Scholar Publishing Group
Kinetic Mechanical Engineering
https://doi.org/10.38007/KME.2021.020202
ISSN 2790-2676 Vol. 2, Issue 2: 9-17

Object Recognition and Grasping Detection Method of Construction Machinery Robot Relying on Deep Learning

Zabolotny Simons[*]

Univ Chem & Technol Prague, Prague 16628, Czech Republic

[*]*corresponding author*

Keywords: Deep Learning, Convolutional Neural Network, Construction Machinery Robot, Grasping Detection System

Abstract: At present, robots are widely used in the field of machinery manufacturing, which improves the production speed on the basis of ensuring the safety of workers. As a result, the traditional mode of enterprises based on manual operation has begun to be transformed into industrial robots as the main body. Construction machinery robots need to capture information about the surrounding environment in production operations. Traditional object recognition methods cannot adapt to complex working environments. Therefore, how to effectively identify target objects and successfully grasp objects has become a challenge for robots. In addition, the grasping detection(GD) method required by the robot to complete the task also relies on the known information of the target object and cannot effectively deal with the complex and changeable unknown environment. To this end, this paper designs a robot GD system based on deep learning, constructs a GD model and system framework through a convolutional neural network(CNN), and introduces a multi-target object grasping recognition algorithm to improve the grasping accuracy of the robot. The simulation experiment of the GD system proves that the accuracy rate of the system successfully grasping objects is over 95%.

1. Introduction

When a robot grabs an object, the robot fixture must move to the position of the object, and then the robot can grab the object. This situation is mainly used in industrial production, which can standardize the production process and improve production efficiency. In order to enable the robot to cope with the uncertain working environment, computer vision and recognition algorithms are applied to the robot system, through the vision system to perceive the external environment, identify and locate the target object, and then control the robot to move to the corresponding position to achieve object grasping.

With the rapid development of information technology and manufacturing, scholars and experts in the field of robotics have produced a variety of intelligent robots with grasping functions. For example, some researchers propose to apply feature learning to robot grasp detection, and directly detect grasp points on input image data. When the robot grabs an object, the user can guide the robot to approach the target point, then segment and detect the object on the plane, and finally implement the grab. The device does not have the ability to recognize objects in an unstructured environment, and can only detect objects placed on a plane in a single background, so it cannot autonomously complete the grasping operation of arbitrarily placed objects [1]. Before the rapid development of deep learning, traditional visual detection methods were mainly used for robot grasping position detection. This method generally requires a complete 3D model, is applicable to fewer application scenarios, and has low flexibility. Researchers have begun to gradually turn to the development of convolutional neural models [2]. Some scholars propose to describe the grasping position as 6-dimensional grasping frame coordinates, and use two convolutional neural networks to solve the problem of GD of target objects [3]. However, the robotic grasping technology still lacks the ability of autonomy and adaptability to the environment, which seriously affects the popularization and application of robots.

This paper first introduces the network structure of CNN and proposes a multi-target object grasping and recognition algorithm, and then uses CNN to design a robot GD system. The introduction of CNN GD algorithm and target recognition algorithm can improve the accuracy of robot object recognition and grasping. Finally, the grasping time and detection success rate of the system for grasping a single object and multiple objects are analyzed through simulation experiments.

2. Deep Learning and Grab Recognition Algorithms

2.1. Deep Learning Model - Convolutional Neural Network

Deep learning is a very hot research field in recent years. Deep learning has good feature extraction ability, and can use multi-layer network to extract different levels of feature information [4]. Convolutional Neural Networks are one of the most commonly used network structures in deep learning models.

At present, the commonly used convolutional neural network is generally composed of five layers of structure, and each structural layer plays a different role in the process of information transmission to achieve different functions [5]. The input layer processes the input data information into a specified format; the convolution kernel in the convolution layer mainly implements the feature sampling function of the input data; the activation layer is generally bound to the convolution layer, and the activation function is used to output the convolution layer. Each value is delinearized to enhance the linear separability of features; the pooling layer can not only complete the downsampling operation of the feature map, but also reduce the amount of data calculation by sparse processing of the feature map; the fully connected layer will Calculate the probability density of the features sampled from the first few layers, reduce the loss of feature information, and output the possibility that the object belongs to the category; the output layer filters the probability value of each category according to the probability value from high to low, and finally outputs the first one The class corresponding to the probability [6-7]. The structure of convolution group-full connection layer can obtain image features layer by layer, complete category classification, and improve image recognition rate.

2.2. Multi-Target Object Grasping and Recognition Algorithm Based on Multi-Level Feature Fusion

(1) Grasp the mathematical representation of the identification variable

Some studies have pointed out that the plane grasping pose of an object represented by a 5-dimensional vector in a plane image can be mapped to a three-dimensional space and expressed as a 7-dimensional vector to guide the robot to complete the grasping operation of space objects. The research on object grasping and recognition algorithms in complex scenes based on 2D visual information promotes the development of robot intelligent grasping technology [8].

In order to realize the object grasping operation, it is necessary to establish a mathematical model of the robot grasping, abstract the actual physical problems, use the specific mathematical parameters to represent the grasping target information of the robot, and then determine the GD network structure and the composition of the loss function [9]. When building an end-to-end grasp detection network, the forward prediction operation inputs the RGB image of multiple objects in the scene, and then directly outputs the grasp parameters of the object, while the grasp depth of the object is obtained from the depth map of the object; Take the target grasping pose obtained by the system for coordinate transformation, and then control the gripper at the end of the robotic arm to complete the grasping operation [10-11].

The object grasping recognition algorithm studied in this paper is mainly aimed at the situation of robot grasping. The grasping rectangle frame contains the coordinates of the center point of the grasping area of the object, the width of the mechanical finger, and the length and direction of the opening of the mechanical claw. Its mathematical expression is shown in formula (1).

$$U = \{x_u, y_u, h_u, w_u, k_u\} \tag{1}$$

(x_u, y_u) is the coordinate of the center point of the optimal grasping rectangular frame of the target object; h_u is the length of the opening of the mechanical finger; w_u is the width of the mechanical finger; k_u is the angle between the opening direction of the mechanical finger and the horizontal axis of the image.

(2) Multi-target object grasping and recognition network model

The network model of the multi-target object grasping recognition algorithm based on feature fusion uses a deep neural network with strong feature learning ability [12]. The grasping recognition model can be divided into four modules: feature extraction, object detection, grasping pose measurement and joint reasoning according to its functions. The network feature completes the object detection and grasping pose measurement of complex scenes, and finally calculates the best grasping pose of each target object through the joint reasoning module. In addition, in order to make full use of multi-level and different scale network features, the feature information fusion part adopts the attention mechanism algorithm [13].

3. Construction Machinery Robot GD System Based on CNN

3.1. CNN-Based Grasp Detection Model

Before deep learning methods became popular, the construction of traditional object detection algorithms mainly relied on artificial feature design. The design of these handcrafted features is mostly based on statistical or empirical knowledge, and detection objects have great limitations and little effect. Due to the lack of effective image feature representation, researchers have to constantly design complex feature representation methods and use a series of accelerated computing methods to maximize the use of limited hardware computing resources [14].

In order to improve the recognition rate of the target, CNN draws on the ResNet network design idea. This operation is equivalent to readjusting the arrangement of the feature map, so that the shallow information can be transmitted to the subsequent network layers without loss of feature information. Therefore, the deep network can acquire more fine-grained features [15]. CNN uses batch normalization (BN) to improve the convergence speed of the network. BN improves the convergence effect of the network without adding any regularization methods. Its essence is to insert a normalization layer between each input layer and output layer of the network, and each layer of the network has a d-dimensional input:

$$X = \left(x^{(1)} \cdots x^{(d)}\right) \tag{2}$$

The normalization layer is a learnable network layer with parameters (y, β). Assuming that the output of the previous layer of the network is normalized and then input to the next layer of the network, in order not to destroy the features learned by the upper network, the learned parameters y and β can be introduced to restore the features learned by the input layer, k ∈ [1, d]. The formula is as follows:

$$y^{(k)} = \sqrt{Var[x^{(k)}]}\,\hat{x}^{(k)} + \beta^{(k)} \tag{3}$$

3.2. Framework of Grab Detection System Based on CNN

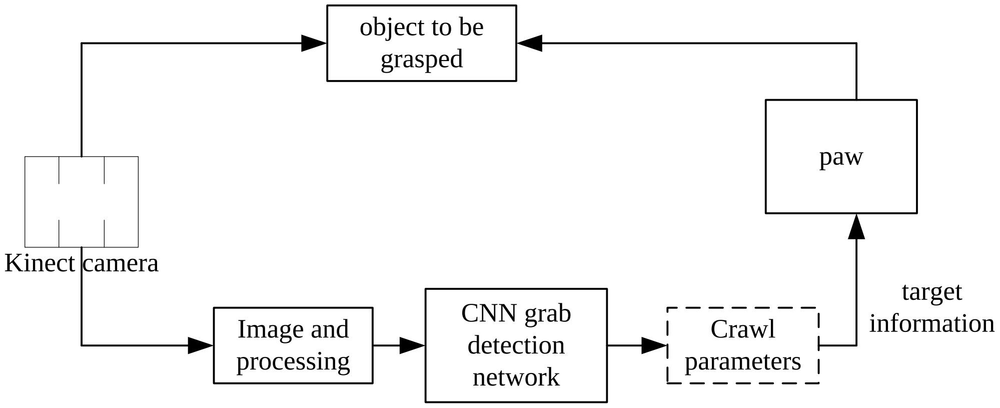

Figure 1. Grasp detection system framework

The GD system framework is mainly composed of three parts: Kinect visual perception, CUNN GD network, robotic arm and gripper. The workflow of the system is as follows: the Kinect camera first collects the three-channel RGB image and depth map of the object, and transmits the collected RGB image to the upper computer for image preprocessing; then the capture detection network outputs the category and capture parameters; the upper computer integrates The depth data of the depth map transmits the category label and target grasping pose of each object to be grasped to the industrial computer, and then controls the construction machinery robot to complete the corresponding grasping operation [16-17]. The overall framework is shown in Figure 1.

3.3. The Realization Process of Robot Grasping

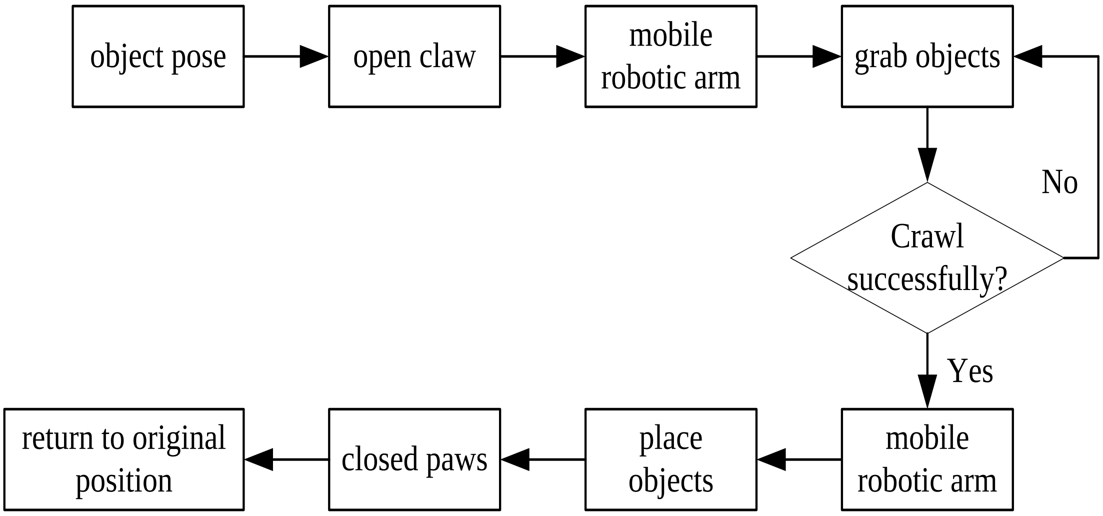

Figure 2. Flowchart of robot grasping objects

Figure 2 is a flow chart of the robot grasping objects. The robot first changes its gripper from a closed state to an open state to facilitate subsequent grasping of objects. The robot executes motion commands according to the specified position and posture of the gripper calculated by the computer. When the claw moves to the specified position and posture, the object is grasped by closing the claw, and after lifting the object, it is judged whether the object is grasped by detecting whether the object is lifted. If the object is not grabbed, grab it again; if the object has been grabbed, continue to move the robot to the designated area. After the object is moved to the designated area, it is necessary to open the gripper to separate the object from the gripper, and then close the gripper and move to the initial position [18-19].

4. Application of Robot GD System

4.1. Performance Test on Multi-Target Object Grasping Dataset

The multi-object grasp recognition model based on multi-pole feature fusion of attention mechanism is denoted as MGD_HFA. In order to test the contribution of each functional module of the grasping recognition model to the grasping recognition algorithm, the object detection accuracy, grasping pose measurement accuracy and recognition speed of grasping recognition models such as MGD_SF38, MGD_SF19, MGD_HF and MGD_HFA were compared.. Among them, MGD_SF38 and MGD_SF19 represent multi-target object grasping and recognition methods based on single-stage network features. The network features of their grasping pose measurement modules are the output tensors y and z of the convolution modules Conv4_3 and Conv7, respectively. MGD_HF represents a multi-target object grasping and recognition algorithm based on simple multi-level feature fusion. The difference from the MGD_HFA method is that the MGD_HF method does not use an attention mechanism when performing multi-level feature fusion.

Table 1. Recognition speed test results of grasp recognition algorithm

Grab recognition method	Recognition speed(fps)
MGD_SF19	37.65
MGD_SF38	36.94
MGD_HF	36.21
MGD_HFA	34.79

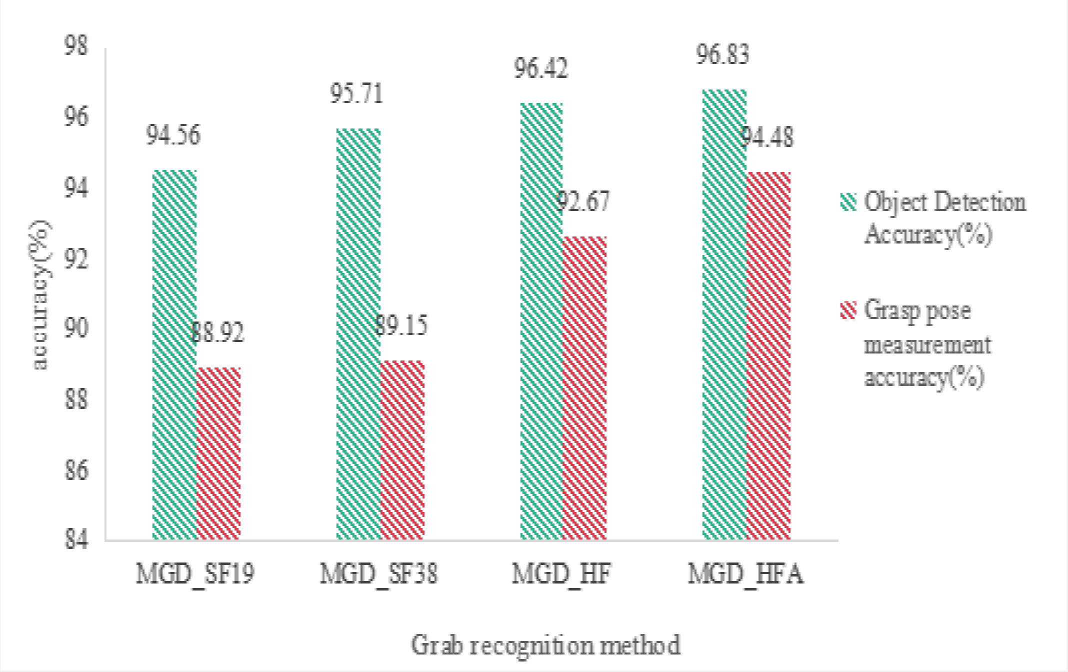

Figure 3. Accuracy (%) of the grab recognition algorithm

Comparing the performance test results of each grasping recognition method in Figure 3 and Table 1, it can be seen that, compared with the multi-target object grasping and recognition method based on single-level features, the multi-target object grasping and recognition algorithm based on multi-level feature fusion has better performance in object detection. Excellent recognition results have been achieved on both functional modules and grasp pose measurement, especially the MGD_HFA algorithm. The grasping pose measurement accuracy of the MGD_HFA grasping recognition algorithm is 1.81% higher than that of the MGD_HF method, and the object detection and grasping pose measurement accuracy of the MGD_HFA method are 5.56% and 5.33% higher than those of the MGD_SF19 and MGD_SF38 methods. It can be seen that the multi-level feature fusion algorithm based on the attention mechanism can make full use of the useful information in the image, reduce the interference of noise signals, and then improve the recognition accuracy of the grasping recognition model.

4.2. Simulation Experiment Verification of Robot GD System

The experiment was divided into two groups. The first group carried out a single-object environment grasping experiment. The position and posture of a single object were arbitrary, and 30 experiments were carried out. The second group conducted the unstructured environment grasping experiment, placing multiple objects without touching each other, and conducted a total of 50 experiments. The result record is shown in Table 2, which includes the time of image segmentation, the total time to complete the grab, the probability of successful target detection and the probability of successful experiment.

Table 2. Experimental Results

	Average duration of contour extraction (s)	Average time to complete crawl (s)	Total crawls	Target detection experiment success rate	Experiment success rate
Single object	1.236	6.3	30	98.75%	97.24%
Multiple objects	1.471	6.5	50	96.34%	95.82%

From the results in Table 2, it can be seen that when experimenting on a single object, the average duration of image segmentation is 1.236s, the complete grasping time is 6.3s, the target detection success rate is 98.75%, and the grasping success rate is 97.24%; When grabbing is implemented, the average duration of image segmentation is 1.471s, the complete grabbing time is 6.5s, the target detection success rate is 96.34%, and the grabbing success rate is 95.82%.

On the whole, if the object is correctly identified, the robot can basically successfully grasp the object. When there are multiple objects, the average processing time of the image increases. The preliminary analysis reason is that there are more objects in the scene, some object contours overlap, the image segmentation time increases, and the target recognition and pose estimation time increases. Experiments show that when the position of the target object is given, the grasping time of the robot is the same, which indicates that the robot trajectory planning control is stable.

5. Conclusion

In this paper, the research on the robot grasping system based on deep learning is carried out for construction machinery robots, and a robot GD system is built based on CNN. The position information of the object provides the grasping pose and motion trajectory for the robot. It is hoped that the application of this system in industrial production can improve the adaptability of robots to new environments and reduce the workload of production personnel.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Ulrich M, Follmann P, Neudeck J H. A comparison of shape-based matching with deep-learning-based object detection. Tm - Technisches Messen, 2019, 86(11):685-698. https://doi.org/10.1515/teme-2019-0076

[2] Weber I, Bongartz J, Roscher R. Artificial and beneficial - Exploiting artificial images for aerial vehicle detection. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 175(8):158-170. https://doi.org/10.1016/j.isprsjprs.2021.02.015

[3] Hwang P J, Hsu C C, Wang W Y. Development of a Mimic Robot-Learning From Demonstration Incorporating Object Detection and Multiaction Recognition. IEEE Consumer Electronics Magazine, 2020, 9(3):79-87. https://doi.org/10.1109/MCE.2019.2956202

[4] Rouhafzay G, Cretu A M, Payeur P. Biologically Inspired Vision and Touch Sensing to Optimize 3D Object Representation and Recognition. IEEE Instrumentation and Measurement Magazine, 2021, 24(3):85-90. https://doi.org/10.1109/MIM.2021.9436099

[5] Tsuru M, Escande A, Tanguy A, et al. Online Object Searching by a Humanoid Robot in an Unknown Environment. IEEE Robotics and Automation Letters, 2021, PP(99):1-1.

[6] Rosenberger P, Cosgun A, Newbury R, et al. Object-Independent Human-to-Robot Handovers Using Real Time Robotic Vision. IEEE Robotics and Automation Letters, 2021, 6(1):17-23. https://doi.org/10.1109/LRA.2020.3026970

[7] Cortez W S, Oetomo D, Manzie C, et al. Technical Note for "Tactile-Based Blind Grasping: A Discrete-Time Object Manipulation Controller for Robotic Hands". IEEE Robotics and Automation Letters, 2020, PP(99):1-1.

[8] Puente P, Fischinger D, Bajones M, et al. Grasping objects from the floor in assistive robotics: real world implications and lessons learned. IEEE Access, 2019, 7(99):123725-123735. https://doi.org/10.1109/ACCESS.2019.2938366

[9] Bottarel F, Vezzani G, Pattacini U, et al. GRASPA 1.0: GRASPA is a Robot Arm graSping Performance BenchmArk. IEEE Robotics and Automation Letters, 2020, 5(2):836-843. https://doi.org/10.1109/LRA.2020.2965865

[10] Hasegawa S, Yamaguchi N, Okada K, et al. Online Acquisition of Close-Range Proximity Sensor Models for Precise Object Grasping and Verification. IEEE Robotics and Automation Letters, 2020, PP(99):1-1.

[11] Bunis H A, Rimon E D. Toward Grasping Against the Environment: Locking Polygonal Objects Against a Wall Using Two-Finger Robot Hands. IEEE Robotics and Automation Letters, 2019, 4(1):105-112. https://doi.org/10.1109/LRA.2018.2882865

[12] Ishak A J, Mahmood S N. Eye in hand robot arm based automated object grasping system. Periodicals of Engineering and Natural Sciences (PEN), 2019, 7(2):555-566. https://doi.org/10.21533/pen.v7i2.528

[13] Sarakon P, Kawano H, Shimonomura K, et al. Improvement of Shrinking CNN Architecture Using Weight Sharing and Knowledge Distillation for Tactile Object Recognition. ICIC Express Letters, 2021, 12(7):627-633.

[14] Takeuchi M, Kawakubo H, Saito K, et al. ASO Visual Abstract: Automated Surgical Phase Recognition for Robot-Assisted Minimally Invasive Esophagectomy Using Artificial Intelligence. Annals of Surgical Oncology, 2021, 29(11):6858-6859.

https://doi.org/10.1245/s10434-022-12006-0

[15] Brosque C, Fischer M. safety qua1ity schedu1e and cost impacts of 10 construction robots. Construction Robotics, 2021, 6(2):163-186. https://doi.org/10.1007/s41693-022-00072-5

[16] Bobkov V A, Kudryashov A P, Inzartsev A V. Object Recognition and Coordinate Referencing of an Autonomous Underwater Vehicle to Objects via Video Stream. Programming and Computer Software, 2021, 48(5):301-311. https://doi.org/10.1134/S0361768822050024

[17] Ko J H. Robot Vision System based on 3D Depth map and Object Recognition. Journal of the Institute of Electronics and Information Engineers, 2020, 57(3):101-105. https://doi.org/10.5573/ieie.2020.57.3.101

[18] A Kovács, Erds F G, Tipary B. Planning and Optimization of Robotic Pick-And-Place Operations in Highly Constrained Industrial Environments. Assembly Automation, 2021, 41(5):626-639. https://doi.org/10.1108/AA-07-2020-0099

[19] Hanafusa M, Ishikawa J. Mechanical Impedance Control of Cooperative Robot during Object Manipulation Based on External Force Estimation Using Recurrent Neural Network. Unmanned Systems, 2020, 08(03):239-251. https://doi.org/10.1142/S230138502050017X