The Realization of Moving Target Tracking in Monitoring Video based on Deep Learning

Qiong Ren [a, *], Juming Chang [b], Hai Han
School of Mathematics and Computer Science, Jianghan University, Hubei Wuhan 430056, China
[a]qren@163.com, [b]JmChang@163.com

***Abstract*—The purpose of this research is to expand the application field of deep learning, to apply deep learning to target tracking problems, to conduct an in-depth study on the monitoring video detection and recognition methods of suspect targets. In this research, from the perspective of feature extraction, the deep learning algorithm that can better reflect the essential features of objects was deeply understood. Then, the stack noise reduction self-encoder was trained with the offline training method without supervision. On this basis, support vector machine (SVM) classifier was added to improve the accuracy of target tracking, and the trained self-encoder was used to extract object features from the deep network. Finally, the deep learning algorithm of "offline training + online fine-tuning" was used to track the moving target in monitoring video, and the robustness performance of the algorithm in this experiment was compared with other kinds of the experimental algorithm through the visual tracker benchmark (VTB) data set. The results showed that the accuracy rate and success rate of all sequences of target tracking algorithm based on deep learning were more than 60%, and their accuracy rate and success rate were higher than other algorithms. And the accuracy and success rate of inter-robustness evaluation of target tracking algorithm based on deep learning were much higher than other algorithms. The tracking algorithm proposed in this research had a higher success rate and accuracy rate than many excellent algorithms such as direct linear transformation (DLT), circulant structure kernel (CSK) and tracking learning detection (TLD) in one-time pass evaluation (OPE) and time robustness evaluation (TRE), and the performance of video sequence attribute algorithm was obviously higher than other algorithms. Target tracking algorithm based on deep learning had good accuracy, robustness, and tracking effect, so it was worth further exploration and research.**

Keywords—Deep learning; Target tracking; OPE; Robustness; Accuracy

I. Introduction

In recent years, with the development of science and technology and the rise of the camera industry, target tracking technology has been applied in various fields [1]. As the main content of video analysis, motion tracking target has been accepted by more and more people. There is no doubt that motion target tracking technology will bring unexpected benefits to humans [2]. Just because of this, most scholars at home and abroad have invested a lot of energy and manpower to conduct in-depth research on this technology, which also makes the target tracking technology develop rapidly. With the rapid development of science and technology in the 21st century, the concept of SIFT feature has been proposed [3]. The appearance of SIFT feature completely solves the problem of little information and slow speed in feature extraction before. The SIFT features have the characteristics of scale scaling and constant brightness [4], which provides a new idea for the later study of region-based feature descriptors. In 2010, someone applied filtering correlation theory and technology in the tracking of moving objects, which started a new wave of traditional tracking methods based on it. In recent years, target tracking technology based on deep learning has been successfully applied in practice and has a broad application prospect. It is an important field of computer vision research in the future [5].

In the process of generation and development, deep learning continues to achieve incredible success, and the fundamental reason lies in its strong feature extraction ability [6]. The task of identifying, detecting and analyzing objects with powerful feature extraction ability of data is twice the result with half the effort. Deep learning is such a structure [7] that can fully reflect the essential characteristics of objects. Compared with the low-learning model, the hierarchy of deep learning is more powerful in describing and summarizing highly unstructured and complete data [8]. On the one hand, it is concerned with the deep level of the network, and the number of hidden layers is basically five or more. On the other hand, it focuses on the characteristics of learning. Each layer is an expression of input, and the extracted features are often transformed, which is more conducive to classification or prediction [9].

Based on the above background, in order to expand the application field of deep learning, deep learning was applied to target tracking. In this study, the trained self-encoder was used to extract object features from deep network, and the target tracking algorithm based on deep learning was deeply studied. It was found that the algorithm had good accuracy and robustness, which provided a new theoretical basis for the future research of tracking algorithm.

II. Methods

A. Machine learning model

Deep learning has a very powerful feature learning ability. In this research, stack noise reduction self-encoder was firstly trained, and SVM classifier was added on this basis to improve the accuracy of target tracking. Figure 1 shows the most basic machine learning model. The algorithm architecture proposed in this research was consistent with this structure. The feature part was extracted by the deep automatic intelligent learning algorithm, and a linear support vector machine was added to the classifier. If the classification problem was divided into many types, the Soft Max classifier was used. Compared with logistic

978-1-7281-2598-5/19/$31.00 ©2019 IEEE

regression classifier, linear support vector machine can effectively improve the classification accuracy of deep learning, so as to greatly improve the accuracy of target tracking. In this research, there were two main reasons for choosing the noise reduction auto-encoder. Firstly, compared with the formation of noiseless and non-damaged training data, the noise generated by the damaged data after training was relatively small. Secondly, after adding noise, the damaged data can effectively reduce the difference between training data and test data, so as to improve the robustness of training network.

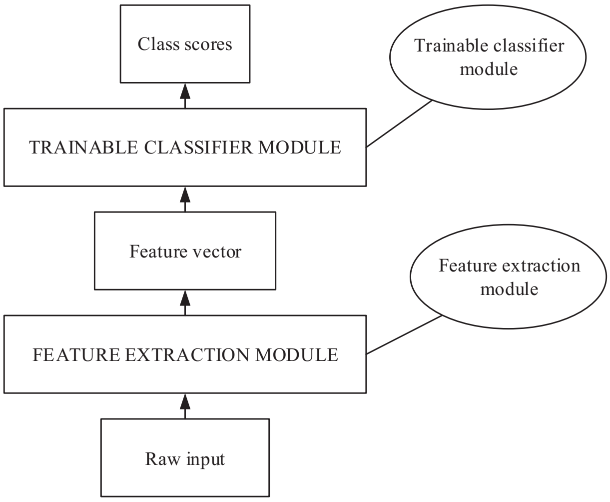

Figure. 1 Basic architecture of machine learning model

B. Deep automatic encoder

Both network theory and performance are gradually becoming mature, but in terms of target tracking, it still has such problems as slow convergence speed, poor generalization ability, easy to fall into local minimum and limited network generalization ability. Therefore, the concept of deep learning neural network and its training strategy come into being, and because of the defect of BP neural network, the deep automatic encoder has also been developed rapidly. The deep automatic encoder can process large and complex data. The tool is used to extract the feature parts, which not only achieves high-level data expressions, but also reduces the workload of manual design features to a certain extent. Automatic encoder is an unsupervised learning method based on learning method, which enables it to deeply understand the essential characteristics of data provided by most unregistered data and some mark samples. As an important deep learning model, the research and development of deep automatic coding is of great significance to the development of neural networks.

Deep autoencoders extract useful features by compressing input information at each level. It belongs to the typical unsupervised learning network and a deep network model. Each layer consists of one layer of automatic encoder, and the output of the following layer is the input of the upper layer. The essence of the automatic encoder is equivalent to learning from an equal function, that is, the network input is equal to the output of the reconstructed network, and the training and optimization process of parameters is the reproduction input of the output. Figure 2 shows the principle of automatic encoder.

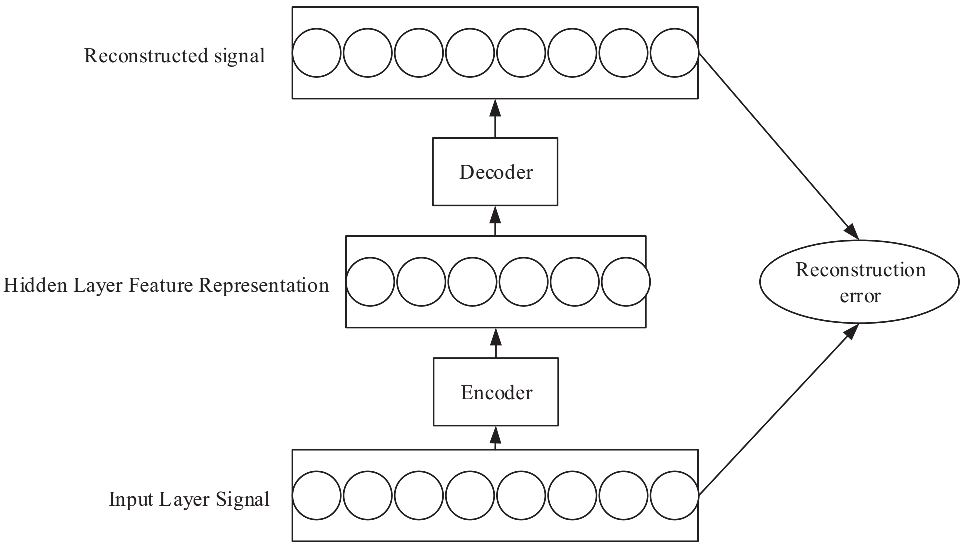

Figure. 2 Schematic diagram of self-encoder

C. Noise reduction automatic encoder

Noise reduction automatic encoder is a kind of automatic encoder which takes damaged data as input data and forecasts original data as output through training. Noise reduction autoencoder not only refers to how to learn noise reduction, but also to learn a good internal representation method. The learned representations can be used to perform deeper training without monitoring the network. The function of the de-noising automatic encoder is to learn the original data of superimposed noise, whose features are almost the same as the data without superimposed noise. However, the features learned in the input superimposed noise are more robust than those learned in the non-superimposed noise. Another feature is that the denoising automatic encoder can learn the same eigenvalues.

According to the statistical theory, the essence of denoising automatic encoder is to disturb the original input with noise with certain rules, thus destroying the original input, and then input the destroyed original input data into the network, so as to obtain the representation of hidden layer. Depending on the nature of the noise, the decoder must contain the original data completely and eventually produce the closest original input. This is the premise of how to eliminate interference and conduct more in-depth training without monitoring the network. The general expression of gaussian noise is shown in equation 1

$$\widetilde{x} = x + \varepsilon, \varepsilon \sim N(0, \sigma^2 I) \quad (1)$$

Among them, x represents the original input data without noise interference, $\widetilde{x}$ indicates the data after being disturbed by noise, and σ indicates the degree of regularization.

D. SVM classifier

In this experiment, in order to improve the classification accuracy, the SVM linear classifier was selected. SVM linear classifier carries out binary classification, which takes the target and background in the tracking process as positive and negative samples to supervise the formation of the whole network, so as to achieve the purpose of fine-tuning the network. Once the fine-tuning network is complete, the new network can not only express the original goal, but also classify it.

The training sample is set to 1 and $n = 1,...,N, x_n \in R^D, y_n \in \{-1,1\}$, and the general expression of unconstrained SVM optimization is shown in equation 2.

$$\min_{W} \frac{1}{2} w^T w + C \sum_{n=1}^{N} \max(1 - w^T x_n y_n, 0)^2 \quad (2)$$

In the above equation, C represents the rule parameter.

Equation 3 shows the class marker prediction for sample x:

$$\underset{y}{\operatorname{argmax}} (w^T x) y \quad (3)$$

BP neural network algorithm is used to fine tune the whole network. The above equation (2) is taken as the objective function, the activation value of the penultimate layer of the network is taken as the input, and the optimization formula is shown as follows.

$$\frac{\partial l(W)}{\partial h} = -2Cy_n w(\max(1 - w^T h_n y_n, 0)) \quad (4)$$

E. Target tracking algorithm based on deep learning

Deep learning is widely used in object detection and recognition. However, few people have conducted further research on the target tracking. Deep learning is the effective learning of a large number of targeted training data. Target tracking training data only contains the target information of the first frame. By contrast, it can be found that deep learning is difficult to be applied in the field of target tracking because of the lack of training data, which makes it difficult to establish a depth model based on target tracking training.

Because the data of target tracking training are limited, it is absolutely necessary to reduce the dependence on target tracking training data. This requires the application of other non-target data to establish the depth model in advance, so that the established depth model has the general representation ability of object features. By combining the depth model established by this method with online development, the target tracking task can not only overcome the problem of insufficient target samples, but also make the model have better classification effect.

The deep learning algorithm of "offline training + online fine-tuning" is used to track the single target task, as shown in figure 3.

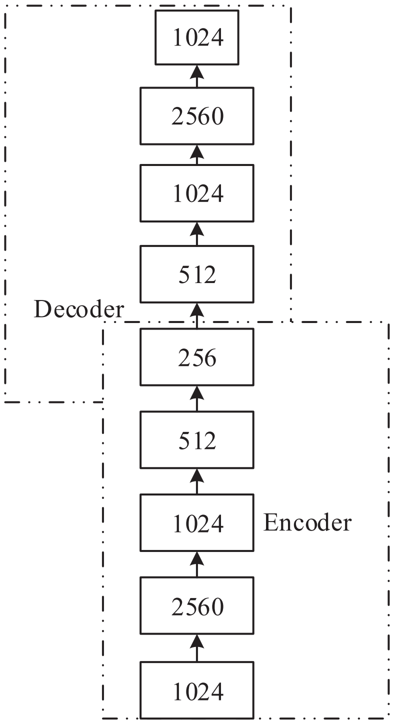

Figure. 3 Network Structure Diagram

F. Evaluation platform

The VTB dataset pack contains more than 50 fully annotated sequences that are commonly used to evaluate target tracking algorithms. More than 50 fully labeled video sequences cover a variety of possible situations during the tracking process, such as rapid movement, complex background, dim optical fiber, etc. In order to better analyze and compare the advantages and disadvantages of target tracking methods, more than 50 video sequences were labeled into 11 unique attributes in this experiment.

Figure 4 is the attribute distribution diagram of 11 video sequences in VTB data set. As can be observed from the figure, a sequence usually contains multiple properties. The frequency of properties OPR and IPR is the highest. Therefore, multiple video sequences have these two properties. Since the VTB data platform can provide a subset of each attribute, in addition to analyzing and comparing the algorithm performance of the entire data set, the algorithm performance can also be evaluated under specific conditions. For example, the OCC subset consists of 29 sequences, which can be used to analyze the ability of tracking algorithm to deal with occlusion.

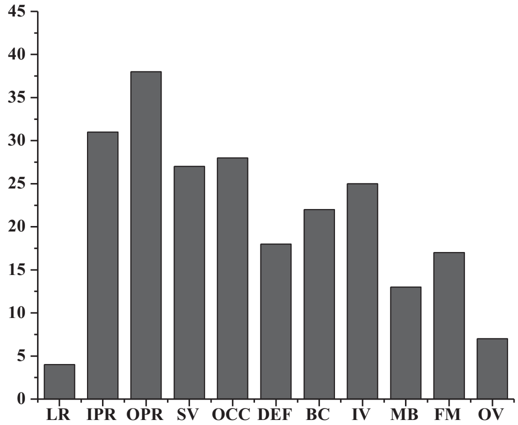

Figure. 4 Attribute distribution map

III. Results and Discussion

Quantitative analysis of the results of the moving target tracking of the algorithm:

In the unified test of standard VTB platform, the numerical values of precision rate and success rate were used to reflect the advantages and disadvantages of this algorithm compared with other algorithms. The tracking results of target tracking algorithm based on deep learning were quantitatively analyzed. The analysis results are as follows. The data recorded in table I is the comparison data of the evaluation results of each algorithm at one time. It can be observed from table 1 that the accuracy rate and success rate of one-time assessment of all sequences of target tracking algorithm based on deep learning were above 60%, and their accuracy rate and success rate were higher than other algorithms.

TABLE I Comparison of robustness one-off pass assessment results

Algorithm	Ours	TLD	CSK	LSK	DLT	OAB	Frag	KMS	CT
Accuracy rate	65.2	60.3	54.8	49.9	46.5	51.0	46.7	43.5	41.1
Success rate	61.3	53.2	44.7	46.2	44.8	43.3	40.5	32.7	33.6

Table II records the comparison data of robustness evaluation results of each algorithm time. Similarly, it can be observed from the figure that the accuracy rate and success rate of inter-robustness evaluation of target tracking algorithm based on deep learning were much higher than other algorithms.

TABLE II Comparison of time robustness evaluation results

Algorithm	Ours	TLD	CSK	LSK	DLT	OAB	Frag	KMS	CT
Accuracy rate	65.9	63.0	62.2	57.8	55.7	56.1	54.5	46.6	48.1
Success rate	61.5	52.5	51.7	52.5	52.5	47.7	47.0	38.4	41.3

From all the test results, it can be observed that the tracking algorithm proposed in this research had a higher success rate and accuracy rate than many excellent algorithms like DLT, CSK and TLD in terms of OPE and TRE, and the performance of video sequence attribute algorithm was also significantly higher than other algorithms. Therefore, the tracking performance used in this experiment had achieved good results, which was worthy of further study.

IV. Conclusion

The application of deep learning method in monitoring video target tracking field effectively improved the robustness performance of the tracking method. In this research, from the perspective of feature extraction, the deep learning algorithm that can better reflect the essential features of objects was deeply understood. Then, the stack noise reduction self-encoder was trained with the offline training method without supervision. On this basis, SVM classifier was added to improve the accuracy of target tracking, and the trained self-encoder was used to extract object features from the deep network. Finally, the deep learning algorithm of "offline training + online fine-tuning" was used to track the single target task, and the robustness performance of this algorithm was compared with other kinds of the experimental algorithm through the VTB data set. According to the results, the proposed algorithm had good accuracy and robustness in various complex situations such as fast motion, complex background, and dim optical fibre. Therefore, the tracking performance used in this research had achieved good results, which was worthy of further study.

Although good results have been obtained in this experiment, due to the limited time, there are still many shortcomings and worthy of improvement. At present, there are still many problems in using recursive neural network to monitor target tracking in video and applying it to general tracking and target tracking in open environment. Of course, the characteristics of recursive neural network can't be ignored. This neural network can effectively constrain the filter and deeply explore the correlation relationship. Therefore, most scholars believe that this neural network will make a breakthrough in the field of target tracking. It can be expected that the application of recursive neural network can help the positioning system to predict under the constraints of specific targets and fixed lenses. It is a space worthy of further exploration to study trajectory and prevent position movement from the perspective of memory.

Acknowledgments

Scientific Research Program Guiding Research Project of Hubei Provincial Department of Education, (Project No. B2016281); research fund of Department of Housing and Urban-Rural Development of Hubei Provincial (2016).

References

[1] Cheng S, Sun J X, Cao Y G, et al. Target tracking based on incremental deep learning[J]. Optics & Precision Engineering, 2015, 23(4):1161-1170.

[2] Li, Yundong, Zhang, Xueyan, Li, Hongguang, et al. Object detection and tracking under Complex environment using deep learning-based LPM[J]. IET Computer Vision, 2019, 13(2):157-164.

[3] Zhong B, Pan Shengnan, Wang Cheng, et al. Robust Individual-Cell/Object Tracking via PCANet Deep Network in Biomedicine and Computer Vision[J]. Biomed Research International, 2016, 2016(12):1-15.

[4] Wang L, Liu, Ting, Wang, Gang, et al. Video Tracking Using

Learned Hierarchical Features[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2015, 24(4):1424-35.

[5] Ren N, Junping D U, Zhu S, et al. Robust visual tracking based on scale invariance and deep learning[J]. Frontiers of Computer Science, 2017, 11(2):230-242.

[6] Johnny L. Chen, Jason E. Summers. Deep neural networks for learning classification features and generative models from synthetic aperture sonar big data[J]. Acoustical Society of America Journal, 2016, 140(4):3423-3423.

[7] Luo H, Xu L, Hui B, et al. Status and prospect of target tracking based on deep learning[J]. Infrared & Laser Engineering, 2017, 46(5):502002.

[8] Wang Z, Wang Lijia, Zhang Hua. Patch Based Multiple Instance Learning Algorithm for Object Tracking[J]. Computational Intelligence and Neuroscience, 2017, (2017-02-22), 2017, 2017:1-7.

[9] Cardenas C E, McCarroll, Rachel E, Court, Laurence E, et al. Deep Learning Algorithm for Auto-Delineation of High-Risk Oropharyngeal Clinical Target Volumes With Built-In Dice Similarity Coefficient Parameter Optimization Function[J]. International Journal of Radiation Oncology Biology Physics, 2018, 101(2): S0360301618302451.