

# 基於深度學習之工件自動辨識與取料之應用

## Automatic Workpiece Recognition and Gripping Based on Deep Learning

黃成凱<sup>1\*</sup>、黃緒哲<sup>2</sup>、嚴健榮<sup>3</sup>、古有彬<sup>4</sup>、楊淳宜<sup>5</sup>、陳俊皓<sup>6</sup>

<sup>1</sup> 工研院智慧機械科技中心智慧機械技術組 群聚產業創新部 工程師

<sup>2</sup> 國立台灣科技大學機械工程系控制組 特聘教授

<sup>3</sup> 國立台灣科技大學機械工程系控制組 研究生

<sup>4</sup> 工研院智慧機械科技中心智慧機械技術組 群聚產業創新部 研究員

<sup>5</sup> 工研院智慧機械科技中心智慧機械技術組 群聚產業創新部 副工程師

<sup>6</sup> 工研院智慧機械科技中心智慧機械技術組 群聚產業創新部 經理

**摘要：**本文使用深度學習進行工件辨識，並整合機械手臂進行物件夾取。利用 GPU 建立 3D 視覺辨識系統，藉由深度攝影機擷取影像資訊，辨識物件種類及座標，使機械手臂可夾取特定之物件。

視覺系統由 GPU 搭配深度影像之函式庫，分別進行影像資料擷取、深度資訊運算、座標轉換、影像輪廓搜尋和卷積類神經網路模型訓練等。本文採用 YOLOv2 判別目標物體之種類和中心點並利用輪廓搜尋方法找出物體之角度資訊，作為機械手臂操作目標點，並透過座標轉換的方式將相機座標轉為機械手臂座標，由 TCP/IP 通訊傳至運動控制系統進行物件夾取。

**Abstract :** Deep learning is used for workpiece identification in this paper, and the robot arm is integrated for object gripping. The Graphic Processing Unit (GPU) is used to create a 3D visual recognition system. The depth camera captures the image information and identifies the object types and coordinates. Robot arm is used to capture the specific objects.

The vision system is composed of a GPU and a deep image library for data capture, depth information calculation, coordinate conversion, contour retrieval, and convolutional neural network model training. In this paper, YOLOv2 is used to identify the types and center point of the target object. The method uses the contour retrieval method to find the orientation angle which will be used to guide the robot arm, and to convert the camera coordinate to the mechanical arm coordinates. The coordinate conversion information will be transferred to the motion control system for object picking through TCP/IP.

**關鍵詞：**深度學習、卷積類神經網路、目標檢測

**Keywords :** Deep learning, Convolution neural network, Target detection

### 前言

隨著科技的進步，人工智慧的應用愈來愈普遍，而目前機械學習大多應用於影像識別或語音辨識；機械學習大致分為兩種：監督式學習 (Supervised Learning) 及非監督式學習

(Unsupervised Learning)；在非監督式學習中，Cockburn 等人 [1] 將此方法應用於力量感測器之物件夾取，並辨識物件抓取之穩定性，藉由感測器之特徵值訓練，夾取成功率將可達到 83.70%；而在監督式學習的架構中，JaehyunYoo 和

Johansson[2] 將道路的九個路標作為特徵值訓練，並建立路標辨識之模型，讓機器人去執行定位的辨識，並且得到良好的辨識率。

深度學習 (Deep Learning) 是機械學習中準確率極高的一種方法，而工業界與深度學習方法之搭配有許多應用，如：應用於機械手臂之視覺物體辨識和取物等 [3]；在生活上的應用，如：應用於機器人之道路偵測以及巡航 [4]。在深度學習當中，捲積類神經網路 (Convolution Neural Network) 更是廣泛被使用，因為其本身對於視覺分類有著非常大的強健性。Peng[5] 等人利用影像物體的輪廓以及物體不同的姿態去進行捲積類神經網路 (CNN) 之訓練，在 3D 物件辨識也有不錯的效果。

在物件辨識方面，捲積類神經網路當中，有非常多的架構及演算法應用於模型強化及其應用：Ren[6] 等人提出了 Faster R-CNN 的方法，應用於 Region Proposal Networks 的架構，對於物體的辨識上有著不錯的辨識率，而對於物體的重疊的情況下，必須利用影像切割去辨識重疊的物體，Long[7] 等人提出 FCN(Fully Convolutional Networks) 的方法，對影像中的物體進行切割動作，有著很好的強健性，對於即時系統的物體辨識，Redmon[8] 等人提出了 YOLO(You Only Look Once) 的方法搭配捲積神經網路之應用，有著極快的速度與極高的穩健性可滿足即時系統之需求。

本研究主要的目的為透過深度學習建立影像模型，再整合影像處理方法，讓此系統能夠分類出物體的種類及物體的位置座標，最後以機械手

臂進行物件夾取。

## 捲積類神經網路系統

本文採用捲積類神經網路模型 (CNN) 進行深度學習，本節將說明捲積類神經網路系統之架構及運算方式，捲積類神經網路的示意架構圖，如圖 1 所示。輸入影像資料進入此模型時，必須先經過捲積層及池化層的運算，等運算完畢後，再將運算完畢之資料作為類神經網路之輸入，進行模型訓練後即可取得輸出。

### 1. 捲積層、池化層與全聯接層之架構及運算方式

在進入神經網路訓練之前，首先須先將彩色圖片分別轉換成三個 RGB 二階陣列，並必須藉由捲積類神經網路之捲積層以及池化層進行影像資訊之特徵擷取，分別對輸入影像執行捲積運算以及池化運算，其中圖片可同時對多個捲積層進行運算：捲積運算就是將下圖兩個  $3 \times 3$  的矩陣作相乘後再相加，以下圖為例： $0 \times 0 + 0 \times 0 + 0 \times 1 + 0 \times 1 + 1 \times 0 + 0 \times 0 + 0 \times 0 + 0 \times 1 + 0 \times 1 = 0$ 。

在進行捲積層運算時，首先必須選擇所要的濾鏡大小，而濾鏡的大小可為  $(n,n)$  之單一或多個矩陣，如圖 2 所示。此架構使用了  $(3,3)$  的矩陣架構來當作濾鏡的大小，而  $(4,4)$  為我們輸入資料之矩陣，而完成捲積運算後輸出之矩陣資料稱為特徵地圖。

由於捲積計算會造成資料衰減，讓原來的圖像資料解析度縮減，並失去部分特徵。為了不使

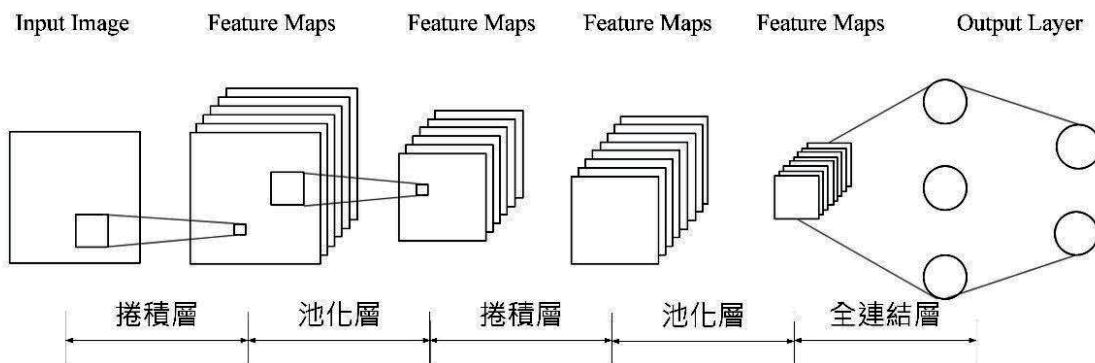


圖 1 捲積類神經網路架構示意圖

## 更完整的內容

詳見 ■ 機械工業雜誌 ■ · 428 期 · 107 年 11 月號

---

機械工業雜誌 · 每期 **220** 元 · 一年 12 期 **2200** 元

劃撥帳號：07188562 工業技術研究院機械所

匯款帳號：兆豐國際商業銀行新竹分行(代號 017)，帳號/ 203-07-02288-0

訂書專線：03-591-9339

傳 真：03-582-2011

機械工業雜誌 · 官方網站：[www.automan.tw](http://www.automan.tw)

機械工業雜誌 · 信箱：[jmi@itri.org.tw](mailto:jmi@itri.org.tw)