KEY TECHNOLOGIES OF SMART FACTORY MACHINE VISION BASED ON EFFICIENT DEEP NETWORK MODEL

Fangfang Zhang*

School of Fine Arts, Weifang College, Weifang, Shandong, 261061, China

20110738@wfu.edu

Kunfan Wang

College of Grammar, Northeast Forestry University, Harbin, Heilongjiang, 150040, China

Reception: 12/11/2022 **Acceptance:** 26/12/2022 **Publication:** 10/01/2023

Suggested citation:

Z., Fangfang and W., Kunfan (2023). **Key technologies of smart factory machine vision based on efficient deep network model**. *3C Empresa. Investigación y pensamiento crítico, 12(1)*, 15-35. https://doi.org/10.17993/3cemp.2023.120151.15-35

ABSTRACT

Most of the existing smart space machine vision technologies are oriented to specific applications, which are not conducive to knowledge sharing and reuse. Most smart devices require people to participate in control and cannot actively provide services for people. In response to the above problems, this research proposes a smart factory based on a deep network model, which is capable of data mining and analysis based on a huge database, enabling the factory to have self-learning capabilities. On this basis, tasks such as optimization of energy consumption and automatic judgment of production decisions are completed. Based on the deep network model, the accuracy of the model for image analysis is improved. Increasing the number of hidden layers will cause errors in the neural network and increase the amount of calculation. The appropriate number of neurons can be selected according to the characteristics of the model. When the IoU threshold is taken as 0.75, its performance is improved by 1.23% year-on-year. The residual structure composed of asymmetric multiple convolution kernels not only increases the number of feature extraction layers, but also allows the asymmetric image details to be better preserved. The recognition accuracy of the trained deep network model reaches 99.1%, which is much higher than other detection models, and its average recognition time is 0.175s. In the research of machine vision technology, the smart factory based on the deep network model not only maintains a high recognition accuracy rate, but also meets the real-time requirements of the system.

KEYWORDS

Smart factory; Deep network model; Machine vision; Image processing; Model optimization

https://doi.org/10.17993/3cemp.2023.120151.15-35

PAPER INDEX

https://doi.org/10.17993/3cemp.2023.120151.15-35

1. INTRODUCTION

With the development and innovation of computer science and communication technology, the degree of intelligence of computing is also getting higher and higher. Smart space is a typical distributed system, so the acquired situational data has distributed characteristics [1, 2]. Context-aware applications need to synthesize distributed context data and perform appropriate storage management for the collected context data. On this basis, the data is analyzed to provide usable information for the users of the smart space [3,4]. Smart factory is a typical application of smart space, which realizes the application of situational awareness in smart factory [5]. Its main task is to proactively eliminate the fault and prompt the cause of the fault when a fault occurs in the operation of the smart factory. Introduce the sources of contextual data and use ontology-based hierarchical modeling methods to model application scenarios [6]. Finally, the rule setting of situational reasoning is explained, and the reasoning result is analyzed. A smart factory is defined as a factory that can intelligently perceive situational information and use situational awareness technology to help people and machines perform tasks. Industrial robots have the characteristics of high work efficiency, high work quality, and high repeat positioning accuracy. It is a development trend for industrial robots to replace manual labor. However, the robot does not have the developed human brain and visual system, and cannot adjust the movement trajectory according to the specific environment. These problems can be well solved by robot vision, and machine vision technology is the key to the transformation from traditional production to intelligent production [7].

The information of the entire production process is collected by sensors and smart devices in the smart factory, and transmitted to the upper computer for development and utilization [8]. Zhang et al. [9] used the graph-optimized SLAM algorithm to construct a two-dimensional grid map of the physical simulation environment of the printing smart factory, and designed a combined navigation scheme combining the SLAM algorithm and the dynamic window method (DWA). The simulation results verify the accessibility of global path planning and the feasibility of local path obstacle avoidance. , Jerman et al. [10] proposed a case study on the impact of Industry 4.0 on the business model of smart factories for the key elements of smart factories. Its main objective will be that the roles of employees working in production will be primarily to express their creativity, to make emergency interventions and to perform process custody. The key factors influencing smart factory business models are top management and leadership orientation, employee motivation, collective intelligence, creativity and innovation. The study provides useful guidance for the strategic management of innovative companies in the early stages of the decision-making process. SHANTHIKUMAR et al. [11] have a limited level of operational control in scheduling for each factory, assigning different types of products to different factories. In a dedicated factory, only similar products are produced in the same factory. A multi-server, multi-job-class queuing model is established, and the model is used to form a centralized factory to handle similar tasks. In intelligent manufacturing, machine vision is the "eye" of the machine and the "brain" of vision. The machine vision module is composed of cameras, light sources, brackets, computers, etc., and can meet the needs of rich vision applications such as visual positioning, measurement, detection and identification. Machine vision is used to collect images, which can carry out continuous acquisition and external control acquisition. [12] expounded the role and importance of machine vision systems in industrial applications. The intended function of a vision machine is to exploit and impose the environmental constraints of the scene, capture images, analyze the captured images, and identify certain objects and features in each image. and initiate follow-up actions to accept or reject the corresponding object. Application areas include automated visual inspection (AVI), process control, part identification, and play an important role in the guidance and control of robots. Looking ahead to the development of manufacturing, it is possible to improve reliability, improve product quality, and provide technical support for new production processes. Chen et al. [13] studied the hardware and software requirements and recent developments of machine vision systems, focusing on the analysis of multispectral and hyperspectral imaging in modern food inspection. Future trends in the application of machine vision technology are discussed. Robie et al. [14] investigated the development of machine vision techniques for the automated quantitative analysis of social behavior, greatly improving the scale and resolution at which we analyze interactions between members of the same species. Several components of machine vision-based analysis are discussed: high-quality video recording methods for automated analysis, video-based tracking algorithms for estimating the position of interacting animals, and machine learning methods for identifying interaction patterns. The applicability of these methods is very general, reviewing the successful application to biological problems in several model systems with very different types of social behavior. [15] studied the Lambert diffuse model for computational vision applications, and the Lambertian model can be shown to be a very inaccurate approximation of the diffuse component. The brightness of Lambertian surfaces is independent of the viewing direction, whereas the brightness of rough diffuse surfaces increases as the observer approaches the source direction. The model takes into account complex geometric and radiation phenomena such as masking, shadowing and inter-reflection between surface points. The resulting reflectance measurements are in strong agreement with the model-predicted reflectance values. Davies et al. [16] studied the application progress of machine vision in the field of food and agriculture since 2000. It involves applying different wavelengths of radiation to the material, not only looking at the surface but also the internal structure. With its powerful feature learning capabilities, deep networks have been widely used in machine vision such as face recognition [17], semantic segmentation [18], and human pose detection [19]. Among them, convolutional neural network has become one of the most successful image analysis models and is widely used in the field of computer vision. In addition to training deep neural networks in a single feed-forward manner, recurrent neural networks, a deep model that captures temporal information, are more suitable for prediction of sequence data of arbitrary length. Agarwal et al. [20] proposed a new deep neural network-based approach that relies on coarse-grained sentence modeling. Use convolutional neural network and recurrent neural network (RNN) models combined with specific fine-grained word-level similarity matching models. The proposed deep paraphrase-based method achieves

https://doi.org/10.17993/3cemp.2023.120151.15-35

good results in both types of text and is thus more robust and general than existing methods. Tai et al. [21] made great breakthroughs in the field of computer vision by using deep learning for the obstacle avoidance problem of mobile robots, especially in recognition and cognitive tasks. Taking the original depth image as input, generating control commands as network output and realizing model-free obstacle avoidance behavior. [22] proposed an end-to-end trainable deep network structure based on a Gaussian Conditional Random Field (GCRF) model for image denoising. The proposed deep network explicitly models the input noise variance and is able to handle a range of noise levels. Experiments on the Berkeley segmentation and PASCALVOC datasets show that the proposed method produces the same results as the state-of-the-art without training a separate network for each individual noise level. Bai et al. [23] studied a novel comprehensive solution for the compression and acceleration of visual question answering systems. The algorithm uses long and short-term memory to compress convolutional neural networks to improve processing speed. To further compress this parameter, a tensor shrinking layer is used to compress the feature flow between layers.

To sum up, in the traditional fault diagnosis, the production efficiency of the factory is reduced due to the poor communication of information. It is difficult for the management of the factory to get the fault information of the factory and make a response at the first time. Smart factories can provide an intelligent fault detection service that supports dynamic collection of abnormal event information, real-time transmission, and abnormal response-level monitoring. The smart factory machine vision technology built with the help of efficient deep network model provides a guarantee for troubleshooting events and improving the operating efficiency of the factory. When equipment in the smart factory fails, the factory control center can quickly obtain abnormal information. At the same time, relevant maintenance personnel can obtain real-time warnings and fault-related information through mobile devices such as mobile phones.

2. DEEP NETWORK MODEL

Deep learning, which attempts to capture high-level abstract features through multiple nonlinear transformations and hierarchical representation structures, has become a hot research topic. According to the input of the network and the function of the network, it is divided into four categories: convolutional neural network, recurrent neural network, graph convolutional network and binary network. The input of the first three networks is static image or vector, sequence and graph data respectively. The binary network is based on the existing deep model, and further quantifies the network weights and activations.

2.1. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are currently one of the most successful image analysis models and are widely used in various computer vision tasks [24, 25]. The design of its model was first inspired by biologists' research on the biological functions of human brain nerve cells. Convolutional neural network has made great progress in the field of deep learning and is the mainstream in the field of image related. Different from the traditional neural network, its original image can be directly input into the convolutional neural network for calculation. Simplified image preprocessing is a feedforward neural network. The main structure of the convolutional neural network layer is: input layer, convolutional layer, pooling layer and fully connected layer.

In practical applications, the experience of human-computer interaction is often at the millisecond level. Only by solving the efficiency problem of convolutional neural networks can convolutional neural networks be more widely used in various mobile devices. The usual method is to perform model compression, that is, to perform compression on the already trained model, so that the network carries fewer network parameters. The mathematical formula for the convolutional layer is as follows:

$$X_j^L = f\left(\sum_{i=1}^{M} W_{ij}^L * X_i^{L-1} + W_b\right) \tag{1}$$

where X_j^L is the output feature of the convolutional layer, f(•) is the activation function, W_{ij}^L is the convolution kernel weight matrix, X_i^{L-1} is the input image feature, and W_b is the bias value.

2.2. RECURRENT NEURAL NETWORK

There are many sequential data in real-world problems, such as text, speech, and video, which have obvious temporal correlations [26]. The output and input of traditional neural networks (such as CNN) are independent of each other, have no memory ability, and cannot use the correlation in time series. For this reason, as early as 1986, the Jordan network proposed by Bilski et al. directly feeds the final output of the entire network back to the input layer of the network through the delay module, so that all layers of the entire network are recursive. Recurrent Neural Network (RNN) is a network designed for sequence signal processing and is suitable for predicting sequence data of arbitrary length. The state of the hidden layer inside the RNN is cyclic, and the state of the hidden layer depends not only on the current input but also on the state of the previous hidden layer.

$$\begin{aligned} \boldsymbol{o}_t &= g\left(\mathbf{V}\boldsymbol{s}_t\right) \\ \boldsymbol{s}_t &= f\left(\mathbf{U}\boldsymbol{x}_t + \mathbf{W}\boldsymbol{s}_{t-1}\right) \end{aligned} \tag{2}$$

where V is the output weight matrix, U is the weight matrix of the input x_t, W is the weight matrix of the input s_{t-1} at the previous moment, and f(·), g(·) are the activation functions.

Each memory unit module uses three gate structures: input gate, output gate and forget gate to control the hidden state at different times, and to model the temporal correlation in the sequence. Great success in automatic speech recognition, music creation, grammar learning, and even image tagging. At present, LSTM has been

https://doi.org/10.17993/3cemp.2023.120151.15-35

widely used in classification-based tasks, and combined with convolutional neural networks, images can be automatically annotated.

2.3. DEEP GRAPH CONVOLUTIONAL NEURAL NETWORKS

Existing convolutional neural networks utilize a large amount of trainable data to efficiently accelerate computing resources (GPU).

The unique network structure and efficient training mechanism can effectively extract data feature representations from Euclidean space data such as images, texts and videos for various tasks [27, 28]. Graph convolutional networks well-designed for different tasks continue to emerge, including attention graph neural networks, graph autoencoders, graph generative networks, and graph spatiotemporal networks. The attention mechanism has proved useful in multiple tasks, and the essence of the attention map neural network is to assign weights to different neighbors when aggregating feature information. The goal is to embed node feature representations into a low-dimensional vector space through neural networks, and utilize decoders to reconstruct the nearest neighbor statistics of nodes. Use a graph neural network to generate probabilistic dependencies between nodes and edges of a graph. The input to each node changes over time, and the graph spatiotemporal network captures both the temporal and spatial dependencies of the spatiotemporal graph. The spectral domain-based graph convolution and the spatial domain-based graph convolution introduce convolution kernels from the perspective of graph signal processing to define graph convolution. The feature extraction of nodes is realized by the feature decomposition of the Laplacian matrix of the graph, and the information of adjacent nodes in the graph is aggregated.

Spatial features in graph data have the following characteristics:

1) Node characteristics: each node has its own characteristics;

2) Structural features: Each node in the graph data has structural features, that is, there is a certain relationship between nodes and nodes. Graph data needs to consider both node information and structural information. Graph convolutional neural networks can automatically learn not only node features, but also the association information between nodes.

Suppose there is a set of graph data, which has N nodes (nodes), each node has its own characteristics. Let the features of these nodes form an N×D-dimensional matrix X, and then the relationship between each node will also form an N×N-dimensional matrix A. The way it propagates from layer to layer is:

$$H^{(l+1)} = \sigma\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(l)} W^{(l)}\right). \tag{3}$$

Among them, A wave=A+I, I is the unit matrix, D wave is the degree matrix of A wave, H is the feature of each layer, and for the input layer, H is X$_\sigma$ is the nonlinear activation function.

https://doi.org/10.17993/3cemp.2023.120151.15-35

2.4. MODEL TRAINING AND EVALUATION METRICS

When testing the performance of deep learning models, the smaller the memory occupied by the model, the faster the detection speed and the higher the accuracy, the stronger the performance. The evaluation index is a quantitative index for the performance of the model. One evaluation index can only reflect part of the performance of the model. If the selected evaluation index is unreasonable, wrong conclusions may be drawn. Therefore, different evaluation indexes should be selected for specific data and models. For the trained model, a model evaluation coefficient is required to measure the accuracy of the model. The mean absolute error MAE and the coefficient of determination are widely used in the evaluation of model accuracy, and the expressions are:

$$MAE = \frac{1}{m}\sum_{j=1}^{m}\left|y_j^{'} - y_j\right| \quad (4)$$

$$R^2 = 1 - \frac{\sum_{j=1}^{m}\left(y_j - y_j^{'}\right)^2}{\sum_{j=1}^{m}\left(y_j - \overline{y}\right)^2} \quad (5)$$

In the formula, $y_j^{'}$ and y_j represent the predicted value and the true value of the j th sample respectively; $\overline{y}$ represent the average value of the true value of the sample; m represent the number of samples. The smaller the value of the mean absolute error MAE, the smaller the error of the model and the better the effect: the coefficient of determination R^2 is a value from 0 to 1, which indicates the goodness of fit, and the closer its value is to 1, the better the model fitting effect. the better.

In target detection, IOU is used to represent the coincidence area of the predicted detection frame and the real frame. The larger the value, the more accurate the algorithm positioning is. The mathematical formula is as follows:

$$\mathrm{IOU} = \frac{A \cap B}{A \cup B} \quad (6)$$

3. SMART FACTORY VISION TECHNOLOGY APPLICATION

A smart factory is a complex production system that requires the high precision, immediacy and reproducibility of machine vision perception control technology [29-30]. The systematic scheme of machine vision perception control in smart factories relies on the actual production needs to design the intelligent vision imaging system and the automatic image acquisition part in sequence. Images of objects are collected. The second-step processing is performed on the obtained clear and high-quality images. The image content is judged and classified and screened according to the pre-set information base to complete the identification, inspection, measurement and calculation of the detection object. The information obtained in the processing of the image allows for appropriate optimization control.

3.1. MACHINE VISION DEVELOPMENT PRINCIPLES

The machine vision system performs the detection function of the measured object based on the computer, and the functional composition of the machine vision system is divided into three parts[31-32]. Collect images of the object under test, process and analyze the images, and output or display the test results, respectively. The imaging system of machine vision is mainly composed of a light source, a lens and a camera, which is the basis for the system to perform the detection function. Therefore, the final imaging quality of the measured object of the machine vision imaging system has a direct impact on the detection result of the system. Due to the influence of factors such as light intensity, surrounding environmental factors and the image acquisition equipment itself, there must be many interference signals and noises in the images captured by the book sorting system. These disturbances and noises make the image mainly show the imbalance of light and dark contrast, and the noise drowns out some important information. In order to improve the signal-to-noise ratio of the image, the image enhancement method is used to process the machine vision image. Due to the high real-time nature of spatial filtering, the spatial filtering method is used to realize the filtering processing of the collected images. The principle of spatial filtering is shown in the following formula.

$$G(x,y)=\sum_{a}\sum_{b}F(x,y)K(x-a,y-b) \tag{7}$$

Among them, F(x, y) is the original gray value; G(x, y) is the gray value after filtering; K is a filter kernel function used in image processing.

Morphological operations can remove image noise and also highlight the localization of regions of minimum and maximum values in the image. Image erosion is to find the local minimum value of the image, the main purpose is to make the area of minimum value cover the area of other maximum value. The expansion of the image is to find the local maximum value of the image, the main purpose is to make the area of maximum value cover the area of minimum value. Table 1 is a comparison of the standard deviation of pixel values of different algorithms.

Table 1 Comparison of standard deviations of pixel values in different algorithms

Filtering algorithm	a	b	c	d
Mean filter	49.83	49.09	48.03	48.83
Gaussian filter	49.83	47.56	47.83	48.65
Median filter	46.48	46.98	45.46	46.78
Bilateral filtering	44.41	44.58	45.48	44.32

The specific implementation steps are as follows:

(1) Image edge detection: First, edge detection is performed on the tilt-corrected book image. Complete detection of image edges facilitates analysis of object detection, localization, and recognition.

(2) Image binarization: The problem in the previous step can be solved by the operation of image binarization. All pixels in the image are judged and screened to complete the retention of the image part.

(3) Morphological processing: According to the obtained binarized image, the barcode area becomes the largest rectangular area through erosion and expansion processing. Other areas in the image background are basically removed, and the largest area of the image can be selected.

3.2. RESEARCH ON MACHINE VISION BASED ON DEEP NETWORK

With the continuous development of deep learning, especially the emergence of deep neural networks based on ensemble learning. Complex network models emerge in an endless stream, and although the prediction accuracy of the models continues to improve, it is difficult to apply them to real-world scenarios.

The image distortion correction of the machine vision system based on the deep network proposed in this paper is divided into two parts: the solution of the image nonlinear distortion model and the image correction. Firstly, a large number of images with different degrees of distortion of the machine vision system are simulated based on MATLAB, and then the structure of the deep network is designed, and the obtained distorted images are used as the data set to train the deep network. When correcting the distorted images of the system, the trained deep network model can be applied to machine vision systems with different degrees of image distortion. Solve the distortion parameters of its image nonlinear distortion model.

In a deep network, the input feature image is convolved, pooled, and the output feature image of each hidden layer is obtained. Each hidden neuron corresponds to a region of its input feature map, then this region is the receptive field of the corresponding neuron. The local receptive field of the feature image is used as the input of the lowest layer of the deep network structure, and the information of the local receptive field is transmitted to different layers in turn. Each layer obtains the relevant features of the data through the convolution operation of the convolution kernel, and integrates the global information at the high level. This mode can reduce the number of network parameters and the calculation speed is faster. A lens in a machine vision system consists of a set of lenses, when light enters the lens parallel to the main optical axis. The point where all rays converge is the focal point, and the distance between the focal point and the center of the lens is the focal length.

The conversion relationship between the image coordinate system and the pixel coordinate system is:

https://doi.org/10.17993/3cemp.2023.120151.15-35

$$\begin{cases} u = \dfrac{x}{dx} + u_0 \\ v = \dfrac{y}{dy} + v_0 \end{cases} \quad (8)$$

The conversion of the image coordinate system and the camera coordinate system is:

$$\begin{cases} x = \dfrac{fXc}{Zc} \\ y = \dfrac{fYc}{Zc} \end{cases} \quad (9)$$

The distance from the origin of the camera coordinate system to the image plane. The above relationship is expressed as a matrix in homogeneous coordinates as:

$$Zc\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} Xc \\ Yc \\ Zc \\ 1 \end{pmatrix} \quad (10)$$

The transformation of the camera coordinate system and the world coordinate system is:

$$\begin{pmatrix} Xc \\ Yc \\ Zc \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & t \\ 0^T & 1 \end{pmatrix}\begin{pmatrix} Xw \\ YW \\ Zw \\ 1 \end{pmatrix} \quad (11)$$

The generation of the data set is mainly by changing the distortion parameters in the nonlinear model to randomly generate images with different degrees of distortion. Therefore, using the regression function of the deep network, the distorted images in the dataset do not need to be classified.

Based on Matlab, the program to generate the data set was programmed, and the LABEL function was used to record the distortion parameter matrix corresponding to the distorted image. Table 2 shows the matching time and matching degree under different parameters.

Table 2 Matching time and matching degree under different parameters

MinScore	Match time (ms)	Can match
0.5	24	Yes
0.6	19	Yes
0.7	14	Yes
0.88	10	Yes
0.96	/	No

3.3. REALIZATION OF VISION TECHNOLOGY IN SMART FACTORY

The deep learning network structure designed in this paper consists of an input layer, four convolutional layers, four pooling layers, two fully connected layers and an output layer, and each convolutional layer applies a relu activation function. After this series of layers, the output of the last pooling layer is flattened and fed into the fully connected layer. The final output result is each distortion parameter of the image nonlinear distortion model, which is the complete training model. Use the Flatten layer to convert the feature image into one-dimensional features and then input them into the fully connected layer. Deep learning is a special kind of neural network that can remember spatial information, so flattening the feature image input to the fully connected layer does not affect the positional relationship between the features.

Smart factory is a complex system engineering, using the results of visual inspection and identification as well as the results of positioning and attitude determination as a reference. Smart factories can control robots to accurately complete very complex tasks such as positioning, picking, and classification. The visual control rate is defined by visual error to confirm the control amount. Control the robot to move, and then achieve the specified work tasks. The manipulator is used in the smart factory to serve the products. During the work of the manipulator, the manipulator may overheat (overHeat), run out of battery (batteryDie) and stop working (breakDown). There are three levels of failures in the factory: low, medium, and high, corresponding to the three conditions described above.

3.4. VISION APPLICATION OF SMART FACTORY BASED ON DEEP NETWORK

Machine vision systems are widely used in the field of industrial product quality inspection, especially in workpiece size measurement and appearance defect detection. By systematically measuring the size of large-scale workpieces and inspecting workpieces for defects, the system distinguishes between good and non-conforming products. The image distortion correction algorithm for machine vision system based on deep network proposed in this paper is suitable for installed machine vision systems with different degrees of image distortion. In this paper, the

https://doi.org/10.17993/3cemp.2023.120151.15-35

standard template is used as the measured object of the system in the scene demonstration of the machine vision system, and the obtained distorted template image is input into the trained deep network model. The distortion parameters of the system are obtained, and then the system is used to capture an image of the workpiece to be measured, and the distortion image of the workpiece is corrected based on MATLAB. The PTZ module is a support device specially used to install and fix the camera, and it can also expand the scanning range of the camera. Automatically adjust the camera's level, up and down angles on the PTZ, and only need to adjust the mechanism to make the camera achieve the best working position and posture. The system is used to test and identify 120 different images containing components, and these 120 images have different tilt angles. During the test, the book images are divided into 4 categories according to the different tilt angles, with an average of 30 book images for each category. The collected images of books with different inclination angles are one in the range of 0~90°, 90°~180°, 180°~270°, and 270°~360°, respectively, from left to right. Table 3 is the image recognition system test.

Table 3 Image recognition system testing

Angle	Number of tests	Unrecognized	Recognition rate	Processing time	Recognition time
0~90	30	0	100%	1.43s	0.57s
90~180	30	1	96.67%	1.31s	0.62s
180~270	30	0	100%	1.60s	0.44s
270~360	30	0	100%	1.53s	0.59s

4. RESULTS AND DISCUSSION

4.1. THE EFFECT OF THE NUMBER OF NEURONS ON THE PERFORMANCE OF DEEP NETWORKS

The sample images in the dataset need to be preprocessed, and the head poses corresponding to the images should be marked. There are some parameters in the network that need to be set manually. All network parameters are set as follows: the number of iterations is 80, the total number of batch training samples is 1000, and the learning rate η is 0.0002. Figure 1 shows the training performance of the neural network under R2 to obtain a high coefficient of determination R2. When the number of neurons in the hidden layer increases from 5 to 10, MAE gradually decreases and R2 increases first, indicating that increasing the number of neurons is conducive to improving the performance of the neural network; when the number of neurons increases from 10 to 15 , MAE gradually increases and R2 gradually decreases, indicating that excessively increasing the number of neurons will cause overfitting of

https://doi.org/10.17993/3cemp.2023.120151.15-35

the neural network, thereby reducing the performance of the neural network. As shown in Figure 1, the neural network has the best performance when the number of neurons is 10. At this time, R2 is 0.981 and MAE is 25. Increasing the number of hidden layers will cause errors in the neural network and increase the amount of computation, which is not conducive to model training. Therefore, it is necessary to select the appropriate number of neurons according to the characteristics of the model.

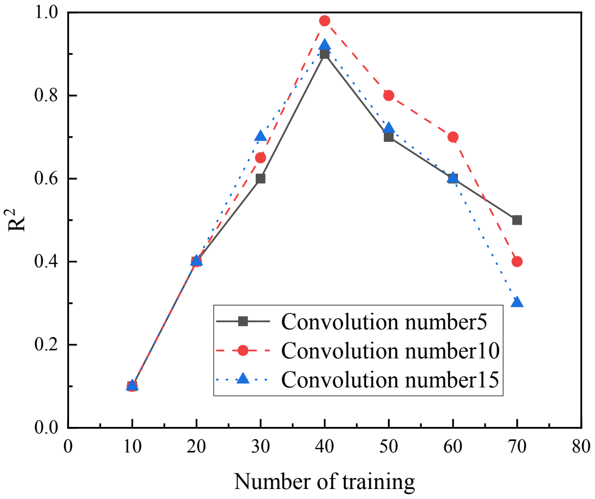

Figure 1 R^2 trained with different numbers of neurons

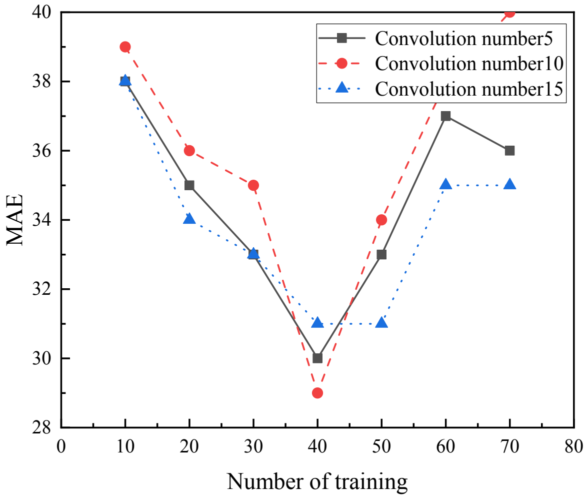

Figure 2 MAE under training with different numbers of neurons

4.2. DEEP NETWORK MODEL OPTIMIZATION FOR SMART FACTORIES

According to the application scenario of the factory, the data in the factory is dynamically obtained in real time, the amount of data generated is large and most of the data is in the normal range. For the factory inspection dataset, a total of three inspection models are trained. The model trained by the original convolutional network is used as the base model. First replace the convolution default regression loss function MSE loss with the GIoU loss function. Then use the GIoU loss as the regression loss, and the last group uses the

https://doi.org/10.17993/3cemp.2023.120151.15-35

The proposed deep network algorithm completes the bounding box regression task. When testing the performance of the model, the mAP value when the IoU thresholds were 0.5 and 0.75 were mainly calculated as the model evaluation index. The model trained using the MSE loss function is used as the benchmark, although using the GIoU function under the AP75 standard improves the model performance by 0.15%. When the IoU value is less than 0.6, the detection model based on AIoU loss function is slightly lower than the detection accuracy based on MSE loss function. However, when the value is greater than or equal to 0.6, its model performs better, and from the overall trend, the model test results using the AIoU loss function are significantly better than the other three groups. When the IoU threshold is taken as 0.75, its performance improvement is the highest, which is improved by 1.23%. The residual structure composed of asymmetric multiple convolution kernels not only increases the number of feature extraction layers, but also allows the asymmetric image details to be better preserved. And it can normalize the image features, thereby ensuring the stability of the network learning process.

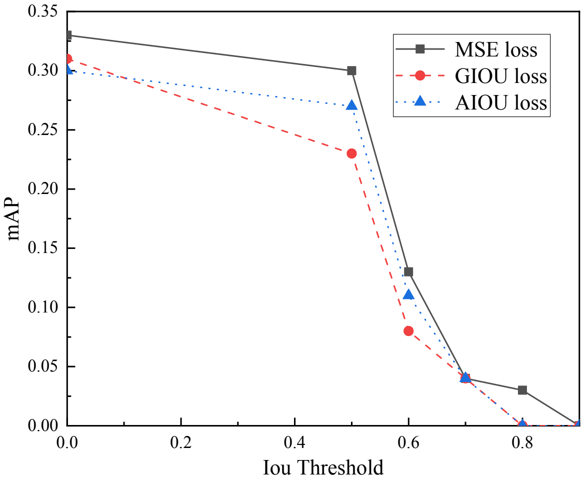

Figure 3 mAP of different detection models

4.3. DEVELOPMENT OF KEY VISUAL TECHNOLOGIES FOR SMART FACTORIES

The deep network is an algorithm designed for image processing. The number of input samples is in picture format, and the sample data provided in this paper is the node data collected by the system camera. The feature data (numerical type) extracted by the operation, so in order to meet the requirements of the convolutional neural network framework, there will be no lack of dimension when the data enters the operation of each convolutional layer and pooling layer. By setting the learning rates to be 0.1, 0.01, and 001 for training, the loss of all 8 learning rates decreases rapidly during the training process. But relative to the learning rate of 0.01 and 0.001, when the learning rate is 0.001. The training loss decreases the fastest, that is, the learning speed of the model is the fastest, so the training model in this experiment chooses a learning rate equal to 0.001. Continuously collect 300s image node data to construct

https://doi.org/10.17993/3cemp.2023.120151.15-35

600 detection samples. Input the trained deep network model and compare it with the detection results of thresholding, support vector machine, strong separator, decision tree, k-nearest neighbor model. Its recognition accuracy reaches 99.1%, which is much higher than other detection models. The average recognition time is 0.175s, which is much faster than several other machine learning detection algorithms. The detection model not only maintains a high recognition accuracy rate, but also meets the real-time requirements of the system.

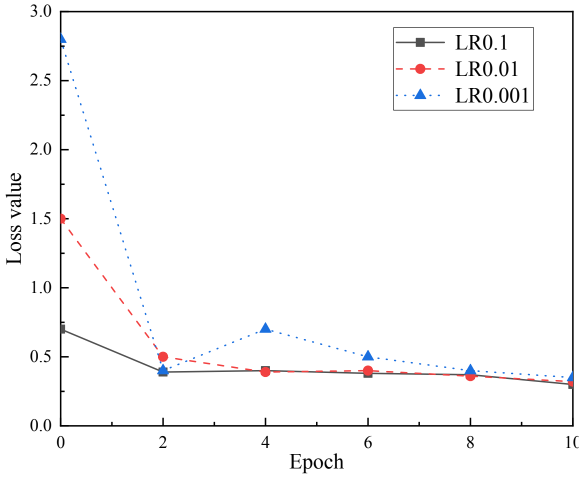

Figure 4 Training loss at different learning rates

The quality of the saliency map can be measured by the target localization evaluation. By extracting the maximum point from the saliency map to observe whether the point falls within the target bounding box, the extraction of the maximum point is extended to the entire saliency map. Determine how much of the saliency map can fall within the target bounding box. Table 4 shows that the method in this paper is significantly better than the comparison method in the target localization performance. The numerical results show that more than 60% of the pixels in the saliency map obtained by the method in this paper fall within the target detection frame. It shows that the target noise of the saliency map of the method in this paper is small, which is consistent with the previous subjective visualization analysis results. First, the image and the target category bounding box are binarized, where the inner region is assigned a value of 1, and the outer region is assigned a value of 0. It is then multiplied point-by-point with the generated saliency map and summed to get the energy in the target bounding box. The larger the Location value, the better the localization performance of the saliency map.

Table 4 Target positioning evaluation comparison test results

Method	Locatioin
Grad-CAM	0.45
Grad-CAM++	0.47
Score-CAM	0.54
XGrad-CAM	0.57

https://doi.org/10.17993/3cemp.2023.120151.15-35

Ablation-CAM	0.52

5. CONCLUSION

In the traditional fault diagnosis, the production efficiency of the factory is reduced due to the poor communication of information. It is difficult for the management of the factory to get the fault information of the factory and make a response at the first time. Smart factories can provide an intelligent fault detection service that supports dynamic collection of abnormal event information, real-time transmission, and abnormal response-level monitoring. This research proposes a deep learning algorithm, and this research proposes a smart factory based on a deep network model, which is capable of data mining and analysis based on a huge database, enabling the factory to have self-learning capabilities. Based on the deep network model, the accuracy of the model for image analysis is improved. In the research of machine vision technology, the smart factory based on the deep network model not only maintains a high recognition accuracy rate, but also meets the real-time requirements of the system. It has great development prospects and determines that the deep network model has a significant impact on smart factories. and came to the following conclusions:

(1) When the number of neurons in the hidden layer is 10, the increase of R2 indicates that increasing the number of neurons is beneficial to improve the performance of the neural network. When the number of neurons is 15, the gradual decrease of R2 indicates that excessively increasing the number of neurons will cause overfitting of the neural network, thereby reducing the performance of the neural network. Therefore, it is necessary to select the appropriate number of neurons according to the characteristics of the model.

(2) When the IoU value is less than 0.6, the detection model based on AIoU loss function is slightly lower than the detection accuracy based on MSE loss function. When the value is greater than or equal to 0.6, its model performs better, and from the overall trend, the model test results using the AIoU loss function are significantly better than the other three groups. When the IoU threshold is taken as 0.75, its performance improvement is the highest, which is improved by 1.23%.

(3) The deep network is an algorithm designed for image processing. The number of input samples is in picture format, and the sample data provided in this paper is the node data collected by the system camera. Continuously collect 300s image node data to construct 600 detection samples. Its recognition accuracy rate of 99.1% is much higher than other detection models, and the average recognition time is only 0.175s.

6. DATA AVAILABILITY

The data used to support the findings of this study are available from the corresponding author upon request.

7. CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

8. ACKNOWLEDGMENTS

The writing process was strongly supported by other teachers.

REFERENCES

(1) Selvarajah K, Zhao R, Speirs N. **Building Smart Space Applications with Pervasive Computing in Embedded Systems (PECES) Middleware[J]**. *GSTF Journal on Computing (JoC)*, 2014, 1(4):12-15.1.

(2) Shi Y, Xie W, Xu G. **Smart remote classroom: Creating a revolutionary real-time interactive distance learning system[M]**. *Advances in Web-Based Learning. Springer Berlin Heidelberg*, 2002: 130~141.

(3) Berhe G, Brunie L, Pierson J M. **Content Adaptation in distributed multimedia system[J]**. *Journal of Digital Information Management*, 2005, 3(2): 95~100.

(4) DengL,YuD. **Deep learning: methods and applications[J]**. *Foundations and trends® in signal processing*, 2014, 7(3–4):197-387.

(5) J Chen H, Perich F, Finin T, et al. **Soupa: Standard ontology for ubiquitous and pervasive applications[C]**. in: *Mobile and Ubiquitous Systems: Networking and Services*, 2004: 258~267

(6) Wang, Georgette and Yi-Ning Katherine Chen. **Collectivism, relations and Chinese communication**. (2010):1-9.

(7) FangT, FaureGO. **Chinese communication characteristics: A Yin Yang perspective[J]**. *International Journal of Intercultural Relations*, 2011, 35(3):320-333.

(8) LeCunY, BengioY, HintonG. **Deep learning [J]**. *nature*,2015,521(7553):436-444.

(9) Zhang Y F, Zhang W, Liu S H, et al. **Research on AGV Navigation Simulation in Printing Wisdom Factory[C].** *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE*, 2021, 5: 2312-2316..

(10) Jerman A, Bertoncelj A, Erenda I. **The influence of critical factors on business model at a smart factory: A case study[J]**. *Business Systems Research: International journal of the Society for Advancing Innovation and Research in Economy*, 2019, 10(1): 42-52.

(11) SHANTHIKUMAR J G, Xu S H. **Strong Asymptotic Optimality Of Focused Factory[J]**. 1999.

(12) Golnabi H, Asadpour A. **Design and application of industrial machine vision systems[J]**. *Robotics and Computer-Integrated Manufacturing*, 2007, 23(6): 630-637.

(13) Chen Y R, Chao K, Kim M S. **Machine vision technology for agricultural applications[J]**. *Computers and electronics in Agriculture*, 2002, 36(2-3): 173-191.

(14) Robie A A, Seagraves K M, Egnor S E R, et al. **Machine vision methods for analyzing social interactions[J]**. *Journal of Experimental Biology*, 2017, 220(1): 25-34.

(15) Oren, Michael, and Shree K. Nayar. **Generalization of the Lambertian model and implications for machine vision**. *International Journal of Computer Vision* 14.3 (1995): 227-251.

(16) Davies E R. **The application of machine vision to food and agriculture: a review[J]**. *The Imaging Science Journal*, 2009, 57(4): 197-217.

(17) HaoX, ZhangG, MaS. **Deep learning[J]**. *International Journal of Semantic Computing*, 2016,10(03):417-439.

(18) KimP. Matlab deep learning[J]. **With machine learning, neural networks and artificial intelligence, 2017, 130 (21 AFRAMA, JANABI-SHARIFIF. Review of modeling methods for HVAC systems[J]**. *Applied Thermal Engineering*, 2014,67(1-2):507-519.

(19) GulliA, PalS. **Deep learning with Keras[M].** *Packt Publishing Ltd*, 2017.

(20) Agarwal B, Ramampiaro H, Langseth H, et al. **A deep network model for paraphrase detection in short text messages[J]**. *Information Processing & Management,* 2018, 54(6): 922-937.

(21) Tai L, Li S, Liu M. **A deep-network solution towards model-less obstacle avoidance[C].** *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE*, 2016: 2759-2764.

(22) Vemulapalli R, Tuzel O, Liu M Y. **Deep gaussian conditional random field network: A model-based deep network for discriminative denoising[C].** *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016: 4801-4809.

(23) Bai Z, Li Y, Woźniak M, et al. **DecomVQANet: Decomposing visual question answering deep network via tensor decomposition and regression[J]**. *Pattern Recognition*, 2021, 110: 107538.

(24) Schmidhuber J. **Deep learning in neural networks: An over view[J].** *Neural networks*, 2015,61:85-117.

(25) WangH, RajB. **On the origin of deep learning[J]**. *arXiv preprint arXiv*: 1702.07800,2017.

(26) RenM, ZengW, YangB, etal. **Learning to reweight examples for robust deep learning[C]**. *International conference on machine learning.* PMLR, 2018:4334-4343.

(27) Lamport L. **On interprocess communication[J]**. *Distributed computing,* 1986, 1(2): 86-101.

(28) Luhmann N. **What is communication?[J]**. *Communication theory*, 1992, 2(3): 251-259.

(29) Chen G M. An **introduction to key concepts in understanding the Chinese: Harmony as the foundation of Chinese communication[J]**. 2011.

https://doi.org/10.17993/3cemp.2023.120151.15-35

(30) Smith T W, Colby S A. **Teaching for deep learning[J]**. *The clearing house: A journal of educational strategies, issues and ideas*, 2007, 80(5): 205-210.

(31) Horani M. O., Najeeb, M., y Saeed, A. (2021). **Model electric car with wireless charging using solar energy**. *3C Tecnología. Glosas de innovación aplicadas a la pyme*, 10(4), 89-101. https://doi.org/10.17993/3ctecno/2021.v10n4e40.89-101

(32) Chen Shuang & Ren Yuanjin.(2021). **Small amplitude periodic solution of Hopf Bifurcation Theorem for fractional differential equations of balance point in group competitive martial arts**. *Applied Mathematics and Nonlinear Sciences* (1). https://doi.org/10.2478/AMNS.2021.2.00152.