

國立臺南大學

資訊工程學系

碩 士 論 文

藉由合成資料訓練深度網路進行無紋理物體姿態估測

Training Deep Networks with Synthetic Data for Textureless
Object Pose Estimation

指導教授：朱明毅 博士

研 究 生：賴靖融

中 華 民 國 一 零 九 年 一 月

藉由合成資料訓練深度網路進行無紋理物體姿態估測

Training Deep Networks with Synthetic Data for Textureless
Object Pose Estimation

By

Jing-Rong Lai(賴靖融)

A thesis submitted in partial fulfillment of the requirements

for the Master of science degree

in Computer Science and Information Engineering

in the College of Science and Engineering

at the National University of Tainan

Tainan, Taiwan

Advisor: Dr.Ming-Yi Ju(朱明毅 博士)

January 2020

中華民國一〇九年一月

國立臺南大學碩士學位論文考試

審定書

資訊工程學系碩士班

研究生 賴靖融 所提之論文

藉由合成資料訓練深度網路進行無紋理物體姿態估測

經本委員會審查，符合碩士學位論文標準。

學位考試委員會

召 集 人 黃國勝 簽章

委 員 陳昱仁

朱明毅

指 導 教 授 朱明毅 簽章

系 主 任 高學財 簽章

中華民國 109 年 1 月 2 日

藉由合成資料訓練深度網路進行無紋理物體姿態估測

學生：賴靖融

指導教授：朱明毅 博士

國立臺南大學資訊工程學系

中文摘要

目前的產業型態逐漸轉向工業 4.0，而當中機器視覺結合機器學習演算法驅使機器手臂完成任務是發展智慧工廠的核心技術之一，在生產線上我們需要頻繁的移動機器手臂對生產中的物體進行搬移或是組裝，在傳統的解決方案中，往往需要利用導軌及制具限制物體使演算法能夠正確計算，並且這些影像處理演算法很容易受到光源等因素干擾，不過近年來深度學習方法在圖像處理上的優秀表現為上述問題帶來了新的可能性，我們提出利用工業中常用的電腦輔助設計格式產生模型，產生出大量有著姿態標籤的合成資料，減輕深度學習中在姿態估測上手動標記姿態標籤的負擔，並針對無紋理目標提出一個深度網路架構來估測工作場景中物體的 3D 姿態，用來指導機器手臂抓取目標。

關鍵詞: 深度學習、卷積神經網路、空間姿態估測、視點估測

Training Deep Networks with Synthetic Data for Textureless Object Pose Estimation

Student: Jing-Rong Lai

Advisor: Dr.Ming-Yi Ju

Department of Computer Science and Information Engineering National
University of Tainan, Taiwan (R.O.C).

ABSTRACT

As the manufacturing today has moved toward the Industry 4.0 era, the integration of machine vision with machine learning algorithms for controlling robot arms to accomplish various tasks is one of the core technologies for the development of smart factories. Robot arms usually carry out production or assemble parts in automated production lines. To ensure the manufacturing system works fine, the use of guide rails and jigs to provide constraints on relative workpiece poses are necessary for traditional solutions. Besides, interference from external light sources also affects the correctness of image processing. Fortunately, state-of-the-art deep learning methods have shown excellent performance in the field of image processing and bring hope to overcome the above problem. The proposed data generating method leverages the computer-aided design data and model to generate a large number of labeled synthetic data for training deep neural networks to alleviate the burden of obtaining hand-labeled datasets. A deep network structure is also proposed to estimate the 3D poses of textureless objects in the workspace to guide robot arms for workpieces grasping.

Keywords: Deep learning, Convolutional neural network, 3D pose estimation, Viewpoint estimation

目 錄

論文摘要	I
Abstract	II
目 錄	III
表目錄	VI
圖目錄	VII
1 緒 論	1
1.1 研究背景	1
1.2 研究目標	2
1.3 論文架構	3
2 相關研究	4
2.1 傳統方法	5
2.2 深度學習方法	5
2.2.1 分類方法	5
2.2.2 迴歸方法	6
2.3 相關研究總結	7
3 研究架構	9
3.1 CAD 模型模擬器	10
3.1.1 光源及陰影	11
3.1.2 模型材質及背景	11
3.1.3 CAD 模擬器介面	12

3.2	異質編碼器架構	14
3.3	多階段姿態預測	17
3.3.1	物體定位網路	18
3.3.2	類別預測網路	20
3.3.3	特徵圖切割	21
3.3.4	類別決策	22
3.3.5	平面旋轉估測網路	23
3.3.6	姿態估測網路	24
3.4	模型訓練	25
3.4.1	訓練資料產生	26
3.4.2	訓練標籤前處理	27
3.4.3	模型訓練流程	28
3.5	姿態預測流程	30
4	實驗結果	32
4.1	異質編碼器架構濾除陰影效果	32
4.2	多階段預測之必要性	34
4.3	切割特徵圖的效益	36
4.4	整體模型預測實驗	37
4.4.1	定界框測試結果	39
4.4.2	平面旋轉估測網路測試結果	40
4.4.3	姿態估測網路結果	42
5	結論與未來展望	44

5.1 結論	44
5.2 未來展望	45
參考文獻	46
附錄一：真實測試資料產生流程	49