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

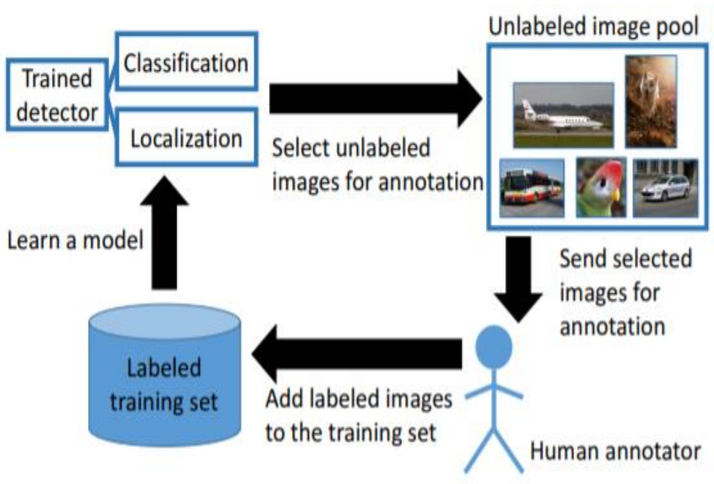
Introduction

為了節省照顧老人的人力，我們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有需要幫助。

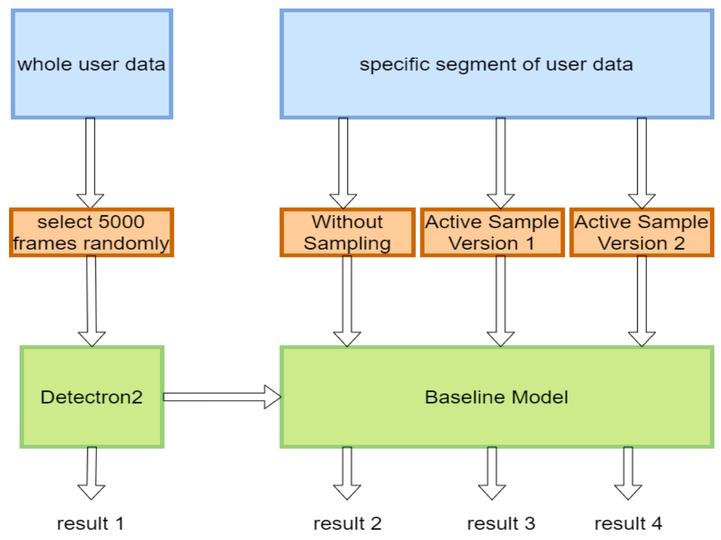
我們參考了[1][2][3][4][5]的作法，並融合這幾篇論文中的概念來發想。利用主動學習(Active Learning)讓模型進入新環境後，可以在更短時間內或更少的資料量下有顯著的進步。而主動學習就需要從現有的資料中，挑出模型比較不會的地方，讓模型能加強學習效率，因此我們定義了Active Sample的演算法，去挑出現有資料中的的偽陽性(false positive)和偽陰性(false negative)的影格，並重新將這些相對重要的影格丟回模型中，因為對模型而言連續的影片中並不是每張影格都能對模型有顯著幫助，特意挑出影格就能在較少的資料量下，以較高的效率去增強模型尚未學好的部分。

Method and System Design

為了節省標記資料的人力成本，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，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。



▲圖一、主動學習流程圖



▲圖二、系統架構

Algorithm 2: Active Sample Algorithm (including area)

Input: A video sequence S, prediction result produced by model P
Output: Valuable frames V
Parameter:
ϕ_h, θ_h for FP, FN human thresholds respectively;
$\theta_{opt}, \theta_{iou}, \theta_{area}$ (optical, IoU, area) thresholds respectively;
$F_P, F_N \leftarrow$ empty list; // Store FP, FN results
while *S is not over* **do**
  $F_i \leftarrow$TakeFrame(S); // Take a frame among S
  $P_i \leftarrow$FindPred(P, F_i); // Find prediction box from F_i
  $B_i \leftarrow$Apply background subtraction on F_i;
  $O_i \leftarrow$Analyze optical flow from F_i; // for False Negative use
  foreach *element p of* P_i **do**
    $M_{box} \leftarrow$ Find motion box by B_i;
    Find motion region ratio R_{motion} through B_i;
    if $R_{motion} > \phi_h$ **then**
      Append P_i to F_P;
      Draw bounding box information on V;
    end
    Save V as a figure and put V back to training set;
  end
  foreach *element q of* O_i **do**
    $M_{box} \leftarrow$ Find motion box by B_i;
    Find motion region ratio R_{motion} through B_i;
    Find optical flow region ratio $R_{optical}$ through q;
    $A \leftarrow$Calculate area of M_{box};
    if $R_{motion} > \theta_h$ *and* $R_{optical} > \theta_{opt}$ *and* $A > \theta_{area}$ **then**
      $R_{iou} \leftarrow$ Find iou result between motion box and P_i;
      if $R_{iou} > \theta_{iou}$ **then**
        Append M_{box} to F_N;
        Draw bounding box information on V;
      end
    end
    Save V as a figure and put V back to training set;
  end
end

▲圖三、演算法設計

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。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。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。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。

圖三則是我們改良過後的演算法，針對光流法產生的結果來去和background subtraction的資訊去建立關係，並計算FN box的面積當作篩選之參考，額外多加入一段FN box的面積的閥值來進行篩選，如此一來代表篩選條件更為嚴格，有機會能篩選出更關鍵之影格。

Quantitative Analysis

	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%)**
APm	0.483	0.685(+41.8%)	0.649(+34.4%)	**0.711(+47.2%)**
APl	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%)**
ARm	0.704	0.746(+6.0%)	0.719(+2.1%)	**0.769(+9.2%)**
ARl	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，Baseline的AP僅有0.493，由於pre-train的資料是隨意挑選的，因此對於辨識能力的提升相當有限。

第二行數據(without sampling)是直接將30分鐘的影片放入模型訓練後的效果，比起Baseline模型，在各項指標都有飛躍提升，但額外多使用了約略10,000多張的影格，非常浪費資源。

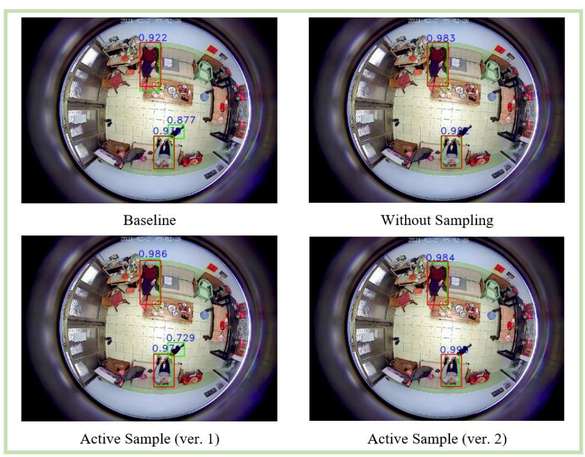
第三行數據，是利用原始的Active Sample所產生的輸出進行訓練，在我們設定之閥值下產生出約500張左右的影格，效果整體都不如前一種方法。

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%，訓練所需的影格量卻只需要250張左右，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。

Qualitative Analysis

畫面中央的黑狗在Baseline當中被框上綠框了，也就是模型預測為人，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，但這無悖於表一的數據。

而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ver. 1的0.661到ver. 2的0.727，足足有13%的長進。



▲圖四、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。)

Conclusion

經過我們的實驗，發現多考慮一項「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，有效地節省了標記數據(data labeling)的成本。

我們的方法在本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，如果也可以對FP進行優化，或許模型的表現還可以更上層樓。

Reference:
[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