

深度學習應用於機器人視覺之最佳化(2/5~5/5)



計劃主持人：賴尚宏

共同主持人：許秋婷、邱瀟德、李哲榮、周志遠、李濬屹

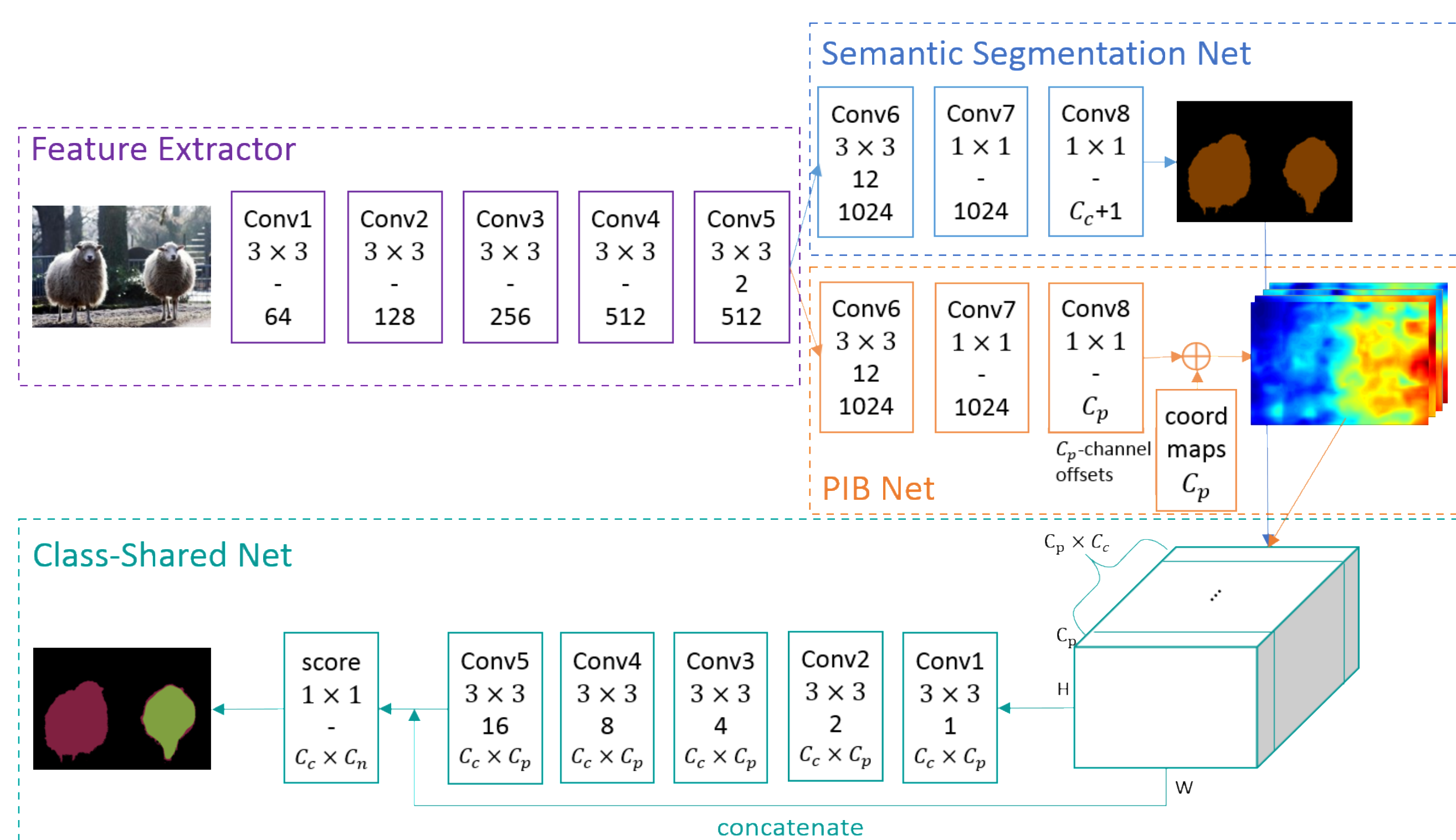
執行單位：國立清華大學 資訊工程學系

計劃摘要

隨著近來大型資料庫的取得越來越普及，深度學習的技術在人工智慧許多領域已獲得革命性的進展。本計畫之研究主題為研究深度學習技術在機器人視覺應用，將著重在智慧機器人視覺的相關應用，包括場景及人物的辨識、動作及異常事件的偵測，我們不僅研究深度學習應用至智慧機器人所需的影像視訊辨識技術，我們也將研究如何利用深度學習進行三維場景重建及三維定位的技術。為了解決深度學習在網路模型訓練及實際應用時所需大量的計算，我們在此計畫也會研發如何更有效利用GPU運算來加速深度學習網路中大量的卷積運算的效率以及利用雲端平台的平行運算架構來加速深度學習中的訓練過程，另外，我們也會研究適用於電腦視覺的深度學習網路的高效能硬體架構，並考量系統低功耗的要求，從適用於深度學習網路之嵌入式系統的角度以及從低功耗晶片設計的角度進行研發。

場景辨識與三維重建定位

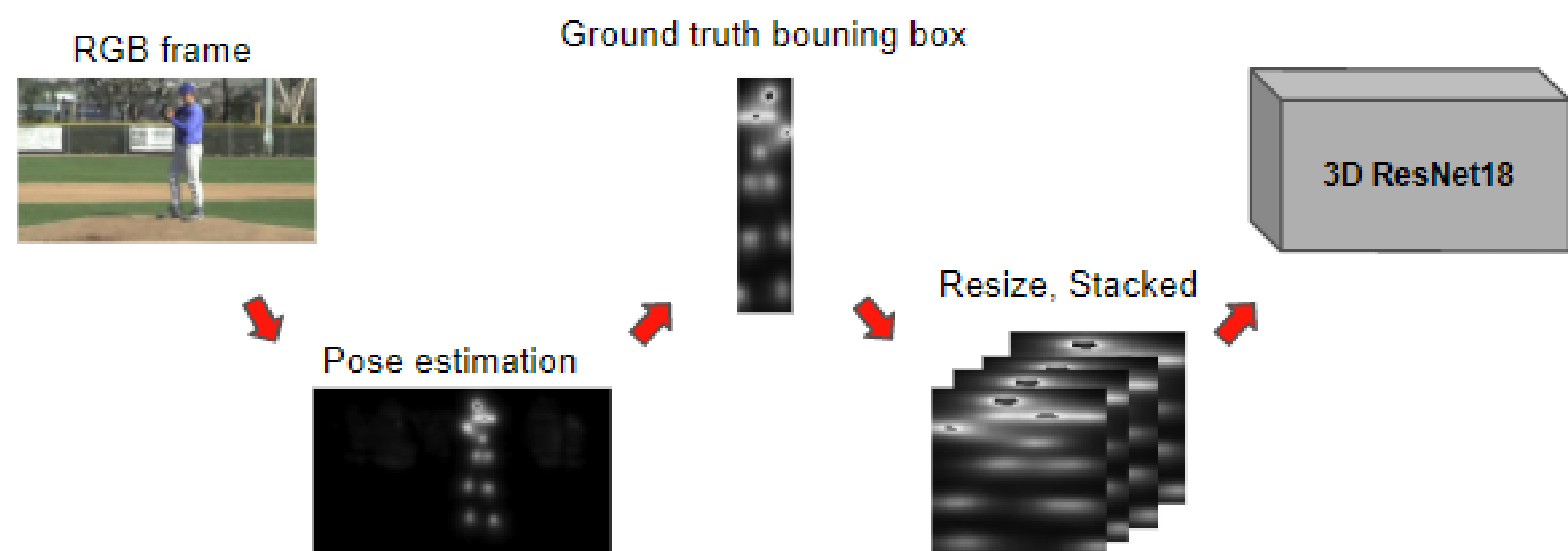
在機器人視覺應用中，為使移動機器人藉由電腦視覺技術，以了解周圍的場景與其中的人物行為，本計畫擬藉由近年來深度學習在電腦視覺領域的進步，透過語意分割 (Semantic Segmentation) 技術，自動地進行場景的分析與辨識。同時也需要透過 SLAM (Simultaneous Localization and Mapping 同步定位與地圖建構) 技術的支持，讓機器人或是無人機可以在未知環境中運動與定位。上個年度的計畫中，本團隊也針對場景辨識提出一個不依賴 bounding box 或 object proposal 的 instance segmentation 方法，並且設計了新的深度學習網路去結合 instance segmentation 與 PIB 的特徵，此模型在 MSCOCO 與 Pascal VOC2012 dataset 上都能達到 state-of-the-art 的表現。



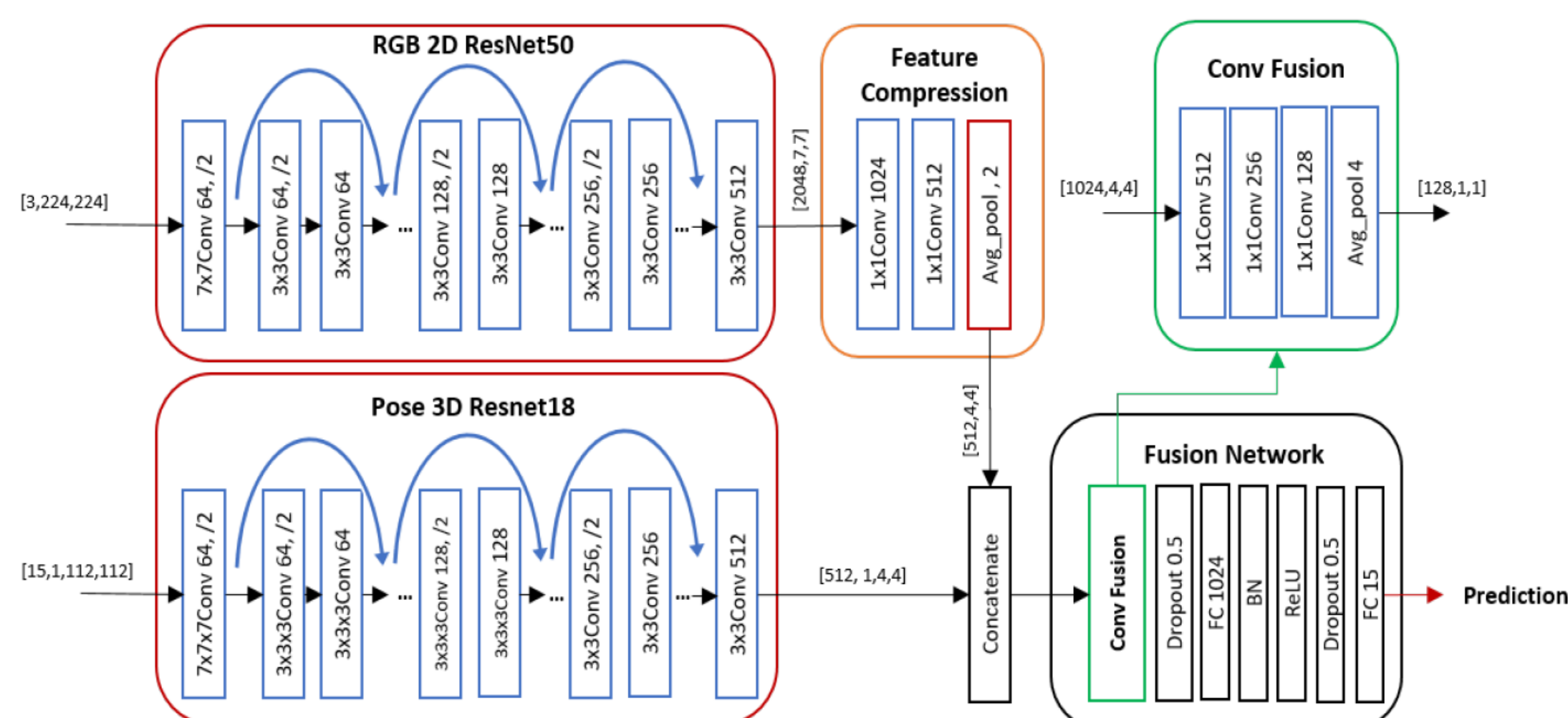
本團隊提出的 instance segmentation 方法

人物與行為辨識

要達到機器人視覺中人物辨識的功能，基本系統需要有物體偵測、以及人的身分及行為辨識等功能。機器人要能進行人物的辨識，首先要從影像偵測出人與物體的位置，深度學習方法對於物體偵測的精確度及數度有很顯著的提升，近年來較具代表性的方法如 Faster R-CNN[1]、YOLO[2]等。不少研究也將深度學習用於動作辨識，各方摸索如何取得更好的辨識成功率或者增進辨識的效率，例如 FastTAP[3] 和 TURN TAP[4]。近期本團隊在動作辨識上也有了新的成果，我們開發了新的 Deep Learning Model，其使用 3D CNN 來辨識影像中的 temporal human pose feature，同時我們也提出了新的 Multi-dimensional Fusion Network，準確度在 Sub-JHMDB 和 PennAction dataset 上都有很好的表現。



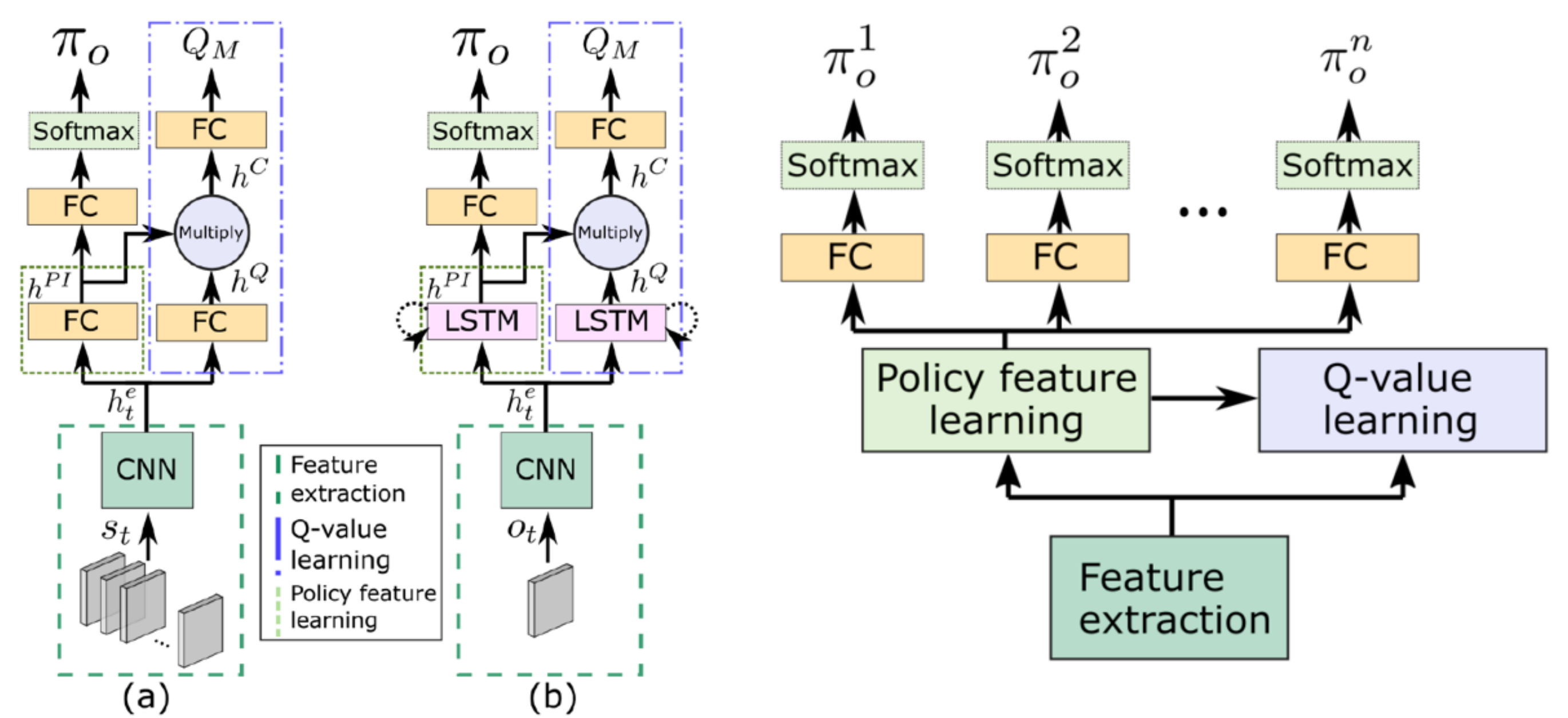
Pose-Based 3D CNN model 流程圖



Multi-dimensional Fusion Network 流程圖

創新深度學習技術

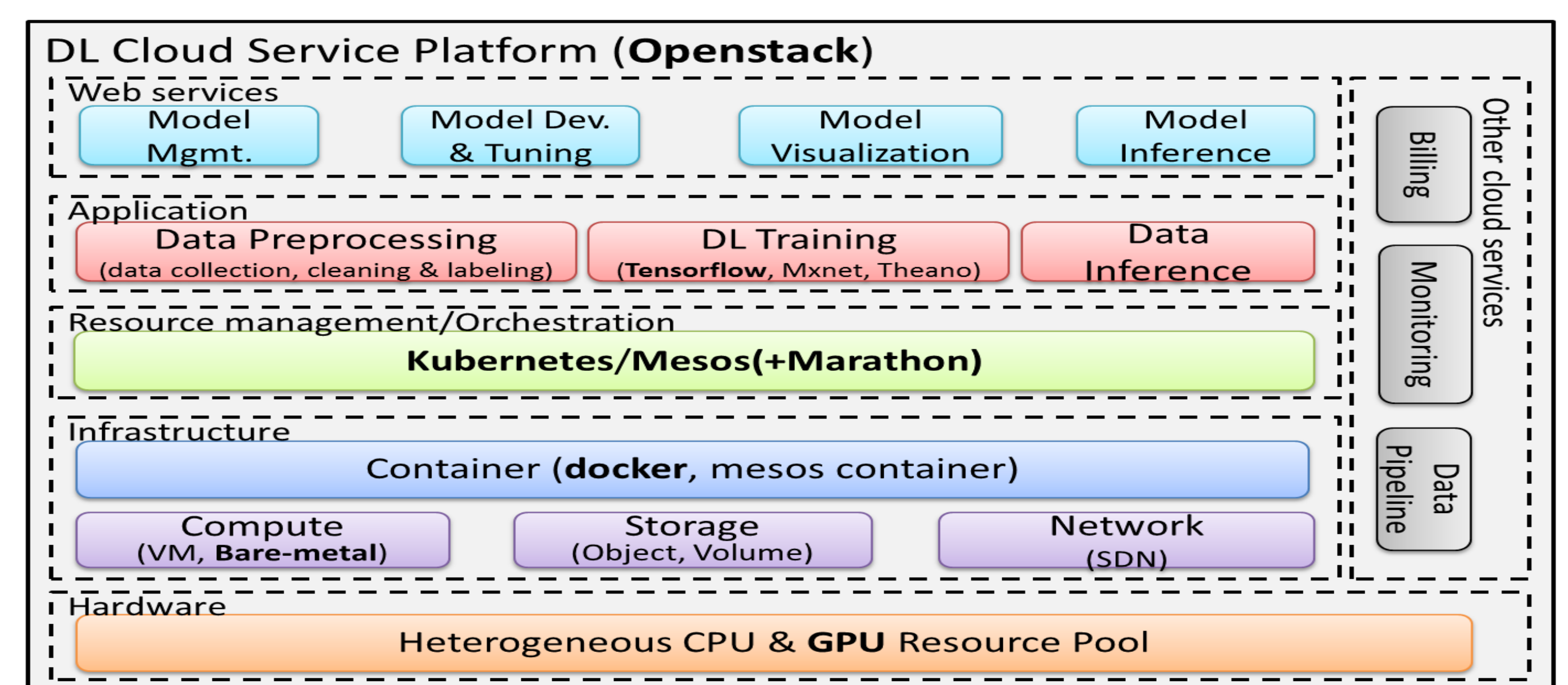
本計畫中擬針對有標記的訓練影像資料取得不易的限制，探討幾個研究方向，包括了弱監督式學習之語意分割、非監督式域自適應學習、資料擴增，和基於生成對抗網路之額外模擬資料。同時我們在此計畫預計開發多機器人協同合作技術，初步研發的系統架構圖如下圖所示，本系統架構名為 Deep Policy Inference Q-Network (DPIQN) 及 Deep Recurrent Policy Inference Q-Network (DRPIQN)。



DPIQN 與 DRPIQN 系統架構圖

深度學習計算優化

因為行動機器人本身不能裝備太大的計算資源，而在目前的主流深度學習架構下，即便是推論的模型也需要一定程度的運算能力，才能滿足行動機器人的即時運作能力。深度學習的訓練速度或是推論速度一直都是深度學習中的一大問題。新的運算架構讓深度學習的速度有顯著進步，例如 Nvidia 和 Google 提出的 TPU 架構等；然而即便有了 TPU，深度學習整體的運算還有很多可以加速的空間，雖然人工神經網路模型的開發難度因此減低，但是模型的訓練仍然需要大量的計算資源和時間，而且整個應用開發過程還包含了除了模型設計與訓練之外的許多其他計算服務需求。因此一般開發者仍遇到計算資源不易取得或管理，且欠缺完整應用開發環境的問題，本計畫的目的是掌握利用開源碼建立一套深度學習雲計算服務平台的關鍵技術，研究分散系統與虛擬化資源下深度學習計算的服務部署及資源分配管理問題，藉此提升系統整體資源使用率和計算效能，並降低使用者計算成本和時間。



深度學習雲計算服務平台架構圖

預期貢獻

- 本計畫研發的智慧視覺辨識技術是許多智慧機器人及視訊監控應用的底層關鍵技術，可經由技轉或產學合作協助國內產業界開發智慧視訊應用產品，預計在計畫四年期間與8家公司進行技轉或產學合作。
- 本計畫研發多項深度學習技術並應用至智慧機器人視覺的實際問題，預期對學術界可產生重要的貢獻。
- 本計畫研發的視覺深度學習技術，可以作為底層模組，提供業界開發相關的前端應用。
- 本計畫整合基礎數學、電腦視覺、機器學習、晶片設計與計算系統研究，培育相關人才。
- 本計畫研發的雲平台系統，可以作為業界成立公司，或是提供相關服務的系統。

參考文獻

- [1] S. Ren, K. He, R. Girshick, and J. Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks". In Advances in neural information processing systems, pages 91–99, 2015.
- [2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. "You only look once: Unified, real-time object detection". arXiv preprint arXiv:1506.02640, 2015.
- [3] FastTAP: Fast Temporal Activity Proposals for Efficient Detection of Human Actions in Untrimmed Videos, CVPR2016
- [4] TURN TAP: Temporal Unit Regression Network for Temporal Action Proposals, arXiv2017