

專題報告摘要

考生姓名： 徐顥瑄，畢業學校科系組： 國立陽明交通大學 機械工程學系

1. 專題名稱： MVLA: Toward Pre-Contact Material-Aware Force Regulating for Vision-Language-Action Models

指導教授： 王傑智，專題成績： 非正式選課

參與成員： 獨力完成

專題研究期間： 2024 年 9 月至 持續研究中

2. 研究內容簡介：(限300字以內)

現行 Vision-Language-Action (VLA) 模型雖已應用於機器臂與人形機器人操作，但在高接觸性任務中仍面臨挑戰：依賴被動位置控制，無法於接觸前預測材料特性或調節接觸力。近期如 ForceVLA 雖提升反應能力，仍需物理感測器輔助，且僅能於接觸後調整。

本研究提出 Material-Vision-Language-Action (MVLA) 模型，具備類人接觸前認知能力，可透過視覺進行材料分類與物理特性估算，預測所需接觸力，實現預測性力調節。實驗顯示，MVLA 在處理易碎物體時表現優異，如下方 Fig3、Table III 所示。MVLA 在 **辨識準確率與特定任務表現皆優於 OpenVLA(SOTA)**。我於大學階段以此為起點，結合多模態感知、語意融合與模擬決策，透過跨域課程與實作，逐步建立感知與推理能力，為未來推進具人類認知的機器人系統奠定基礎。

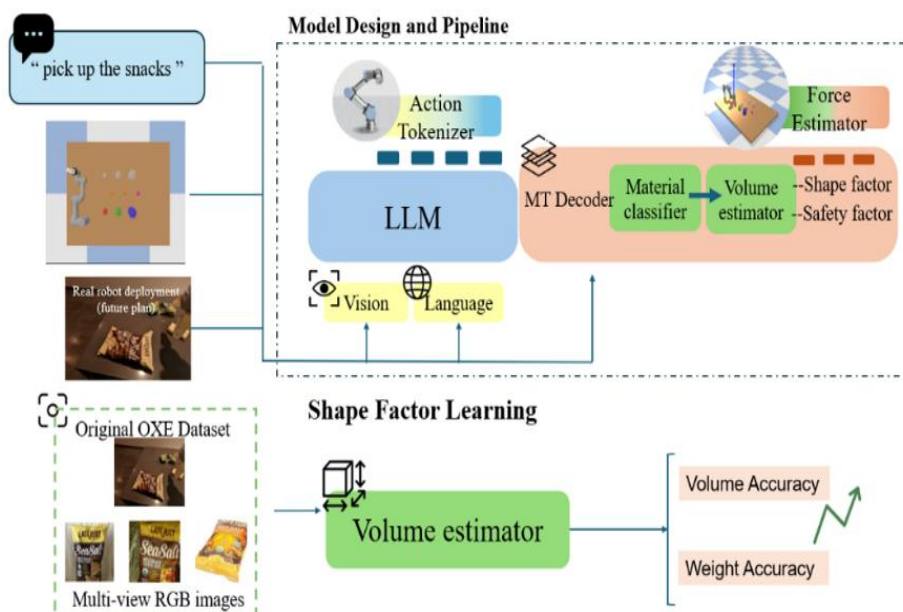


Fig.1 MVLA 系統架構圖

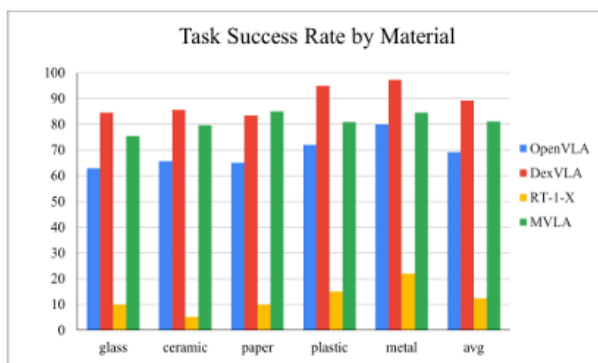


Fig. 3: Comparison of Task Success Rates Across Different Object Materials for the Models

TABLE III: Damage Rate comparison among 5 materials: glass, ceramic, paper, plastic, and metal

| Model | glass | ceramic | paper | plastic | metal | avg |
|---------|-------|---------|-------|---------|-------|--------|
| OpenVLA | 28.3% | 25.4% | 24.4% | 15.2% | 8.7% | 20.4% |
| DexVLA | 13.7% | 13.3% | 9.2% | 7.2% | 7.5% | 10.18% |
| RT-1-X | 64.3% | 59.2% | 65.4% | 53.4% | 33.2% | 55.1% |
| MVLA | 11.7% | 11.2% | 12.2% | 9.7% | 7.2% | 10.4% |

Fig. 3、Table III 為目前 MVLA 計畫與 SOTA 的表現比較

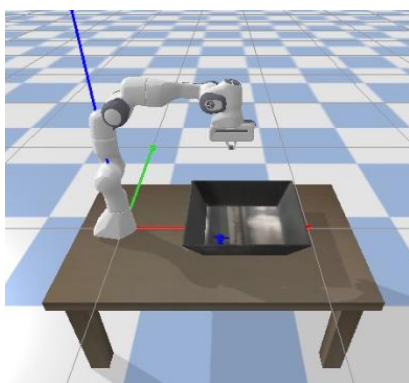


Fig. 4 初步以 Pybullet 檢查手臂在 MVLA 的基本夾取動作

3. 貢獻（參與）程度（若有多人一起研究，說明分工情形）：
本研究由本人獨立完成，涵蓋模型設計、模組整合、實驗規劃與結果分析，貢獻度達100%。

主要工作項目：

- 設計 VLA 模型的嵌入架構，修正訓練集
- 建立**模態感知融合**機制（視覺、語言、材質辨識）
- 整合開源 VLA 模型，建構**可微調訓練框架**
- 規劃並模擬多元場景下的任務實驗(參考 Fig. 5、Fig. 6)
- 分析系統在不同精細任務的安全性與效率表現



Fig. 5 在 Nvidia Isaac Lab 環境中模仿倉庫情景(手臂為 Franka Panda)



Fig.6 在 Nvidia Isaac Lab 環境架設廚房情景，使其更貼近居家

4. 研究心得：(限300字以內)

本研究始於對語言模型與機器人技術的初步探索，隨著研究深入，我逐步建立起**跨模態整合**的系統性架構，並在實作中不斷驗證理論與方法的可行性。從材料辨識到力調節模組的設計，每一階段都伴隨著技術挑戰與邏輯推演，幫助我培養出**獨立思考與問題拆解**的能力。尤其在模擬環境中進行接觸前力預測時，我學會如何在不依賴感測器的前提下，透過視覺與物理推論達成精準控制，這不僅提升了我對機器人操作安全性的理解，也拓展了我對**人工智慧應用邊界的認知**。

此外，在撰寫報告與規劃投稿過程中，我更體會到研究不僅是技術的累積，更是學習**溝通、表達與持續精進**的過程。這段歷程讓我深刻理解「研究」的本質是探索與突破，也奠定了我未來在電控領域持續發展的信心與動力。

5. 其它說明（後續參加比賽、論文發表、入圍得獎等）：

- 本研究已以前瞻性概念投稿至 ICRA 2026，並規劃進一步優化模型性能，期望於 IROS 2026呈現更完整成果。
- 本研究主要是訓練 VLA 在接觸前預測材料特性或調節接觸力，這是 VLA 模型理解物理世界的開始，在實際的應用，VLA 模型會面臨更複雜或是陌生的環境，以 World models 來增強 VLA 對物理世界的認識，會是未來研究的主題

專題報告指導老師說明函

被指導的學生姓名：徐顥瑄

被指導的學生專題報告名稱：MVLA: Toward Pre-Contact Material-Aware Force Planning for Vision-Language-Action Models

1. 學生做專題的起始時間

113年度上學期至今

2. 學生對此專題報告的貢獻程度

學生自113年度上學期起主導本專題，展現出高度的自主研究能力與技術創新精神，在研究構想、技術實作、方法設計與學術表達等層面皆有卓越貢獻。

首先，在研究構想與問題定義方面，學生主動提出以視覺語言動作

（VLA）模型為核心的研究方向，針對機器人在接觸前階段的力規劃挑戰進行深入探討，展現對觸覺感知與語意理解整合的前瞻性思維。在技術整合與實作層面，學生獨立完成 MVLA 模型的設計，並成功整合至 Nvidia Isaac Sim 模擬環境，同時規劃感測器配置與資料收集流程，展現高度的系統整合能力。

進一步地，他在專題中引入具材料辨識能力的視覺語言模型，並結合力控制策略，提出具備預測性與適應性的操作規劃機制，具原創性與實務應用潛力。此外，學生具備良好的中英文技術溝通能力，能清楚表達研究動機、方法與成果，並積極參與指導討論與技術文件撰寫，展現成熟穩健的研究態度。

整體而言，學生在本專題中扮演關鍵角色，其多面向的貢獻有效推動研究進展，並為團隊注入創新能量。

3. 學生參與指定的研究課題深入程度

在參與指定研究課題的過程中展現出高度投入與深入參與的態度，在理論深化、技術實作與跨域整合等層面皆有卓越表現。

他不僅充分理解研究核心目標，更主動延伸問題範疇，提出「Pre Contact Material-Aware Force Planning」的創新子題，強化 MVLA 模型

在實體操作前的感知與規劃能力，展現出前瞻性的研究視野。

在跨領域整合方面，學生能有效融合視覺語言模型、力控制策略與模擬環境，掌握機器人感知、語意理解與動作規劃的核心技術，並針對材料特性進行語意引導與力道調整，展現深厚的跨域能力。技術實作層面亦表現亮眼，深入參與 Isaac Sim 模擬環境建置與 MVLA 模型部署，涵蓋感測器配置、資料收集流程設計及雲端資料自動化整合，具備完整的系統建構能力。

在理論理解方面，學生對 VLA 模型架構與演化具深入掌握，能進行模型分析與優化，並提出具體改進建議。此外，他持續追蹤相關領域最新研究進展，並將所學應用於專題中，展現穩定且積極的研究節奏，為專題成果累積奠定堅實基礎。

4. 其他（若有不足，請惠予另紙說明）

具備高度的研究熱忱與自我驅動能力，能在指導下獨立完成跨模態模型整合與模擬實驗設計，並展現優異的系統思維與問題解決能力。其專題成果具備延伸為研究論文或實體驗證計畫之潛力，亦可作為進一步攻讀或參與國際研究合作之良好基礎。整體而言，學生在本專題中表現傑出，已展現具備成為優秀研究人才之潛力與特質。

王傳智

2025/11/20