CNN-Based Model for Pose Detection of Industrial PCB

Li Haochen[1,2], Zheng Bin*[1], Sun Xiaoyong*[1,3], Zhao Yongting[1]

1. Chongqing Institute of Green and Intelligent Technology, CAS, Chongqing, 400714, China;
2. University of Chinese Academy of Sciences, Beijing, 100049, China;
3. Chongqing De ling Technology Co. Ltd., 400713, China, 400713, China

zhengbin@cigit.ac.cn; suncq@cigit.ac.cn

Abstract—For applications in robot manipulate with object, get the pose of objects is very important for controller's subsequent operations, especially in PCB feeding and blanking field, the grasp success rate will be enhanced if robot can get a exact pose of objects that relative to end manipulator. So in this paper we utilize the CNN model to build on a neural network for 3 tasks: object recognition, location and pose detection. This model treat pose detection as a classification problem and try to combine recognition, location at the same level. To validate the performance of the multi-task detection model, experiments and analysis of the model performance was carried out by the real-time PCB detection test. In the experiment, we use the PCB dataset comprised of 3 types which contains different poses made by ourselves as train/test samples. The number of object pose categories was divided into 8bins, 12bins and 36bins according to pose detection precision. We analysis the effect of the non-uniform datasets on training process and the final detect results shows that this CNN-based detection model can achieve high accuracy on PCB pose detection.

Keywords- PCB; Vision System; Deep Learning; Pose

I. INTRODUCTION

The concept of deep learning was proposed by Hinton in 2006 [1]. In this work, he used the fast neural network training algorithm on the deep-belief-network and obtain a good effect. It means that with the help of high performance computing and algorithm optimization, researchers can overcome the inadequate of traditional neural network and utilize the generalization ability of deep learning model in vary of fields. As research continues, variety deep learning model have been developed, such as CNN(Convolution Neural Network) that is the most commonly used in image processing, RNN(Recurrent Neural Network) which has the ability to deal with time-series data and so on [2].

LeCun et al [3] developed a handwriting character recognition system based on CNN and put it into commercial application. It was the first time that CNN showed the performance on feature extraction and classification. Different deep learning model have shown the powerful performance in many complex application field, such as face recognition, multi-target detection, speech recognition and so on [4] until now.

In the development of image processing, object detection and classification have become hotspots in current research, and scholars have presented variety models for object recognition while many applications need to perceive the pose, type and the location of the objects at the same time, including human activity recognition, industrial robot grasping and manipulation in general [5]. By means of multidimensional objects information, scholars can take the high-level sematic analysis. While classical method that used to detect the pose is to find the key-points of objects or add the pose as a part of DPM(Deformable Part Model) [6]. Our method is similar with the 3D object detection model proposed by Jincheng Yu et al [7], we both use the neural network to realize the multi-task detection with object and want to apply the model to robot manipulation, but we utilize the multi-task framework and have a better performance in distinguishing foreground and background. Joseph Redmon et al [8] presented an real-time CNN model for robot grasping detection, they build their model from AlexNet and extend it to grasp box detection.

In this paper, we propose a CNN detection model for the detection of industrial PCB, which included 3 tasks : the PCB recognition, location in the image and the pose detection. By means of experiment and PCB dataset, the performance of this detection model were obtained, which provided a new solution for the real-time PCB feeding and blanking process and verified the feasibility of multi-task(>2) model in the industrial detection filed.

II. POSE DETECTION MODEL

Object detection is the first step for robot visual servo control. After identify the object, robot can use hand-eye calibration to make sure the coordinate between robot and object, so the grabbing, moving and placing operations can be performed by end effector mounted on robot. But robot cannot adjust the rotation angle of end effector according to the pose of objects, so sometimes the success rate of grasping is not high.

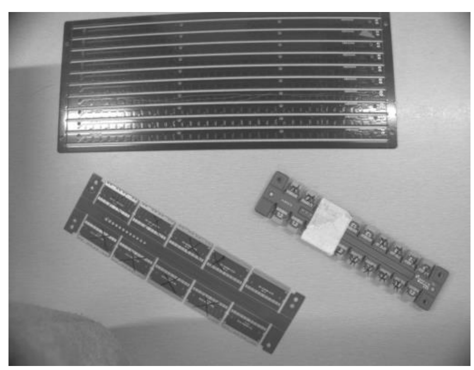


Figure 1. The PCB Datasets consists of 3 Types

In view of this problem, we use the visual model to implement the pose detection and choose PCB datasets as training data to validate the model. The PCB datasets are shown in Figure 1.

978-1-5386-1230-9/17 $31.00 © 2017 IEEE
DOI 10.1109/ICICTA.2017.93

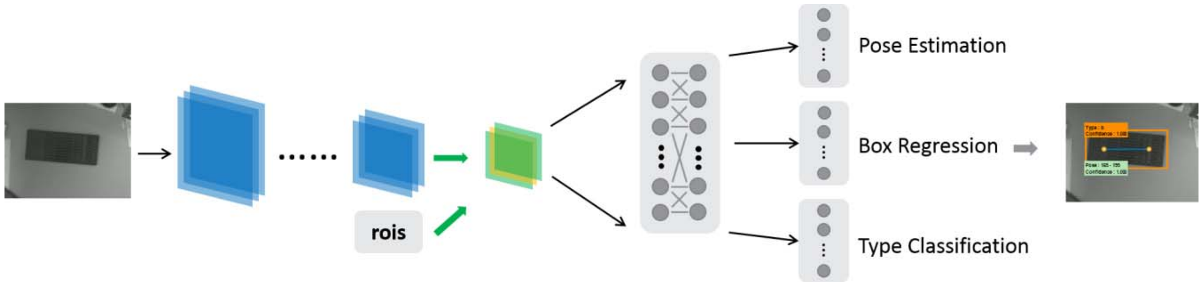


Figure 2. The Full Architecture of our detect model

CNN has currently shown its powerful ability on feature extraction. The quality and quantity of the automatic extracted feature outperform the traditional artificial designed items. Many scholars take advantage of CNNs to do object recognition and solve problems in machine vision. So we utilize the extensive ability of a multi-task CNN to make a global object pose detection. Our detection model has 5 convolutional layers, 2 fully connected layers with 3 output layers which output the box coordinate regression value, the probability of every type class and pose class. After parsing the output-matrix, we can get the detection-info about current object.

A full illustration of the model architecture is shown in Figure 2. Our simple model was derived from the powerful CNN which was a winner in ILSVRC 2013 and presented by Zeiler and Fergus [9].

The functions of network was enhanced by tweaking the architecture of ZF-Net, in particular by expanding the output task layer to make pose detection and utilize the appropriate loss function(large-margin-softmax-loss) to make back propagation computation corresponding to pose detection task.

III. EXPERIMENTS RESULTS AND PERFORMANCE ANALYSIS OF MODEL

A Description of Datasets and Detection System

The main purpose of this experiment is to test the performance of our CNN model within 3 aspects: object recognition, object location and the detection of pose for object.

While in the experiment, we use PCB datasets made by ourselves which contains 3 types of pcb with different poses on the 2D planes, including 731 images without flip for training/testing sample and every image is labelled with ground truth horizontal rotation angle, box coordinate and object type.

Every pcb object may has multiple labelled pose classes corresponding to the pose detect precision.

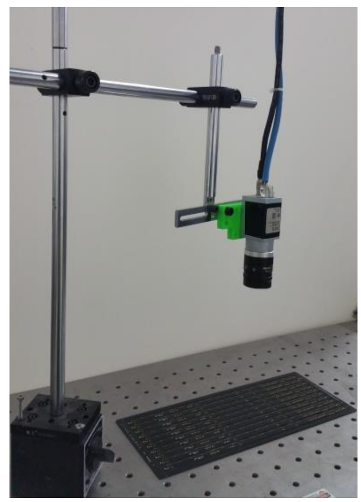


Figure 3. The Illustration of This Detection System

We use the Basler-acA 1600-20gm as our camera. The Detection System is shown in Figure3. The Basler Industrial Camera is mounted on the bracket which the lens is parallel with PCB. After the model has been trained with the datasets, we rotate the PCB as different poses to observe and record the results of our model during the course of the experiments.

Table 1 Test Results of The Model With Image Flip

Detect Precision	Average Time/s	Number of Test Sample (With Flip)	Pose Accuracy (Rotate Angle)/%	Detection Accuracy(IOU) /%	Recognition Accuracy/%
8 bins(45 degrees)	0.045	1096	97.5	95.5	95.5
12 bins(30degrees)	0.044	1096	93.6	96.4	96.4
36 bins(10 degrees)	0.049	1096	89.7	96.9	96.9

B The Experiment Result : Pose Detection Output

In the training phase, we discretized the pose space into several uncorrelated bins, the type of bins can be divided into 3 cases including with 8, 12 and 36 bins so that we can formulate the pose detection problem as a classification task. In the labeled images, there are several different pose classes correspond to each pcb object under the 3 cases. The visualization of s-class pcb detect result on 36 bin is shown in Figure 4. The top and bottom annotation of bounding-box shows the type and the range of rotate angle of current object following with its confidence value. The higher confidence value, the higher probability of representing a correct detection. A line with two endpoints was used to represent the direction of pose (depicted as the line in each bounding-box from Figure 4 to Figure 6). It will rotate a few degrees corresponding to outputs of network with the pose changing. Results on 36 bin object recognition and pose detection on m-class and b-class is shown in Figure 5 and Figure 6 with the range of pose is 125-135 degrees and 175-185 degrees respectively. They both have high confidence value.

We also present the evaluated results on different metric including average detect time, pose, detection and type recognition accuracy in Table 1. The corresponding detect precision of above cases is 45, 30 and 10 degrees. As a result, with the improvement of precision, the pose detect result will approximate to real pose state more easily which is beneficial to subsequent robot control but will increase difficulty in network training and convergence of model.

C The Relationship between Samples in Training Stage and Detect Precision

CNN-based deep learning is considered not suitable to be widely used in industry, the reason is that it’s hard to obtain enough and appropriate dataset that matches the requirements of a industrial task. Accordingly, the different pose-precision model was tested several times in view of the number and distribution of samples. We analyzes the effects of using different sample distribution with and without image flip for training through the experiments and get the pose detect accuracy contrast curve under the above mentioned pose precision (8, 12 and 36 bins) . See Figure 7(c).

The rotate angle distribution of original pcb dataset is uneven (depicted as 15 degrees interval with a range from 0 to 180 degrees in Figure 7(a)). Lacking of pose samples which between 30 to 60 degrees accounts for the lower accuracy as shown in Figure 7(c) with “circle-line” curve and prevents the network from converging.

Based on this foundation, we utilize the flip operation to randomly selected image samples so that we can expand the original dataset and the distribution of new dataset will be more uniform. The “triangle-line” curve indicates clearly that the uniformed dataset can be found in favor of CNN training and can produce a better effect.

In Figure 8, we show the training curves of model that corresponds to the 12bins case. Both accuracy and loss value change rapidly within 5000 iterations but will cost 7x time to converge the network.

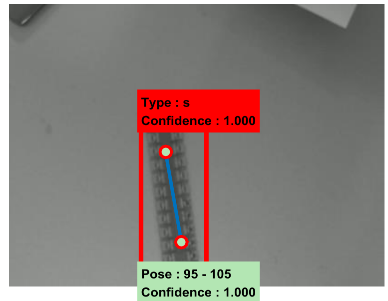


Figure 4. The Pose Detect Result of PCB:s

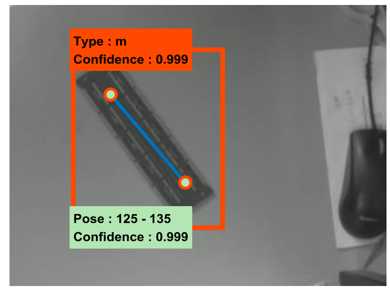


Figure 5. The Pose Detect Result of PCB:m

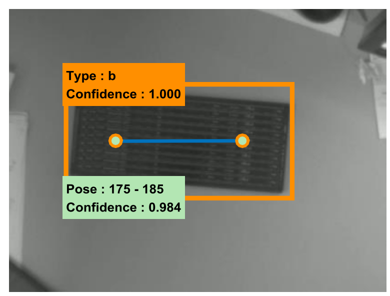
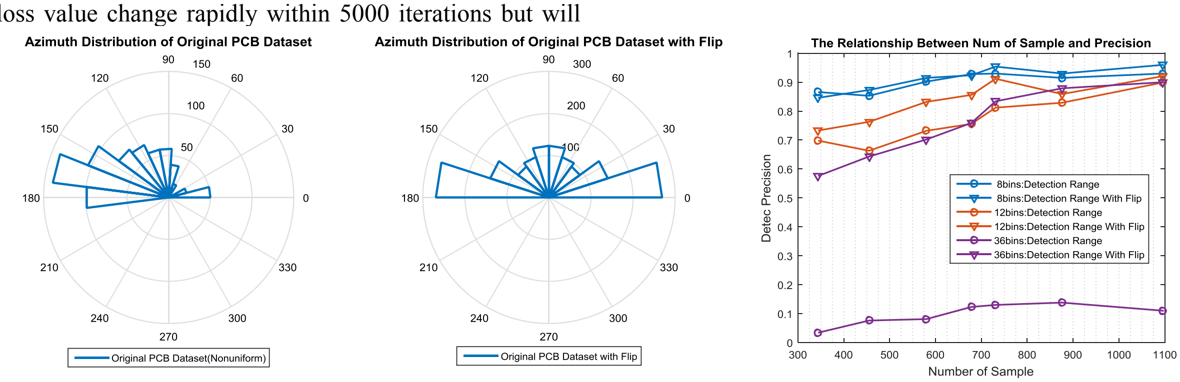


Figure 6. The Pose Detect Result of PCB:b

The loss of multi-task can decrease rapidly down to 0.2 and less but spend 40000 iterations to reach closer to 0. Increasing the number of iterations appropriate will be helpful for the model to reach global optimum.



(a) (b) (c)

Figure 7. The Relation ship Between Number of Sample and Precision

IV. Conclusions

In this paper, a CNN-based multi-task model is presented. The extended model can be used to detect object pose, identify object and locate it in the image at the same level. In practice, the model performs well in multi-task detection while the pose detection of pcb is our focus in this paper. As Table 1 shows, we take 3 experiments according to the 3 cases (8, 12 and 36 bins) respectively, the detect accuracy will decrease as the number of pose space bin goes up, but the accuracy of pose detection almost stay above 85% that illustrates our model is universal and practical and can be used to make pose detection through automatic features learning without construct 3D models for object.

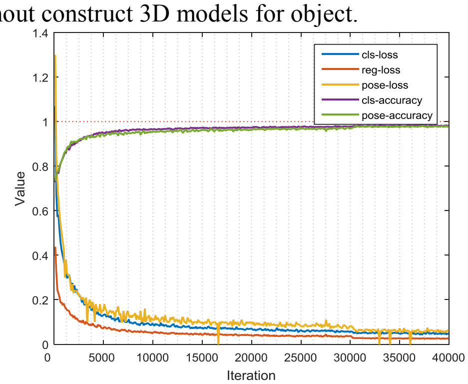


Figure 8. The Training Curve of This Detection Model

CNN models that require a large amount of training samples are not suitable to apply to some machine vision fields which need self-designed station or custom algorithm, such as defects detection of various parts, target location matching and so on. But in the area of pcb feeding and blanking, the station only requires accurate classification and efficient completion of loading and unloading operations, therefore our model is suitable for this industrial task. We will integrate the multi-task CNN model proposed in this paper to robot vision servo control in the next step of research to realize grasping pcb.

ACKNOWLEDGEMENTS

This work is supported by Chongqing Youth Science and technology personnel training program (Grant Nos. cstc2014kjrc-qnrc70001).

REFERENCES

[1] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural computation, vol. 18, no. 7, pp. 1527-1554, 2006.

[2] J. Sanchez-Riera, K.-L. Hua, Y.-S. Hsiao, T. Lim, S. C. Hidayati, and W.-H. Cheng, "A comparative study of data fusion for rgb-d based visual recognition," Pattern Recognition Letters, vol. 73, pp. 1-6, 2016.

[3] Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," Neural computation, vol. 1, no. 4, pp. 541-551, 1989.

[4] A. Wang, J. Lu, J. Cai, T.-J. Cham, and G. Wang, "Large-margin multi-modal deep learning for RGB-D object recognition," IEEE Transactions on Multimedia, vol. 17, no. 11, pp. 1887-1898, 2015.

[5] P. Poirson, P. Ammirato, C. Y. Fu, W. Liu, J. Kosecka, and A. C. Berg, "Fast Single Shot Detection and Pose Estimation," in Fourth International Conference on 3d Vision, 2016, pp. 676-684.

[6] B. Schiele, "Teaching 3D geometry to deformable part models," in Computer Vision and Pattern Recognition, 2012, pp. 3362-3369.

[7] D. Liang, K. Weng, C. Wang, G. Liang, H. Chen, and X. Wu, "A 3D object recognition and pose estimation system using deep learning method," in IEEE International Conference on Information Science and Technology, 2014, pp. 401-404.

[8] J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," in IEEE International Conference on Robotics and Automation, 2014, pp. 1316-1322.

[9] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision, 2014, pp. 818-833: Springer.