencing system has gradually become a mainstream and important way to disseminate multimedia information. In this paper, we use the convolutional neural network model to improve the video quality problems encountered in video conferencing, aim to establish a video quality diagnosis system based on convolutional neural network to improve the quality of video conferencing, so as to achieve the purpose of improving efficiency.

2 Convolutional Neural Networks in Brief

In recent years, convolutional neural networks have become more and more complex as the number of researchers in the field related to convolutional neural networks has increased and the technology is changing day by day. From the initial 5-layer, 16-layer, to the 152-layer ResNet proposed by MSRA [3] or even thousands of layers networks have become commonplace by a wide range of researchers and engineering practitioners.

A simple convolutional neural network is composed of various layers arranged in a sequence, each layer in the network uses a differentiable function to pass data from

J. C. Hung et al. (Eds.): IC 2023, LNEE 1045, pp. 591–597, 2023.
https://doi.org/10.1007/978-981-99-2287-1_85

one layer to the next. Convolutional neural networks consist of three main types of layers: convolutional layers, pooling layers, fully connected layers. By stacking these layers together, a complete convolutional neural network can be constructed. As shown in Fig. 1.

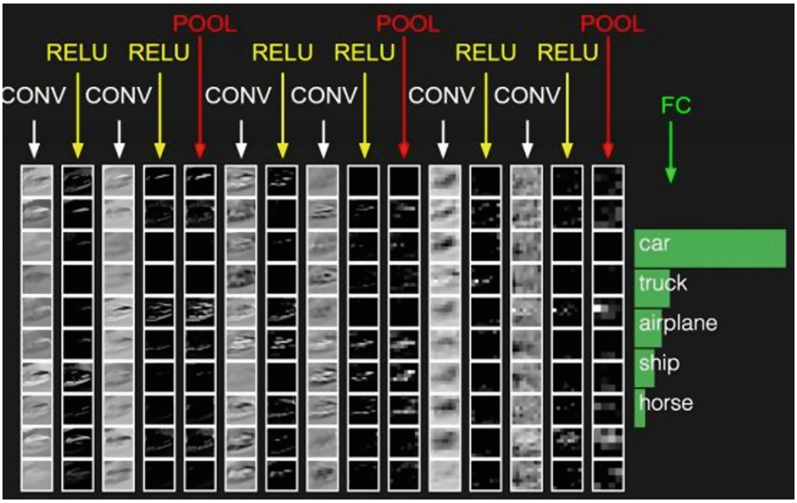

Fig. 1. Complete convolutional neural networks

3 Video Quality Diagnosis System

In recent years, deep learning convolutional neural networks have developed rapidly in the field of machine vision, which can build very complex models by simulating the human nervous system to analyze and interpret data, have powerful expression capabilities to handle complex practical application scenarios, especially convolutional neural networks, which are now widely used in the field of pattern recognition, have shown superior performance in various tasks of computer vision, their unique deep The unique deep structure can effectively learn the complex mapping between input and output. Therefore, the video quality diagnosis algorithm based on deep learning convolutional neural network can extract more features of abnormal video [4], adapt to more complex rules, detect and classify the abnormal types more accurately is completely feasible. Deep learning convolutional neural network algorithm is based on a large amount of data, so expect deep learning convolutional neural network algorithm applied to the field of video quality diagnosis, the following points need to be done (Fig. 2).

First, collect a comprehensive video quality database to form a deep learning convolutional neural network dataset, which is a prerequisite to ensure the generalization performance of the algorithm; second, the video database is organized into a dataset that conforms to the deep learning convolutional neural network model, then a set of schemes is developed, then each video in the video quality database is labeled with

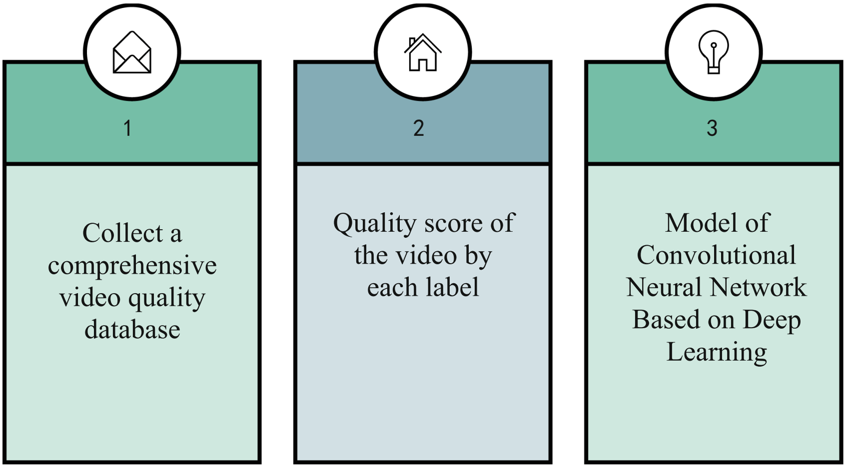

Fig. 2. Focus of building a convolutional neural network video diagnosis system

abnormal types, then the video is scored for quality according to each type of label, the process can be be called subjective evaluation of video; third, establish a deep learning convolutional neural network model based on which can accurately predict the video abnormality types and their video quality scores [5], the predicted results should be consistent with the manually labeled results, the process can be called an objective evaluation method of video quality.

4 Video Quality Diagnosis Algorithm Based on Convolutional Neural Network

4.1 Pool Layer of Convolutional Neural Network in Video Quality Diagnosis

Usually, a pooling layer is inserted periodically between successive convolutional layers. It serves to gradually reduce the spatial size of the data body, which in turn reduces the number of parameters in the network, making it less computationally resource intensive and also effective in controlling overfitting, as shown in Fig. 3. The pooling layer usually uses the MAX operation, which operates independently on each slice of the input data body to change its spatial dimensions. The most common form is to use a filter of size 2×2 to downsample each depth slice in steps of 2, discarding 75% of all activation information in it. Each MAX operation takes the maximum value from 4 numbers (i.e., in some 2×2 region of the depth slice) [6]. Note that the number of channels in the data body remains constant during the pooling process.

4.2 Convolutional Neural Network Video Diagnosis Algorithm in This Paper

The algorithm in this paper is a multi-task multi-label deep learning network model implemented based on the VGG-16 convolutional neural network as a prototype, using the convolutional network for feature map extraction of multi-frame images, which are

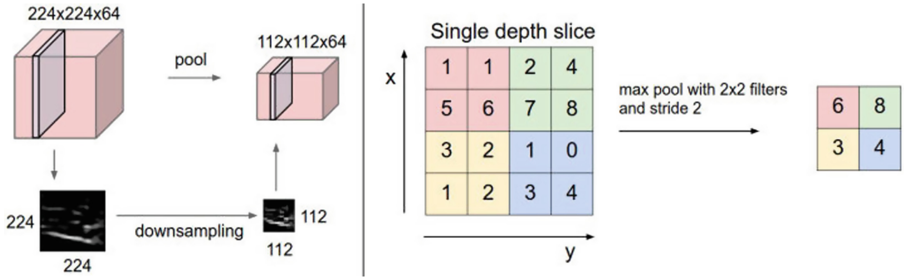

Fig. 3. Role of the pooling layer

then connected together to do anomaly type classification and quality scoring regression tasks. To extend the training set and testing environment, pyramid pooling (SPP) is introduced, the size of the input video images can be unrestricted. Different size images will get different size feature maps after convolutional layers. Since the number of parameters of the fully connected layer is fixed [7], it cannot be connected with the fully connected layer, the different size feature maps can be converted into feature vectors of the same size by means of pyramid pooling. The network structure block diagram is shown in Fig. 4.

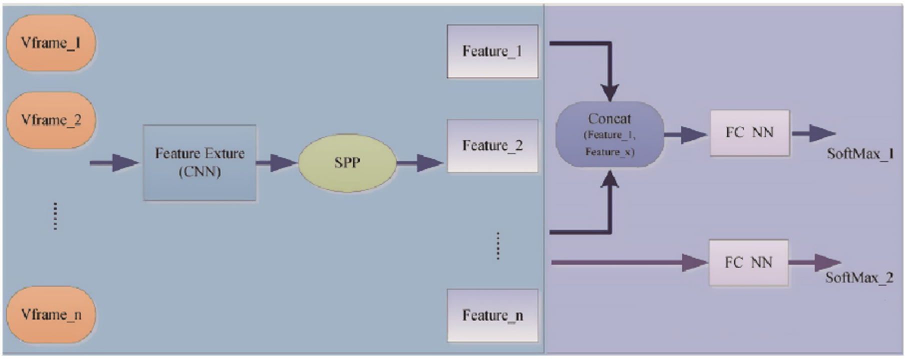

Fig. 4. Network structure block diagram

During the training process, the anomaly type classification is trained first, on this basis the quality score is trained based on the classification results. When using the trained model for prediction, video sequences of different sizes can be input, the abnormality type of the video can be obtained after model processing. If multiple consecutive frames have the same abnormality, it can be assumed that there is some kind of fault in the system, the quality score corresponding to each frame can be obtained by statistical analysis of the relative quality of each frame of the video.

5 Overall System Design Scheme

5.1 Overall System Architecture

The video quality diagnosis system adopts a modular design, including five modules: video frame interception module, OpenCV image processing module, image abnormality detection module, abnormality recording and display module, abnormality alarm module. The workflow of the video quality diagnosis system is mainly as follows: firstly, the required video frames are obtained from the stored surveillance video and saved as Mat entities: secondly. With the help of 0penCV image processing technology. Get the spatial domain structure information that can represent the image content: Finally, give the OpenCV processed image spatial domain structure information to the designed image abnormality detection algorithm for different faults to realize the automatic detection of image abnormality [8], the main functions of each module of the video quality diagnosis system are as follows.

(1) Intercept video frame module: The function of this module is mainly to intercept the stored video image frames, the acquired color image signal is color-separated, separately amplified and corrected to obtain RGB. And the basic image information and pixel data are encapsulated. as the image data to be detected.
(2) OpenCV image processing module: This module makes full use of the image processing library provided by OpenCV to pre-process the image to be detected, such as converting grayscale images, segmenting image channels, color clustering, image space conversion. And obtaining image pixel channel values and other required image structure information.
(3) Image anomaly detection module: this module is the core module of the whole video quality diagnosis system, the main function is to detect the image data information after OpenCV processing, determine whether the detected video frames have abnormal abnormalities including video signal lack of clarity, brightness noise, snow, streaks, color bias, screen freeze, PTZ motion out of control and other abnormalities each image anomaly detection is done by (1) independent algorithm class to complete the input set of video frames for sequential detection[9], return a variety of anomaly detection results.
(4) anomaly logging and display module: this module accepts the return value of the algorithm class of the image anomaly detection module, specifies the description and stores it in the log, so that it can be called when querying the detection results.
(5) abnormal alarm module: this module provides feedback on the detection results to detect abnormal images of the camera according to the abnormal type of mark to remind remind maintenance personnel to solve the problem in a timely manner.

5.2 Functional Module

Video quality diagnosis system mainly includes 6 modules.

(1) each quality diagnosis algorithm uses a single frame or two frames of video frame images at close moments to complete various diagnoses, which do not depend on

the background image, so there is no need for background modeling and the update of kenjing, reducing the false detection caused by unreasonable background model.

(2) Video quality diagnosis is completed by multiple servers, since each server is equally configured, the diagnosis task is equally distributed to each device. With the corpse only need to set up the pre-program of polling detection, LOTUS will start the task according to the start time set in the pre-program, without the need for manual intervention.

(3) Diagnostic results are stored for each recent detection, regardless of whether the camera is working properly or not. Users can query detection records by region [10], fault type.

(4) Fault information stores all the historical fault information of the problematic camera, also contains screenshots of the video frames when the fault is detected, which is easy for the user to view visually. The corpse can query the fault information records of a certain time period by camera, region, fault type. At the same time can be based on the stored video screenshots to determine whether the system is misdetected, can allow misdetected cameras to learn to reduce the probability of misdetection.

(5) In addition, users can count the number of failures and failure rate of cameras in different areas and brands in different time periods according to their own needs, which can be displayed in different forms to facilitate users to understand the operation status of cameras.

(6) In the pre-program management part, the user can set the inspection start time, detection items. In terms of algorithm parameter setting [11], at the early stage of system operation, the algorithm parameters of each camera are set according to the unified threshold; after the system runs for a period of time, the threshold of each detection algorithm of each camera will be adjusted due to various reasons such as equipment quality, service life, site environment and power supply, transmission line., that is, each camera has its own optimal algorithm threshold. In addition, users can set the thresholds of algorithm parameters applicable to special weather such as rain and snow to cope with these bad weather.

6 Conclusion

With the development of the times, video surveillance technology has been widely used, but in the process of its quality diagnosis and control, human operators are required to monitor it. The design of video image detection system based on convolutional neural network described in this paper can avoid the difficulties of traditional video diagnosis as much as possible, provide a new way for video quality diagnosis.

Acknowledgments. Application of neural network in human action recognition.

References

1. MLeCun, Y., et al.: Backpropagation applied to handwritten zip code recognition. Neural Comput. **1**(4), 541–551 (1989)

2. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
4. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
5. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
6. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: ICML (2015)
7. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. arXiv preprint arXiv:1512.00567 (2015)
8. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inceptionv4, inception-resnet and the impact of residual connections on learning. arXiv:1602.07261 (2016)
9. Huang, G., Liu, Z., Weinberger, K.Q., Maaten, L.: Densely connected convolutional networks. In: CVPR (2017)
10. Xie, S., Girshick, R., Dollar, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: CVPR (2017)
11. Hu, J., Shen, L., Sun, G.: Squeeze-and-Excitation Networks
12. Howard, A.G., et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
13. Zhang, X., Zhou, X., Lin, M., Sun, J.: ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices