Control method of shaft and hole mating based on convolution neural network in assembly building prefabricated components

To cite this article: K Zhang et al 2018 IOP Conf. Ser.: Mater. Sci. Eng. 399 012061

View the article online for updates and enhancements.
Control method of shaft and hole mating based on convolution neural network in assembly building prefabricated components

K Zhang¹, S Tong¹*, J Zhao¹ and Q Wang¹
¹College of mechanical engineering, Shenyang Jianzhu University, China.

* E-mail: tongshenghao@sjzu.edu.cn

Abstract. In order to further improve the automation degree of assembly building construction, realize prefabricated component intelligentized positioning, this paper applies the convolutional neural network in positioning control in the field of prefabricated component of assembly building, and puts forward a probability search into target region to establish a new method for loss equation algorithm, the method based on multi label classification loss equation modelling with probability search algorithm, in the training process to continuously improve the positive incentive probability score. The simulation results show that the designed convolution neural network can achieve a correct prediction of the assembly location control instruction, and the proposed method can improve the prediction accuracy and convergence of the network model.

1. Introduction
In prefabricated building construction, after the prefabricated component is hoisted to a designated area, the alignment of the connection hole at the bottom of the prefabricated component and the embedded reinforcement is a special “shaft-hole” assembly problem. When assembling and positioning, the multiple connection holes at the bottom of the prefabricated wall panel must be aligned with the steel bars at the same time, which makes it more difficult. Therefore, it has become the most important issue that restricts the intelligent construction of prefabricated building construction. In recent years, scholars have proposed a variety of theories and methods for automated assembly technology, mainly including the Cartesian impedance control theory, the force/position integrated control assembly theory, the digital assembly technology based on laser tracking, the vision-based servo control theory, and assembly movement. Jokesch [1] proposed a general algorithm for axle hole assembly based on Cartesian impedance control, which was applied to the automatic connection and insertion of an electric vehicle charging device and charging pile, and completed seven pins and holes with asymmetrical positions aligned at the same time, and evaluated the algorithm from different starting positions. Touati [2] proposed an automatic assembly method for force/position comprehensive control based on artificial neural network. First, the neural observer used the data obtained during the contact process of the assembly as an input for off-line training; then, the training parameters were applied. In the neural network with online learning function, the assembly closed-loop control system has good performance. Song [3] proposed a complex shape part assembly control strategy. It controls the forces during assembly movements based on visual geometry information and
CAD models, using cameras to track the position and orientation of the parts. In the assembly process, the impedance control method is used to control the contact force and avoid the occurrence of excessive contact force.

The above control method uses a large number of sensors and requires accurate modelling of the space position of the shaft. Therefore, the amount of calculation is very large, and requires a higher assembly environment, and system parameters need to be readjusted when the assembly environment changes [4]. Obviously, the above control method is not applicable to construction sites with complex construction environment and changeable weather conditions. Therefore, a prefabricated component assembly method, based on the convolution neural network, is proposed in this paper. The relative position image of the bottom connection hole of the prefabricated component and the top of the pre-embedded steel bar is used as the input, and the control instruction of the prefabricated component six degrees of freedom hoisting platform is used as the output, and the neural network is built and trained. This method can reduce the degree of dependence on sensors, reduce control variables and increase the intelligence of assembly of the prefabricated component. Convolution neural network has been applied in various fields in recent years [5], and has made breakthroughs in speech recognition, image recognition, general object recognition, motion analysis, Natural Language Processing and even brain wave analysis. For the first time, this paper combines the big data of construction site with the predictive classification advantages of convolutional neural networks, and proposes a prediction method of prefabricated building in-situ control instructions, based on convolutional neural networks. It can be stated that the deep learning was applied to an exploratory research in the field of building construction. The specific operation is shown in figure 1.

![Flow chart](image)

**Figure 1. Flow chart.**

2. Algorithm and structure

2.1. Algorithm

Because neural network is only interested in some areas of input image, how to search and locate the target area and give feedback in time is an important factor of the neural network. The existing method of visual target recognition usually relies on training data with a large number of annotated images [6].

The image labeling method based on the bounding box is complex and involves many human activities, which takes time and reduces the accuracy of identification. This paper presents a probabilistic search into the target region to establish a new method for loss equation algorithm, this method in multi label classification loss equation, based on adding probability search algorithm, the network training process to continuously improve the positive incentive probability score, can further improve the model prediction accuracy and convergence. This algorithm adds a gradient factor on the basis of the original algorithm, reduces the search time of the significant area of the image, and makes the network model converge quickly. The specific algorithms are as follows:

If target region output $f_k^{(t-1)}(x)$ at $t-1$ time, corresponding probability model is $P[f_k^{(t-1)}(x)]$, the increment is $\Delta f_k^{(t)}(x)$ at time of $t$, the Image features can be represented as:

$$\phi[f_k^{(t)}(x)] = \phi[f_k^{(t-1)}(x) + \Delta f_k^{(t)}(x)]$$  \hspace{1cm} (1)

The probability model of the corresponding time target region can be expressed as:

$$P[f_k^{(t)}(x)] = P[f_k^{(t-1)}(x) + \Delta f_k^{(t)}(x)]$$

$$= \sum_{i=1}^{k} \omega_i \phi[f_k^{(t-1)}(x) + \Delta f_k^{(t)}(x)] + b_i$$  \hspace{1cm} (2)

$$= P[f_k^{(t-1)}(x)] + P[\Delta f_k^{(t)}(x)]$$

$\omega_i$ the formula comprises: the corresponding weight value, $b_i$ the bias, $\phi_i$ the i eigenvalue of the pixel point.

The loss equation is the sum of all two logic regression loss functions [7], from which we could get:

$$L[f_k^{(t)}(x), y_k] = \sum_k \log(1 + e^{-y_k f_k^{(t)}(x)})$$  \hspace{1cm} (3)

$f_k^{(t)}(x)$ is the output of the neural network to the input $x$ at time $t$, $y_k \in (-1, 1)$ represents the output target category $k$ of input $x$.

2.2. Network Structure

Different from the traditional convolutional neural network structure, in order to adapt to the algorithm idea of this paper, we have redesigned the next full-connected layer of the single-channel convolutional neural network according to the training characteristics and training objectives adopted. Most DCNNs use a softmax layer to calculate the classification results after they have fully connected layers [8], and this paper calculates the target output directly through the fully connected layer. Deep network structure and more nodes or filters can improve the performance of neural networks, but it also increases the complexity and training difficulty of the network [9]. Based on the data size of the training set, this paper weighs the relationship between network performance and training difficulty. We use a network structure that consists of 3 convolution layers and 2 full connection layers. After the full connection layer, each active action corresponds to a separate output layer.

Input layer: 128 × 128 RGB image as input to the network.

Convolutional layer: The network consists of three convolutional layers. The first layer uses a convolution filter size of 7×7, and the other two layers have a filter size of 3×3. The step size is set to 1×1.

Pooling layer: 3 activation function layers followed by 3 pooling layers, all using maxpooling, filter size 3×3, step size 2×2.

Fully connected layer: 3 convolutional layers followed by 2 fully connected layers and 2 active function layers.

Output layer: After two fully connected layers are output layers for each valid action corresponding to a single output.
3. Experimental analysis

3.1. Experimental data
The experimental data and images are from the V101150.5 industrial building component lifting installation automatic control digital simulation system. The system has a complete prefabricated library including PC board, PCF board, composite floor, and more than 10 thousand assembly building construction method pieces of information [10]. The sample image is the relative position image of the embedded hole at the bottom of the precast member and the embedded steel bar joint on the floor, which is used to simulate the real lifting condition and enlarge the sample capacity at the same time, adjust the brightness and contrast of the image, as shown in figure 2. 5000 sample images are used for training, and 1000 sample images are used for testing. The 5000 samples were divided into 50 batches, each containing 100 samples.

![Sample image processing](image)

**Figure 2.** Sample image processing.

3.2. Analysis of test results
After the training is completed, the convolution neural network is used to classify the test set [11], and there are 12 action instructions in the output terminal, as follows: movement of the X axis positive direction, movement of the X axis negative direction, movement of the Y axis positive direction, movement of the Y axis negative direction, movement of the Z axis positive direction, movement of the Z axis negative direction, clockwise rotation of X axis, counter clockwise rotation of X axis, clockwise rotation of Y axis, counter clockwise rotation of Y axis, clockwise rotation of Z axis, counter clockwise rotation of Z axis. The reference coordinate system is built on the basis of the building XY plane. The output of the neural network is grouped by the direction of the control instruction [12], and compares the output of the neural network with the expert instruction after grouping. The same decision is correct. The correct output instruction number is divided by the total number of instructions of the group to determine the correct rate of the control instruction of the group. The correct rate of each group is calculated, and then the average correct rate of average output is determined. The result of the experiment is compared with the result of expert labelling, the average correct rate is 90.1%, and the specific results are shown in table 1. The table illustrates that the prediction accuracy of movement of the Y axis negative direction is the highest, and the prediction...
accuracy of clockwise rotation of X axis and counter clockwise rotation of the Y axis is lowest. The prediction accuracy of the mobile output along three axes is higher than that of its revolving around the axis.

**Table 1. Prediction results.**

| Action command                          | Correct rate |
|-----------------------------------------|--------------|
| Positive direction of X axis            | 91.5%        |
| Negative direction of X axis            | 93.6%        |
| Positive direction of Y axis            | 93.1%        |
| Negative direction of Y axis            | 93.2%        |
| Positive direction of Z axis            | 94.8%        |
| Negative direction of Z axis            | 94.9%        |
| Clockwise rotation around the X axis    | 84.5%        |
| Reverse clockwise rotation around the X axis | 84.6%    |
| Clockwise rotation around the Y axis    | 85.2%        |
| Reverse clockwise rotation around the Y axis | 84.5%    |
| Clockwise rotation around the Z axis    | 90.2%        |
| Reverse clockwise rotation around the Z axis | 90.6%    |
| Average value                           | 90.1%        |

In order to explore the effects of different distances of the prefabricated component bottom and the top of rebar on the output performance of the network model, and then to do the model simulation experiment, grouped by distance of the prefabricated component bottom and the top of rebar from the packet input samples to test the network, there is need to obtain the network prediction accuracy with distance of prefabricated component bottom and the top of rebar. As shown in figure 3, the correct rate of prediction increases with the distance of the prefabricated component bottom, and the top of rebar increased, when the distance of prefabricated component bottom and the top of rebar is up to 7.8cm, the network model to predict the average highest correct rate is 96.5%, then the correct rate gradually decreases with the increase of distance.

![Figure 3](image-url)  
*Figure 3. Influence of the distance between the bottom of the prefabricated component and the top of the rebar on the predictive performance of the network in the sample.*
4. Conclusion
Considering the difficulties raised by the position control of the assembly building prefabricated component, we have designed the multi-layer convolution neural network, which proposes a probability search into the target region to establish a new method for loss equation algorithm. Our experiment shows that:
(1) the convolution neural network designed in this paper can realize the correct prediction of the position control instruction of the prefabricated component of the assembly building;
(2) this method improves the prediction accuracy and convergence of the network;
(3) the influence of the distance between the bottom of the prefabricated component and the top of the rebar on the prediction performance of the network is studied, which lays the foundation for the follow-up research.

Acknowledgments
1. National Key R & D Program of China “Research on automatic control technology and system for installment and installation of industrial building components” 2017YFC0703903.
2. Shenyang science and technology program project F16-096-1-00.
3. National Natural Science Fund, 51675353.

Reference
[1] Jokesch M, Suchý J, Winkler A, et al. 2016 Generic Algorithm for Peg-In-Hole Assembly Tasks for Pin Alignments with Impedance Controlled Robots Robot 2015: Second Iberian Robotics Conference (Springer International Publishing) p 1949
[2] Touati Y, Amirat Y and Saadia N 2007 Artificial Neural Network-based Hybrid Force/Position Control of an Assembly Task International IEEE Conference on Intelligent Systems (IEEE)
[3] Song H C, Kim Y L and Song J B 2014 Automated guidance of peg-in-hole assembly tasks for complex-shaped parts ieee/rsj International Conference on Intelligent Robots and Systems (IEEE) p 4517
[4] Tompson J, Stein M, Lecun Y, et al. 2014 Real-Time Continuous Pose Recovery of Human Hands Using Convolutional Networks Acm Transactions on Graphics 33 (5) 1
[5] Mnih V, Kavukcuoglu K, Silver D, et al. 2015 Human-level control through deep reinforcement learning Nature 518 (7540) 529
[6] Cireşan D, Meier U, Masci J, et al. 2012 Multi-column deep neural network for traffic sign classification Neural Networks - Official J. of the Int. Neural Network Society 32 (1) 333
[7] Wei Y, Liang X, Chen Y, et al. 2016 STC: A Simple to Complex Framework for Weakly-supervised Semantic Segmentation IEEE Trans. on Pattern Analysis & Machine Intelligence 39 (11) 2314
[8] Vallance R R, Morgan C and Slocum A H 2004 Precisely positioning pallets in multi-station assembly systems Precision Eng. 28 (2) 218
[9] Wang X G, Sun Y and Tang XO 2015 From unified subspace analysis to joint deep learning: progress of face recognition in the last decade Com. of the CCF 11 (4) 8
[10] Oquab M, Bottou L, Laptev I, et al. 2015 Is object localization for free? - Weakly-supervised learning with convolutional neural networks Computer Vision and Pattern Recognition (IEEE) p 685
[11] Dauphin Y N, De Vries H and Bengio Y 2015 Equilibrated adaptive learning rates for non-convex optimization J 35 (3) 285
[12] Bellemare M G, Veness J and Bowling M 2012 Investigating contingency awareness using Atari 2600 games Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI Press) p 864