The Realization and Optimization Technology of Recognition Algorithm Based on Tensorflow Deep Learning Mechanism

Wencai Xu
Jiangxi Teachers College, Yingtan 335000, Jiangxi, China
*Corresponding author e-mail: xuwencai@jxtc.edu.cn

Abstract. With the rapid development of today's technological society, recognition algorithms have received more and more attention. In addition, in recent years, deep learning algorithms have developed rapidly at the theoretical level, and related new technologies have also been applied to various industries. TensorFlow is a deep learning framework that performs well in all aspects. The purpose of this article is to study the realization of recognition algorithms based on TensorFlow's deep learning mechanism and their optimization techniques. The target detection algorithm used in the system in this paper combines deep learning technology to replace the traditional method based on convolutional filtering. The paper is based on the TensorFlow deep learning framework. TensorFlow is an open source software library for machine intelligence. The learning software library of the network learning framework. This article uses a semi-automatic labeling method combined with an incremental learning algorithm to label the data set. After labeling the data, the parameters are set, the model is trained, and the model is finally trained and applied to the detection system. Studies have shown that: in the recognition algorithm, only the single sub-analysis stream is considered, and the short video sequence analysis stream can get the most excellent accuracy. Compared with the second best long video sequence analysis stream, it can also increase by about 3%.

Keywords: Tensorflow, Deep Learning Mechanism, Behavior Recognition, Algorithm Optimization

1. Introduction
With the continuous development of computer science, there are more and more applications in the field of artificial intelligence. Behavior recognition is an important research field in artificial intelligence [1-2]. In the field of computer vision, the research of behavior recognition is of great significance. Commercial fields such as security monitoring, human-computer interaction, video retrieval, and medical and health have huge application potential. In the natural environment, target detection often encounters many detection difficulties, such as differences between classes and within classes, different conditions of image acquisition, differences in semantic understanding, computational adaptability and complexity [3-4]. Due to these difficulties, there are many traditional target detection algorithms that have used different detection techniques to achieve relatively good
results in the field of detection. There are still many scientific researchers studying these algorithms [5-6].

In the 20th century, the research on behavior recognition has entered a new stage of rapid development. Many well-known traditional algorithms and data sets have also been proposed and established. Numerous schools of algorithms have appeared one after another in the accuracy and effectiveness of behavior recognition. The performance and usability have been greatly improved compared with before [7-8]. In terms of frequency, the obtained local feature visual description words are counted to obtain the histogram feature, and then the histogram feature is further used as the classification feature. The support vector machine, Bayes classifier and other classifiers in the machine learning field are used for classification learning. Such methods can achieve good classification results at that time. However, this type of bag-of-words model, which is simply based on the quantification of feature descriptors and performs preliminary statistics on the frequency, discards the original unique spatial distribution information of feature points in the video sequence, making this type of behavior feature lack the physical attributes of the behavior itself. It has a unique internal overall structure expression [9-10]. Domestically, Hu Qiong et al. also proposed a similar idea, using local temporal and spatial feature points adjacent to other related features to calculate an adjacent feature histogram, and then using this statistical description as a feature expression of the feature point to improve this middle-level feature description ability of similar algorithms [11-12]. In addition, there are also some methods similar to the above methods to improve the salience of the upper-level feature expression through how to effectively use the contextual information of time and space.

This article firstly understands the theory of TensorFlow deep learning mechanism, and on this basis, extends the research of recognition algorithm, and then through the analysis of the current existing recognition algorithm, to discuss the realization of recognition algorithm based on TensorFlow deep learning mechanism and its Technical optimization.

2. Research and Analysis of the Realization of Recognition Algorithm Based on Tensorflow Deep Learning Mechanism and Its Optimization Technology

2.1 Deep Learning Framework Tensorflow
TensorFlow programs are generally organized into an execution phase and a construction phase. In the execution phase, the op in the session execution graph is generally used. After the graph is created, a Session needs to be created. The purpose of creating a Session is to run the created graph to implement logic. If there are no parameters when creating it, TensorFlow will enable the default Session. After the session is finished, a Tensor will be returned. Session is divided into two categories, one is ordinary session, and the other is interactive session.

2.2 Multi-Analysis Processing Traffic Behavior Recognition Algorithm
The multi-analysis processing behavior recognition algorithm is an important part of the behavior recognition system. The multi-analysis and processing behavior recognition algorithm used in the behavior recognition system in this article includes five sub-analysis processing streams, including still frame analysis stream, optical flow analysis stream, short video sequence analysis stream, long video sequence analysis stream, and audio analysis stream. Finally, each analysis stream is fused for behavior recognition and classification. The five sub-analysis and processing streams can extract different levels of dimensional features for various types of behavior features in the video sequence. Each sub-analysis and processing flow can better complement each other and complement each other in terms of characteristics in each dimension for different behaviors.

2.3 Implementation and Improvement of Behavior Recognition System
The behavior recognition system based on deep learning algorithms described in this article is mainly divided into two subsystems: model training system and model reasoning deployment system, which are inseparable from each other. The model training subsystem is a part that is more closely connected
with the algorithm design, and is responsible for the establishment of the entire training process, from the collection, labeling, and management of video data to the training of the specific use of the machine, as well as the preprocessing and cleaning of the data. The model reasoning and deployment subsystem is the main part of the specific behavior recognition and classification reasoning that is specifically deployed on the cloud server. It will include the decoding processing of the user's incoming video stream, the video sequence extraction and conversion part, and the preprocessing of the data into sub-components. Analyze and process parts of the required format of the stream.

The improvement and optimization of the system are mainly aimed at the different characteristics of the model training and inference parts, and targeted optimization is carried out. For the training phase, it mainly runs in a single-platform graphics card cluster. For computationally intensive graphics card computing scenarios, methods such as optimizing communication between graphics cards and optimizing convolution kernels can be used to speed up training, and provide special training for the training phase. For the reasoning stage, there are generally flexible requirements for deployment on multiple platforms and multiple architectures, and calculations are generally run in the main processor. For the reasoning of the model, the hot spots of the targeted deployment code are used for targeted optimization and speeding. And for model reading, some intensive memory operations can be tuned in a targeted manner, so there is also a framework code version specifically for deployment, emphasizing flexibility.

2.3.1 Still frame analysis flow. In the field of behavior recognition, analysis based on a single still frame has a long history. In real life, many behaviors have a strong correlation with specific backgrounds or objects, so it is reasonable to use a single still frame in a video sequence for behavior recognition. The recognition of backgrounds and objects in pictures is one of the more important sub-problems in the field of computer vision, and many algorithms have also achieved good results on this problem. The recent results of deep learning are also reflected in object recognition first. Therefore, combined with the analysis of still frames, there are more models and ideas that can be used for reference.

2.3.2 Optical field flow analysis flow. In the field of behavior recognition, it is one of the more common methods to extract the features of the light field flow of the video sequence for the behavior classification. In the last two decades, many analysis algorithms based on the light field flow of the video sequence have continuously emerged, and Good classification results have been achieved in different periods. In terms of behavioral characteristics, there are generally various types of body movements, and the movement structure objectively has a displacement in the spatial domain in the same pixel point between two adjacent frames in the video sequence, and this displacement can be in the spatial domain. The above is described in the form of a vector.

2.4 The Establishment of the Data Set
The public data set used in this article is the GTSRB (German Traffic Sign Recognition Benchmark) data set published at the International Joint Conference on Neural Networks (IJCNN, the International Joint Conference on Neural Networks) in 2011. The data set samples in GTSRB are very rich. It was collected in the real traffic environment in Germany, including more than 50,000 samples under extreme lighting, angle rotation, blurring and other harsh conditions. There are a total of 43 types of signs, including about four Three-thirds of the training images and one-quarter of the test images.

2.5 Analysis of Performance Requirements
2.5.1 Data set annotation. For the labeling module, the coordinates of the object need to be saved according to the frame selection position. In this process, the process of generating the labeling file is required to be smooth, and there are no other performance requirements.
2.5.2 Model training. For the training module, the number of training iterations may reach tens of thousands or even hundreds of thousands of times. Therefore, the training time of each step in this system should not exceed 3s. Other operations are smooth and there are no other performance requirements.

2.5.3 Target detection. In the target detection module, for the video reading part, this system supports common video formats, such as avi, mp4, wmv, etc., and supports common compression formats, including xvid, mpg, wmv9, etc. For the target classification method, the system supports target detection from multiple angles. For the same type of target, it supports the detection of angles including front-side position, front-side position, back-side position, front-side position, etc., and supports user detection conditions including target color, Target type, detection position range, detection time range, etc.

2.6 Implementation of Training Module
The detection module involves two classes: configuredialog class and configuredialog class. The parameters in the configuration file can be set through the configuredialog class. The function name is void initDialog(intindex,QStringconfigName), where configName represents the configuration file name. The parameters that can be set include whether to set dropout, number of iterations, number of categories, and the middle of saving Maximum number of files, saving interval, input picture length or width, learning rate, etc.

2.7 System Optimization
For the behavior recognition system, the optimization of the system is a very important part. The field of behavior recognition is different from other computer vision sub-fields, such as image recognition, object recognition, etc. The input dimension is larger. Compared with algorithms with similar ideas, the computational complexity will be closer as the video sequence data grows. This is also one of the important reasons why behavior recognition algorithms in the industry are less deployed than image recognition algorithms. Therefore, in order for the system to have higher availability, on the one hand, the prediction accuracy of the system meets the use requirements, and on the other hand, it meets a certain economic efficiency in the operating cost, so that the accuracy and the cost can be better balanced, so that the use Reach the user's acceptable range.

2.8 Training Curve
The training curve includes loss function curve and MAP function curve. The output result of the input signal may be different from the actual output. The error is propagated layer by layer in the gradient descent algorithm, and the parameters of the network are also updated layer by layer. Suppose the loss function of the sample (x,y) is L(W,b;x,y):

\[ L(W,b;x,y)=1/2*||hw,b(x)-y||2 \]  

In the application, the L2 norm will be added to reduce overfitting:

\[ L(W,b)=L(W,b;x,y)+\lambda / 2 * w^T w \]  

3. Experimental Research on the Realization of Recognition Algorithm Based on Tensorflow
Deep Learning Mechanism and Its Optimization Technology

3.1 Subjects
In order to make this experiment more scientific and effective, the analysis flow experiment of this test all uses the verification set data as the test data, and the training data used is the UCF101 data set designated by the competition. The linear support vector machine classifier in the multi-stream fusion stage is only obtained by training with the UCF101 data set. Because not all videos in the UCF101
data set have audio information, the audio analysis stream is not included in the test. The details of some parameters in this experiment are different from then.

3.2 Research Methods

(1) Literature research method
This article reads a lot of previous research literature and the latest industry data reports, and collects the most cutting-edge industry background information. These research results not only provide a large amount of data support and theoretical basis for the topic selection of this article, but also provide sufficient theoretical and data support for the final prediction conclusion of this article.

(2) Comparative method
Aiming at the particularity of algorithm recognition, this article compares the recognition algorithm based on TensorFlow's deep learning mechanism with traditional algorithms, sorts out the data obtained, and judges the pros and cons between the two.

(3) Mathematical Statistics
Use software to perform statistical processing on relevant data and analyze relevant data.

4. The Realization of the Recognition Algorithm Based on the TensorFlow Deep Learning Mechanism and the Experimental Analysis of Its Optimization Technology

4.1 Single analysis flow experiment
First, each sub-analysis flow in the multi-analysis flow algorithm uses the validation set data for independent testing. For the model based on deep neural network, the outputs of the last three fully connected layers are used as the high-dimensional features of the video, and the corresponding classifiers are trained by using this feature. The experimental results are shown in Table 1:

| Fully connected layer | Still Frame Analysis Stream | Optical field flow analysis flow | Short video sequence analysis stream | Long video sequence analysis stream |
|-----------------------|-----------------------------|----------------------------------|-------------------------------------|-----------------------------------|
|                      | Fc17                         | Fc17                              | Fc17                                | Fc17                              |
|                      | Fc18                         | Fc18                              | Fc18                                | -                                 |
|                      | Fc19                         | Fc19                              | Fc19                                |                                    |
| mAP                  | 43.0                         | 32.7                              | 59.0                                | 55.7                              |

**Figure 1.** Single analysis flow results
It can be seen from Figure 1 that only considering the single sub-analysis stream, the short video sequence analysis stream can get the most excellent accuracy, and compared with the second best long video sequence analysis stream, it can also be improved by about 3%.

4.2 Analysis of system accuracy
The system is tested with default quality parameters when accepting the accuracy test. At this time, it is mainly processed by a single-channel short video sequence analysis stream. In this mode, system performance and cost can be better compromised, which is more in line with actual users. Option, the results obtained are shown in Table 2:

| Test Cases | Positive Cases | Negative Cases | True Positive | False Positive | True Negative | False Negative | Precision | Recall | F-score | Accuracy |
|------------|----------------|----------------|---------------|----------------|---------------|----------------|-----------|--------|---------|----------|
| 12995      | 10211          | 2784           | 7520          | 2695           | 2345          | 448            | 94.4%     | 73.6%  | 82.7%   | 76%      |

![Figure 2. System accuracy test data](image)

It can be seen from Figure 2 that more than 73% of the fragments in the positive samples can be correctly identified, the false alarm rate in the negative samples is less than 16%, and the false alarm rate in the positive samples is less than 7%, which is generally accurate. The rate reached 76%, and the actual test result reached the original design index.

5. Conclusion
Deep learning is an important branch based on the field of machine learning. Its main advantage lies in capturing highly complex data features and realizing complex nonlinear mapping. Now it has become the mainstream machine learning method. But the research on programming language processing is still immature. Program algorithm recognition is a research hotspot in the field of software engineering. Through the recognition of program functions, it provides a way to evaluate algorithm behavior, program functions, and system complexity, in terms of software module reuse, system maintenance, and software development efficiency improvement. This aspect has very important significance. However, programming languages have rich and rigorous structural features, and traditional natural language processing methods cannot be used to effectively train them. At the same time, during the training process of neural network, due to the existence of gradient dispersion and over-fitting problems, the network model cannot extract useful program structure features, and the
algorithm recognition effect is not ideal. Therefore, how to optimize the existing program recognition algorithm to the algorithm. The improvement of recognition effect is crucial.

References

[1] Cai J. Robust Filtering-Based Thinning Algorithm for Pattern Recognition. Computer Journal, 2018(7):887-896.

[2] Larrain T, Bernhard J S, Mery D, et al. Face Recognition Using Sparse Fingerprint Classification Algorithm. IEEE Transactions on Information Forensics & Security, 2017, 12(7):1646-1657.

[3] Zhi-qiang, Zhao, Meng, et al. Research on a fusion gait real-time recognition algorithm. Journal of Physics: Conference Series, 2019, 1187(4):42014-42014.

[4] Rampasek L, Goldenberg A. TensorFlow: Biology's Gateway to Deep Learning?. Cell Systems, 2016, 2(1):12-14.

[5] Abadi M. TensorFlow: learning functions at scale. Acm Sigplan Notices, 2016, 51(9):1-1.

[6] Saxena A. Convolutional neural networks: an illustration in TensorFlow. Crossroads, 2016, 22(4):56-58.

[7] Wang L, Cao X, Li M. English Letter Recognition Based on TensorFlow Deep Learning. Journal of Physics: Conference Series, 2020, 1627(1):012012 (6pp).

[8] Heinz A, F Schlagenhauf, Beck A, et al. Dimensional psychiatry: mental disorders as dysfunctions of basic learning mechanisms. Journal of Neural Transmission, 2016, 123(8):1-13.

[9] D Wang, Mu C. Developing nonlinear adaptive optimal regulators through an improved neural learning mechanism. ence China Information ences, 2017, 60(5):1-3.

[10] Bj?Rn L, Tobler P N. Incidental ostracism emerges from simple learning mechanisms. Nature Human Behaviour, 2018, 2(6):405-414.

[11] Cirella S, Canterino F, Guerci M, et al. Organizational Learning Mechanisms and Creative Climate: Insights from an Italian Fashion Design Company. Creativity and Innovation Management, 2016, 25(2):211-222.

[12] Hu Q, F Sun, Fan G, et al. The high-quality genome of Brassica napus cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. The Plant Journal, 2017, 92(3):452.