Robot Machine Vision by Deep Machine Learning

Hong LU
Beijing Information Technology College, Beijing, China

Keyword: Machine vision, Image recognition, Robot.

Abstract. This paper propose the approach of robot machine vision image recognition. In the paper, describing the architecture of robot machine vision, the robot image capture method, image transfer method, image recognition method and machine vision control method, introducing. Emphasis, introducing the method of robot image recognition based on deep machine learning through convolution neural network. In the paper, discussing and providing some methods of optimizing image recognition model.

Introduction
At present, there are many methods to achieve robot machine vision, but often meeting some problem, such as low recognition efficiency, low accuracy of recognition. This paper use deep machine learning that can improve the accuracy of image recognition more and more. This paper give a good performance solution for robot machine vision, which include image capture, image transfer, image recognition and machine vision control by voice.

Related Work
Robot Machine Vision Image Capture
We employ OpenCV framework to obtain image capture, different camera need to set different arguments. Image recognition need that one picture compare with other one. So first we need intercepting a frame image from video streaming.

Robot Machine Vision Image Transfer
This robot machine vision’s architecture is Client/Server mode, robot is client that take charge image capture, server take charge image recognition. The architecture’s advantage is that high speed server can enhance image recognition’s speed. Therefore it needs to implement image transfer from robot to server. Meanwhile, image recognition ID is transferred from server to robot. We use Socket communication protocol to complete the image file transfer.

Robot Image Recognition
The robot image recognition employ deep machine learning by convolution neural network, Figure 1 show the architecture of the convolution neural network, it is eleven layers network, containing five convolution layers, three pooling layers, three fully connected layers and some hide layers. The first convolution layer has $227 \times 227 \times 3$ input image, 96 kernels of size $11 \times 11$, basic learning rate is 1, basic decay is 1, stride is 4, weigh filler type Gaussian, bias filler type constant, value is 0, there is a pooling layer which uses Maximum pooling method[3]. The second convolution layer has 256 kernels of size $5 \times 5$, bias filler type constant, value is 1, there is a pooling layer. The third convolution layer has 384 kernels of size $3 \times 3$, bias filler type constant, value is 0, there is not a pooling layer. The fourth convolution layer has 384 kernels of size $3 \times 3$, there is not a pooling layer. The fifth convolution layer has 256 kernels of size $3 \times 3$, there is a pooling layer. Finally, there are three full-connected layers to be responsible for outputting whole image and features of image.

Furthermore, robot machine vision image recognition need building the model of image recognition, we train the model through deep machine learning. The model training needs some steps. First, we need preparing some training images, test images and valid images. Second, putting label on these images purpose for dividing different image. Third, the model is learnt training
images to compare with valid images through convolution neural network, the accuracy of image recognition depend on getting the feature vectors of the image, enhancing the fitting of the image feature is by iteration. In addition, test images use for checking accuracy of the model.

Figure 1. Eleven layers convolution neural network structure.
Robot Machine Vision Control

Robot machine vision is controlled by voice, voice instruction drive robot image capture program to begin capturing image, then one frame image of capturing is transferred to server implementing image recognition, recognition ID will be transferred to robot after image recognition, robot speak the name of the objective through voice recognition program.

Experiment Results

In the section, we evaluate the accuracy of image recognition in different convolution network structures, different algorisms and different machine learning parameters. The objective of experiment is drink that is recognized by robot. Firstly, we collect a large number of drink dataset about 20000 images of drink from different sides of drink, and 20 classes drink. The image is divided training_set about 16000 images, valid_set about 3000 images, test_set about 1000 images. Machine learning parameter is set learning rate 0.001, batch size 10, weight decay 0.0005 and momentum 0.9.

Increase Dropout

The drink image recognition accuracy is 67.23% initially, the accuracy is not improved that increasing iteration from 5000 to 30000. Deep learning always meet over-fitting, we used Dropout to reduce over-fitting. When increasing Dropout layer, the drink image recognition accuracy become 87.65%, comparing with the accuracy of image recognition before using Dropout and after using Dropout in Table 1.

Choosing ReLU activation function is for better performance through decreasing the parameters computation scale.

<table>
<thead>
<tr>
<th>Parameters of deep machine learning</th>
<th>Accuracy of image recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before use Dropout</td>
<td>Learning rate:0.001, weight decay 0.0005 and momentum 0.9</td>
</tr>
<tr>
<td>After use Dropout</td>
<td>Learning rate:0.001, weight decay 0.0005 and momentum 0.9</td>
</tr>
</tbody>
</table>

Five Convolution Layers vs. Six Convolution Layers

The drink image recognition accuracy of five convolution layers is 87.65%, we make the convolution layers from five to six, the accuracy become 98.26%. The result show on Table 2.

<table>
<thead>
<tr>
<th>Parameters of deep machine learning</th>
<th>Accuracy of image recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Five convolution layers</td>
<td>Learning rate:0.001, weight decay 0.0005 and momentum 0.9</td>
</tr>
<tr>
<td>Six convolution layers</td>
<td>Learning rate:0.001, weight decay 0.0005 and momentum 0.9</td>
</tr>
</tbody>
</table>

Discussion

In this section, discussing the factors of influence for robot machine vision image recognition through deep machine learning. Machine learning parameters influence recognition effect, learning rate set too small to high speed learning and generating local optimum, if setting learning rate too big will lead to the accuracy down, it is better for learning rate 0.001, weight decay set 0.0005 that benefit weight adjusting. Increasing normal layer can reduce computation scale for rising
computation speed, alpha 0.0001, beta 0.75. Adding convolution layers is best way to boost the accuracy of image recognition.

**Summary**

In this paper, we proposed a six-layer convolution neural network architecture as well as increasing some hide layers, such as Dropout layers and normal layers. Activation function use ReLU. The results show that the accuracy of robot machine vision image recognition reach 98.26%. Deep machine learning has better performance for robot machine vision.

In addition, robot machine vision is controlled by voice that is better way, it is personification’s method, achieving multiple modes recognition of robot.

**References**


