MA3113\* Topics in Mathematical Image Processing

**Final Written Report**

Wei-Zhi Chen & Po Jui Huang

Department of Mathematics

National Central University

**2024.06.12**

1. 研究問題

純音聽力檢查(Pure Tone Audiometry, PTA)圖表是醫師對耳科病患的一項檢測，而我們這組的主題是分析的檢測中所產生的PTA圖表結果，藉由表中symbols的辨識，輸出成一個數據形式的excel檔案，協助醫師進行診斷以及後續的分析。

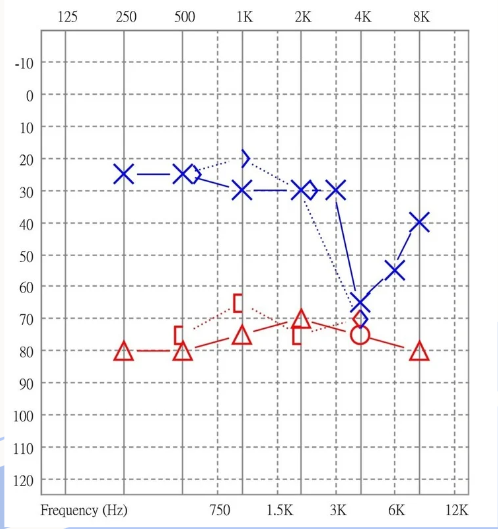
 

圖1：輸入圖表 圖2：輸出格式

圖表內容一共有藍紅2色，分別為左耳以及右耳的檢測，而隨著不同形式的檢查，圖表上一共具有8種symbols，而symbols之間會具有排擠(圖四)、重疊(圖五)、特殊標記(圖六)的清況發生。排擠會發生於在滿足特定條件下，symbol可能不會出現在它該出現的位置，而是被排擠至左右兩邊的其中一側，重疊則是會發生於當不同symbol位於相同座標點上時，但就目前觀測結果都為藍色叉號和紅色圓圈的重疊，不太確定是否有其餘情況，特殊標記則為醫生在做測試時，患者沒有反應的座標點紀錄，可能會出現在任何symbol上，而這三種情況是有可能同時發生在同一點上的。

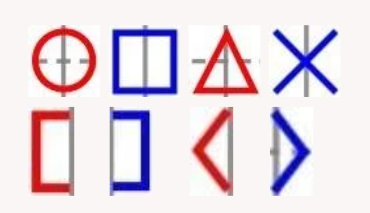


圖3：Symbols



圖4：排擠



圖5：重疊



圖6：特殊標記(以叉號為例)

1. 最終採用方法

最終版本：version 5.1

Ver.5 版本偵測判定依據（也適用於Ver.5.1）：

第5版本的演算法改用覆蓋正確率作為判讀依據，其中利用了圖表不會有大面積RGB色彩塊[[1]](#footnote-1)的特性，當覆蓋正確率相當高的時候能直接引導出有正確的目標圖形這個結果（覆蓋正確率相當高的時候只會有兩個可能性，一為確實有目標圖形，二為有大面積色塊出現在那個區域，好比說將整張圖都改為比#505064彩度更高的藍色，那麼全部的覆蓋正確率都會是100%。）

Pseudo Code (Version 5.0):

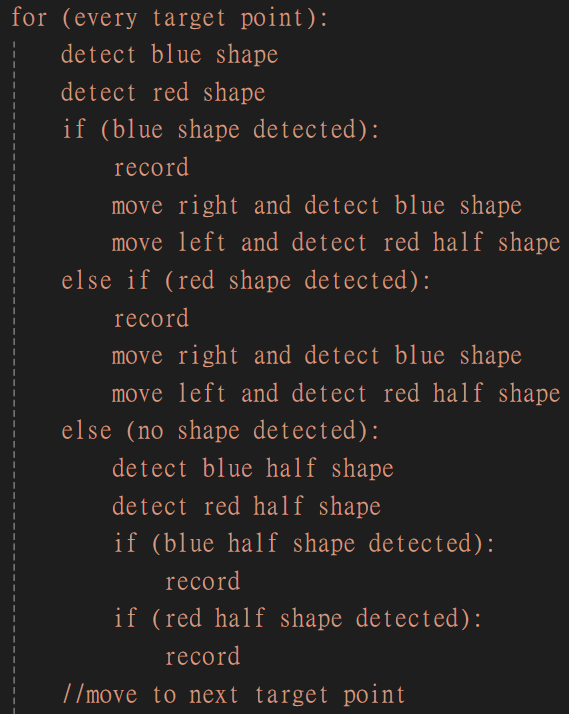


圖7：5.0 Pseudo Code

名詞解釋：

主圖形：圓形、三角形、正方形、叉叉

副圖形：大於、小於、左中括弧、右中括弧

和Ver.4 不同的地方是，偵測到紅色或藍色的主圖形時不只需要往左/往右偵測對應同色的副圖形，但因為後來發現有其他的測資出現了藍色主圖形排擠了紅色副圖形的情況，因此予以修正。

Ver.5.1版本特點：

5.1主要為程式優化以及輸出功能實裝，程式優化包含程式碼縮減以及減少運行時間，輸出功能實裝則為xlsx檔案的輸出，使用者也可以輸出任何MATLAB支援的Table轉換檔案。

Pseudo Code (Version 5.1):

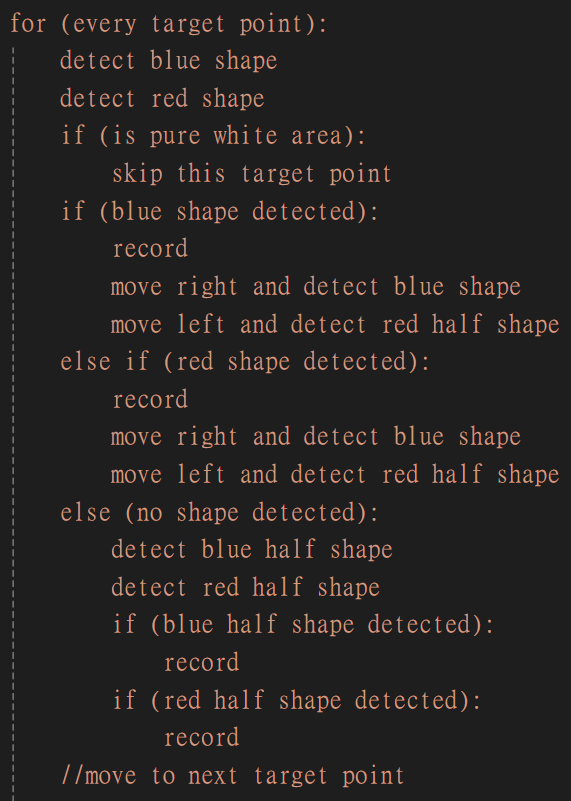


圖8：5.1 Pseudo Code

因MATLAB的其中一個缺點就是其For迴圈運行速度非常慢，所以在程式優化方面以減少For迴圈運行次數為最優先項目。5.1版本中我們增加了跳過全白區的部分，若是偵測區域幾乎都是白色（目前參數設定為白色占比超過99.5%），即該區域不具有任何圖形，相對的也不會有任何副圖形被排擠，因此可直接跳出迴圈，因圖表中總共有260個偵測區域，使用這個方法就能使真正需要運行的偵測區域減少至約60~80個。

1. 實驗結果與討論

結果如下

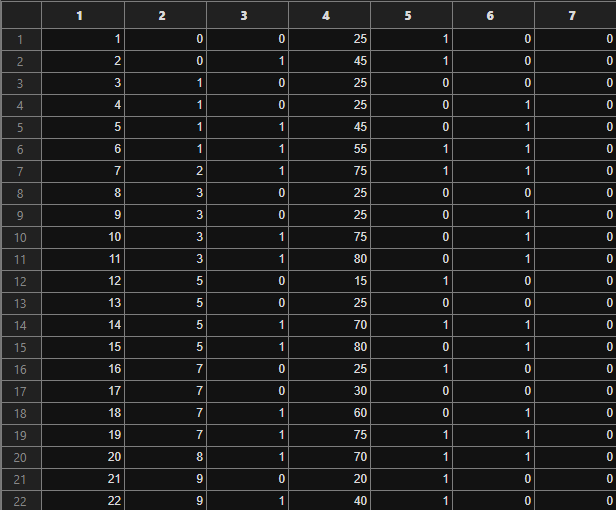


圖9：偵測結果

從Ver.4更新至Ver.5之後成功率有大幅提升

不過目前衍生了另一個問題是因為我們讓程式偵測到帶有箭頭的圖形時就強制中斷偵測，直接將有箭頭的圖形記錄下來，所以可能會有無箭頭判斷成有箭頭的情形發生。

或許我們可以使用偵測全部相同顏色的圖形之後再輸出吻合程度最大的圖形，這樣也可以迴避掉同一個偵測點輸出了兩個圖形的潛在問題。

目前的準確率大約落在90%出頭，一張包含了21～24個圖形的圖表大約會有0～3個圖形偵測不到或是判斷錯誤。

另外，醫院那邊突然和我們說他們的機器其實是可以針對醫師的需求放大或是縮小圖表的所佔版面甚至是位置，醫院那邊表示他們想使用YOLO先抓出圖表所在位置，再放入我們的模型進行偵測。

1. 結論

我們的模型相較於YOLO的結果準很多，但缺點是能夠處理的圖片限制相當高，因為是直接進行絕對比對的關係，要是圖表的位置或是大小和預設有差距的話就一定偵測不到。

1. 相關文獻

我們參考了以下這兩篇paper作為參考依據:

INTERPRETING AUDIOGRAMS WITH MULTI-STAGE NEURAL NETWORKS

Shufan Li, Congxi Lu, Linkai Li, Jirong Duan, Xinping Fu, and Haoshuai Zhou

Audiogram Digitization Tool for Audiological Reports

FRANÇOIS CHARIH AND JAMES R. GREEN , (Senior Member, IEEE)

其中第二篇是作為第一篇的延伸研究，影像處理流程圖如下所示：

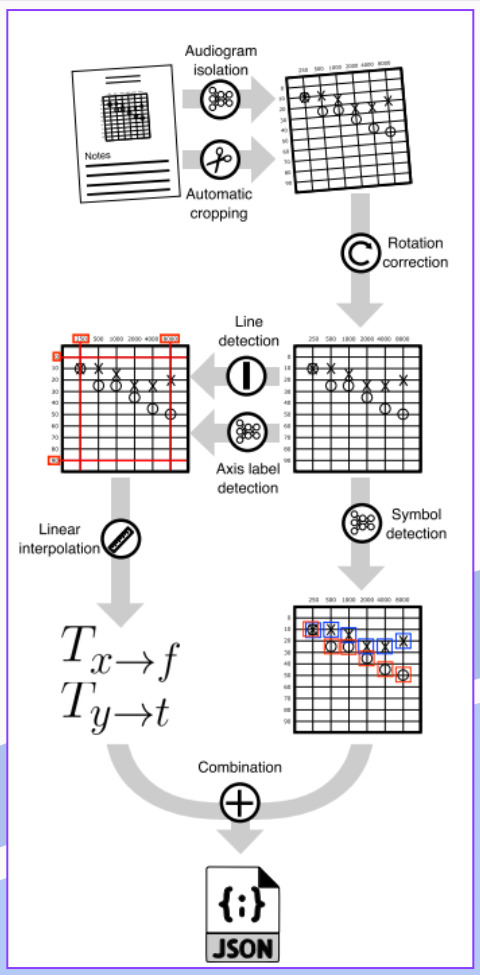


圖10：流程圖

在流程方面，先利用Hough transform 作為尋找圖表的工具，找出圖表在文件上的位置，並透過斜率來判斷是否將圖表進行轉正，透過這個處理方式來整理參差不齊的文件，隨後在進行座標線和symbols 的辨識，由於他們的資料來源是相機拍攝的，和我們的圖表來源都是同個機器和格式的產物不同，我們並不需要前述整理圖表的操作。不過我們曾經有想過利用Hough transform作為消除輔助線的依據，在此簡單敘述一下：

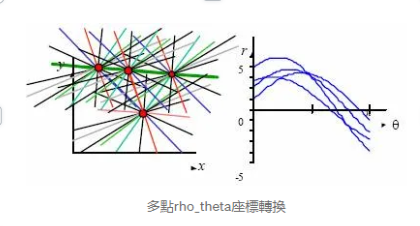


圖11：Hough transform

如圖片所示，一個點可以通過無限條的直線，我們可以將這些直線在極座標上記錄成一條正弦曲線，而多個點的圖像，將會形成具有數個交點的曲線，每個交點都是代表特定的直線，下圖則為例子，左圖中的黃點即代表右圖的黃色直線。

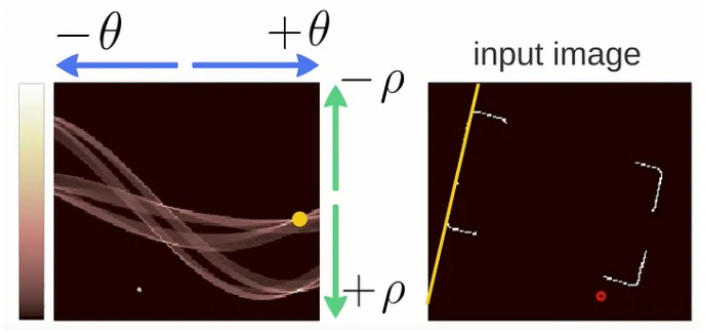


圖12：Hough transform例子

而在後續的辨識上，他們分別採用了YOLOv5和FASTER-RCNN他們都屬於多層的卷積神經網路，用於訓練辨識物件的工具，由於卷積神經網路並非我們所採用的的演算法，在此不多做太多的贅述，不過我們發現了一個有趣的訓練集，Common Objects in Context(COCO) Dataset，它以數以萬計日常生活中的照片，加上許多特徵標記，讓人們減少了許多收集資料，訓練資料的麻煩，只要再另外再加入一些針對專案需要的特殊圖形標記，就能快速的建立一個數量龐大且完善的偵測資料庫，算是以後可以參考的方法。

最後，透過以上方法，第一篇的作者群成功找出圖表，並發表作法MAIN(Multi-stage Audiogram Interpretation Network)，但受到誤差和圖表的侷限性，也只能找出特定兩種圖形，並不支援多功能偵測，如下圖所示。

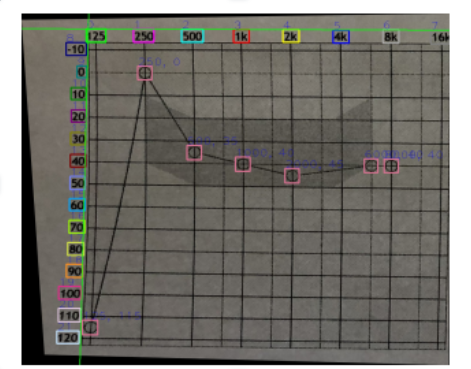


圖13：Multi-stage Audiogram Interpretation Network 結果

而另一團隊則是挑戰了手寫圖表的辨識，並以MAIN做為參考及比較依據，可以發現，就準確率而言比MAIN還清楚，但手寫的複雜度，可以預見它會更加的不穩定，有時會直接錯過需要尋找的圖形。

1. 在HEX色碼中RGB三碼數值有明顯差異的顏色，即彩度高的顏色 [↑](#footnote-ref-1)