Analysis on the Current Status and Development of Industrial Instrument Equipment Detection Technology Based on Convolutional Neural Network

Youkang Zhu

Watford College, Nanjing University of Information Science and Technology, Nanjing City, 210000, China
202383930029@nuist.edu.cn

Abstract. With the fast development of the Industry 4.0 program and the increasing popularity of high-level automation, the need for the identification and monitoring of industrial instruments grows as well. Convolutional neural networks can provide the precision and accuracy required for monitoring and recognition because of their powerful computing power and feature lock extraction, which can become a key technology to solve this problem. In this paper, the current status of digital automation monitoring of industrial instruments is reviewed by analyzing the real-life application cases of classical convolutional neural network models such as CRNN and YOLOv5 in industrial scenarios. According to the research, the application monitoring accuracy of CNN in complex industrial scenarios is excellent, but more research and results are needed in the lightweight of modules and the adaptability of multi-instrument equipment. In the future, the lightweight and highly efficient design of modules, multi-modal fusion technology, and self-supervised learning will be the main methods to help the industry system to be more intelligently advanced. This paper aims to offer technology suggestions to academia and industry, and motivate the industrial vision to improved.

Keywords: Convolutional Neural Networks, Attention Mechanism, Feature Extraction Capabilities, Industrialization and Intelligence

1 Introduction

With the fast development of the industry, blindly using a large number of labor has become an inefficient way to keep the industrial production and operation. It is not only inefficient but also high cost. Hazards can occur while workers are working. So high degree-automation of digital recognition and monitoring has become an inevitable trend and demand.

Convolutional neural networks have undergone a series of major developments and changes [1]. While deep learning is developing fast, nowadays the industry has widely used CNN to replace traditional image text recognition [2].

According to the research before, from the first LeNet-5 module which can recognize hand-writing figures [1] to the AlexNet module which greatly enables the CNN

J. Iqbal (ed.), *Proceedings of the 2025 2nd International Conference on Mechanics, Electronics Engineering and Automation (ICMEEA 2025)*, Advances in Engineering Research 272,
https://doi.org/10.2991/978-94-6463-821-9_47

application in the image recognition [2]. Additionally, according to An End-to-End Trainable Neural Network for Image-Based Sequence[3], the YOLO models[4] of CNN can be more powerfully used in character recognition. Now, depending on the AMR technology in industrial production research[5], CNNs can be used on a large scale in industry.

It has lots of advantages over the traditional methods and modules. CNN can achieve joint training of detection and recognition, which massively approves the efficiency of the CNN application and training.

However, although CNN has been used in automation and industry development, there are still some flaws of the module. For example, the lightweight problem of modules, the robustness problem in complex industrial environments, the adaptation challenges of various instruments, and the real-time accuracy requirements under the resource constraints of edge devices.

This research has the theoretical and practical two main purposes. In the theory level, this paper aims to offer suggestions for cutting-edge technologies such as multimodal fusion and self-supervised learning. In the practical level, this paper is going to give advice to improve the automation level of digital monitoring of industrial instruments to reduce labor costs increase factory safety factors, and provide technical support for intelligent upgrading for all automation industries.

2 The Development of CNN

Convolutional neural networks, which are called CNN, are a kind of deep-learning model. It has been widely used in text and video recognition, especially in computer vision and data reading, greatly improving the development process of automation and intelligence. CNN is composed of three parts, the convolutional layer, the pooling layer, and the fully connected layer.

The first convolutional layer is the core layer of the CNN which mainly performs the convolution operations. The input image is segmented, and then the convolution kernel is used to slide and convolve with the input image to obtain a feature map, thereby realizing the automatic extraction of local features of the image. Notably, the convolutional layer is a multi-level structure. By adding more convolutional layers, the feature reading capability can be enhanced and the depth of the model network can be increased.

The next layer is the pooling layer, which always follows the convolutional layer. It is usually used to reduce the spatial size of the feature map, reduce the amount of computation consumed by the entire model on the image, and enhance the robustness of the model. The common pooling operations are divided into maximum pooling by selecting the maximum value in the local range and average pooling by selecting the average value. Through these two pooling methods, the purpose of compressing features for the model can be achieved. Through pooling technology, the CNN model can compress the feature representation of the target object, improving the efficiency and accuracy of extraction while ensuring that the information is basically not lost.

In the final stage of the CNN model, there is always one or multiple fully connected layers. This layer is used to combine the features that convolution operations and pooling operations together and classify these merged feature values. Each neuron in this layer is fully connected, that is, each neuron is connected to all neurons in the previous layer. Through this process, CNN can effectively perform complex classification and recognition tasks.

The development of the CNN is fast yet stable. First coming is the Le-Net5, which provides the basis for convolutional neural networks and lays a solid foundation[1]. Based on the basic concept of gradient learning, the recognition of 2D handwritten characters is completed. Till now, studies have used the Alec-net as the backbone of the model to improve the handwritten character detection accuracy[2]. These advancements have propelled CNN into widespread applications.

In addition, CNN plays a huge role in the field of industry image recognition. The YOLO model which has the CNN or CRNN as the backbone can immediately detect objects and provide identification feedback [3]. There have been a lot of different versions of the YOLO model. The new YOLOv3 model uses Python and OpenCV can have higher accuracy in object detection. CRNN is also a kind of CNN model and has been widely used in scene recognition [4]. The recent study finds that CRNN can be used to perform a series of character tests instead of blocking them into a single one which greatly enhanced the CNN model's ability to detect scene text.

3 The Application Status

CNN has been widely used in the industry meter reading, and based on it, scientists created many powerful technologies to improve the instrument reading efficiency of digital information of industrial instruments.

AMR is one of these technologies, an automatic meter reading technology based on CNN [6]. Traditional AMR technology usually relies on electricity meters, gas meters and workers to read data. However, in many actual scenarios, due to factors such as poor ambient light, angle changes, and occlusion, it is difficult for traditional methods to obtain accurate data. By applying convolutional neural networks, these problems can be solved well. The digital display image of the instrument is sent to the input of the CNN model through a camera or image sensor. After pre-processing, the image is input to the CNN and gradually passes through the convolution layer, pooling layer, and fully connected layer of the CNN for feature extraction. According to CNN’s trained weights and structure, the model can accurately identify instrument numbers and output the identification results. These results can be used for data collection, real-time monitoring, and remote meter reading. Once the identification results are collected, the results data will be delivered to the center management system. Then combining some network technologies the device can synchronously send the result data to the cloud server for further analysis and processing. Based on CNN, AMR has some advantages over the traditional methods. First, given the automation of the AMR model equipment, it improves the degree of automation and efficiency in industrial production and reduces labor costs. In addition, due to the detail extraction ability of the CNN model, the recognition accuracy is also greatly improved.

Another is the high-precision welding detection and segmentation based on CNN [7]. Welding defect detection is an important part of industrial production. Traditional methods not only consume manpower, but also have low accuracy, poor real-time performance, and cannot be processed synchronously. Therefore, some studies have proposed using CNN as the foundation to build intelligent model detection. The data sources of the equipment are divided into two parts: one is the visual image, which provides the surface information of the welding area, and the other is the laser data, which is used to provide deeper depth information of the welding area. After collecting the two types of data, the data is aligned and fused for preprocessing to provide richer feature information. Combined with the feature fusion strategy, the high-dimensional features of the two data are fused. Then CNN is used to extract features and detect defect areas. Finally, the semantic segmentation network is used to accurately segment the detected defect area. This method significantly improves the accuracy and robustness of industrial defect detection and improves the level of industrial automation.

4 The Difficulty and Fault

Although CNN has been widely used in industrial production, it has not yet achieved perfection in digital reading of industrial instruments.

In modern industrial manufacturing, accurate digital reading is essential for equipment operation status monitoring, parameter control and maintenance decision-making.

Traditional manual reading methods are not only inefficient, but also easily affected by human errors, and it is difficult to meet the needs of industrial intelligent development.

Traditional OCR (optical character recognition) methods perform well in standardized fonts and clear backgrounds, but perform poorly in complex environments [6].

AMR methods perform poorly in complex environments and also perform poorly for multi-scale instrument detection, and real-time performance is not very strong. Here, this article will list some of the shortcomings of the current CNN-based models.

If companies want to improve the detection accuracy of AMR, they need to increase the network depth of the CNN model or the number of parameters of the CNN model to achieve higher computing power. These improvements mean that the computing resources required by CNN models continue to expand, resulting in excessive computing power and difficulty in deploying them on lightweight edge or embedded devices. In addition, the training of a CNN model usually requires a large amount of data set resources to develop sufficient detection and reasoning capabilities, so it is difficult to lightweight the CNN model.

In addition, in the context of modern society, the industrial production environment is very complex and difficult to overcome. Dust often appears in the factory, resulting in low visibility and insufficient lighting, which increases the difficulty of identification. The Gaussian noise generated by the image sensor will further affect the detection accuracy of the CNN model. There is also salt and pepper noise caused by random white or black dots appearing on image pixels due to dust, dirt or hardware defects. Finally, image blur caused by equipment aging or dust and oil pollution will cause image distortion, seriously affecting image quality and reducing the model's effectiveness.

In addition, the types of instruments are not limited to fixed styles but have a wide variety of categories, including but not limited to pointer instruments, digital instruments, LCD instruments, and digital tube instruments. Each type of meter has differences in the digital display format, color, and display layout. This makes it difficult to widely apply the same model to different types of instruments. If models are to be designed specifically for different instruments, the workload will be greatly increased and production efficiency will be reduced. If workers want to design detection functions based on the feature extraction requirements of different types of instruments on the same model, the complexity of model design will increase, requiring more volume resources and computing power, and it will also become difficult to modify, enhance and maintain the model.

Finally, there is the twin problem of real-time and accuracy of models built based on convolutional neural networks. In industrial production, the model responsible for monitoring tasks needs to report in real-time. Although the real-time performance of the model has been improved compared to traditional manual labor records and optical recognition, it is still not enough to cope with larger-scale and more efficient production. If industries pursue accuracy, the response time of the CNN model will be prolonged, and it will not be possible to make correct data transmission in the first place, resulting in a decrease in real-time performance. If industries pursue real-time efficiency, the recognition speed of the model will increase, and the error detection mechanism will be affected and weakened, resulting in a decrease in the accuracy of the model. The decrease in accuracy may lead to the collapse of the production line, causing danger without knowing it, and misjudgment of production data.

How to solve these problems should become the research direction of future CNN-based models.

5 The Future

In recent years, with the development of computer vision technology, convolutional neural networks (CNNs) have shown excellent feature extraction capabilities in the field of image recognition. The problem-solving of CNN models has also been greatly promoted.

Regarding the lightweight of CNN models, model pruning technology and model quantization technology have brought considerable progress. Model pruning technology is similar to the pruning of branches and leaves in reality [8][9]. When pruning a tree, it seems to damage the shape of the tree, but in fact it helps to reduce its burden and is also beneficial to the development of other main parts. The main method is to reduce the number of model network parameters and reduce the model by removing and cutting the neurons that contribute less to the feature extraction ability in the model [8].In most cases, the model clipping technique based on the weight size algorithm can significantly reduce the parameter size of the model while slightly affecting the model accuracy. Another model quantization technology is to reduce storage requirements and computational complexity by converting the 32-bit or 64-bit floating-point numbers originally planned for the CNN model into 8-bit floating-point numbers with lower precision.

About the robustness improvement of CNN model, New research suggests that multimodal fusion technology[10] can be used to improve the robustness and accuracy of regulatory system models. Multimodal fusion technology refers to the use of data from sensors such as temperature, humidity, and sound to monitor the actual industrial production process. The specific strategy steps can be divided into three parts: early fusion, mid-term fusion, and late fusion. Early fusion refers to the merging of multimodal data before feature extraction, merging and centralizing data sent by different sensors. Mid-term fusion refers to the alignment and combination of multimodal data during feature extraction. Late fusion refers to the final decision-level fusion of data. Multi-faceted data verification can significantly improve the robustness and accuracy of model recognition.

In the actual model training, in order to obtain better learning performance, large-scale data labeling is usually required to train the deep neural network of the CNN model, but the cost of doing so is usually high. In order to achieve the purpose of reducing costs, self-supervised learning came into being [11]. It refers to obtaining implicit information in unlabeled data by constructing preliminary training tasks, which can provide efficient feature representation for subsequent large-scale tasks and effectively alleviate the problem of the high cost of industrial data labeling. The methods are generally divided into two categories: image jigsaw reconstruction and contrastive learning. Image jigsaw reconstruction refers to dividing the image into regional blocks, shuffling them, and then training the model to reassemble the original image, which can effectively improve the accuracy and efficiency of the model. Contrastive learning refers to training the model to improve the accuracy of the model by maximizing the similarity between positive samples and minimizing the similarity between negative samples. This type of method can be introduced to train a more robust feature extractor for industrial models, thereby improving the versatility and generalization of the model in a strong noise environment.

The introduction of attention mechanisms (such as SE, CBAM) and multi-scale feature extraction networks (such as FPN), if used properly, can also further improve the recognition performance of CNN-based AMR device models in complex environments.

SE mechanism, the full name is Squeeze-and-Excitation. It is a channel attention mechanism that compresses the spatial information of each channel into a scalar through a global average pooling operation. The overall mechanism then generates weights through the fully connected layer and activation function of CNN to dynamically adjust the weights of each channel in the feature map. After the weight calculation, the mechanism can increase the weight ratio of the feature channel and suppress the response of the less important channel. Finally, the weight is multiplied by the original feature map to obtain a new feature map [12].

CBAM is a further attention mechanism that combines channel attention and spatial attention [13]. Based on the global pooling and function algorithm operation of the SE mechanism to obtain the channel attention weight, the spatial attention weight is generated through convolution operation, and the two weights are weighted and operated in the final feature map. In this way, the purpose of dynamically adjusting the attention weight is achieved, so that the model pays more attention to those important areas and improves the operation efficiency. These two attention mechanisms can significantly improve the feature extraction capability of the CNN model, and significantly help improve the accuracy and efficiency of the model.

Finally, the problem can be solved by introducing FPN training due to the different digital display methods of different instrument types. FPN, the full name of which is Feature Pyramid Network, has the core advantage of its multi-scale feature fusion capability, which can effectively handle targets of different sizes, shapes, and complexities [14]. By constructing a feature pyramid, the feature maps of the low-level (high resolution, low semantics) and high-level (low resolution, high semantics) are fused, which can detect large and small targets at the same time. In addition, in the complex environment of the actual production process, FPN can also capture richer contextual information through multi-scale feature fusion, enhancing the model's robustness to complex backgrounds. Properly used FPN will be an excellent choice for further improving the CNN model.

6 Conclusion

Existing technologies have indeed solved many difficult problems in industrial detection systems based on CNN models, including the joint use of attention mechanisms and feature pyramid networks, but there is still room for improvement in further improving the model. For example, there is still a lot of room for improvement in terms of lightweight and edge deployment of models. Deploying models on edge devices can improve the efficiency of industrial production, while lightweighting can reduce the model's calculation parameters and save more cost resources.As AI becomes more and more advanced, future research can focus on how to integrate transfer learning, reinforcement learning, and edge intelligent deployment into the current CNN model. As AI becomes more and more advanced, future research can focus on how to integrate transfer learning, reinforcement learning, and edge intelligent deployment into the current CNN model. While effectively improving the speed and completion of model training, it can also reduce the transmission delay of model devices to achieve a stronger real-time effect. This requires more efficient algorithms, and future researchers can try to work in this direction. In addition, the automatic updating and adaptation of the model is also a promising research direction. Since the environment in actual industrial production is constantly changing dynamically, the automatic adaptation of the model will be a very powerful function that can significantly reduce resource consumption and maintenance costs. The construction of large-scale industrial data sets is also an effective way to improve model training. The larger the scale, the easier it is to train and test the practicality and completeness of the model. It can also promote the standardization and large-scale application of CNN models in the industrial field. The final direction is to advance the green AI process and promote the sustainable development of industrial monitoring systems by developing low-power, high-efficiency CNN models.

References

1. LeCun, Y., Bottou, L., Bengio, Y., et al.: 'Gradient-based learning applied to document recognition', Proc. IEEE, 1998, 86, (11), pp. 2278–2324

2. Nabi, S. T., Singh, P., Kumar, M.: 'Gender classification from offline handwriting images in Urdu script: LeNet-5 and Alex-Net', J. King Saud Univ.-Comput. Inf. Sci., 2021, 33, (6), pp. 687–697
3. Shi, B., Bai, X., Yao, C.: 'An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition', IEEE Trans. Pattern Anal. Mach. Intell., 2016, 39, (11), pp. 2298–2304
4. Pulipalupula, M., Patlola, S., Nayaki, M., et al.: 'Object detection using You Only Look Once (YOLO) algorithm in convolutional neural network (CNN)', Int. J. Adv. Comput. Sci. Appl., 2021, 12, (3), pp. 1–8
5. Wang, Y., Li, X., Liu, Z., et al.: 'A deep learning approach for industrial meter digit recognition', J. Intell. Manuf., 2020, 31, (4), pp. 899–908
6. Ladarat, W., Naetiladdanon, S.: 'PEA automatic meter reading system: Progress and lessons learned', 2020 Int. Conf. Power Energy Innov. (ICPEI), 2020, pp. 1–4
7. Chen, G., Li, J., Chen, Q., et al.: 'High accuracy vision-laser post-welding defect detection and segmentation method based on convolutional neural networks', J. Manuf. Processes, 2021, 64, pp. 1513–1523
8. Han, S., Mao, H., Dally, W. J.: 'Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding', 4th Int. Conf. Learn. Represent. (ICLR), 2016
9. Jacob, B., Kligys, S., Chen, B., et al.: 'Quantization and training of neural networks for efficient integer-arithmetic-only inference', Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 2704–2713
10. Baltrušaitis, T., Ahuja, C., Morency, L.-P.: 'Multimodal machine learning: A survey and taxonomy', IEEE Trans. Pattern Anal. Mach. Intell., 2019, 41, (2), pp. 423–443
11. Jing, L., Tian, Y.: 'Self-supervised visual feature learning with deep neural networks: A survey', IEEE Trans. Pattern Anal. Mach. Intell., 2020, 43, (11), pp. 4037–4058
12. Hu, J., Shen, L., Albanie, S., et al.: 'Squeeze-and-excitation networks', Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7132–7141
13. Woo, S., Park, J., Lee, J. Y., et al.: 'CBAM: Convolutional block attention module', Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 3–19
14. Lin, T. Y., Dollár, P., Girshick, R., et al.: 'Feature pyramid networks for object detection', Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2117–2125