

2021 人形機器人魔術秀：硬幣魔術

孫煜翔¹ Hanjaya Mandala² Jacky Baltes³ Saeed Saeedvand⁴

¹ 國立臺灣師範大學 電機工程學系

E-mail: 60975030h@ntnu.edu.tw

² 國立臺灣師範大學 電機工程學系

E-mail: 60775042h@ntnu.edu.tw

³ 國立臺灣師範大學 電機工程學系

E-mail: jacky.baltes@ntnu.edu.tw

⁴ 伊朗大不利茲大學 電機與計算機工程學院

E-mail: saeedvand@tabrizu.ac.ir

摘要

硬幣魔術一直深受魔術師喜愛且持續了數十年以上之久。一個著名的把戲是讓志願者把硬幣藏在背後，讓魔術師猜出硬幣在哪隻手上。

我們受到了專業魔術師 Chris Ramsay 的啟發，他的訣竅步驟如下：首先，志願者取一枚硬幣放在你的背後，接著，雙手向前且交叉雙臂，我們可以通過閱讀隱藏硬幣的人的肢體語言來猜測答案。因此，人型機器人需要掃描志願者的肢體語言並預測硬幣藏在哪隻手上。

本論文展示了人形機器人如何在硬幣魔術中對於猜測隱藏硬幣有著高度的準確性。在第一階段，我們嘗試使用光流法(optical flow)，但並未得到預期的結果。在第二階段，我們因應人機互動的環境提出了一個名為 Crossed Arm Voxel Network (CAVN) 的深度學習架構模型，該模型使用機器人上的光學雷達 (LiDAR) 掃描而來的點雲數據來完成我們的分類任務。我們在 THORMANG-Bear 人形機器人上演示了我們所提出的算法，THORMANG-Bear 是 THORMANG3 人形機器人改良的上半身版本。THORMANG-Bear 的每條機械手臂包括 12 個自由度 (DOF)，總共有 27 個自由度。此方法在 IROS 2019 RM 比賽期間的現場表演進行了完全測試，準確率超過 90%。

關鍵詞：深度學習、卷積神經網路、電腦視覺、人形機器人、機器人作業系統(ROS)、點雲。

Abstract

Coin magic trick is one type of magic that always been very fascinating to the audience and has been performed by magicians for centuries. One famous trick is for a volunteer to hide a coin behind their back, and for the magician to guess in which hand the coin is.

Our implementation of this trick inspired by Chris Ramsay, a professional magician. His trick is as follows: first, the volunteer takes a coin and put in your back, next, bring your hands forward, finally, cross your arms. Therefore, the answer can be guessed by reading the body language of the person who hides the coin. So, a humanoid robot needs to scan the body language of the person and predict which hand is hiding the coin.

The technical approaches of our coin magic trick including two phases. In the first phase, we deliberate our results using a heuristic approach optical flow as the fails attempt to perform this magic trick. In the second phase, we propose a robust magic trick in interactive environments through a learning-based algorithm. We propose a deep learning architecture model entitled Crossed Arm Voxel Network (CAVN) that uses point clouds data from the robot's light detection and ranging (LiDAR) scanner for our magic classification task. We implemented our proposed algorithm on the THORMANG-Bear humanoid robot which is a modified upper body version of THORMANG3 humanoid robot. THORMANG-Bear includes 12 degrees of freedom (DOF) at each arm with 27 DOF in total. This approach was thoroughly tested on the live performance during the IROS 2019 RM competition with over 90% accuracy.

Keywords : Deep Learning、Convolutional Neural Network、Computer Vision、Humanoid Robot、Robot Operating System (ROS)、Point Cloud.

1. 前言

現場魔術秀之盛行已經持續數十年之久，通過一些欺騙性的技巧成功地給人們留下了深刻的印象，使其依然為現代盛行的娛樂項目之一。將機器人應用於魔術秀的表演可能是人型機器人最有趣的應用之一，機器人可以依其能力、智慧和技巧為大家帶來驚喜。如果使用人形機器人表演魔術，我想可能會有人問“為什麼要讓人形機器人表演魔術呢？”。我們認為，如果人形機器人能夠執行非常複雜和艱鉅的任務，例如通過學習導向的演算法執行魔術，必然也能夠執行些簡單的日常任務，例如物件的放置與撿取，這對於其他領域的應用想必是有所幫助的。

總而言之，人形機器人應用於魔術表演是非常具有挑戰性的任務，機器人需具備高速視覺、物件的快速操縱和人機互動 (HRI) 等多項能力。例如：在紙牌魔術中，人形機器人必須得在撲克牌高速移動下確保其準確性，以及與觀眾的持續溝通互動，才能製作有吸引力的娛樂節目。

硬幣魔術有許多不同的種類。在 [7] 中，迄今為止已經有 46 種有趣的硬幣魔術。我們需要一台擁有著複雜操控能力的機器人才能施展這樣的魔術。另一方面，開發能夠執行在線任務的智能機器人需要在各種情況下都有著穩健的性能。如今，有著穩健的實時環境且最實用、最強大的人工智能算法，基於深度學習 (DL) 所研發之算法 [8] 想必是很好的選擇。DL 中最著名的核心之一是卷積神經網絡 (CNN)，它是電腦視覺應用中最為成熟的 DL 算法 [9, 10]。因此，在本文中，我們將展示我們提出的使用 DL 算法通過人型機器人執行硬幣魔術。

2. THORMANG 機器人系統概述

THORMANG (Tactical Hazardous Operations Robot) 是由 (ROBOTIS, Inc) 開發的全尺寸開放平台人形機器人[1]。THORMANG 的主要應用是救援任務和多個科學領域的研究。

然而，基於我們在本文中介紹的硬幣魔術，我們不需要機器人的運動功能。因此，我們只使用名為 THORMANG-Bear 的 THORMANG3 的上半身改良版本。在本節的其餘部分，我們將把 THORMANG-Bear 機器人的細節分為硬體和軟體來介紹。

2.1. 硬體介紹

THORMANG3 機器人與 THORMANG-Bear 機器人的主要區別在於機器人的手腕部件。由於原廠的 THORMANG3 所設計的手腕部件為爪子型式，我們將其替換為由 Seed Robotics 開發的等身人形機器手 [15]。修改後，機器人的尺寸改變如下描述，如圖 1(a) 所示 THORMANG-Bear 總共有 27 個自由度：頭部 2 個自由度，軀幹上 1 個自由

度，每個手臂上 12 個自由度。機器人上則配備了兩台用於運動控制和感知的小型電腦、一台網絡鏡頭、一台激光雷達掃描儀和一台揚聲器。圖 1 詳細顯示了 THORMANG-Bear 規格細節。

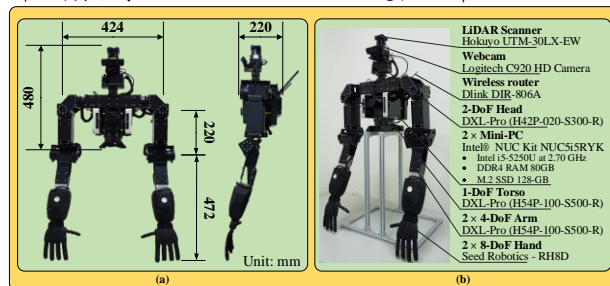


圖 1. THORMANG-Bear 硬體說明。

(a) THORMANG-Bear 機器人尺寸。

(b) THORMANG-Bear 規格。

2.2. 軟體介紹

整個機器人的軟體分佈在三台電腦上 (見圖 2)。感知計算機 (PPC) 接收來自其他兩台電腦設備傳入的感官信息數據，其中內置的網絡攝像頭提供相機畫面和音源輸入，以及來自光學雷達 (LiDAR) 掃描來的點雲數據。運動計算機 (MPC) 處理機器人的運動學模組，該模組通過將為位置輸入執行器從而計算每個關節的運動角度。同時，為了處理 PPC 的所有輸入，以及處理通過 MPC 所執行的運動軌跡，還需要一台計算機來統籌這兩個硬件。最後，第三台電腦則稱為操作電腦 (OPC)。此外，為了將多台 PC 集成到一個同步處理單元中，我們使用了機器人操作系統 (ROS)。使用 ROS 的主要優點是在多計算機通信系統下更容易開發消息傳遞功能。

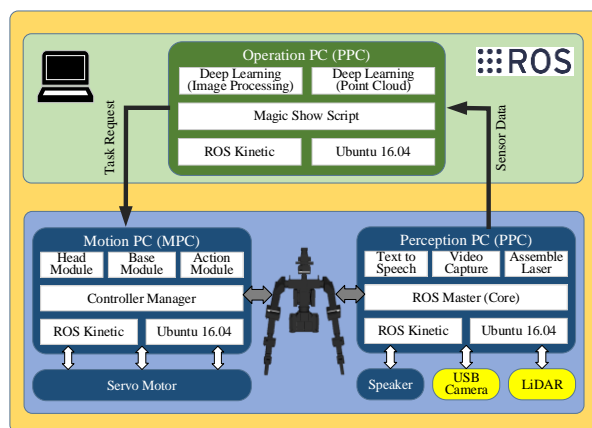


圖 2. THORMANG-Bear 的軟體架構

3. 硬幣魔術秀

Chris Ramsay 是一位有輒高超技術的著名魔術師，在他的 YouTube 頻道上有多個魔術教學，其中包含“硬幣在哪隻手？！- 學習如何用肢體語言找到它！” [16]。影片中教導你如何準確地猜測硬幣在觀眾的哪隻手上。他的解釋如下，首先，取一枚硬幣放在你的背後，然後將你的手向前。他引入了三種肢體語言來準確預測隱藏的硬幣，(i) 第一個技巧是觀察眼睛。自願者的眼睛會微妙地向著握著硬幣的手抽搐。(ii) 觀察自願者的鼻子。舉例來說，鼻子為中心點，你可以從頭部偏動的方向看出硬幣可能在哪隻手上。(iii) 第三種方法是從交叉手臂的情況下判別。這種方法的原理是我們快速地要求志願者將手放在前面並要求“直接交叉雙臂”，交叉於下面的手中有硬幣的機率很高。其原因為人們在心理上會反射性的保護秘密。由於第一種和第二種方法其線索實在過於細微，這意味著成功率較低，所以我們選擇了第三種方法（交叉雙臂肢體語言）作為魔術技術上的基礎，且能夠使用 THORMANG-Bear 機器人來完成。圖 3 為 Chris Ramsay 的 YouTube 影片截圖。

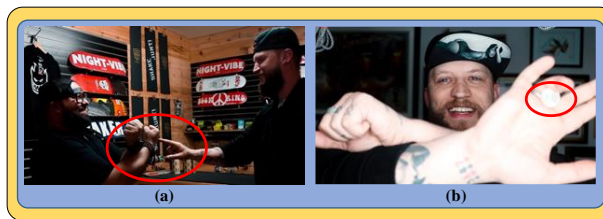


圖 3. Chris Ramsay 展示隱藏硬幣魔術教學。

(a) 他猜自願者把硬幣藏在右臂上。

(b) 他解釋人們會下意識的把硬幣藏在下面。

藉由這些心理微妙之處，他聲稱我們將能夠通過觀察自願者的肢體語言來猜測答案。他聲稱，此肢體語言最為重要的因素就是當們要求交叉雙臂時，大多數人會將硬幣放在另一隻手下方。這是人想保護秘密的反射動作。我們則在沒有告訴測試者訣竅的情況下對 50 個不同人做了實際測試（不使用機器人的情況下），藉此實驗此種方法的準確性，實驗結束後我們得到了 92% 的準確率。綜上述所說，機器人的最終目標就是要找到哪之手是在哪之手的下方。如圖 4 所示，我們演示了機器人尋找人手中隱藏硬幣的魔術流程。

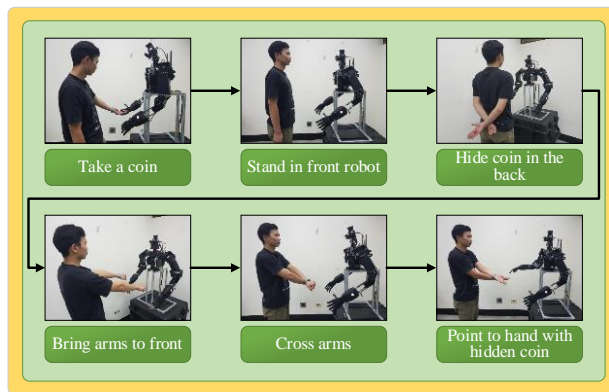


圖 4. 硬幣魔術流程

圖 4 所示的硬幣魔術表演的步驟解釋如下，首先，機器人要求志願者取一枚硬幣。然後，志願者站在機器人面前，一隻手拿著硬幣。接下來，機器人會指示志願者將硬幣放在他/她的背後，決定好要將硬幣藏於哪之手。之後，志願者將雙手向前伸出，在這情況下機器人會要求交叉雙臂。最後，機器人會預測並指出硬幣在哪隻手上。

3.1 光流法

光流(optical flow)是一種視覺上的物件追蹤方法，用於在兩個連續幀之間找到由物件或相機的運動所引起的明顯軌跡。其為一個 2D 的向量場，會顯示從上一幀至這一幀特徵點的移動向量。為了自動決定使用哪些特徵點，S. Jianbo 與 Tomasi(1994) [17] 提出了一種如何自動選取優良的特徵點來追蹤的方法。此外，我們還採用了 Lucas-Kanade (LK) 光流演算法來跟蹤這些特徵點。值得一提的是，LK 算法只計算稀疏特徵集的光流，這意味著它只跟蹤你所給定的特徵點。



圖 5. 跟蹤手臂的光流圖

圖 5 所示跟蹤手臂的光流圖包括四個步驟，描述如下。首先，使用角檢測(corner detection) [17]自動選取好的特徵來跟蹤。其次，我們過濾了畫面的底部區域，剩餘特徵點則為手的重要特徵。第三，我們使用 k-平均演算法(k-means clustering)將特徵點分類為左右手，左側和右側的點則分別用綠色和藍色表示。最後，Lucas Kanade 光流演算法在手臂運動期間跟蹤這些特徵點。為了檢測哪隻手是在下面，我們使用一種算法來計算各側成功移動的特徵點的總數。查看圖 5 最後一步驟，左臂位於頂部位置，由於移動時手臂交叉，位於下方的特徵點會在移動的過程中被覆蓋掉，以至於在最後下方的特徵點會少於上方的特徵點，從而得知哪之手是在下面。

基於光流的視覺跟蹤方法因其迅速以及穩定性而被廣泛應用，但它在非等速運動的物件中很容易發生錯誤，像是檢測不同的人群交叉手臂。並且在現場表演時，很難在手上識別出足夠數量的特徵點。此外，光流法無法判別是物件在移動還是相機在移動，意思是說，相機必須是靜止的，且連續幀下物件的像素須是固定的，這些缺點皆容易造成特徵點的遺失。在這個魔術中必須有更為穩定的依據來做特徵點。因此，我們使用來自光學雷達的三維數據，並採用基於深度學習的演算法來解決光流法中的這些缺點。

3.2 深度學習

上章我們根據視覺數據來檢測哪隻手在另一隻手臂的下方。來自機器人相機的二維 (2D) 影像可以提供有關此問題的答案。然而，影像的信息在高度和寬度上則受到了限制，手臂的深度距離則無從得知。與 2D 影像不同，點雲 (point cloud) 數據則提供了豐富的三維 (3D) 數據。我們可以使用機器人的光學雷達獲取點雲數據。這種方法的主要目的是使用光學雷達掃描擺出交叉雙臂姿勢的自願者從而來進行分類。眾所周知，捲積神經網絡 (CNN) [10] 應用於 2D 影像的物件識別有著顯著的效果。同樣的概念也已通過使用 CNN 應用於 3D 點雲數據 [18, 19]。因此，我們藉由光學雷達的點雲數據上提出了一種基於 CNN 的分類算法，下圖說明了我們所提出的深度學習 (DL) 算法的流程圖。

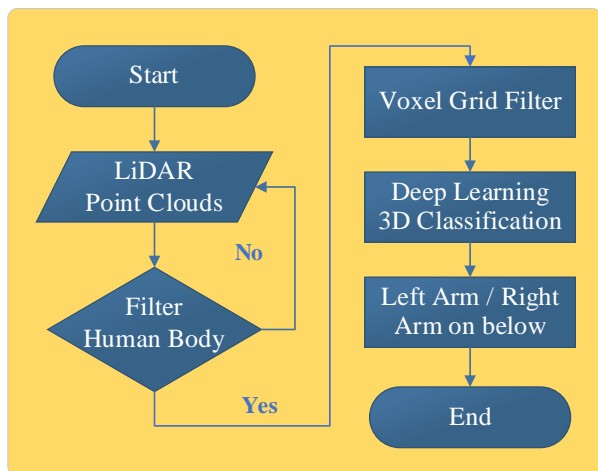


圖 6. 交叉手臂分類深度學習算法流程圖

下面將解釋 DL 交叉手臂分類的工作過程。首先，藉由轉動機器人頭部馬達來帶動光學雷達掃描從而獲取點雲數據。如圖 7(b) 所示，點雲數據除了自願者的身體外還包括而外的環境信息，要解決這個問題，須對點雲數據進行過濾，我們則應用了一種啟發式算法來從雜亂的環境中過濾和提取身體訊息，如圖 7(c) 所示。在這個過程中，提出了基於歐幾里德距離來提取身體訊息的方法，如果機器人前面沒有任何人的話，則重新進行掃描直至收集到點雲數據。

基於 DL 算法的結構，其需要固定數量的輸入大小才能輸入模型，然而由於每次光學雷達掃

描結果給出的點數沒有固定的形式，不適合用於模型輸入，為了解決這個問題，點雲數據需轉換為常規的立體像素網格 (VG)。過濾後的人體點雲在空間上離散為 $32 \times 32 \times 32$ 的二進位體素 [見圖 7(d)]。接下來，固定大小 ($32 \times 32 \times 32$) 的體素數據輸入到 DL 的模型中。在此過程中，DL 模型執行數學計算來處理 3D 分類任務。最後，DL 模型的輸出將根據最高機率給出預測的答案，以識別哪條手臂在另一條手臂下方。

利用體素數據進行 3D 的形狀識別，再應用於 DL 的方法變得愈發流行 [18, 19]。最為明顯的 3D 形狀識別方法，將 VG 與 [18] 提供的監督式 3D CNN 相結合。在 [18] 中，作者介紹了 VoxNet，作為對二進制體素數據的 3D CNN 多類分類任務，具有簡單的網絡架構，可實現實時性能。除了解釋此網絡的優點之外，作者還對各種 3D 數據集上的結果作比較：(i) LiDAR 點雲 [20]，(ii) RGBD 點雲 [21]，以及 (iii) 電腦輔助設計 (CAD) 模型 [22]。

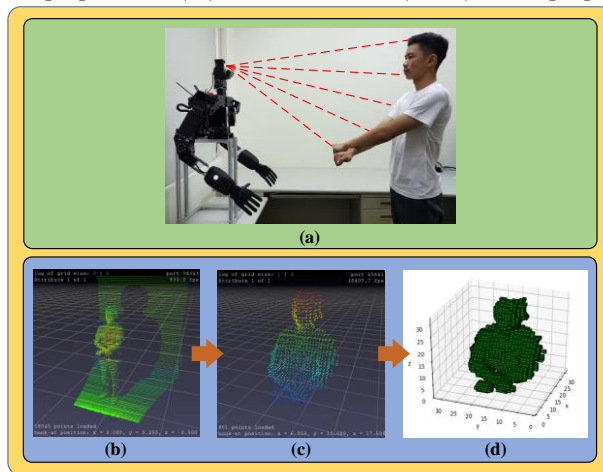


圖 7. 點雲轉換為立體像素表示。

- (a) 掃描過程示意圖。
- (b) 光學雷達掃描結果 (原始點雲數據)。
- (c) 過濾後的點雲數據。
- (d) 立體像素數據表示。

[19] 中則介紹了另一項關於 3D 物件多類分類的研究。在 [19] 中，作者僅使用了標準的 3D CAD 模型數據集 [22]，並為體積數據引入了兩種不同的 CNN 模型 [22]。其中的一種方法，他們提出了一個輕量級的 Volumetric-CNN1 (V-CNN1) 模型。在此方法中，3D 物件的體積被以 2D 空間卷積圖像 (稱為特徵圖) 表示，特別的是，作者沒有使用 3D 卷積層，而是使用 2D 卷積層對 3D 體積 VG 進行卷積，並且訓練的過程更快了 (因為使用了更少的參數)。儘管 2D 卷積在 2D 圖像分類中的表現優於 [9, 10]，但結果還是不如 3D VG [19]。而在 V-CNN1 上，與相同 3D 數據集上的 VoxNet [18] (使用 3D 卷積) 相比，分類精度反而略有下降。關於上一段中提到的兩個模型，它們在 3D 體積 VG 中的多項分類上表現出色，然而，現有模型存在如下缺點。在 VoxNet 模型 [18] 中，作者只考慮了一個非常小的網絡，它只包含兩個 3D 卷積層和兩個全連接層。在這情況下，淺層網絡的架構會導致模型缺乏對數據的泛化 (在各層上學習更多

特徵) [23]。另一方面，在 V-CNN1 [19] 中，作者使用了更深的網絡架構 (5 層)，但未能從 3D 卷積的好處中建立與 3D 數據之間的關係。因此，在本文中，我們提出了一種稱為 Crossed Arm Voxel Network (CAVN) 的新模型，作為一種強大的學習方法來解決之前模型所遇到的問題。CAVN 參考 VoxNet 和 V-CNN1 概念，使用更深的網絡架構以及 3D 卷積層。對於我們的魔術問題，將其輸出簡化為二元分類任務，以哪隻手在另一隻手的下面來進行分類 (見圖 8)。

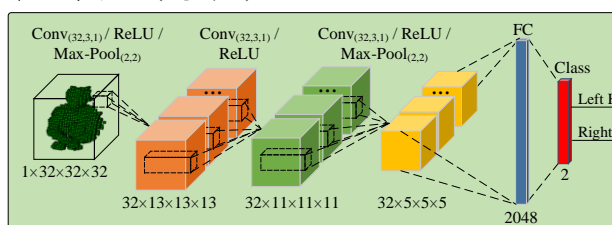


圖 8. CAVN 的網絡架構。

如圖 8 所示，我們提出的 CAVN，使用體素網格來計算，其大小為 $32 \times 32 \times 32$ 。該模型由三個 3D 卷積層組成，全部具有 32 個大小為 3 和跨步為 1 的濾鏡，卷積層如下表格由三個 ReLU 激活函數和兩個最大池化層組成。ReLU 層是通過激活函數在模型中引入非線性，藉此活化模型。ReLU 之後的池化層確保神經元學習中，來自各個局部空間的體素集的冗餘信息不會對模型做出貢獻。模型的最後部分有兩個全連接層，其中最後一個全連接層為 SoftMax 函數分類器，用以歸一化左臂和右臂的概率分佈。在訓練過程中，我們使用概率為 0.8 的 dropout 來防止過度擬合(overfitting)，使用基本學習率為 0.0006 的 Adam 優化器來更新模型權重，CAVN 模型的詳細信息如表 1 所示。

表 1. CAVN 模型架構

Layer type	Filter size / Dropout rate	Stride	Output Size	Number of parameters
Convolution 3D	$3 \times 3 \times 3$	$1 \times 1 \times 1$	$32 \times 30 \times 30 \times 30$	896
ReLU	-	-	$32 \times 30 \times 30 \times 30$	-
Max pooling 3D	$2 \times 2 \times 2$	$2 \times 2 \times 2$	$32 \times 15 \times 15 \times 15$	-
Convolution 3D	$3 \times 3 \times 3$	$1 \times 1 \times 1$	$32 \times 13 \times 13 \times 13$	27680
ReLU	-	-	$32 \times 13 \times 13 \times 13$	-
Convolution 3D	$3 \times 3 \times 3$	$1 \times 1 \times 1$	$32 \times 11 \times 11 \times 11$	27680
Max pooling 3D	$2 \times 2 \times 2$	$2 \times 2 \times 2$	$32 \times 5 \times 5 \times 5$	-
Dropout	0.8	-	$32 \times 5 \times 5 \times 5$	-
Fully connected	-	-	2048	8194048
Fully connected	-	-	2	4098

我們收集了 2000 份不同人的各種類型的交叉手臂數據，並在我們的模型上做訓練（數據集可在 <https://bit.ly/2FFW1CT> 獲得）。由於查看頂部的手臂更容易再實驗時做區分，我們使用這種本能來標記我們收集的每一項數據。在 DL

模型預測的最後，我們多添加了邏輯非來反轉結果並分類哪個手臂在下方。最後，我們在數據集上使用 VoxNet [18]、V-CNN1 [19] 和 CAVN 比較了模型的性能。

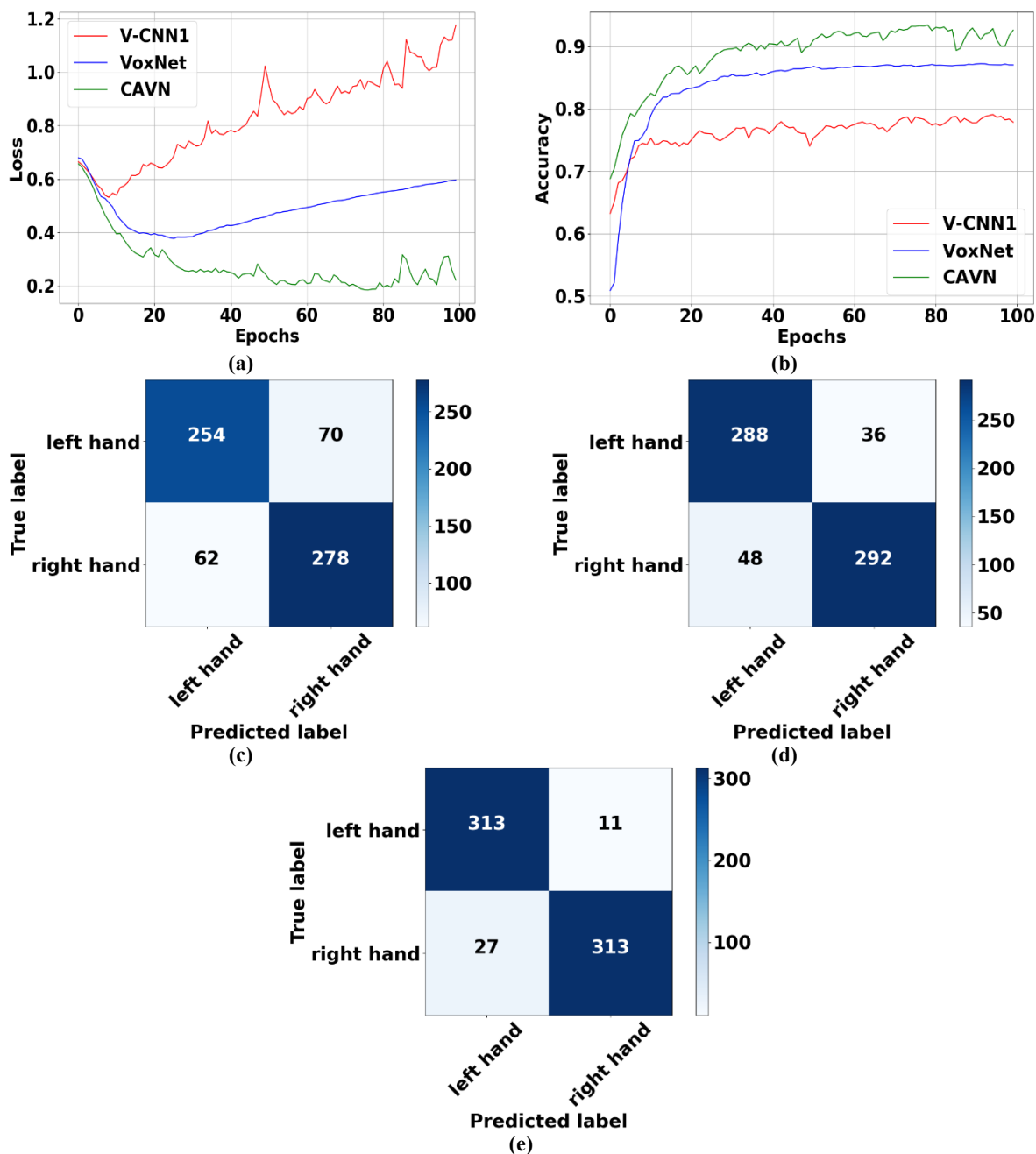


圖 9. 提出的 CAVN 與 V-CNN1 和 VoxNet 模型之間的性能比較。評估結果基於我們的自定義手臂數據集。數據集中訓練進度結果 (a) 損失值。(b) 精準值。混淆矩陣 (c) V-CNN1。(d) VoxNet。(e) CAVN 模型。

4. 結論

THORMANG-Bear 機器人雖然只是 THORMANG3 成人型機器人的上半身改良版，但改良後的機器人其手部部件使它有著強大的功能及可能性。我們提出了一個名為 CAVN 的深度學習模型來對硬幣魔術進行分類，並實時實現了 94% 的準確率。在現場表演的情況下，我們的研究展現了一個穩健的魔術，它可以在沒有人類協助的情況下表演。通過開發魔術表演用的人形機器人，我們還參與支持善待動物組織 (PETA)，停止在娛樂節目（如動物馬戲團）中暴露和虐待動物。

5. 參考文獻

- [1] ROBOTIS Inc. "THORMANG3 Full size open platform humanoid." http://en.robotis.com/model/page.php?co_id=prd_thormang3#(accessed December 19, 2019).
- [2] H. Moon, Y. Sun, J. Baltes, and S. J. Kim, "The IROS 2016 competitions [competitions]," *IEEE Robotics & Automation Magazine*, vol. 24, no. 1, pp. 20-29, 2017.
- [3] J. Baltes, Y. Sun, and H. Moon, "2017 Competitions: Magical, Manipulating, Mercurial Robots [Competitions]," *IEEE Robotics & Automation Magazine*, vol. 25, no. 2, pp. 8-15, 2018.
- [4] K. J. Morris, J. Anderson, M. C. Lau, and J. Baltes, "Interaction and learning in a humanoid robot magic performance," in 2018 AAAI Spring Symposium Series, 2018.
- [5] K. J. Morris, V. Samonin, J. Baltes, J. Anderson, and M. C. Lau, "A robust interactive entertainment robot for robot magic performances," *Applied Intelligence*, vol. 49, no. 11, pp. 3834-3844, 2019, doi: 10.1007/s10489-019-01565-7.
- [6] J. Yang, J. Jeong, and J. Baltes, "Humanoid Robot Magic: various responses and communication," in *The 25th International Symposium on Electronic Art*, Gwangju, Korea, 2019, vol. 25, pp. 73-78.
- [7] "Cool Coin Magic Tricks." <https://www.goodtricks.net/coinmagic.html> (accessed December 26, 2019).
- [8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436-444, 2015.
- [9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
- [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
- [11] J. Baltes, K.-Y. Tu, S. Sadeghnejad, and J. Anderson, "HuroCup: competition for multi-event humanoid robot athletes," *The Knowledge Engineering Review*, vol. 32, 2017, doi: 10.1017/S0269888916000114.
- [12] L. Yu-Ren, G. H. Galelli Christmann, R. B. Grando, R. Da Silva Guerra, and J. Baltes, "Dr. Eureka: a humanoid robot manipulation case study," *The Knowledge Engineering Review*, vol. 34, p. e27, 2019, Art no. e27, doi: 10.1017/S0269888919000213.
- [13] G. H. Galelli Christmann, L. Yu-Ren, R. da Silva Guerra, and J. Baltes, "Can a Robot Shoot an Olympic Recurve Bow? A preliminary study," presented at the The 16th International Conference on Automation

- Technology Paper Award Competition, Taipei, Taiwan, 2019.
- [14] C. G. Atkeson et al., "What Happened at the DARPA Robotics Challenge Finals," in *The DARPA Robotics Challenge Finals: Humanoid Robots To The Rescue*, M. Spenko, S. Buerger, and K. Iagnemma Eds. Cham: Springer International Publishing, 2018, pp. 667-684.
- [15] Seed Robotics Inc. "RH8D Human-inspired, Adult-size Robot Hand." <https://www.seedrobotics.com/rh8d-adult-robot-hand> (accessed December 19, 2019).
- [16] C. Ramsay. "Which hand is the coin in? - Learn How to find it with body language!" <https://www.youtube.com/watch?v=qvPJokZtwak&t=22s> (accessed December 19, 2019).
- [17] S. Jianbo and Tomasi, "Good features to track," in 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 21-23 June 1994 1994, pp. 593-600, doi: 10.1109/CVPR.1994.323794.
- [18] D. Maturana and S. Scherer, "Voxnet: A 3d convolutional neural network for real-time object recognition," in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015: IEEE, pp. 922-928, doi: 10.1109/IROS.2015.7353481.
- [19] V. Hegde and R. Zadeh, "Fusionnet: 3d object classification using multiple data representations," arXiv preprint arXiv:1607.05695, 2016.
- [20] M. De Deuge, A. Quadros, C. Hung, and B. Douillard, "Unsupervised feature learning for classification of outdoor 3d scans," in *Australasian Conference on Robotics and Automation*, 2013, vol. 2, p. 1.
- [21] S. Song, S. P. Lichtenberg, and J. Xiao, "Sun rgb-d: A rgb-d scene understanding benchmark suite," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 567-576.
- [22] W. Zhirong et al., "3D ShapeNets: A deep representation for volumetric shapes," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7-12 June 2015 2015, pp. 1912-1920, doi: 10.1109/CVPR.2015.7298801.
- [23] B. Chakraborty, B. Shaw, J. Aich, U. Bhattacharya, and S. K. Parui, "Does Deeper Network Lead to Better Accuracy: A Case Study on Handwritten Devanagari Characters," in 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), 24-27 April 2018 2018, pp. 411-416, doi: 10.1109/DAS.2018.72.
- [24] D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar, and A. I. Rudnicky, "Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices," in 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, France, 2006, vol. 1:IEEE,pp.I-I,doi: 10.1109/ICASSP.2006.1659988.

6. 致謝

本研究部分由國立台灣師範大學“國語教學中心”和“高等教育深耕計劃”、台灣教育部、台灣科技部資助，以下為科技部授權碼，109-2221-E-003-019 - & 109-2634-F-003 -009 - 。