

# Application of image recognition in workpiece classification

Advances in Mechanical Engineering  
2021, Vol. 13(6) 1–9  
© The Author(s) 2021  
DOI: 10.1177/16878140211026082  
journals.sagepub.com/home/ade  


Hsin-Yi Chien , Yu-Chen Wang and Guan-Chen Chen 

## Abstract

With the rapid development of information technology and widespread use of the Internet of Things, machine intelligence will undoubtedly emerge as a leading research topic in the future. The main purpose of the present research is to incorporate an image recognition system into a robotic arm motion to achieve automatic classification. First, we upload captured images to a PC for classification process and use chess patterns to conduct a sampling test. Next, when the system identifies these patterns as proper chess patterns, the robotic arm grabs the objects and moves them to designated locations. The project is divided into two main sections: image recognition and robotic arm motion. In the image recognition section, we use Keras and the Tensorflow open source learning machine to build a convolutional neural network model. Then, we use a learning model network that is a considerably more compact variant of the VGGNet network in the image recognition system. With this model, we achieve a recognition accuracy of 95%. In the robotic arm section, we use a five-axis robotic arm and an Arduino Uno board as the controller. We design the Denavit–Hartenberg parameters of the arm and calculate the direct (inverse) kinematics parameters to plan its trajectory. Thereafter, we use MATLAB software to simulate prototype processes, such as grabbing, moving, and placing. Finally, we import the program into the controller so that the robotic arm can execute classification based on the chess pattern.

## Keywords

CNN, VGGNet, image recognition, robotic arm, Denavit–Hartenberg

Date received: 8 March 2021; accepted: 21 May 2021

Handling editor: James Baldwin

## Motivation

With the continuous vigorous development of information technology, the prevalence of the Internet of Things, and the intelligence of machines, it is the future trend. Many industries are constantly seeking new profit models and new technology upgrades, and the industry hopes to obtain higher prices at the least cost. The main way to deal with the profits of the industry is to reduce labor costs. Nowadays have many works that using mechanical arm with image recognition system for classification work, for instance, using image recognition to implement recycling task, like smart trash can; Using the distinction of plant image recognition and classification

The aim of the present study is to combine the use of image recognition techniques with robotic arm operation by uploading captured images onto a PC for classification. We use a chess pattern to conduct the sampling test, and after the system recognizes it as a

---

Department of Mechanical and Electro-mechanical Engineering, Tamkang University, Taipei

### Corresponding author:

Guan-Chen Chen, Department of Mechanical and Electro-mechanical Engineering, Tamkang University, No. 151, Yingzhan Road, Tamsui District, New Taipei 25137.  
Email: gcchen@mail.tku.edu.tw





**Figure 1.** Piranha LED automatic thermal test sorting machine.

proper chess pattern, the robotic arm grabs the objects and moves them to the designated location.

## Literature review and discussion

### *Automatic classification*

Automatic classification techniques have been applied in many domains. For example, sorting machines have been used to classify materials based on their chemical compositions, and image recognition has been used to develop intelligent trash cans; Using image processing technology to filter the leukocyte from image, subsequently distinguish the nucleus from its cytoplasm. Based on their nucleus, and features of leukocyte use various eigenvector to accomplish Nucleus image recognition process, then achieved the purpose of body fluid automatic blood cell classification.<sup>1</sup>

The Piranha LED Automatic Thermal Test Sorting Machine produced by Chang Yu Co., Ltd., as displayed in Figure 1,<sup>2</sup> uses heated-up products to determine product quality.

Superbin, an AI recycle bin developed by the Korean Recycling Service Corporation,<sup>3</sup> uses image recognition technology to identify whether the garbage items disposed by the people have some value, such as plastic bottle, glasses, and tin cans. Moreover, it presses them into flakes for recycling. However, these machines cannot classify garbage items, which it also to become this project concern. If we can integrate three-dimensional (3D) image recognition with automatic classification, we could enhance the development of automatic factories.

### *Image recognition*

Image recognition is a combination of artificial intelligence, machine learning, neural network, and other

technologies. The basic requirement of this technology is the ability to automatically identify the desired objects from a given image. This process involves extracting the features of these objects based on similarity. Machine learning can be classified into three types: supervised learning, unsupervised learning, and reinforcement learning. In image recognition systems, supervised learning is more commonly used than two other types because this type of learning provides questions and answers and teaches the computer to recognize. Unsupervised learning does not give the machine a standard answer; instead, it only identifies its own rules from a given data set. Reinforcement learning does not require proper input and output values, nor does it require precise correction. However, it does require a thorough environment and experience modification.

To increase recognition accuracy, deep learning, a specific machine learning method, is usually implemented in image recognition systems. Deep neural networks are mainly used to apply deep learning methods in machine learning. It requires the construction of a sufficiently multilayered model with adequate parameters that can be changed and strong learning capabilities. Convolutional neural networks (CNNs) represent one type of deep neural networks. Based on this feature, a model is constructed that can learn the characteristics of the objects in a picture and identify the model. Convolution is the process of extracting features from an image to identify the best features and classify them.

A CNN can be divided into the following three parts:

- Convolution layer, which is responsible for extracting object features;
- Pooling layer, which pools a picture into smaller pictures and finally obtains the same number of pictures with smaller pixels; and
- Fully connected layer, which uses the extracted features to realize classification.

**Convolutional layer.** In CNNs, multiple convolution kernels are used to extract different features from an image for identification. The extracted data composed of features are collectively called a feature map. When the parameters of an entire neural network are changed through a learning process, the weight of the convolution kernel is updated to extract more accurate features to facilitate the classification in the fully connected layer, as illustrated in Figure 2.<sup>4</sup>

**Pooling layer.** After the convolutional layer, a pooling layer is occasionally added to reduce the amount of image data, retain important information and highlight features, and perform a maximum or average

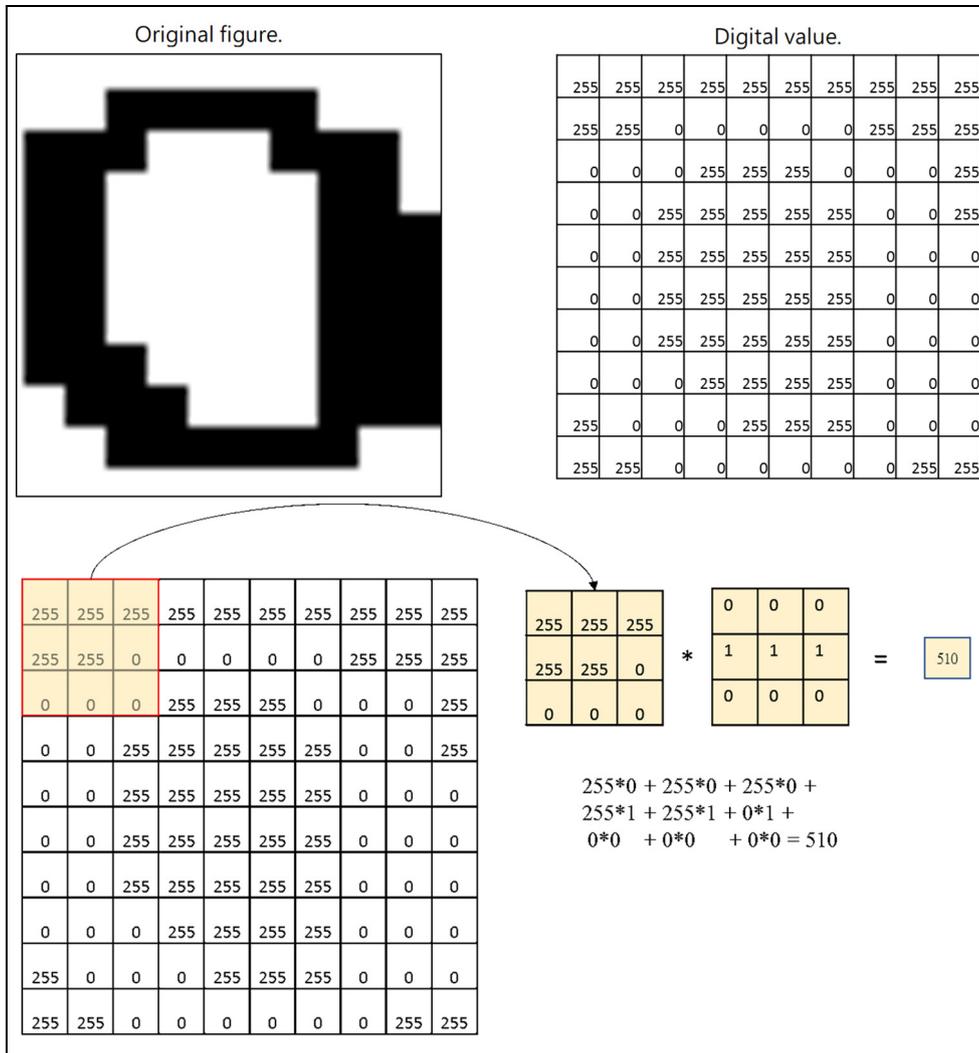


Figure 2. Convolutional layer.<sup>4</sup>

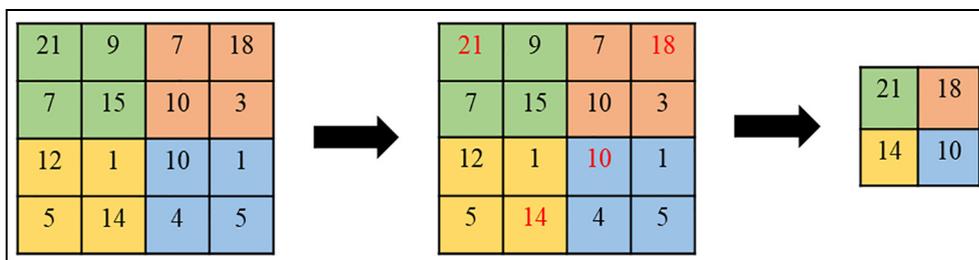


Figure 3. Pooling layer.

dimensionality reduction computation on the original data, as illustrated in Figure 3.

Pooling and convolution reduce the data set size and the computation load. Because extraction is performed with respect to the maximum value range, we are simply extracting the most representative features. Therefore, the amount of data is reduced, even as characteristics of the objects in images are retained. Moreover, it has a good antinoise function.

**Fully connected layer.** The fully connected layer is executed after flattening the feature map extracted by the convolutional layer (Flattening). Then, we connect them with the neural network, as illustrated in on Figure 4. After flattening, every feature is weighted. Different nodes assign different weights to different features, meaning that for every object feature that a node is responsible for identifying, the more obvious or representative feature, the higher is the weight assigned to

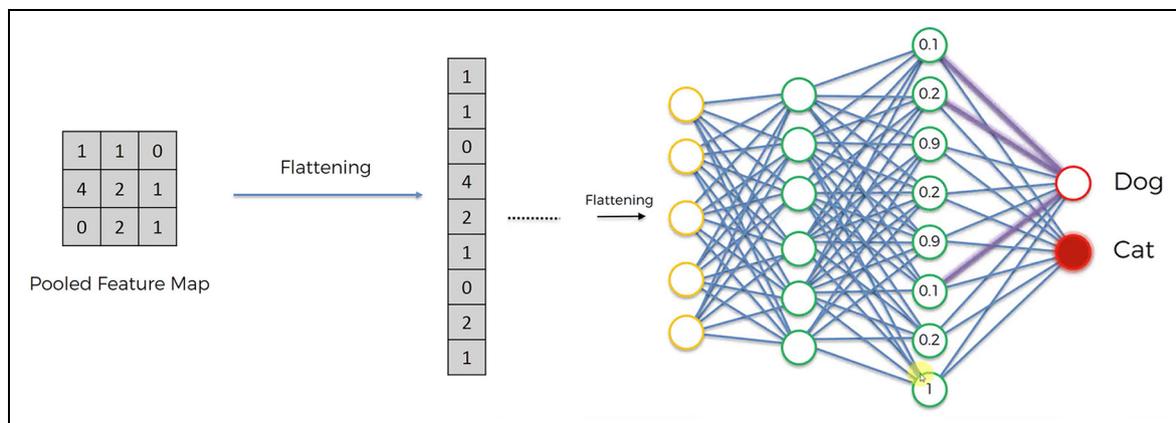


Figure 4. Fully connected layer.<sup>5</sup>

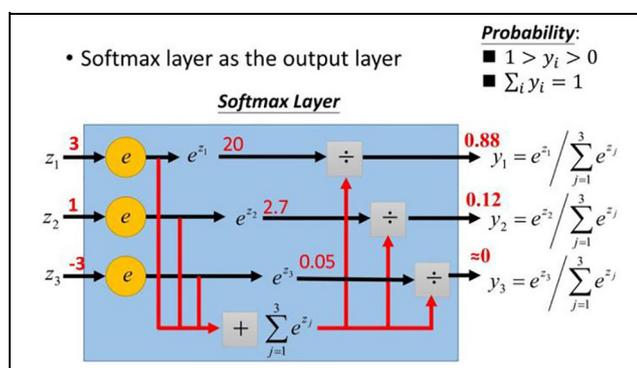


Figure 5. Softmax layer.<sup>6</sup>

it by the node. By contrast, the less obvious the feature, the lower is the weight assigned to it by the node. Therefore, when the extracted feature value and the weight corresponding to the feature are large, the output value of the node will be higher than those of the other nodes.

Finally, the Softmax layer comes into play in the last layer to input the sum of each neuron, as the output of the neuron, and all values range between 0 and 1. The larger the output value of a neuron, the higher is the probability that the category corresponding to the neuron is the real category.

Softmax regression uses the Softmax operation to make the total probability distribution of the output of the last layer be 1. The Softmax function is usually placed in the last layer of the neural network, the output of the last one of all nodes by an exponential function (exponential function), and the result is adding as the denominator. It is output as individual molecules.

The entirely operation illustrated in Figure 5.<sup>6</sup>

### Related technologies required for image recognition

**Tensorflow.** Tensorflow is a machine learning framework developed by Google that supports programming

languages such as Python and C++. This framework has been used to successfully implement automated image captioning software. For example, Deepdream and Python 3.6.8.

**VGG-16.** VGG convolution neural network model is proposed in the literature “Very Deep Convolution Networks for Large Scale Image Recognition” by Simonyan and Zisserman, abbreviations Oxford visual geometry group (Visual Geometry Group) whose name comes from the author’s.

VGG convolution kernel sizes according to different layers and convolution purpose, can be divided into A, A-LRN, B, C, D, E arranged a total of six (ConvNet Configuration), which D, E more commonly used in two configuration. They are called VGG16 and VGG19.<sup>7</sup>

VGG16 contains:

- 13 convolutional layers (Convolutional Layer), respectively represented by conv3-XXX
- 3 Fully connected Layers, respectively represented by FC-XXXX
- 5 pooling layers (Pool layer), respectively represented by max pool

One part of the convolutional layer and the fully connected layer have weight coefficients, so they are also called weight layers. The total number is  $13 + 3 = 16$ , which is the source of 16 in VGG16.

**Keras.** Keras, an advanced deep learning library based on Tensorflow, is scripted in Python and provides concise and highly modular APIs. Users can use more concise and more readable code to build various complex deep learning models and can directly load the pertained VGG-16 model.

The application of image recognition is combined with not only automatic classification but also with workpiece (screw) classification and object (animals, flowers, shapes, etc.) classification. The project will study the combination of image recognition and robotic arms.

## Research methods and steps

The present research is divided into three parts for discussion: establishment of the image recognition system, kinematic control of the robotic arm, and integration and construction of a vision-based workpiece classification system.

### Establishment of image recognition system

**Establishment of data set.** First, we place each chess on a different background and at a different position to capture images. The photo specification is 1:1. Thereafter, we classify the captured images to place the chess with the same font into the same code and sort in the folder. Finally, we have 14 varieties of chess pawns and approximately 14,000 images.

**Training model.** After inputting the dataset into the system, the system initializes the Python programming language, Tensorflow deep learning library, OpenCV library for image processing of chess characters, and CNN model training. Then, all the characteristics of the data set are evaluated and analyzed, system is compared and tested according to the test samples, and recognition rate after the test is corrected until the recognition rate reaches a stable level. Tensorflow is used to establish a CNN network architecture and an enhanced learning mechanism. At the outset, commands such as (Pip install tensorflow == 1.18.0 and Pip install keras) are entered into the command prompt or terminal to install Tensorflow and Keras. As the installation is completed, to build a CNN model, we need to define the parameters (Weight variable) and (Bias variable), respectively, to define the two-dimensional convolutional layer and the pooling layer. After one pass of filtering and compression, respectively, the two-dimensional convolutional layer and the pooling layer are defined; thereafter, after another pass of filtering and compression is performed to realize layer-by-layer calculation, that is, complete CNN modeling. Among them, Keras provides a model for constructing deep learning architecture. In the 2014 paper “Ultra-Deep Convolutional Network for Large-scale Image Recognition” by Simonyan and Zisserman, the image recognition system uses a smaller and more compact architecture than the network layer of the deep learning model of VGGNet. It uses  $3 \times 3$

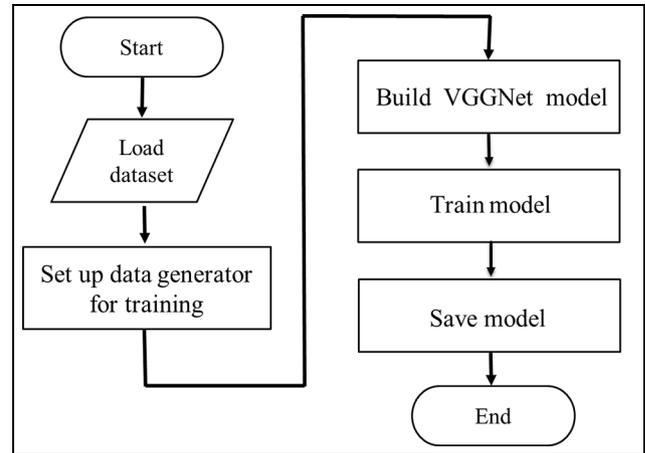


Figure 6. Flowchart for training the VGGNet model.

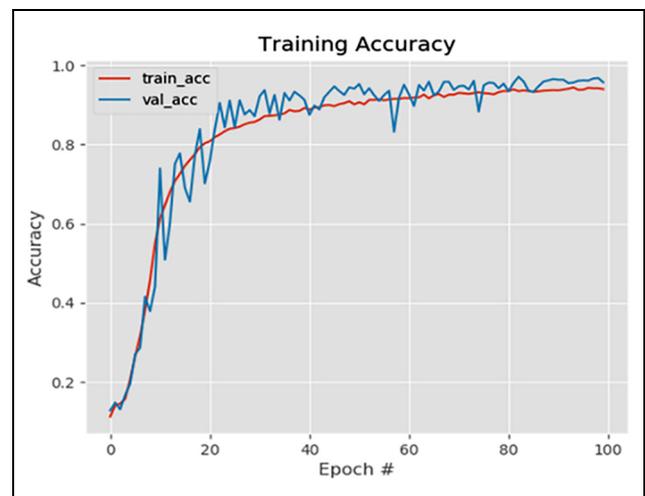


Figure 7. Training model result.

convolutional layers to increase depth in a continuous superposition, reduces the size of the volume layer through maximum pooling, and places the fully connected layer at the end of the network layer and before the Softmax classifier, see Figures 6 and 7.

### Kinematic control of robotic arm

**Choice of robotic arm.** A five-axis robotic arm robotic arm is used in this work. The requirements outlined for selection of the robotic arm were as follows: lightweight, affordable, and working range that can cover a chessboard measuring  $(32.7 \times 23.7)$ . Therefore, we selected the six-degrees-of-freedom metal robotic arm ROB-003492 (shown in Figures 8 and 9)<sup>8</sup> manufactured by Taiwan Smart Sensing Technology Co., Ltd.

Power is supplied to the robotic arm by using six MG996R servo motors. The MG996R motor is an

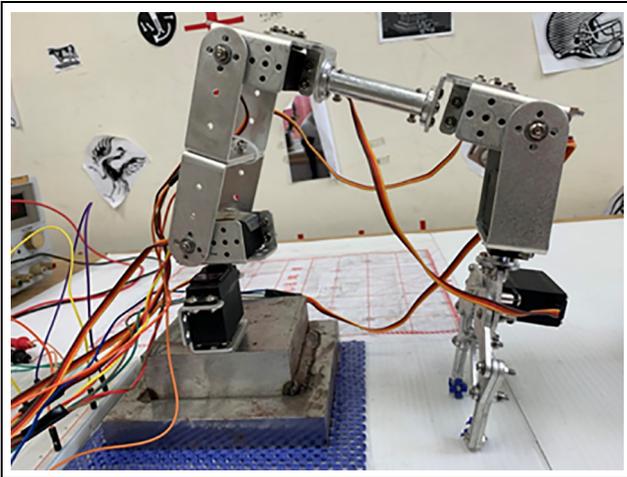


Figure 8. ROB-003492.

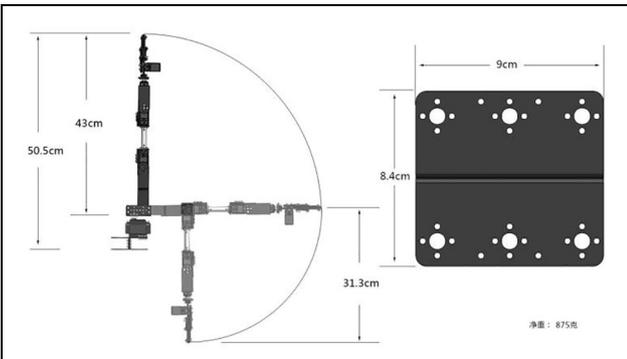


Figure 9. Working range.<sup>8</sup>

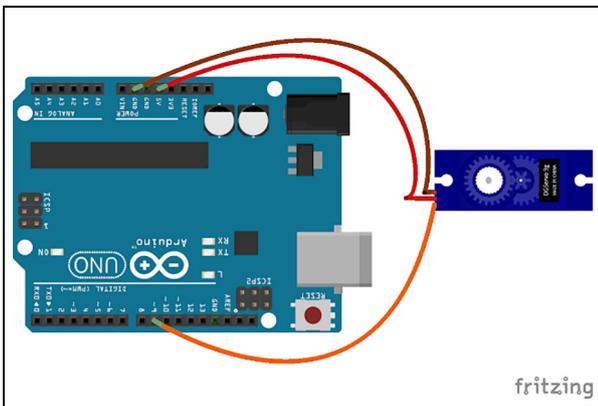


Figure 10. Arduino Uno board.<sup>9</sup>

upgraded version of the original MG995 motor. The speed, pulling force, and accuracy of the upgraded version are higher than those of the original version. This motor is currently one of the most cost-effective large-torque motors in the market. Three different connecting wires are used, namely orange, red, and brown.

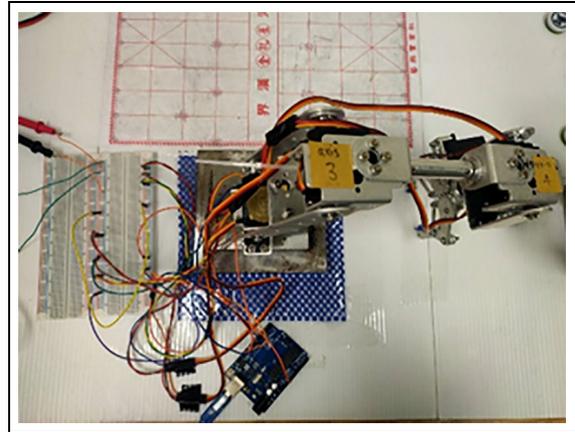


Figure 11. Overall system wiring.

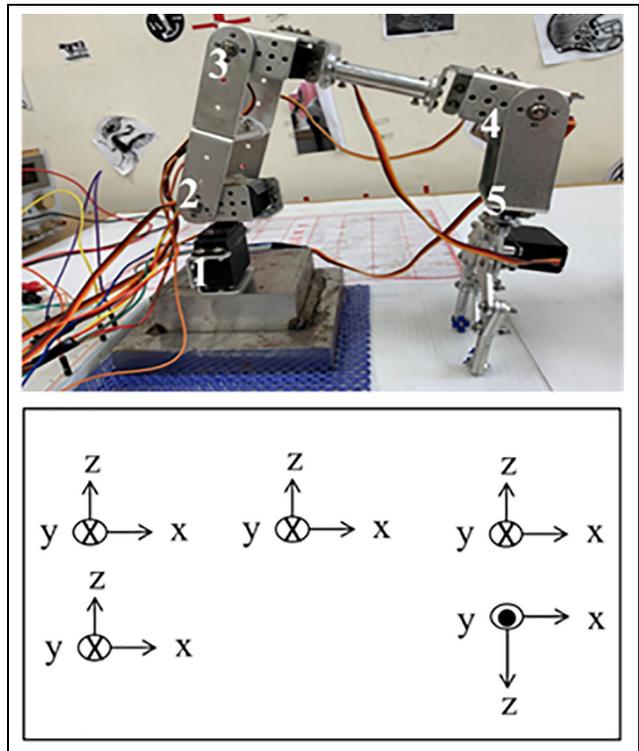


Figure 12. Coordinate system of the robotic arm.

Their respective functions are PWM connection, positive connection, and negative connection. These three wires are connected to an Arduino Uno board, as illustrated in Figure 10.

### Determining the designated location

*Denavit–Hartenberg Model and analysis of forward and inverse kinematics.* See Figure 11 for Overall system wiring, Figure 12 for the coordinate system of the robotic arm, and Figure 13 for Index of DH model.

First, we establish the Denavit–Hartenberg (DH) model of the robot arm and then use the DH model

$d$  : offset along previous to the common normal.  
 $\theta$  : angle about previous  $z$ , from old  $x$  to new  $x$ .  
 $a$  : length of the common normal.  
 Assuming a revolute joint, this is the radius about previous  $z$ .  
 $\alpha$  : angle about common normal, from old  $z$  axis to new  $z$  axis

**Figure 13.** Index of DH model.<sup>10</sup>

Link	$d(m)$	$a(m)$	$\alpha$	$\theta$
1	0.0207	0	$-\pi/2$	0
2	0	0.1064	0	0
3	0	0.15	0	0
4	0	0	$\pi/2$	0
5	-0.112	0	0	0

**Figure 14.** DH model table.

table, illustrated in Figure 14, in the MATLAB software environment to model the machine, as illustrated in Figure 15. After confirming that the model machine

is the same as the robot arm, we input the DH model values and substitute them in the general solution formula, as illustrated in Figure 16, to analyze the forward and inverse kinematics. We press the model machine into the specified position and import its position information into the forward and inverse kinematics program. In addition, we calculate the angle required for each motor position and bring it into the prototype machine, see Figure 12.

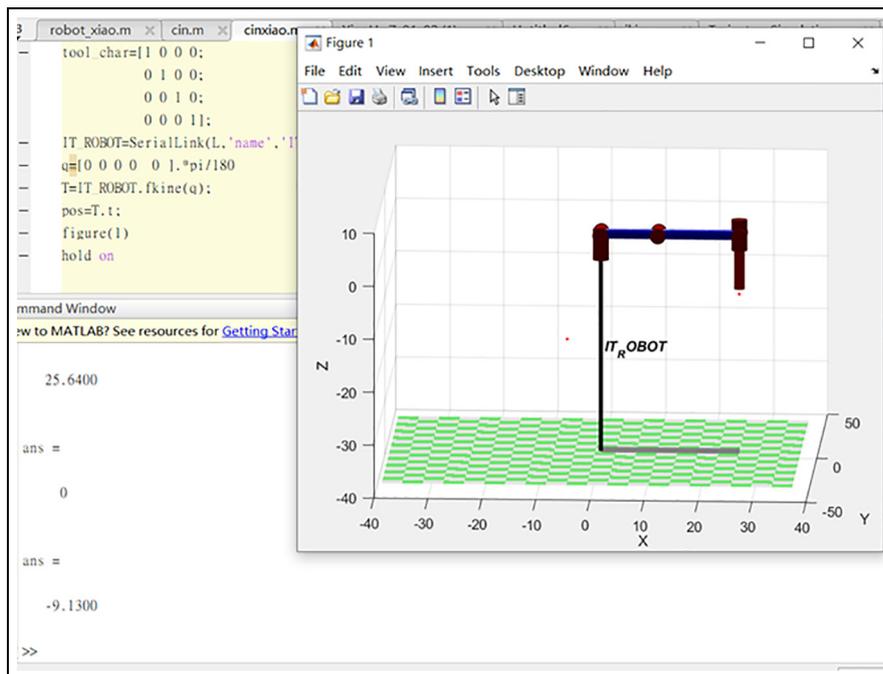
**Integration and construction of vision-based workpiece classification system**

First, it's the entirely system in the Figure 17. The left is the PC which is major of dealing with program, in the middle is robot arm and the right is the camera.

Second, in this chapter will divide into two part to discuss.

**Image capture.** We install a camera at a designated location to shoot the target, and then, we capture an image and import it into the computer for identification.

**Image recognition.** When the CNN model is established, the captured image can be input into the CNN model, and the chess image is recognizable.



**Figure 15.** Graph of prototype machine.

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 & a \\ \sin(\theta)*\cos(\alpha) & \cos(\theta)*\cos(\alpha) & -\sin(\alpha) & -\sin(\alpha)*d \\ \sin(\theta)*\sin(\alpha) & \cos(\theta)*\sin(\alpha) & \cos(\alpha) & \cos(\alpha)*d \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 16. General solution of forward and inverse kinematics.<sup>10</sup>

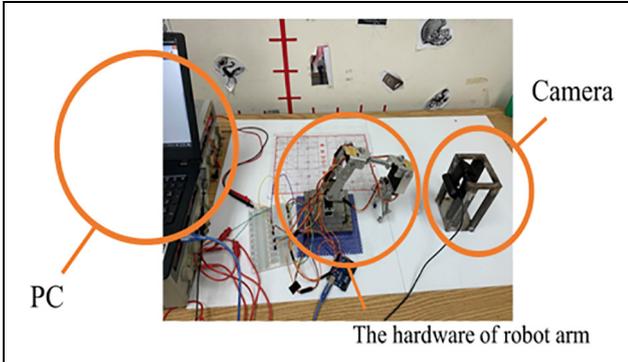


Figure 17. The entirely system.

### Conclusion

In this work, we used Tensorflow and CNN deep learning to design an image recognition system. When the recognition rate is higher than 95%, the computer transfers the recognition result to the Arduino Uno board through a serial port. Then, it controls the robotic arm move to a specific location. The entire process is shown in Figure 18. At this time, the robotic arm grabs the chess and moves it to a specific location. In this work, the computer links the camera with the robotic arm. The process of taking pictures to moving the robotic arm and placing chess is longer than other transmission methods. To ensure that the robotic arm moves to the specified position stably, each of the designated positioning programs is equipped with a delay to ensure that the arm moves to the correct position. Therefore, in this experiment, it is expected that the recognition rate can reach 95% to reduce the waiting time, and 90% of the chesses tested were recognized at the first instance. The recognition rate was higher than 95%, and the system moved the pieces to the desired positions.

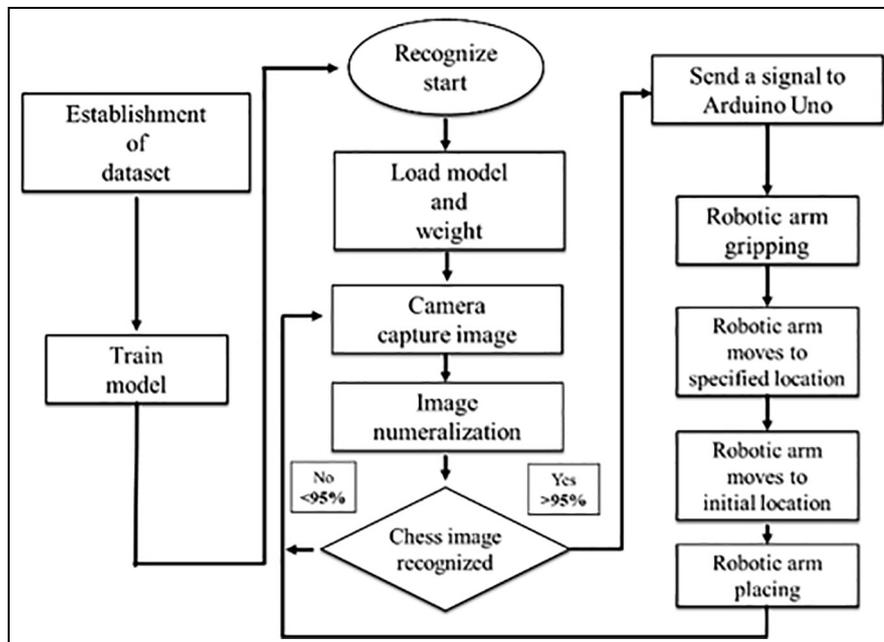


Figure 18. Flowchart of the entire process.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Hsin-Yi Chien  <https://orcid.org/0000-0002-5306-1152>

Guan-Chen Chen  <https://orcid.org/0000-0003-1208-6713>

### References

1. Hsu Yi-C. *An efficient and effective approach to recognition and classification of leukocytes in peripheral blood and body fluid*. Taiwan: National Chung Hsing University, 2005.
2. Chang-Yu Technology Co., Ltd. Online referencing 2(3), <http://www.cyttw.com.tw/List.asp?id=47> (2016, accessed 30 April 2021).
3. SuperBin Co. Korean. Online referencing, <https://www.storm.mg/lifestyle/442679> (2018, accessed 30 April 2021).
4. Hung T. Online referencing 2(3), <https://bit.ly/2In9thG> (2018, accessed 30 April 2021).
5. James Y. Online referencing 2(3), <https://bit.ly/39rI1Lw> (2017, accessed 30 April 2021).
6. Airwaves. Online referencing 2(3), <https://reurl.cc/LbXMRK> (2018, accessed 30 April 2021).
7. Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *ICLR conference on computer vision and pattern recognition*. San Diego, 7–9 May 2015, paper number.ar-Xiv:1409.1556, pp.3–7. 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA.
8. Taiwan Smart Technology Co., LED. Online reference 2(3), <https://bit.ly/3cyBII4> (2018, accessed 30 April 2021).
9. Ceiling TSAI. Online referencing 2(3), <http://reurl.cc/ynVao2> (2017, accessed 30 April 2021).
10. Wikipedia. Online referencing 2(3), <http://reurl.cc/a5yDz3> (2021, accessed 30 April 2021).