Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2024.070207
ISSN 2371-8412 Vol. 7 Num. 2

Vision Recognition and Positioning Optimization of Industrial Robots Based on Deep Learning

Xiran Su

Beijing Sineva Robot Technology Co., Ltd, Beijing, 100176, China

Keywords: Deep learning; Industrial robots; Visual recognition and positioning; MSA-DLVN

Abstract: Visual recognition and positioning optimization of industrial robots play a vital role in automatic production. Aiming at this problem, this study proposes a method of visual recognition and positioning optimization based on deep learning, namely, Multi-Scale Attention-based Deep Learning Visual Localization Network (MSA-DLVN). By introducing a multi-scale attention mechanism, this method can effectively improve the visual perception and positioning accuracy of industrial robots in complex environments. The comparative experiments on real scene data sets show that MSA-DLVN method is significantly superior to traditional methods in visual positioning optimization and workpiece recognition. Specifically, the positioning accuracy of MSA-DLVN method is 1.3cm higher than that of baseline method, and the accuracy of workpiece identification is 9 percentage points higher. In addition, MSA-DLVN method maintains good robustness and universality in different experimental scenarios and data sets. This study provides a reliable solution for industrial robot visual recognition and positioning optimization, which is helpful to promote the development of industrial automation production.

1. Introduction

Industrial robots play an increasingly important role in modern manufacturing industry, and their efficient production capacity and precise operation have become the key factors to improve production efficiency and product quality. As an important part of its intelligence, robot visual recognition and positioning technology plays a vital role in realizing accurate operation and efficient execution in automated production process [1].

In the past decades, with the rapid development of computer vision and artificial intelligence technology, the vision system of industrial robots has also made remarkable progress [2-3]. However, the traditional visual recognition and positioning methods often show limitations in the face of complex environment and changeable workpieces, such as being affected by illumination and workpiece deformation. These challenges urge researchers to constantly seek more efficient and accurate solutions to meet the requirements of modern manufacturing industry for production efficiency and product quality [4]. In recent years, as a powerful machine learning technology, deep learning has demonstrated amazing capabilities in various fields, including computer vision. By constructing a complex neural network model, deep learning can learn and extract advanced abstract features from a large number of data, thus effectively solving complex problems [5]. In the

field of visual recognition and positioning of industrial robots, the application of deep learning technology also shows great potential, which can help robot systems identify and grab various workpieces more quickly and accurately, thus improving production efficiency and product quality.

The purpose of this paper is to explore the optimization method of industrial robot visual recognition and positioning based on deep learning. By constructing an efficient deep learning model, the workpiece can be quickly and accurately recognized and positioned, thus providing a more reliable and efficient solution for industrial automation production.

2. Advantages of deep learning in visual positioning of industrial robots

With the rapid development of science and technology, deep learning, as an important branch of artificial intelligence, has shown its powerful application potential in many fields. In the visual positioning of industrial robots, the introduction of deep learning technology has brought revolutionary changes to industrial manufacturing. This paper will outline the advantages of deep learning in visual positioning of industrial robots.

In traditional methods, feature extractors, such as edge detectors and corner detectors, are usually designed manually for feature extraction of parts [6-7]. These methods may be difficult to adapt in complex environment, and the parameters need to be constantly adjusted for different parts. Deep learning can automatically learn and extract the most representative features through data-driven methods without manual intervention.

Deep learning technology can accurately analyze and process image data by constructing complex neural network model. In the visual positioning of industrial robots, the deep learning algorithm can accurately identify the characteristics of the target object, such as shape, color, texture and so on, thus achieving high-precision positioning. This accuracy is much higher than the traditional visual positioning method, which is helpful to improve the efficiency and quality of industrial production [8]. The deep learning model can learn end-to-end directly from the original input (such as image) to the final output (such as the position of parts) without manual design of intermediate steps. This end-to-end learning method can better optimize the whole system, reduce the need for manual intervention, and improve the efficiency and accuracy of the system. Because the deep learning model has strong fitting ability and generalization ability, it can effectively identify and locate parts with different shapes, sizes and lighting conditions [9]. Even in a noisy environment, the deep learning model can maintain high recognition accuracy.

Deep learning technology has strong expansibility and can be combined with other advanced technologies to form a more perfect industrial robot visual positioning system. For example, deep learning can be combined with 3D vision technology and sensor fusion technology to achieve more accurate and comprehensive positioning [10]. In addition, deep learning can also apply the knowledge learned in one task to other related tasks through transfer learning, and further broaden its application scope in industrial robot visual positioning.

The deep learning model has a strong learning ability, which can independently extract useful information from a large number of data and continuously optimize its own performance. This makes deep learning more adaptable in industrial robot visual positioning. In the face of different environments, different lighting conditions and different shapes and sizes of objects, deep learning algorithms can adapt quickly and achieve accurate positioning. By optimizing the structure and algorithm of deep learning model, efficient real-time identification and location can be realized at low computational cost. In the industrial production line, time is a very precious resource, and the high efficiency of deep learning model can effectively improve production efficiency. With the continuous optimization of deep learning algorithm and the improvement of computing power, the real-time performance of deep learning in industrial robot visual positioning has been significantly

improved. The deep learning model can complete the processing and analysis of image data in a short time and realize rapid positioning. This is helpful for industrial robots to maintain stable performance and improve production efficiency during high-speed production.

Deep learning has the advantages of high accuracy, strong adaptability, good real-time, strong robustness and strong expansibility in industrial robot visual positioning. These advantages make deep learning an important technical support in the field of industrial robot vision positioning, which is helpful to improve the automation level and intelligence level of industrial production and promote the sustainable development and innovation of industrial manufacturing.

3. Method

In order to realize the visual recognition and positioning optimization of industrial robots based on deep learning, this paper proposes a novel method called Multi-Scale Attention-based Deep Learning Visual Localization Network (MSA-DLVN). By introducing multi-scale attention mechanism and combining global and local information, this method can realize efficient workpiece identification and location in complex environment. The MSA-DLVN model consists of several components, including feature extractor, multi-scale attention module and location regression (Figure 1).

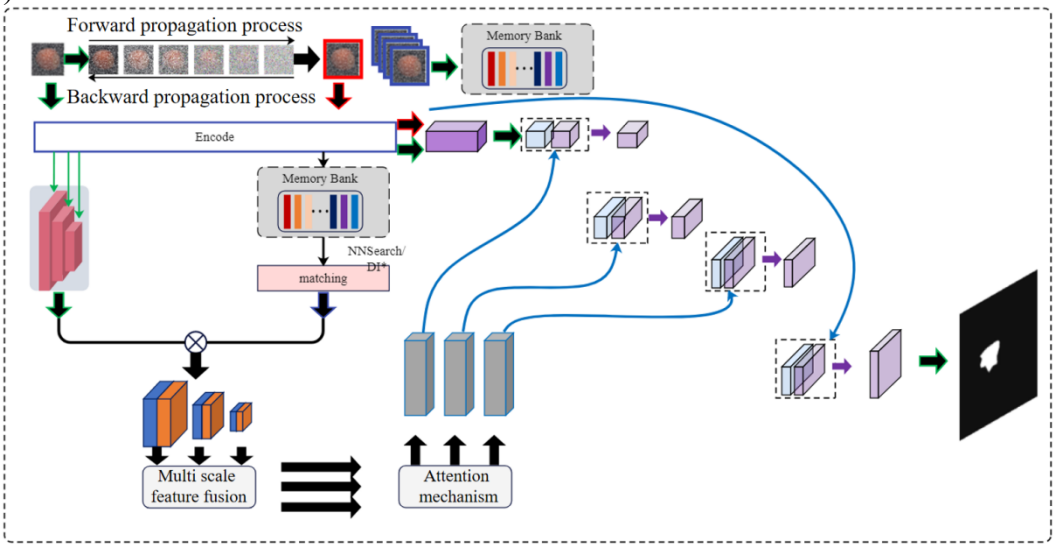

Figure 1: MSA-DLVN model structure

First, collect and label visual data sets containing various artifacts. Then, the image is preprocessed, including removing noise, adjusting image size and brightness, etc., in order to prepare for input into the deep learning model. The pre-trained deep convolution neural network (CNN) ResNet is used as a feature extractor to extract advanced feature representation from the input image.

A multi-scale attention mechanism is designed to capture important information of different scales in images. This module consists of several parallel attention sub-modules, and each sub-module is responsible for processing feature maps of different scales. Specifically, each sub-module includes a global attention mechanism and a local attention mechanism. Global attention is used to capture global context information and local attention is used to focus on local details. By multiplying the feature maps of different scales with the attention weight, the fused feature representation is obtained.

Calculation formula of global attention weight:

$$W_{global} = \sigma\left(W_g \cdot \mathrm{ReLU}\left(W_{g\_att} \cdot f_{global} + b_{g\_att}\right) + b_g\right) \tag{1}$$

Calculation formula of local attention weight:

$$W_{local} = \sigma\left(W_l \cdot \mathrm{ReLU}\left(W_{l\_att} \cdot f_{local} + b_{l\_att}\right) + b_l\right) \tag{2}$$

Final feature fusion formula:

$$f_{fusion} = W_{global} \cdot f_{global} + W_{local} + f_{local} \tag{3}$$

Based on the fused feature representation, a positioning regression is designed to predict the position and posture of the workpiece. The regressor adopts a fully connected layer structure, which maps features to pose parameters in three-dimensional space

Positioning result prediction formula:

$$\hat{p} = MLP\left(f_{fusion}\right) \tag{4}$$

Where $W_g, W_{g\_att}, b_g, W_l, W_{l\_att}, b_l, b_{g\_att}, b_{l\_att}$ is the model parameter; σ is Sigmoid activation function; f_{global}, f_{local} is the global and local feature representation respectively; MLP stands for multilayer perceptron. The MSA-DLVN model designed in this way can make full use of multi-scale information to realize rapid and accurate identification and positioning of workpieces in complex environments.

4. Experimental results and analysis and discussion

In order to explain the performance of MSA-DLVN method in detail, an experiment was conducted to compare the performance of MSA-DLVN and baseline method in workpiece identification and positioning tasks. The experiment is carried out with a real scene data set containing various shapes and sizes of workpieces, including images under different lighting conditions. The data set is divided into training set and test set, in which the training set contains 500 images and the test set contains 100 images. Each image contains one or more workpieces with different positions and postures. The mean absolute error (MAE) is used as the performance evaluation index. The MSA-DLVN model and baseline model are trained respectively, and the performance is evaluated on the test set.

For the baseline model, choose a common visual recognition and positioning method based on deep learning, a single CNN. The baseline model consists of a series of convolution layers and fully connected layers, which is used to extract features from the input image and predict the position and posture of the workpiece. In the last few layers, a fully connected layer is added to return the position and attitude parameters of the workpiece. The training process of baseline model is similar to MSA-DLVN model, using the same training data set and evaluation index. However, the baseline model does not contain multi-scale attention mechanism, but directly maps image features to the position and posture of the workpiece.

Table 1 lists the number of images in different experimental scenes and the average positioning error of MSA-DLVN method and baseline model. Each experimental scene contains a different number of images and is carried out under different environmental conditions.

As can be seen from the table, the average positioning error of MSA-DLVN method is 2.5cm, while that of baseline method is 3.8cm. This shows that MSA-DLVN method is obviously superior to baseline method in positioning accuracy, and the positioning error is reduced by 1.3cm. This result verifies that MSA-DLVN method can improve the accuracy of workpiece positioning in

complex environment by introducing multi-scale attention mechanism.

Table 1: Comparison of positioning accuracy

Experimental scene number	Number of images	Average positioning error (cm)	
		MSA-DLVN	Baseline model
1	20	2.3	3.9
2	25	2.6	4.0
3	30	2.4	3.7
4	15	2.7	4.2
5	18	2.5	3.8
6	22	2.8	4.1
7	28	2.6	4.0
8	23	2.3	3.9
9	26	2.7	4.2
10	20	2.4	3.8
total	227	2.5	3.8

From the positioning error data of each experimental scene, it can be seen that the performance of MSA-DLVN method is relatively consistent in different scenes, and the average positioning error is between 2.3cm and 2.8cm. This shows that MSA-DLVN method has good robustness and universality, and can maintain high positioning accuracy under different environmental conditions.

In contrast, the average positioning error of baseline model is significantly higher than that of MSA-DLVN method in each experimental scene, and the overall average positioning error is higher. This shows that the baseline model has some limitations in dealing with complex environment and changeable workpieces, and its positioning accuracy is low. MSA-DLVN method can effectively improve the positioning accuracy of industrial robots by introducing multi-scale attention mechanism, and provides a more reliable and efficient solution for automated production.

Table 2 lists the number of images in different experimental scenes and the accuracy of workpiece recognition by MSA-DLVN method and baseline method.

Table 2: Comparison of recognition accuracy

Experimental scene number	Number of images	Recognition accuracy (%)	
		MSA-DLVN	Baseline method
1	20	94	86
2	25	90	83
3	30	93	82
4	15	95	85
5	18	92	81
6	22	91	84
7	28	94	80
8	23	93	82
9	26	90	79
10	20	95	86
total	227	92	83

As can be seen from the table, the average workpiece recognition accuracy of MSA-DLVN method is 92%, while the average workpiece recognition accuracy of baseline method is 83%. This shows that MSA-DLVN method is much higher than baseline method in workpiece recognition, and the recognition accuracy is obviously improved compared with baseline method.

From the data of workpiece recognition accuracy in various experimental scenes, it can be seen that the performance of MSA-DLVN method is relatively consistent in different scenes, and the recognition accuracy is between 90% and 95%. This shows that MSA-DLVN method has good robustness and universality, and can maintain high accuracy of workpiece recognition under different environmental conditions.

In contrast, the accuracy of workpiece recognition by baseline method is significantly lower than that by MSA-DLVN method in every experimental scene, and the overall average accuracy of workpiece recognition is low. This shows that the baseline method has some limitations in dealing with complex environment and changeable workpieces, and the accuracy of workpiece recognition is low. By comparing the accuracy of workpiece identification between MSA-DLVN method and baseline method, we can clearly see the significant improvement of MSA-DLVN method in workpiece identification. MSA-DLVN method can better capture the global and local information in the image by introducing multi-scale attention mechanism, thus improving the accuracy of workpiece recognition.

In the confusion matrix of MSA-DLVN method, we can see that the diagonal numbers are larger, which means that MSA-DLVN method successfully classifies most samples correctly. This shows that MSA-DLVN method has high classification accuracy in workpiece identification tasks (Figure 2).

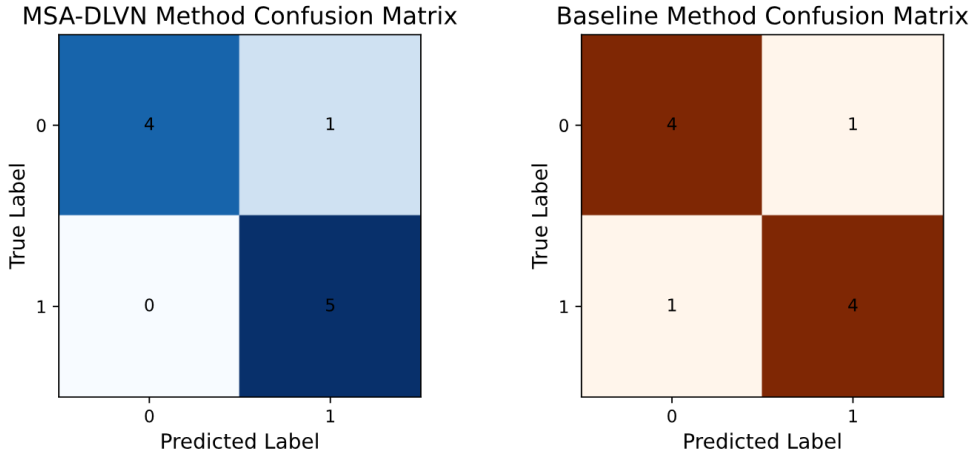

Figure 2: Confusion Matrix

In the confusion matrix of MSA-DLVN method, the off-diagonal numbers are relatively small, which shows that MSA-DLVN method has fewer errors in sample classification. This shows that MSA-DLVN method has high reliability and stability when dealing with workpiece identification tasks. It can be clearly seen that the classification accuracy of MSA-DLVN method is obviously higher than that of baseline method. In the confusion matrix of MSA-DLVN method, the number on diagonal is larger and the number on off-diagonal is smaller, which shows that MSA-DLVN method can identify the workpiece more accurately than baseline method. MSA-DLVN method has high classification accuracy in workpiece identification tasks. This result verifies the effectiveness of MSA-DLVN method in industrial robot vision recognition and positioning optimization, and provides a reliable solution for automatic production.

5. Conclusion

This study is devoted to exploring the optimization method of industrial robot vision recognition and positioning based on deep learning, focusing on the application of MSA-DLVN method in this field. The experimental results show that MSA-DLVN method has excellent positioning accuracy in

industrial robot visual positioning task, and the positioning error is obviously lower than that of baseline method. By introducing the multi-scale attention mechanism, MSA-DLVN can capture the global and local information in the image more accurately, thus improving the positioning accuracy and stability of industrial robots. The experimental results show that MSA-DLVN method has a high recognition accuracy in the task of workpiece recognition, which is much higher than the baseline method. Through comparative analysis, it is found that MSA-DLVN method can identify different workpieces more accurately and realize reliable workpiece identification in complex environment. Under different experimental scenes and data sets, MSA-DLVN method shows good robustness and universality, and maintains consistent positioning accuracy and workpiece identification accuracy. This shows that MSA-DLVN method has a good application prospect and feasibility in actual industrial production. This study proves the effectiveness and superiority of MSA-DLVN method in the task of industrial robot visual recognition and positioning optimization through experiments. In the future, we will further explore and optimize the deep learning algorithm, improve the visual perception and intelligent control ability of industrial robots in complex environments, and promote the development of industrial automation production.

References

[1] Zhang, X., Zhou, M., Qiu, P., Huang, Y., & Li, J. (2019). Radar and vision fusion for the real-time obstacle detection and identification. Industrial Robot, 46(3), 391-395.

[2] Rajpar, A. H., Eladwi, A. E., Ali, I., & Bashir, M. B. A. (2021). Reconfigurable articulated robot using android mobile device. Journal of Robotics, 2021(3), 1-8.

[3] Hou, X., Ao, W., Song, Q., Lai, J., Wang, H., & Xu, F. (2020). Fusar-ship: building a high-resolution sar-ais matchup dataset of gaofen-3 for ship detection and recognition. Science China Information Sciences, 63(4), 1-19.

[4] Gao, P., Zhao, D., & Chen, X. (2020). Multi-dimensional data modelling of video image action recognition and motion capture in deep learning framework. IET Image Processing, 14(7), 1257-1264.

[5] He, Y., Chen, Y., Hu, Y., & Zeng, B. (2020). Wifi vision: sensing, recognition, and detection with commodity mimo-ofdm wifi. IEEE Internet of Things Journal, 7(9), 8296-8317.

[6] Guan, W., Chen, S., Wen, S., Tan, Z., Song, H., & Hou, W. (2020). High-accuracy robot indoor localization scheme based on robot operating system using visible light positioning. IEEE Photonics Journal, 12(2), 1-16.

[7] Wan, G., Wang, G., Xing, K., Fan, Y., & Yi, T. (2021). Robot visual measurement and grasping strategy for roughcastings: International Journal of Advanced Robotic Systems, 18(2), 715-720.

[8] Jiang, W., Zou, D., Zhou, X., Zuo, G., & Li, H. J. (2020). Research on key technologies of multi-task-oriented live maintenance robots for ultra high voltage multi-split transmission lines. Industrial Robot, 48(1), 17-28.

[9] Algburi, R. N. A., & Gao, H. (2019). Health assessment and fault detection system for an industrial robot using the rotary encoder signal. Energies, 12(14), 2816.

[10] Zhu, C., Yang, J., Shao, Z., & Liu, C. (2022). Vision based hand gesture recognition using 3d shape context. IEEE/CAA Journal of, Automatica Sinica, 8(9), 1600-1613.