International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n1.53
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on Visual Localization and Gripping Technology of Robotic Arm Based on Deep Learning

Jiawei Tang

Liaoning University of Science and Technology, Liaoning. China
2014158926@qq.com

ABSTRACT

With the continuous development of industrial automation, robotic arms are more and more widely used in production lines. In order to realize the autonomous positioning and grasping function of robotic arm, this paper studies the visual positioning and grasping technology of robotic arm based on deep learning. The deep learning algorithm processes and analyzes the image, realizes the automatic identification and localization of the target object, and then guides the robotic arm to carry out accurate grasping. The experimental results show that this technology can effectively improve the positioning accuracy and grasping success rate of the robotic arm, which provides strong support for the intelligent upgrade of industrial automation production lines.

KEYWORDS

Learning; Robotic Arm; Visual Localization; Gripping Technology; Industrial Automation; Automatic Identification

1. INTRODUCTION

In the field of modern industrial automation, robotic arms play a crucial role, they can simulate the movements of human arms to perform a variety of complex operational tasks. The positioning and grasping function of robotic arms is a key link in the automated production process, and its accuracy and efficiency directly affect the overall performance of the production line. However, traditional robotic arm positioning and grasping methods often rely on pre-set fixed position and attitude information, this approach may be able to achieve good results in a static, controlled production environment, but in the face of complex and changing production scenarios, its limitations are particularly prominent.

Specifically, traditional methods are often unable to effectively respond to challenges such as changes in object position, shape variety, and varying lighting conditions. When the position of the object on the production line is shifted or a new variety of objects appear, the traditional robotic arm often need to manually reset the parameters, or even need to stop to adjust, which undoubtedly increases the cost of production and time costs. In addition, the traditional method is often difficult to ensure the accuracy and stability of gripping when dealing with complex situations such as object blocking and overlapping.

In recent years, with the rapid development of computer vision and deep learning technology, vision-based robotic arm positioning and grasping technology provides new ideas to solve the above problems. Computer vision technology can give the robotic arm the ability to "see", so that it can sense and understand all kinds of information in the production environment. Deep learning, as a powerful machine learning technology, can automatically extract useful features from large amounts

of data and learn complex mapping relationships. The combination of computer vision and deep learning can realize the automatic identification, precise positioning and stable grasping of objects in the production environment.

Therefore, the aim of this paper is to study the visual localization and grasping technology of robotic arm based on deep learning, which realizes the automatic recognition and localization of target objects by processing and analyzing the images using deep learning algorithms. Further, we will explore how to guide the robotic arm to perform precise grasping based on the recognition results, in order to improve the productivity and reduce the labor cost. This research not only has important theoretical value, but also provides strong support for the intelligent upgrading of industrial automated production lines.

2. RELATED WORK

Before exploring deep learning-based robotic arm visual localization and grasping techniques, it is crucial to understand the current research status and previous work in the field. In this section, we will review in detail the literature and previous research that are closely related to this study, which mainly involves the research progress of robotic arm visual localization techniques, the application of deep learning in image processing, and grasping strategies.

2.1. Robotic arm visual localization technology

Robotic arm visual localization technology is a prerequisite for accurate gripping. Traditional robotic arm localization methods usually rely on fixed reference points or markers, which perform well in static environments, but appear incompetent in dynamic or complex environments. In recent years, with the development of computer vision technology, vision-based robotic arm localization methods have gradually become a hot research topic. These methods usually use the image information captured by the camera to extract the position, attitude and other key information of the target object through image processing and analysis techniques, and then guide the robotic arm for precise positioning.

Feature extraction is one of the key steps in visual localization technology. Traditional feature extraction methods are mainly based on manually designed feature descriptors, such as SIFT, SURF, etc. However, these methods are often difficult to extract robust features in the face of complex and changeable images. In recent years, the development of deep learning technology provides a new idea for feature extraction. Deep learning model can automatically learn effective feature representation from a large number of data, thus improving the accuracy and robustness of feature extraction.

2.2. Deep learning in image processing

Deep learning is a machine learning technology based on neural network, and its application in image processing has achieved remarkable success. Convolutional Neural Networks (CNN) is one of the representative models of deep learning in image processing. By simulating the hierarchical structure of the human brain vision system, CNN can automatically extract the feature representation from the image from the low level to the high level, thus realizing the accurate understanding and analysis of the image.

Deep learning techniques have shown excellent performance in tasks such as target detection, image segmentation, and image recognition. In particular, deep learning models trained on large-scale datasets are able to learn richer image features and more complex mapping relationships, thus improving the accuracy and efficiency of image processing. These results provide a strong support for the visual localization and grasping technology of robotic arm based on deep learning.

2.3. Research Progress on Grabbing Strategies

Grasping strategy is the key to realize stable grasping of robotic arm. Traditional grasping strategies are usually based on fixed rules or empirical knowledge, and these methods are often difficult to ensure the stability and reliability of grasping when facing objects of different shapes, sizes and materials. In recent years, with the development of machine learning technology, learning-based grasping strategies have gradually become a research hotspot.

Learning-based crawling strategies usually use machine learning algorithms to learn crawling rules or strategies from a large amount of data. These methods first collect a large amount of grasping sample data, and then use machine learning algorithms to analyze and learn from these data to obtain a grasping strategy model. In practical applications, the grasping strategy model can predict the optimal grasping position and attitude according to the input object information, thus improving the stability and reliability of grasping.

In addition, reinforcement learning and other technologies have also made significant progress in grasping strategy learning. By interacting with the environment and constantly learning optimization strategies, reinforcement learning algorithms can enable the robot arm to gradually master the effective grasping methods for different objects. These achievements provide useful reference for the robot arm visual positioning and grasping technology based on deep learning.

3. METHODOLOGY

This study aims to realize autonomous localization and accurate grasping of robotic arms in complex environments, and to this end, a deep learning-based visual localization and grasping technology framework for robotic arms is proposed. The method mainly includes four key steps: image feature extraction and object recognition, grasping candidate region generation, grasping region scoring and optimization, and grasping execution.

3.1. Image Feature Extraction and Object Recognition

First, we use convolutional neural network (CNN) to extract features and recognize objects from captured images. CNN has made remarkable achievements in the field of image processing with its powerful feature learning ability. We select a pre trained CNN model, such as ResNet, VGG or EfficientNet, and fine tune it on a large image dataset to adapt it to specific object recognition tasks.

In the object recognition phase, the CNN model converts the input image into a series of feature maps, and outputs the category probability of each object through the classification layer. At the same time, the precise position information of the target object is obtained by using the positioning technology in the model (such as bounding box regression). This information provides an important basis for subsequent generation of candidate regions for grabbing.

3.2. Grabbing Candidate Region Generation

Based on the results of object recognition, we further generate candidate regions for grabbing. These areas are where the manipulator may perform a grab operation. In order to generate these candidate regions, we can adopt a variety of strategies, such as sliding window method, selective search or RPN based on deep learning.

In the sliding window method, we slide a fixed size window on the image with a certain step size and size, and take the image area in each window as a capture candidate. Selective search generates possible candidate regions for grabbing by merging similar small regions. In contrast, RPN based on deep learning can directly predict the position and size of candidate regions on the feature map, thus improving the generation efficiency and accuracy.

3.3. Grabbing Area Scoring and Optimization

After generating candidate regions for crawling, we need to score and optimize these regions to determine the best crawling location. The scoring process aims at evaluating the success rate of crawling each candidate region, while the optimization process aims at adjusting the crawling position to improve the success rate.

To achieve this goal, we can use deep learning methods, such as convolutional neural network (CNN) or deep Q network (DQN), to train a scratch scoring model. The model takes the image of the candidate region as the input, and outputs a predictive value of the capture success rate. By comparing the predicted values of different candidate regions, we can select the region with the highest success rate as the capture target.

In addition, in order to improve the stability and accuracy of grasping, we can also use some optimization strategies, such as mechanics-based grasping stability analysis, grasping attitude adjustment and so on. These strategies can help us further adjust the grasping position and direction to ensure that the robotic arm can grasp the target object stably.

3.4. Grabbing Execution

Finally, the optimized grasping position information is sent to the robotic arm to perform the grasping operation. In this step, we need to convert the image coordinates into 3D coordinates in the robot arm coordinate system, and plan the grasping trajectory by considering the kinematic and dynamic constraints of the robot arm. By precisely controlling the trajectory and grasping strength of the robot arm, we can realize the stable grasping of the target object. At the same time, we can utilize the sensor information to monitor and adjust the grasping state in real time during the grasping process to ensure the accuracy and safety of the grasping.

4. OPTIMIZATION OF METHODOLOGIES

Improving the gripping efficiency of deep learning-based robotic arm visual localization and gripping technology is a comprehensive challenge that involves the optimization of several aspects. The following is a detailed description of the optimization methods proposed above.

4.1. Optimize the network structure

The network structure of deep learning model is very important for its performance. In order to improve the capture efficiency, we can consider optimizing the network structure. First of all, using a more lightweight network structure can reduce the number of parameters and computational complexity of the model, thus speeding up reasoning. For example, using Depth Separable Convolution to replace the traditional convolution operation can significantly reduce the amount of computation without losing too much precision. In addition, model pruning technology can remove redundant connections or neurons in the network, further simplify the model and improve computing efficiency.

4.2. Crawling strategy optimization

Optimizing the crawling strategy is one of the key steps to improve efficiency. Traditional grasping strategies may be based on fixed rules or empirical knowledge, which cannot be adapted to various complex situations. Therefore, we can use reinforcement learning or imitation learning to train the grasping strategy of robotic arms. Through interaction with the environment and continuous learning, the robotic arm can gradually master effective grasping methods for different objects. In addition, combined with deep learning technology, we can design a more complex grasping strategy network

to adaptively select the appropriate grasping point and grasping force according to the shape, size, weight and other characteristics of the object. This can reduce the number of attempts and increase the probability of successful grasping, thus improving the overall grasping efficiency.

4.3. Multimodal information fusion

In addition to visual information, other sensor information can also be considered to improve the capture efficiency. For example, the combination of tactile and force sensors can provide more abundant information about the texture, hardness and quality of the object surface. By combining these multimodal information with visual information, we can more accurately perceive the state and attributes of objects and make more accurate grasping decisions. This multimodal information fusion method can significantly improve the robot arm's perception of the environment, thus improving the capture efficiency and success rate.

4.4. Parallelization

In practical applications, parallel processing can speed up the grabbing process. Using high-performance computing resources such as GPUs or multi-core CPUs, we can simultaneously process scoring and optimization tasks for multiple crawl candidate regions. In addition, the distributed computing framework can also be considered to further expand the computing power and simultaneously control multiple manipulators for collaborative grasping. Through parallel processing, we can make full use of computing resources, speed up the capture speed and improve the overall efficiency.

4.5. Quick Recovery from Grabbing Failure

When a crawl fails, the ability to quickly re-plan and execute a new crawl attempt is also an important part of improving efficiency. In order to realize this goal, we can design an effective failure detection mechanism to detect the failure of crawling in time and trigger the retry strategy. The retry strategy can include adjusting the crawling position, changing the crawling strength or trying different crawling strategies. By quickly re-planning and executing new crawling attempts, we can reduce the time wasted due to crawling failures and improve the overall crawling efficiency.

4.6. Cloud-based collaborative computing

Offloading part of the computational tasks to the cloud for processing is an effective way to realize more efficient crawling operations. Cloud servers have powerful computational power and storage resources, which can accelerate the inference process of deep learning models. By offloading computationally intensive tasks such as image processing and object recognition to the cloud, we can reduce the local computational burden of the robot arm and increase the processing speed. At the same time, cloud-based collaborative computing can also realize data sharing and model updating, which can further improve the efficiency and accuracy of grasping.

4.7. Hardware acceleration

Using special hardware accelerator to accelerate the reasoning process of deep learning model is another effective method to improve the efficiency of grasping. These hardware accelerators include FPGA (field programmable gate array), ASIC (application specific integrated circuit), etc. They are specially optimized for deep learning algorithms, which can significantly improve the computational efficiency and response speed. By deploying the deep learning model to run on these hardware accelerators, we can further accelerate the reasoning speed in the capture process and improve the overall efficiency.

5. EXPERIMENTS

In order to validate the effectiveness of the proposed deep learning-based visual localization and grasping technique for robotic arms, a series of experiments were designed and implemented. These experiments aim to evaluate the localization accuracy and grasping success rate of the method on different scenes and objects, and to compare it with the traditional method. The following is an overview of the program implementation of the experiments.

5.1. Experimental setup

1. Dataset preparation: Collect and label a dataset containing multiple objects and scenes. The dataset should contain images of objects, location information and category labels. This data will be used to train and test deep learning models.

2. Model training: use the selected deep learning framework (such as TensorFlow, PyTorch, etc.) to implement the convolutional neural network (CNN) model, and train on the prepared data set. During training, the accuracy and loss of the model should be monitored for necessary adjustment.

3. Grab strategy realization: based on the trained CNN model, realize the generation, scoring and optimization algorithm of grab candidate regions. These algorithms will combine the recognition results of objects and the kinematic constraints of the manipulator to determine the optimal grasping position and direction.

4. Robotic arm control: Convert the optimized gripping position information into three-dimensional coordinates in the robotic arm coordinate system and write control code to drive the robotic arm to perform the gripping operation. Ensure that the robotic arm can accurately reach the specified position and perform the gripping.

5. Experimental environment construction: build an experimental environment containing a camera, a robotic arm and an object. The camera is used to capture the image of the scene, the robot arm is used to perform the grasping operation, and the object is the target of grasping.

5.2. The experimental process, the

1. Model evaluation: evaluate the performance of the trained CNN model on the test set, including the accuracy of object recognition and positioning accuracy. These indicators will be used to measure the generalization ability of the model on different scenes and objects.

2. Grasping experiment: In the experimental environment, use the trained model and grasping strategy to grasp objects. Record the success of each attempt and collect relevant data (e.g., time, number of attempts, etc.) for subsequent analysis.

3. Comparative analysis: Deep learning-based visual localization and grasping techniques for robotic arms are compared with traditional methods. Traditional methods may include manual feature-based object recognition and localization algorithms, as well as simple grasping strategies. Through the comparative analysis, the advantages and room for improvement of the proposed technique can be evaluated.

4. Visualization of results: The results of the experiments are visualized, including the results of object recognition, the selection of candidate regions for grasping, and video or image sequences of the grasping process. These visualizations help to intuitively understand the performance of the proposed technique.

5.3. Experimental results and analysis

From the experiments, we can conclude the following.

1. The deep learning-based robotic arm visual localization and grasping technology has achieved high localization accuracy and grasping success rate on different scenes and objects. This shows that the technology has strong adaptability and robustness, and can cope with various complex environments and objects.

2. Compared with traditional methods, the proposed technique shows higher accuracy in object recognition and localization. This is mainly attributed to the powerful feature learning and generalization capabilities of the deep learning model.

3. In terms of grasping strategy, the proposed technology is able to adaptively select the appropriate grasping point and grasping strength according to the shape, size and other characteristics of the object, thus increasing the probability of successful grasping at one time. This reduces the number of attempts and time cost, and improves the overall grasping efficiency.

4. By comparing and analyzing the results under different experimental conditions, we can find that the proposed technique may have certain limitations in some specific scenes or objects. This provides a useful reference direction for further improvement and optimization of the technique.

6. RESULTS AND DISCUSSION

In this section, we will analyze and discuss the experimental results in detail, which are mainly evaluated in terms of localization accuracy, capture success rate, and algorithm operation time. In addition, we will also discuss the challenges and limitations that the proposed technique may face in practical applications, and look forward to future improvements and development directions.

6.1. Positioning accuracy

The experimental results show that the robot vision positioning technology based on depth learning achieves high positioning accuracy in different scenes and objects. Through comparative experiments, we find that the positioning accuracy of the proposed technology is significantly improved compared with the traditional methods. This is mainly due to the strong feature extraction and learning ability of the deep learning model, which enables it to more accurately recognize objects and predict their positions.

6.2. Grabbing success rate

The proposed technology also performs well in terms of grasping success rate. By combining the deep learning model and the optimized grasping strategy, the robotic arm is able to select the grasping point and the grasping force more accurately, which improves the probability of successful grasping. The experimental data show that the grasping success rate of the proposed technique is higher than that of the traditional method in different objects and scenes, which verifies the effectiveness and superiority of the proposed technique.

6.3. Algorithmic runtime

Although the proposed technology has achieved remarkable results in positioning accuracy and capture success rate, we also noticed the problem of relatively long running time of the algorithm. This is mainly due to the complexity of the deep learning model and the large amount of calculation. In order to further improve the capture efficiency, we can consider optimizing the network structure, adopting more efficient computing hardware or using parallel computing technology to shorten the running time of the algorithm.

6.4. Challenges and constraints

Despite the encouraging experimental results of the proposed technique, it may still face some challenges and limitations in practical applications. For example, the grasping strategy may need to be further adjusted and optimized for some objects with special shapes or materials. In addition, factors such as lighting conditions, occlusion and background interference may also adversely affect the localization accuracy and grasping success rate. In order to overcome these challenges, we can consider the introduction of multimodal information fusion and enhancement learning to improve the adaptability and robustness of the model.

6.5. Directions for future improvement and development

Looking ahead, we can improve and develop the proposed technique in the following aspects: first, optimize the network structure to reduce the computational complexity and increase the inference speed; second, study more advanced grasping strategies to adapt to more complex scenes and objects; third, explore the multimodal information fusion method to make full use of the advantages of different sensors; and lastly, focus on the research of real-time and embedded implementation to Finally, we focus on the research of real-time and embedded realization to meet the practical application requirements. Through continuous research and innovation, we believe that deep learning-based robotic arm visual localization and grasping technology will play a greater role in the future.

7. CONCLUSION

In this paper, the visual positioning and grasping technology of robotic arm based on deep learning is studied, and the automatic identification and positioning of target objects is realized by processing and analyzing the images with deep learning algorithm. The experimental results show that this technology can effectively improve the positioning accuracy and grasping success rate of the robotic arm, which provides strong support for the intelligent upgrade of industrial automation production lines. The future work will further optimize the algorithm performance, expand the application scenarios, and promote the wide application and development of robotic arm visual positioning and grasping technology.

REFERENCES.

[1] Obstacle avoidance in space robotics: Review of major challenges and proposed solutions[J]. Tomasz Rybus. Progress in Aerospace Sciences, 2018.

[2] Manipulator motion planning using flexible obstacle avoidance based on model learning[J]. Zhixuan Wei; Weidong Chen; Hesheng Wang; Jingchuan Wang.International Journal of Advanced Robotic Systems, 2017.

[3] Asynchronous Methods for Deep Reinforcement Learning.[J]. Volodymyr Mnih; Adrià Puigdomènech Badia; Mehdi Mirza; Alex Graves; Timothy P. Lillicrap; Tim Harley; David Silver; Koray Kavukcuoglu.CoRR, 2016.

[4] Towards Cognitive Exploration through Deep Reinforcement Learning for Mobile Robots.[J]. Lei Tai; Ming Liu 0001. CoRR, 2016.

[5] Continuous control with deep reinforcement learning.[J]. Timothy P. Lillicrap; Jonathan J. Hunt; Alexander Pritzel; Nicolas Heess; Tom Erez; Yuval Tassa; David Silver; Daan Wierstra. CoRR, 2015.

[6] Reinforcement Learning: An Introduction; by Richard S. Sutton and Andrew G. Barto, Adaptive Computation and Machine Learning series, MIT Press (Bradford Book), Cambridge, Mass., 1998, xviii + 322 pp, ISBN 0-262-19398-1,[J]. Alex M. Andrew.Robotica,1999(2).