

A Study on Active Learning for Improving Object Detection Model in Home Care System Using Limited Amount of Data

利用主動學習在有限資料下優化

居家照護系統的目標檢測模型之研究

組別：A85

指導教授：孫民

組員：劉亦傑、謝霖泳

Abstract

In recent years, with the development of technology and medicine, the population of the elderly keeps increasing. The problem of taking care of the elderly living alone gradually arises. Therefore, our lab founded CarePLUS Group with the aim to develop a smart home care system, monitoring the action of the elderly and detect whether they are in danger using computer vision. Besides, the home care system will remind them to take medicine and do exercise regularly. By doing so, we can accompany the elderly living alone using the power of AI instead of labor.

In CarePLUS Group, we are responsible for preprocessing training data, and our goal is to improve the performance of object detection model by finding out frames that are more valuable. With a limited amount of data, we construct “Active Sample” algorithm by Active Learning to achieve this goal. By the videos recorded by the camera and the prediction result from the detection model, it is possible to strengthen some parts that the model tends to misjudge.

The ultimate goal of Active Sample is to enhance the ability of our model to distinguish humans. In this study, we observe the effect of our algorithm by building Detectron2[11] developed by Facebook AI.

The result of our experiments reveals that sampling the frames merely with the change of light is not sufficient. Thus, we take the area of bounding boxes into consideration. The accuracy of revised method increases by 47% in comparison of the baseline model. From the perspective of efficiency, the number of frames needed for our revised algorithm is only 2.5% of the original one, which greatly reduce the cost of data labeling.

In conclusion, we can improve our model efficiently and successfully by finding out the frames using Active Sample. In the future, we expect that our algorithm can be not only applied in offline use but in real time detection.

一、前言

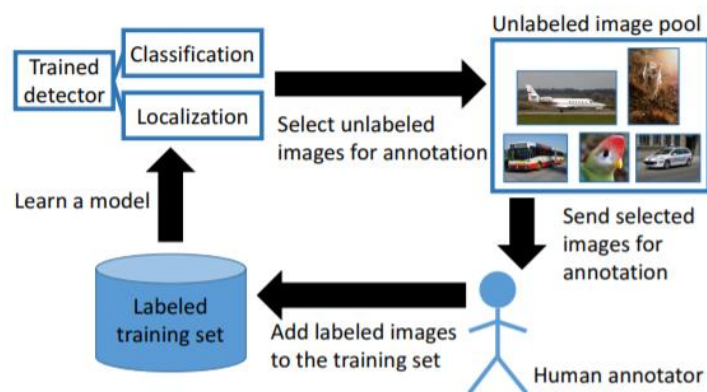
在科技日益發達的今天，人們的平均壽命越來越長，老年人口的比例也越來越多，因此「獨居老人」的照護問題逐漸浮上檯面。但不管是請看護照顧老人，抑或是把他們送到養老院，這些成本都很高，每個月都需要付出一筆龐大的費用。如果單純用監視器觀察老人的行為，可能也要有人 24 小時不間斷地盯著看，否則意外發生時仍難以提供及時協助。

於是我們的實驗室以「智慧居家照護」為宗旨成立了"CarePLUS"團隊，在使用者家中裝上鏡頭，配合深度學習的技術即時監測老人是否有危險，並以聲音互動的方式向老人詢問是否需要幫助。舉例來說，當偵測到老人跌倒或滑倒時，趕快詢問他是否需要幫助；當偵測到老人久坐一整天時，也會發出聲音提醒他起來動一動，並傳送通知給老人的子女讓他們充分了解長輩的健康狀況。

我們參考了[1][2][3][4][5]的作法，並融合這幾篇論文中的概念來發想。利用主動學習(Active Learning)讓模型進入新環境後，可以在更短時間內或更少的資料量下有顯著的進步。而主動學習就需要從現有的資料中，挑出模型比較不會的地方，讓模型能加強學習效率，因此我們定義了 Active Sample 的演算法，去挑出現有資料中的偽陽性(false positive)和偽陰性(false negative)的影格，並重新將這些相對重要的影格丟回模型中，因為對模型而言連續的影片中並不是每張影格都能對模型有顯著幫助，特意挑出影格就能在較少的資料量下，以較高的效率去增強模型尚未學好的部分。

二、原理分析與系統設計

在許多機器學習的領域範疇中，最珍貴的往往就是標記資料所消耗的人力。為了節省這些曠日廢時的人力成本，這次專題中，我們主要用到的方法為主動學習(Active Learning)，它的精神就是挑出模型比較學不會的地方，經過人工 label 之後，讓模型去學習，圖一是從[3]中擷取出來的，說明了主動學習的過程。藉由主動學習，我們就不用將所有資料都進行 label，也就節省了許多人力成本。至於挑出模型比較學不會的部分，我們使用的方法稱為"Active Sample"，也就是「主動抽樣」，挑出影片中模型容易誤認的 FP 和 FN，讓模型加強學習。



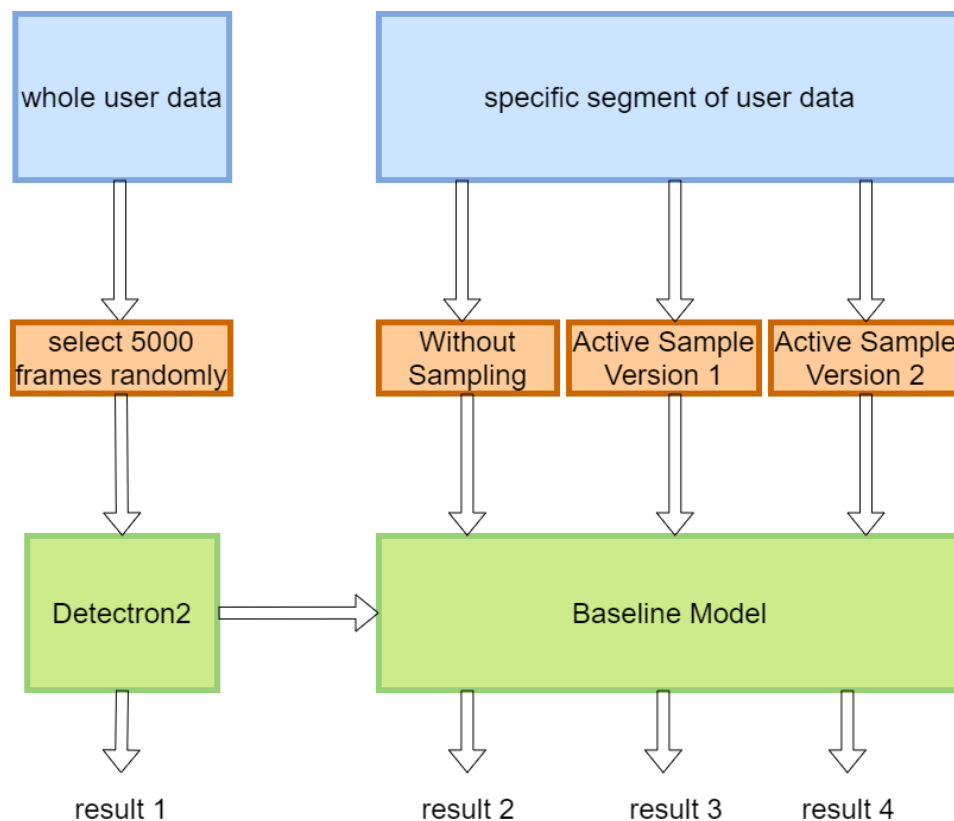
圖一：主動學習流程圖

False Positive(偽陽性，以下簡稱 FP)和 False Negative(偽陰性，以下簡稱 FN)是兩個我們主要會探討的問題，因為這些 FP 和 FN 就是模型比較弱、比較容易犯錯的地方，也就是在 Active Learning 的過程中，需要被 sample 出來給模型學習的地方，因此先定義 FP 和 FN。

FP 就是模型原先預測結果認為有人(positive)，但實際上卻沒有人的情況。首先我們的演算法假設：「人會動，所以人才會造成動作(motion)，而物品都是靜止的。」因此「動作」在演算法判斷有沒有人的過程中扮演著關鍵的角色，至於實際上運用的方法主要是看這個影格的 motion 和前面連續 500 個影格的平均去比較，利用每兩張影格間 pixel 的差異，只要某影格前面連續很多個影格都是幾乎靜止的，但此時突然有很大的 motion，演算法就會判定此影格有人(positive)。

相較於 FP，FN 就是模型認為沒有人，但實際上卻有人的情況。會造成 FN 的主要原因就是人的 motion 太低，導致模型把人和背景融為一體了。因為 CarePLUS 的宗旨就是照顧老人，而老人常常會在電視機前面久坐一整天，這個期間因為幾乎都沒有 motion，因此模型就會認為這是沒有人的情況，稱為 FN。

我們認為 FN 的問題比較需要先被解決，於是多加上「motion 的 response 在圖片上造成的大小」來判斷這張 FN 是否 sample 正確的標準。因此，我們嘗試去篩選這些原始 FN 的資料，在改良版本的演算法設計上多判斷一項面積閾值，判斷藍色框面積符合前述條件的圖片記下來，並在後續的討論中與原始版本的演算法比較其效果差異。



圖二：系統流程圖

圖二是我們整個系統的流程。有了前述 Active sample 演算法後，我們就需要自行訓練出目標檢測模型來進行評估。在這次的專題中，我們所挑選的是 Faster RCNN[7]，背後是基於 ResNet-101[8]的架構，並加上[9]FPN(Feature Pyramid Network)來比較優秀地提取出不同 scale 下的特徵。雖然原始的 Faster RCNN 論文中並無使用 FPN 這項技巧，但在速度和準確度 trade off 之觀點下，FPN 是能有相當不錯的表現，因此選用 FPN 的架構來進行實作，本文將稱此為 Faster RCNN R101-FPN 之模型。

實作上會先載入已經在 COCO dataset 上進行過一輪預先訓練(pre-train)的 Faster RCNN R101-FPN，但經過 COCO 訓練並不會有對於魚眼視角有足夠的辨識能力，因此會先利用部分魚眼資料進行 pre-train，來讓模型先能夠對於魚眼視角的狀況能有基本的辨識能力，也就是建立出一個屬於我們的 baseline 模型。我們用來訓練的資料為 30 分鐘的影片，共分成以下三種情形。

1. 直接將 30 分鐘影片當做新的訓練集放回 baseline 訓練
2. 訓練在本影片經過原始之「Active Sample」演算法所挑選的影格
3. 訓練在本影片經過改良後之「Active Sample」演算法所挑選的影格

以上這三部分，都將會在 baseline 模型上額外去訓練 200 個 iterations，並將於後續實驗結果的部分來和 baseline 模型（也就是總共四種情況）去比較、探討。

三、實驗結果

	Baseline	Without Sampling	Active Sample (ver.1)	Active Sample (ver.2)
AP	0.493	0.699(+41.8%)	0.661(+34.1%)	0.727(+47.5%)
AP ₅₀	0.740	0.950(+28.4%)	0.915(+23.6%)	0.974(+31.6%)
AP ₇₅	0.676	0.914(+35.2%)	0.886(+31.1%)	0.936(+38.5%)
AP _m	0.483	0.685(+41.8%)	0.649(+34.4%)	0.711(+47.2%)
AP _l	0.681	0.810(+18.9%)	0.781(+14.7%)	0.853(+25.3%)
AR	0.709	0.754(+6.3%)	0.732(+3.2%)	0.779(+9.9%)
AR _m	0.704	0.746(+6.0%)	0.719(+2.1%)	0.769(+9.2%)
AR _l	0.748	0.811(+8.4%)	0.824(+10.2%)	0.855(+14.3%)
Extra Frames for Training	0	≐ 10,000	≐ 500	≐ 250

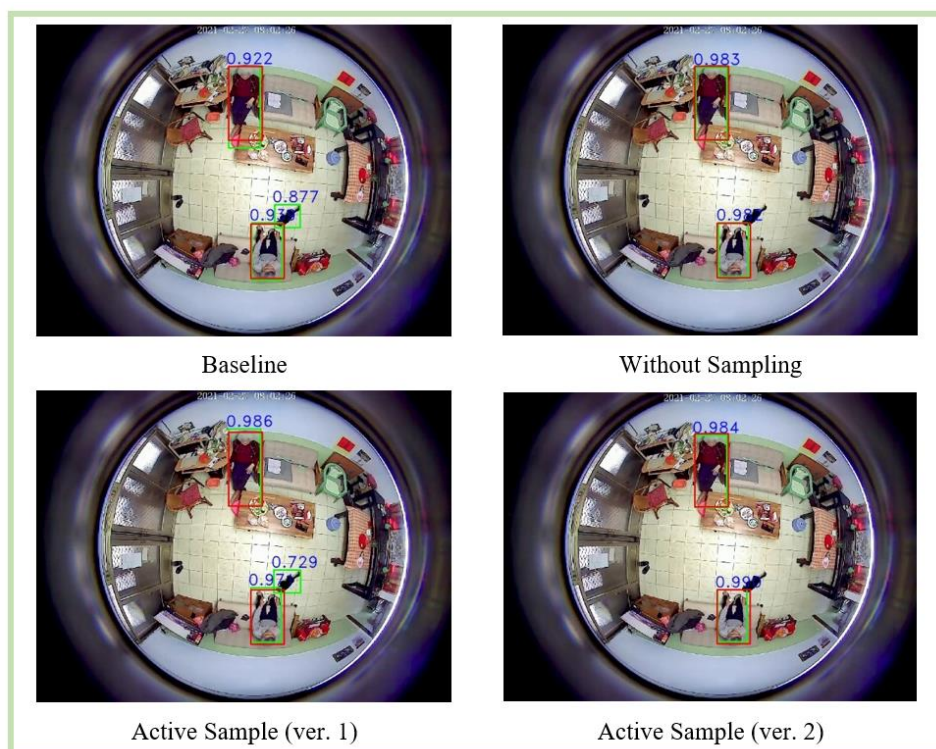
表一：實驗數據

根據表一，可以先針對 Baseline 模型分析。可以看到其 AP 僅有非常不理想的 0.493，換言之其先前所 pre-train 的資料，由於是隨意挑選的，因此對於其辨識能力的提升相當有限。就算是比較寬鬆的指標，如 AP₅₀, AP₇₅，分別只有 0.74 和 0.676。

表一之中，第二行數據(without sampling)是不經過任何演算法，直接將 30 分鐘的影片放入模型訓練後的效果，可以看到此時模型已經比起 Baseline 模型，在各項指標都有飛躍提升，但我們整個模型額外多使用了約略 10,000 多張的影格（圖片）數，若考慮每張進行訓練前都需要進行標記(label)，則是一件非常浪費資源的事情。

第三行數據，是利用原始的 Active Sample 所產生的輸出進行訓練，並從原本 10,000 多張的影格，在我們設定之閾值下產生出約 500 張左右的影格，但根據結果我們會發現到其效果整體都不如前一種方法。但我們所改良的 Active Sample 演算法，效果則為所有方法中最好的，以 AP 為例就足足比 Baseline 進步了 47.5%，AP₇₅ 也有將近 40% 的長進。訓練所需的影格量卻只需要 250 張左

右，比原始的 Active Sample 減少了一半，而這就是我們本專題實驗結果最佳的項目，也是本專題的創新貢獻。



圖三：四種方法的模型預測結果（資料來源為識睿科技股份有限公司所提供使用者資料，僅用於研究使用，不用於其他用途。）

在圖三中，紅色框是 ground truth，也就是人的真實位置，而綠框是模型的 prediction，至於綠框旁邊的藍色數字代表 confidence score，觀察這幾張圖我們可以發現以下幾件事情。

首先，畫面中央的黑狗在 Baseline 當中被框上綠框了，也就是模型認為這是一個人，分數為 0.877，但是在 without sampling 的版本中，狗的綠框不見了，而在原始的 Active Sample 方法中，狗的綠框又出現了，又再次被誤判為人，雖然這不是我們樂見的結果，但這無悖於表一的數據，以 AP 為例，從沒有進行 sample 時有 0.699，但經過原始的 Active Sample 方法之後卻降到現在的 0.661，所以模型又再次把狗誤判成人是合理的。但是，在經過我們改良的演算法之後，狗的綠框又再次消失了，代表說模型學習的更好，對應到 AP 也從 ver1 的 0.661 到 ver2 的 0.727，足足有 13% 的長進。

除了狗的框之外，我們也可以觀察老奶奶（鏡頭中上處）與老爺爺（鏡頭中下處），不管在哪個方法之下，他們兩個都有成功被模型預測到，但分數卻有些許差異。在圖十二與圖十三當中，老奶奶的分數幾乎一模一樣，約為 0.98，而老爺爺的分數則從 0.971 上升到 0.990，證明我們所用的 Active Sample 方法的確有讓模型學習的比原始的方法更好，代表模型對於人的辨識能力有提升。

四、結論

經過我們的實驗，從結果可以知道若採用單純的光影變化去設計演算法來篩選可能會有不足，但若經過我們對演算法進行的改良，也就是多考慮一項「bounding box 面積」的因素去篩選，則能在準確度有大幅提升、所需要的影格數量則是大大減少，我們認為這樣的結果可以非常有效地節省了標記數據 (data labeling) 的成本。

未來我們也希望可以將模型發展到可以做到 real time，一邊收集資料的同時就去分析資料的重要程度，將那些對模型更有提升效果的資料放回模型去訓練，使得目標檢測模型學習的效率提升。

而從各面相來看我們的方法都比原始的作法來的有效，因此在本次專題內已經達成目前最優(State-Of-The-Art)。儘管如此我們認為方法上還是有地方需要加強，像是在本專題中，我們只有針對 FN 進行處理，如果也可以對 FP 進行優化，或許讓模型的表現還可以更上層樓。

五、參考文獻

- [1] Burr Settles. Active Learning Literature Survey, 2010
- [2] Donggeun Yoo and In So Kweon. Learning Loss for Active Learning. arXiv preprint arXiv:1905.03677, 2019
- [3] Localization-Aware Active Learning for Object Detection, arXiv preprint arXiv:1801.05124, 2018
- [4] Towards Human-Machine Cooperation: Self-supervised Sample Mining for Object Detection, arXiv preprint arXiv:1803.09867, 2018
- [5] Scalable Active Learning for Object Detection, arXiv preprint arXiv:2004.04699, 2020
- [6] Efficient adaptive density estimation per image pixel for the task of background subtraction, 2006
- [7] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, arXiv preprint arXiv:1506.01497, 2015
- [8] Deep Residual Learning for Image Recognition, arXiv preprint arXiv:1512.03385, 2015
- [9] Feature Pyramid Networks for Object Detection, arXiv preprint arXiv:1612.03144, 2016
- [10] COCO Detection Evaluation, <https://cocodataset.org/#detection-eval>
- [11] Detectron2, <https://github.com/facebookresearch/detectron2>

六、心得感想

我們一開始加入孫民老師的實驗室做專題，就是希望能夠以實作為主，藉由這次專題去學到修課所學不到的。於是我們時常與博士班的學長請益，去討論我們該做什麼樣的題目，而學長提到說實驗室的 CarePLUS 正在進行一項大型專案，是關於老人照護的題目，希望我們能從這整個大型專案挑選一小部份去實作，一方面能一定程度地協助大型專案的進度，另一方面則是能有更多實作機會而非純理論研究。

由於我們先前對於機器學習、深度學習領域涉略還不夠深，因此上學期前半段主要的進度是讓我們自行去學習相關的線上課程，並於每周向學長們報告課程內所學，並記錄學長給予的方向、建議及回饋。到了後半段，則是開始每周自己去選取深度學習領域的 paper，並於每周開會的時候一樣花費 2 小時到實驗室報告。並於上學期期末的時候，去把想做的題目跟整個系統的流程設計定案。

這一路以來，我們學到許多深度學習相關的知識，也學到許多解決問題的方法。感謝實驗室的學長們，厚寧學長、偉誠學長、冠弘學長、峰榮學長、Victor 學長和其他每一位幫助過我們的學長們，真的非常感謝你們在每週讀書會時，都和我們討論課程或論文較為艱澀的地方，也感謝你們在我們遇到困難的時候，願意花時間與我們討論並提供我們許多協助。若沒有你們的協助，這次專題就無法順利完成。此外，更要謝謝孫民老師提供我們一些實用的想法與建議，讓我們在茫然不知所措之際，能突破瓶頸，勇往直前。

最後，也要感謝去年的現在，那份主動加入專題的勇氣，謝謝那個勇敢的自己！