Mobile Robot Target Tracking System Based on Deep Learning

Xiaoyu Guo[1], Haoyu Liu[2] and Hengzhang Dai[3]

[1]Information Department, Beijing University of Technology, Beijing, 100124, China

[2]School of Statistics, Renmin University of China, Beijing, 100000, China

[3]College of Engineering Science and Technology, Shanghai Ocean University, Shanghai, 201306, China

Keywords: Target tracking, Caffe deep learning framework, twin convolution neural network, GOTURN target tracking algorithm, ROS

Abstract: In order to solve the problem that intelligent mobile robots often lose tracking targets due to changes in the appearance of targets in the tracking process, Caffe deep learning framework and ROS robot operating system are used as the development platform. A mobile robot target tracking system with high accuracy and high real-time performance is designed and studied. Using GOTURN target tracking algorithm based on twin convolution neural network, which is robust to target deformation, viewing angle, slight occlusion and illumination change, the tracking model of offline training is applied to TurtleBot mobile robot in real time through ROS system, and detailed tests are carried out. The target tracking system not only has a feasible design scheme and can effectively track targets in various complex scenes, but also has the characteristics of low cost, high performance and easy expansion.Abstract: In order to solve the problem that the current intelligent mobile robot often loses the tracking target due to the appearance change of the target during the tracking process, we have designed a high accuracy and high Real-time mobile robot target tracking system. We use a GOTURN target tracking algorithm based on twin convolutional neural networks that is robust to target deformation, viewing angle, slight occlusion, and illumination changes. We used the ROS system to apply the offline training tracking model to the TurtleBot mobile robot in real time, and conducted detailed tests. The experimental results show that the target tracking system not only has a feasible design scheme, but also realizes that the mobile robot can effectively track the target in various complex scenarios. In addition, it also has the characteristics of low cost, high performance and easy expansion.

1. Introduction

Robot technology involves research in many fields. Target tracking is not only one of the hot spots in the research field of robot technology, but also the basis of realizing the intelligence of mobile robot. visual target tracking is a very challenging task, because it is very difficult to deal with all kinds of complex and changeable scenes, and feature selection is an important factor affecting the accuracy of target tracking [1] .The convolutional neural network maps the input data by a multi-layer neural network to obtain an effective abstract expression of hierarchical features, which solves the problem of manually designing features in traditional target tracking methods. In addition, it enhances the robot's understanding and modeling of the environment and enables mobile robots to detect and track targets in various scenarios in real time [2].

In this paper, the deep learning framework is used to train the tracking model which is robust to target deformation, visual angle, slight occlusion and illumination change, and adjust it according to the requirements of practical application to make it more accurate to track specific types of targets. The mobile robot can detect and track targets effectively in various scenes. So that it can efficiently assist other robots, automation equipment and related personnel to complete the tasks in the corresponding complex scenes.

2. Mobile Robot and its Control system

2.1 ROS robot operating system

Traditional robot system development usually focuses on the design of the whole function, which makes the coupling ability of each module between robot systems very strong and difficult to separate from each other. It greatly affects the reusability and portability of the code and the efficiency of deployment in actual development [3].

With the increasing requirements of code reuse and modularization for robots, the open source robot operating system ROS, which is well adapted to its requirements, has come to the fore, which is composed of several independent nodes, and each node can communicate with each other through the publish / subscribe message model [4].

ROS supports the cooperation of multiple compiling languages, uses the standard TCP/IP protocol to realize the communication between the nodes within the system, and the interface can regard the third-party components as part of its operating system. These functions can meet the needs of most research and development personnel. Today, the application of ROS has covered the high-tech fields such as manipulator grasping, navigation robot, humanoid robot, mobile robot chassis, unmanned robot ship, unmanned aerial vehicle and so on.

2.2 TurtleBot mobile robot

TurtleBot is a low-cost open source mobile robot based on ROS robot operating system, which can achieve many functions in complex environments and fully meet the needs of most scientific research work.

The Kinect visual depth sensor projects the light of a specific pattern onto the surface of the object, and then uses the camera to receive the deformation of the structured light pattern reflected from the surface of the object to measure the distance between the object. Kinect can provide both color information and depth information, so it is very suitable for mobile robots to track targets, and can avoid the trouble caused by the lack of single information.

3. Caffe Deep Learning Framework

3.1 Convolution neural network

The traditional target tracking obtains the features of different samples through manual design, which not only requires professional knowledge fields, but also the effect is not satisfactory in practical application because of its lack of generalization of features. Therefore, the target tracking accuracy achieved by the traditional target tracking algorithm can not meet the needs of mobile robot engineering applications.

Convolution neural network is an efficient recognition method developed in recent years, which has been widely used in the field of computer vision and has achieved fruitful results. It can learn different levels of feature abstraction of the original input image. It has excellent learning ability and generalization ability for all kinds of computer vision tasks, which lays a foundation for realizing the function of mobile robot target tracking.

Convolution neural network is a kind of feedforward neural network structure, its main body is usually composed of convolution layer and pool layer, and then the extracted feature vectors are classified by full connection layer or classifier, and the classification results are outputted. It adopts the technical means of local perception and parameter sharing to reduce the scale of parameters in the convolution neural network. As a result, a network model with stronger robustness for different detection tasks is obtained.

3.2 Caffe

With the rapid development of deep learning technology in recent years, many universities and scientific research institutions have launched a deep learning development framework, which promotes the process of engineering application of deep learning, and greatly enhances the reusability of deep learning work and the efficiency of researchers actually engaged in deep learning in development. Compared with Torch7, Theano and Pylearn2, Caffe has the following unique advantages:

(1) Caffe has built-in Python and Matlab external interfaces to facilitate subsequent analysis of experimental data.

(2) Caffe has training configuration documents for several basic data sets, including fast training model and complete training model.

(3) Caffe uses GPU parallel technology to enhance the efficiency of code execution and greatly shorten the training time of the model.

(4) Caffe is based on the principle of modularization as much as possible. This makes it easy to extend new data formats, network layers, and lost functions.

Use a convolutional neural network that can learn deep essential features from the data, and use the Caffe deep learning framework to train the model on a large scale to obtain a large amount of representative functional information to achieve the target tracking function of the mobile robot.

4. Implementation of Target tracking algorithm based on convolution Neural Network

4.1 GOTURN target tracking algorithm

Many researchers often try to use neural network to track on-line. However, neural network training is a slow process, which leads to a very slow tracking speed. The tracker with the best performance runs at a speed of 100 FPS on GPU. Therefore, these trackers using on-line training are not suitable for mobile robots that need to track targets at real-time speed.

GOTURN (Generic Object Tracking Using Regression Networks) is a general object tracking algorithm based on regression network, which can train off-line through a large amount of data and use the method of single regression to the position of the target object. It becomes the first deep learning tracking algorithm that can track the target at a speed of 100 FPS. It can track the class samples that have never been seen before, and the tracking effect of the specific class samples is better.

The network structure of GOTURN takes the tracking target area and search area as input, inputs to two convolution neural networks at the same time, and then returns the position of the target through the full connection layer, as shown in figure 1. Because the configuration and parameters of the two convolution layers are the same, it is also called twin neural network.

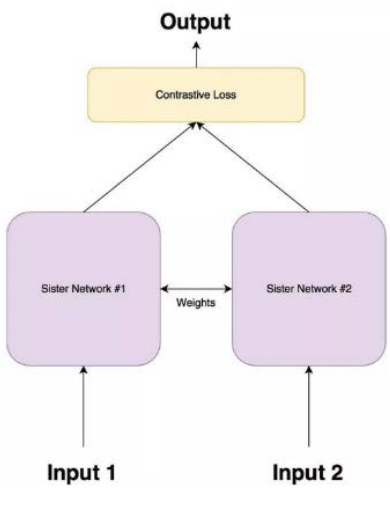

Figure 1. Misjudgment effect of moving target

The convolution layer of GOTURN network is composed of the first five convolution layers and pooling layers of Caffe Net. The output of the convolution layer is connected into three fully connected layers with 4096 nodes, and finally connected to the output layer with four nodes. The output rectangle returns the moving position of the target, indicating the coordinates of the upper left corner and the lower right corner of the target.

In order to know the position of the target in the current frame, the tracking first finds the area of the target to be searched based on the position of the previous frame, and determines the approximate position of the target in the current frame. The purpose of the network is to return to the position of the target in the current search area, select the tracker's search box in the frame, determine the search field, and leave the image block corresponding to the search field.

4.2 Implementation of offline training in GOTURN Network

The target tracking model designed in this paper is trained and evaluated on ILSVRC Challenge and VOT Challenge data sets, as shown in figures 4 and 5. It was created by Li Feifei of Stanford University in the United States. Each frame of the video contains nearly 14 million image frames. Each frame of the video is annotated in the form of occlusion, illumination change, motion, size change, and position movement and so on. The training set of the improved target tracking algorithm is composed of ILSVRC2017 training set, verification set and ALOV training set. The test set of the improved detection model is evaluated on the VOT2014 test set.

Deep convolutional neural networks require massive labeled data for training, so that the trained model can obtain better feature expression from the training image data set.When the data set is insufficient, the model trained by the deep convolution neural network often performs poorly. For the training samples, the existing datasets with real tags contain fewer video sequences. In order to make the models trained by deep learning have stronger generalization ability for different tasks, it is necessary to use larger data sets for training. At present, the way to improve this problem is to enhance the dataset. That is to deal with the existing data and create new data to expand the training set.

Figure 2. ILSVRC challenge data set

The image in the training set is randomly cut into blocks, and then the samples are processed by gradient descent in small batches, so as to ensure that the tracking window must contain at least half of the target, and the excessive stretching and deformation of the target window can be prevented by constraints. This method is used to expand the training samples and greatly improve the generalization ability of the model.

The off-line training process of the GOTURN algorithm is shown in figure 3. It is an important reason why the algorithm is fast enough that the time-consuming calculation process is carried out offline so that the tracking process does not need to be updated online. Because the information in the test data can be obtained at any time, it is possible to search forward and backward on the test data at the same time for global optimization.

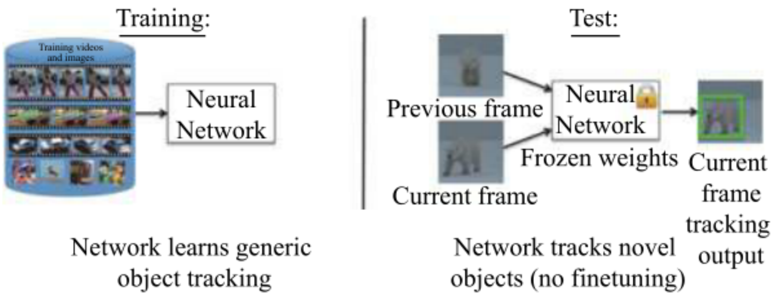

Figure 3. GOTURN offline training and Target tracking Test

The data set and the loss function are changed respectively, and rigorous experiments are carried out on the tracking results of GOTURN. As shown in figure 4, with the increase of training video, the generalization performance of GOTURN algorithm is better. Therefore, if you mark more videos to increase the size of the training set, you can get further benefits.

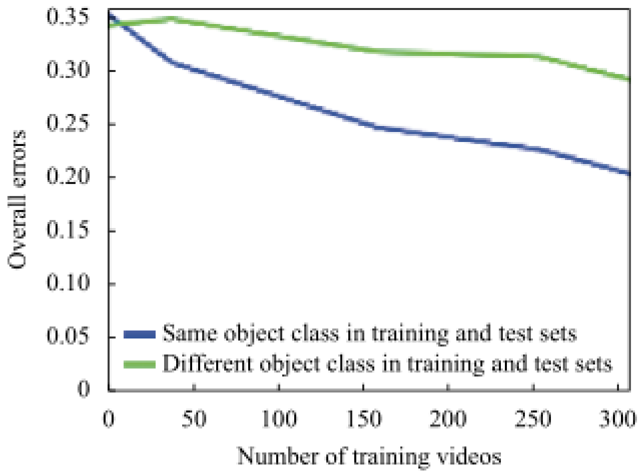

Figure 4. Generalization performance experiment

5. Implementation of Target tracking system for Mobile Robot

5.1 Realization of target tracking program

This system adopts Ubuntu14.04 operating system, ROS system (version is Indigo), CPU is Intel (R) Core (TM) i7-4720HQCPU @ 2.60GHz, and the memory is 16 GB,. Under the development environment of Caffe deep learning framework, the target detection and tracking system is realized by using Caffe deep learning framework, and it is applied to the mobile robot. Thus efficiently assist other related robots, automation equipment and personnel to complete the tasks in the corresponding complex scenes.

The system is composed of nodes of each functional program, and uses the topic mode in the ROS system to realize the communication between each program. The main part of the program is to obtain visual information nodes, tracking nodes and mobile robot target tracking programs. The image transmission is also carried out in message format in ROS, and the communication status between nodes is shown in figure 5.

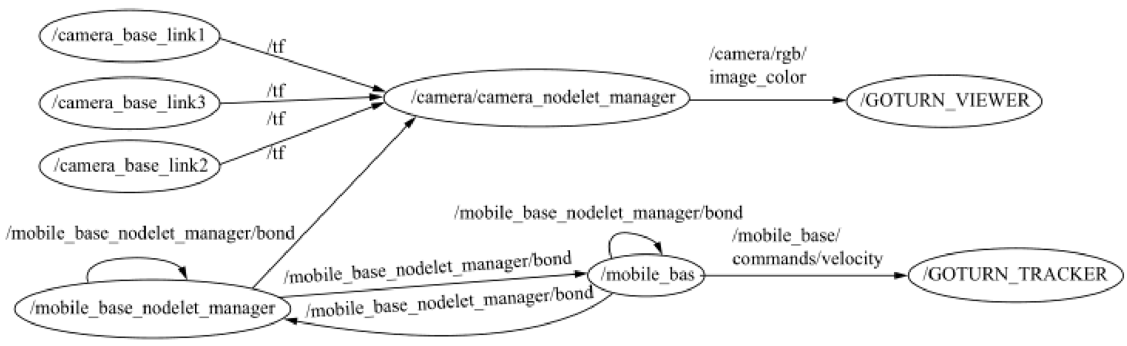

Figure 5. Node communication state diagram of target tracking running program

5.2 Mobile Robot tracking based on GOTURN algorithm

Open the terminal, compile all the files in the package, when the terminal window shows that the compilation of the program is 100% complete, and find that all the files are normal, the program runs normally. Perform the system function operation, start the tracking node, according to the position of the previous frame, find the area of the target to be searched, and determine the approximate position of the target in the current frame. Find and locate the target in the search area of the current frame according to the search radius, as shown in figure 5.

Figure 6. TurtleBot tracking target

The target tracking model is adjusted according to the practical application, so that the model can track specific types of targets more accurately, and the mobile robot can effectively detect and track targets in various scenes. The tracking experiment effect is shown in figure 7.

Figure 7. TurtleBot tracking target

In this system, the robot signal is transmitted in real time, the equipment is simple and easy to operate, each file program in the ROS system cooperates with each other as a node, and the function of each module is reliable.

6. Conclusion

The deep learning framework is used to train the tracking model which is robust to the changes of target viewing angle, deformation and illumination, and adjust it according to the practical application to make the tracking of specific types of targets more accurate. Mobile robots can effectively detect and track targets in a variety of scenes.

The experimental results and performance analysis prove that the GOTURN target tracking algorithm is a target tracking algorithm with high real-time and precision requirements. Its offline training and single-return target position methods enable it to track the target at a speed of more than 100 FPS. Learning target tracking application becomes possible on mobile robots. And it meets the real-time and robustness of mobile robot target tracking, which is an important reason why we choose to use GOTURN target tracking algorithm.

References

[1] Lu Huchuan, Li Peixia, Wang Dong. An overview of target tracking algorithms. Pattern recognition and artificial Intelligence, 201831 (1): 61Mel 76.

[2] Liu Tiantian. Research on Target tracking algorithm based on Kernel correlation filter [Master's thesis]. Wuhan: Huazhong University of Science and Technology, 2017

[3] Li Peng. Research on Target tracking algorithm based on full convolution Twin Network [Master's thesis]. Wuhan: Huazhong University of Science and Technology, 2018.

[4] He Xinhua, Zhao Long. Real-time moving target detection and tracking based on improved Gaussian mixture model. Computer Application Research, 201027 (12): 4768MU 4771.