

# 應用深度學習之半結構化環境物件分類與機械手臂持取系統

## Semi-Structured Object Classification with Deep Learning for Robotic Grasping Systems

陳金聖<sup>1\*</sup>、陳世剛<sup>2</sup>、Luis Cardenas<sup>3</sup>、廖述政<sup>4</sup>、林育信<sup>5</sup>

<sup>1</sup> 國立台北科技大學 自動化科技研究所 教授

<sup>2</sup> 國立台北科技大學 機電科技研究所 博士生

<sup>3</sup> 國立台北科技大學 機械與自動化碩士外國學生專班 碩士生

<sup>4</sup> 旭東機械工業股份有限公司 半導體事業處 副理

<sup>5</sup> 旭東機械工業股份有限公司 電子設備事業處 組長

**摘要：**因應智能工廠之來臨，為了達成智能化物件持取流程，本文提出一高效率之演算法，並結合深度學習演算法，透過物件群集之特徵，解決多重類別物件在半結構化環境中之辨識與分類問題。本研究根據所需的物件建立訓練樣本，訓練深度學習網路，根據輸入的影像資訊，進行物件之識別與類別分類。最後為了要完成機械手臂持取任務，本文透過半結構性環境之約束條件，基於物件位置的估測來規劃持取位置，降低夾爪與其他物件碰撞的情形，提升持取的成功率。透過本文所提出之方法，未來可在無人工廠中，針對工作重複性較高的出入料區，透過機器視覺提供物件資訊，並使用機械手臂進行自動取放的工作，減少人工成本。

**Abstract :** In order to approach the target of the intelligent pick-and-place process in smart factory, this paper proposed an efficient method toward the problem of identifying and classifying of multiple objects from different classes within a semi-structured environment based on deep learning algorithm. A deep neural network is trained to detect and classify objects in the scene and estimate their positions and rotations. An analysis on whether it is better to use clustered or non-clustered items for this specific purpose is provided. Finally, to fulfill a subsequent robotic picking task, a grasping configuration is predicted based on the estimation of object's position, considering the constraints of this semi-structured environment. The proposed intelligent pick-and-place process uses the object information which is detected by the machine vision, and to command robotic arm to pick-and-place the object automatically. The proposed method is a valuable technique in the unmanned factory to replace the work which has highly repeatability.

**關鍵詞：**深度學習、物件分類、夾取配置

**Keywords :** Deep learning, Object classification, Grasping configuration

### 前言

近幾年來，視覺影像在各領域應用多樣化，而隨著電腦視覺科技的蓬勃發展，電腦視覺發展成熟，如今被廣泛的應用於各種領域，例如瑕疵

檢測、物件辨識、影像分類及定位技術 [1,2]。因應智慧機械之蓬勃發展與人工智慧技術日漸成熟，智能化的製造流程伴隨著智能工廠之發展增加，也造就了更多的視覺模組與工業機器手臂的配合需求。對於工業上重複性高、環境因素固定的工

作場合，能提高生產效率以及降低人力成本等等，具備高利用價值。機器視覺整合機械手臂系統常用於，如：物件持取或物件整列等應用場合 [3]。雖然機器視覺已經能夠提供高效能與高精度的影像辨識成果，但是對於整合機械手臂系統仍然是一大挑戰。以物件持取問題為例，仍然會遇到下述幾項需要突破之瓶頸，(1) 雖然現今之視覺系統已經能夠提供物件辨識之位置，但是對於特定物件之樣本建立仍然必須透過人工之方式進行感興趣區域 (ROI) 之選取；(2) 對於物件中心之訂定，則取決於物件邊緣輪廓特性或是機械手臂夾爪之設計；(3) 對於物件在實際空間上之位置，需透過相機座標、世界座標與機械手臂座標等轉換才能得到物件實際上在機械手臂坐標系上的位置；(4) 在半結構化與隨機擺放的環境中會有物件堆疊的情況發生。為了解決上述的問題結合機器學習演算法來進行影像處理的案例越來越多，並且經由深度學習演算法進行物件特徵的學習後，能夠成功識別物件的成功率越來越高，深度神經網路也被提出來結果各式各樣的影像處理問題；透過深度學習演算法來進行物件識別的成功率甚至可以高過於傳統的演算法，因此透過深度學習來對於多物件樣本進行辨識處理的技術越來越被廣泛應用 [4,5]。

本文提出一個基於 Single Shot MultiBox Detector (SSD) 的神經網路影像分類器 [6]，並且針對三種不同的物件樣本進行訓練，物件樣本分別為：積木塊、鈕扣與螺絲。SSD 演算法會先產生出包圍盒包圍物件在影像中的位置，再透過最小矩形的運算，根據物件輪廓特徵產生出包圍物件的最小矩形，並且計算物件在影像中的位置資訊。在計算完成物件位置後，夾爪配置規劃為持取前非常重要的一個步驟，若沒有經過配置的規劃直接進行物件的持取可能導致夾爪碰撞到其他的物件，造成夾取失敗的現象發生，因此經由夾取配置規劃能夠提升機械手臂對於持取的成功率 [7,8]。本文所使用的夾取配置演算法首先根據最小矩形的位置計算夾取配置矩形，並且根據夾取配置矩形

與影像中其他物件透過二值化處理進行碰撞的計算，若有碰撞的情形發生，則逐漸旋轉夾爪直到沒有碰撞的情況發生，最終透過計算完成的夾爪配置來進行物件的實際物件的夾取測試 [9,10]。

過去機器視覺系統使用中央處理器 (CPU) 進行運算，透過傳統演算法進行群集物件之識別與分類花費較短的時間，較適合應用於具有即時性需求之場合；然而伴隨著現今影像處理單元 (GPU) 之運算速度增加，透過神經網路分類器進行影像分類的效果與運算速度逐漸提升。針對群集物件進行影像處理的過程，透過卷積運算 (Convolution) 來進行特徵萃取的效果甚至已經可以超越傳統演算法，並且透過 GPU 進行平行運算使得效能足以應付機器手臂整合機器視覺之即時需求。根據以上所敘述的優勢，透過本文所提出智能化物件辨識與機械手臂持取姿態規劃演算法來進行特定物件之持取工作，可解決傳統方法在物件堆疊時，容易有持取失敗與影像辨識失敗的問題。因此透過智能化的機械手臂系統來解決具有高重工性的工作區域，例如：出入料區、倉儲區與物料籃中的物件持取，相對於以往透過人工方式來進行工作，能夠提供更穩定的工作品質且減少出錯誤機率，並降低人工成本。

## 深度神經網路影像分類器

在一半結構化的環境中，為了要能夠辨識多個不同類別的物件，並且解決物件重疊的問題，本研究透過深度神經網路分類器，透過輸入影像以及預先訓練好的神經網路進行物件辨識，最終透過一邊界盒包圍透過影像辨識所到的物件，並且提供機械手臂系統作為物件持取的依據，其流程如圖 1 所示。因此影像辨識之強健性，即辨識速度以及精度皆須納入考量，而會直接影響以上性能之因素如：特徵擷取器、影像解析度、IoU 閾值 (Intersection of Union)、匹配策略等，皆會在以下章節進行探討。

### 1. 深度神經網路架構

目前常見的類神經網路架構皆能提供良好的

## 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)