

程式說明書

(1) 程式名稱

輔助視障者看世界之裝置

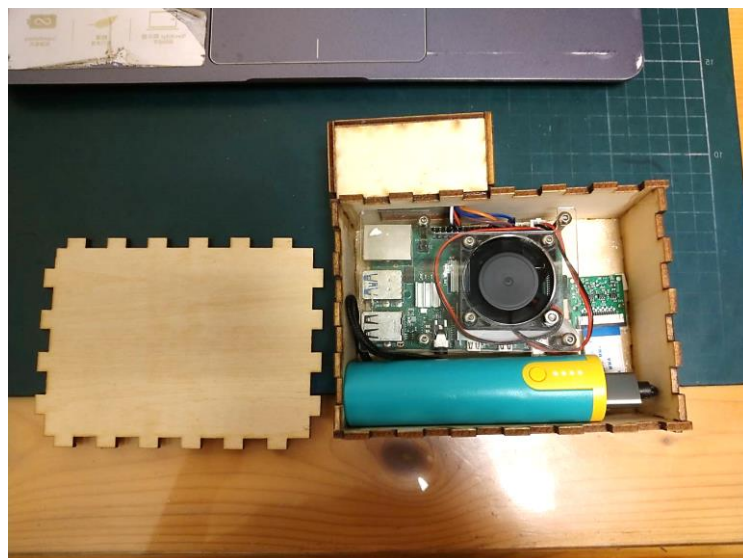
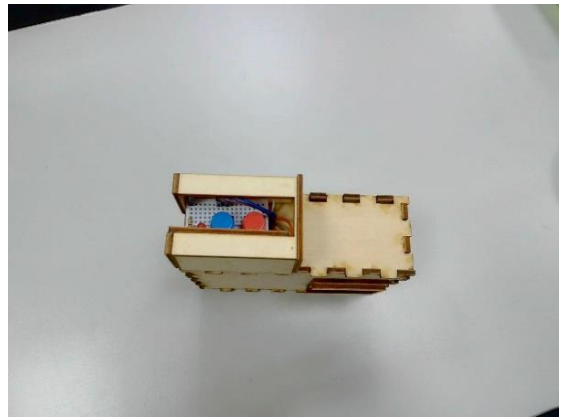
(2) 創作理念

由 2016 Microsoft research project——「Seeing AI」Prototype 為啟發，希望能做一個「輔助視障者看世界」之裝置如下：

使用者以攝像頭拍攝欲看到之畫面後，將影像或影片擷取，輸入進訓練好的 AI 模型，模型會將拍攝到的畫面轉為一句話輸出，而使用者透過藍芽耳機，便可直接聽取該句話，以實現讓視障者實際用於生活的目標。

(3) 功能說明（執行成果畫面圖文說明）

「輔助視障者看世界」之裝置外觀

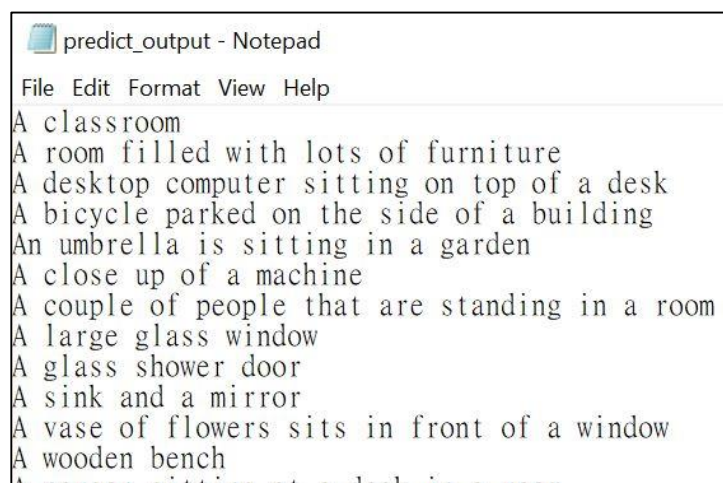


一分鐘 DEMO 影片

（影片中以藍芽音響為示範，實際使用可改用藍芽耳機聽取）

<https://www.youtube.com/watch?v=K1PBLS9UvVs>

此外最後也會產生一 **predict output.txt**(如下)將 output 產生出來的句子逐行紀錄：



(4) 程式流程、資料分析流程或系統架構圖

程式開發過程：

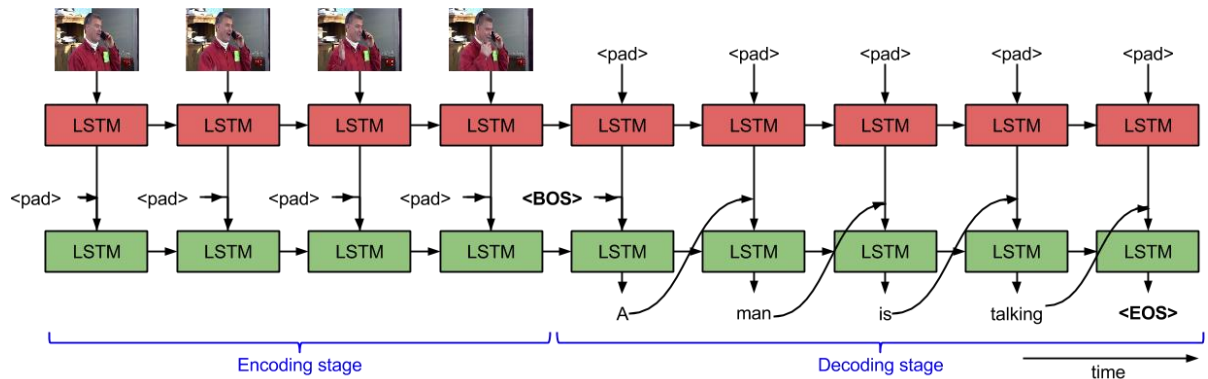
■ 原先：

原先計劃 train 一個「video-caption」之模型(輸入影片進入模型，可以輸出有關該影片內容描述的一句話)，並計畫再 train 好後壓縮以供在樹莓派上使用，模型參考作者 Subhashini Venugopalan 於 2015 年發表的論文〈Sequence to Sequence - Video to Text〉架構(如下圖一)，並以 MSVD dataset(包含 1450 個訓練影片及 50 個測試影片)作訓練，但訓練完拿非測試集的實際影片(以 youtube 上任意之影片作測試)餵進模型後發現，output 的結果(如下圖二)無法足夠精確地描述實際生活情境，檢討後認為可能原因為：

1. 資料量不夠多，或許之後可用 2~3 個 video dataset 下去訓練。
2. 除增加資料集外，欲得到更好的 performance 或許可以限縮使用情境(如:建僅室內的場景資料集)，或者限制使用頻率出現最高的 500 個字詞去作描述(但可能造成描述不精確，或產生模糊籠統的句子)
3. 設計或使用更好的模型。

(圖一) S2VT model

將影片餵進 CNN model(VGG16)，抽取 feature 後，再將 feature 餵進兩層 LSTM cell，model 最終會 output 一句話，即是欲得的結果。



(圖二) 以五個任意 youtube 影片實際測試之結果。

```
klteYv1Uv9A_27_33.avi, a man is riding a motorcycle  
vid819.mp4, a woman is dancing  
vid1493.mp4, a person is running on a bathtub  
vid1545.mp4, a woman is walking  
vid1492.mp4, a man is cooking a piece of meat  
video_2.gif, <unk>  
vid818.mp4, a man is running in a glass
```

■ 後來：

因模型預測結果準確性及穩定度尚不足以應用於生活中，因此後來退而求其次，改用 image caption(模型的 input data 為影像而非影片)，硬體部分則以 Raspberry Pi 4 進行製作(後來發現應該能用更小的硬體，如 Raspberry Pi Zero W 等，可以再縮小裝置的體積)。

使用情境：<https://www.youtube.com/watch?v=K1PBLS9UvVs>

當使用者按下頂端的藍色按鈕後(如下左圖)，將擷取鏡頭拍攝到的畫面，並將該畫面傳給雲端的 microsoft computer vision service，此服務會回傳模型的 output 出來(.json)，將 json 檔中 captions” 的欄位印出，即可得該影像的描述語句，並再透過藍芽耳機，讓使用者可以直接聽到該句話。(*樹莓派設定為開機後可自動登入，並可自動執行 blindless.py 程式)



(5) 程式開發與執行環境說明

python 3.x

Raspberry Pi 4

linux

2021 臺北醫學大學程式設計競賽

資料科學組

智慧應用與服務組

blindless.py

image.jpg

predict_output.txt

程式說明書.pdf

說明：

blindless.py 主程式

image.jpg 程式會將使用者按下藍按鈕拍到的圖片存成此檔，每次按按鈕就會覆蓋此 **image.jpg**

predict_output.txt 會將模型輸出的語句，存在此
predict_output.txt file

(6) 程式碼與註解

詳見 blindless.py