

國立台北大學資訊工程學系專題報告

Book Grade Evaluation Master – Classifying second-hand books and marking problem points through deep learning

書籍書級評鑑大師 – 透過深度學習進行二手書的書況分類並標示問題點

專題組員：林均彥 梁祐銘 池秉宸 吳孟軒

專題編號：PRJ-NTPUCSIE-110-005

執行時間：110 年 9 月 至 111 年 6 月

1. 摘要

以目前二手書店的估價方式來看，我們發現到目前尚未有一套完整的系統去做客觀性的估價，都是以每個店員遵照一些主要原則並以較主觀的方式去估價。查閱了過去的研究成果和文獻也尚未發現類似的估價技術。我們希望使用設計一套基於目前熱門的深度學習方法的自動化系統。藉由系統自主判斷書的定價來減少在定價中人為主觀意識的影響。



書類估價的先驅！

本計畫為了結合了物件偵測 (object detection) 和影像分類 (image classification) 等技術的優點，選用了 Mask R-CNN 當作第一層主幹，它提供更準確的輪廓遮罩，對於偵測微小的物件有很大的幫助。而第

二層主幹選用兼具效率和準確性的 EfficientNet，為整體做一個更完整的評估，希望能提高準確率並預測無法用人工評估的價值。

關鍵詞：物件偵測、影像分類、書類估價

2. 簡介

2.1 背景

現今顧客前往二手書店消費時，能看到的僅有書的標價，顧客無法了解書店是如何決定二手書的價格，而且不同店家對於書況差不多的二手書，也常有差異甚大的定價。而我們也實地訪問過書店店員，目前確實沒有一套完整的系統去對二手書做客觀性的定價。

2.2 目標

我們希望能夠建立出一個系統來以客觀的角度對二手書做定價。而為了確保整套系統不會受

到太多的人為主觀意識所影響，故我們希望使用設計一套基於目前熱門的深度學習方法的自動化系統。藉由系統自主判斷書的定價來減少在定價中人為主觀意識的影響，並最終完成僅只要傳入書籍資訊，機器便能訂出最終合理價格。

2.3 主要預期效益

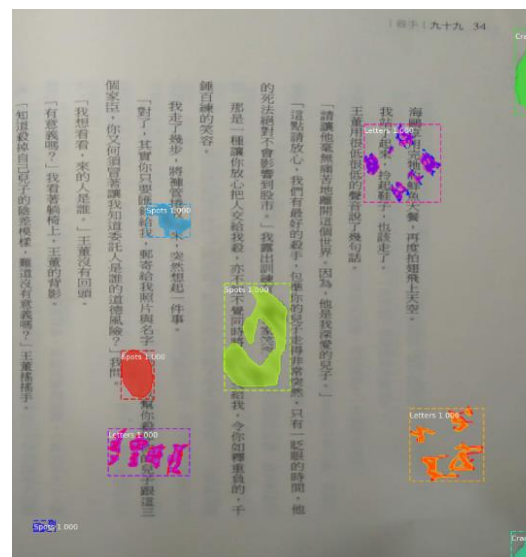
此計畫不僅可以對於未來的書籍二手估價技術有了全新的面貌，可套用在各大二手書店的網站上為二手書店省去不少人力資源在不客觀的評估價值上。

3. 專題進行方式

3.1 系統分析與設計摘要

會影響二手書定價的因素包含書上的黃黑斑、筆跡、破損、摺痕和整體泛黃程度等書況資訊，以及書齡、暢銷度和書籍類別等外在資訊，故如何訓練機器抓出這些資訊便是第一個問題。我們預計設計一基於深度學習的定價系統去處理我們的研究問題，系統將分成三個部分來處理以上資訊：

- (1) 第一部分:先進行物件偵測來讓機器讀取書上的黃黑斑、筆跡、破損等資訊，此部份我們選用較適合偵測微小物件的 Mask RCNN 網路。



- (2)第二部分:透過卷積式神經網路(CNN)來讓機器讀取出，書上的摺痕和整體泛黃程度等較全面的資訊。

- (3)第三部分:定義出一個定價函式，其中參數包含經第一、二部份求出的書況資訊，和使用者的書齡、暢銷度和書籍類別等外在資訊，最終求出書本的定價。

3.2 主要的困難與解決之道

而在建立深度學習網路並訓練的過程中，預期遇到的問題便是訓練資料的收集。因為該主題為新創題目，沒有前人開發與收集之資料集，故沒有現成的資料集可使用，所有的訓練和測試資料都須自行收集與生成，而在自行收集的過程中，自行由二手書本中以拍照方式和人工標誌非常花費人力成本，通常能完成之標記資料量是很少的，故如何提升資料量和提升資料的精準度則是本研究預期遭遇到的一大難題。

$$\begin{matrix} 30 & \times & 14 & \times & 50000 & = & 2.1 \times 10^7 \\ \text{Seconds} & & \text{Pages} & & \text{Books} & & \text{Seconds} \end{matrix}$$

= 243 Days !
IMPOSSIBLE !

因此我們參考了論文[3]，在該篇論文，有提到合成訓練資料相關方法，我們希望由該論文的方法，幫助我們去發想一些概念，從而設計出一套資料集成程式，來合成出我們所需要的資料集，以解決資料量不足之困境。

3.3 實作平台與技術

3.3.1 EfficientNet

EfficientNet 主要從三個層面來改善模型，分別是 Width(Filter 的數量)、Depth(模型的層數)以及 Resolution(圖片的解析度)，當要改善一個模型時，調整其中一個的參數，會發現其他兩個也十分重要，因此作者找出一個新的思路來加強模型，也就是 Compound Scaling Method，將問題以最佳化的方式的方式來敘述，限制式為運算平台擁有的資源，

$$\begin{aligned} \max_{d,w,r} \quad & \text{Accuracy}(\mathcal{N}(d, w, r)) \\ \text{s.t.} \quad & \mathcal{N}(d, w, r) = \bigodot_{i=1 \dots s} \hat{\mathcal{F}}_i^{d \cdot \hat{L}_i}(X_{\langle r \cdot \hat{H}_i, r \cdot \hat{W}_i, w \cdot \hat{C}_i \rangle}) \\ & \text{Memory}(\mathcal{N}) \leq \text{target_memory} \\ & \text{FLOPS}(\mathcal{N}) \leq \text{target_flops} \end{aligned}$$

接著藉由不同的 Depth、Width 以

及 Resolution 使用 $\alpha \beta \gamma \phi$ 作為控制的參數，設定最佳化目標

$$\text{ACC(m)} \times [\text{FLOPS(m)}/T]w$$

$$\text{depth: } d = \alpha^\phi$$

$$\text{width: } w = \beta^\phi$$

$$\text{resolution: } r = \gamma^\phi$$

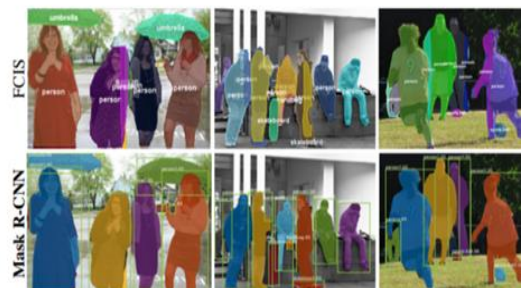
$$\text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

這個模型的方法在各方面都表現相當的好，而且在想法上相當的直觀，運用自動化的方式調整模型，減少大量的人力需求。

3.3.2 Mask RCNN

在物件偵測方面，已經有許多技術呈現在眾人面前，例如:FAST-RCNN、Mask RCNN 及 YoloV4 等，最後我們選擇了 Mask RCNN 作為我們的模型，其一是因為它提供更準確的輪廓遮罩，對於我們二手書籍的分級判斷有很大的幫助。



(2) 圖 2：Mask RCNN 與 FCIS 輸出圖例

此外，Mask RCNN 改良了對於位置的描述，將 ROI Pooling 改良成不會取整的 ROI Align，大大增進了模型的精確度，此模型使用的 LOSS Function 是。

前兩項與 Faster-RCNN 相同，第三項輸出為 $m \times m \times K$ ， $m \times m$ 為 MASK 的解析度， K 則為種類的數量，雖然其沿用了 FCN 的設計，但卻改以 Sigmoid 作為激發函數，Average Binary Cross Entropy 作為損失函數，並對 ROI 給出 K 個 mask，但並不是每個 mask 的 Loss 都會計入第三項中。Mask RCNN 有著很好的 instance segmentation，因為更改了激發函數及損失函數，這個設計避免了原 FCN 作法產生的 Mask 跨類別競爭的不利影響，帶來良好 instance segmentation 的結果，也不會產生 FCIS 對於重疊物件帶來奇怪的疊影物件等系統誤差(見圖 2)。

3.3.3 訓練資料半自動產生

此外，我們也參考了 Handwritten Text Segmentation via End-to-End Learning of Convolutional Neural Networks 這篇論文[3]，研究了訓練資料生成的技術，因為假如我們以人工的方式蒐集訓練所需的資料，這會是一個花費龐大人力成本的作業，因此我們就思考能不能以合

成的方式生成我們所需的訓練資料，論文中原合成方式如圖 3 所示。

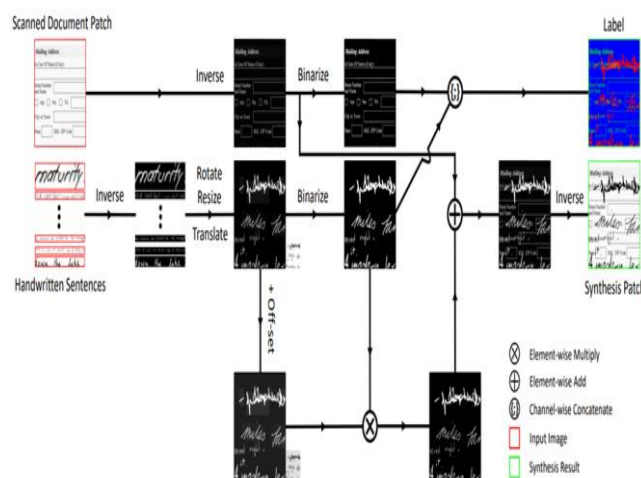


圖 3：論文[3]所提之文字合成流程圖

在我們的系統中，我們會利用 Threshold 汲取所需的文字，貼在所需的背景上，並留下起初留下的 Label，此部分之合成方法會參考此論文生成資料的作法，原因是我們的系統中也需要進行判斷書本中手寫字的任務。

但我們的系統不只有文字需要進行訓練，還有黃黑斑等，我們進行了修改，故我們必須將此方法再行擴充修改，以符合我們的需求，我們將在研究方法小節中敘述我們系統生成合成訓練資料的流程作法。

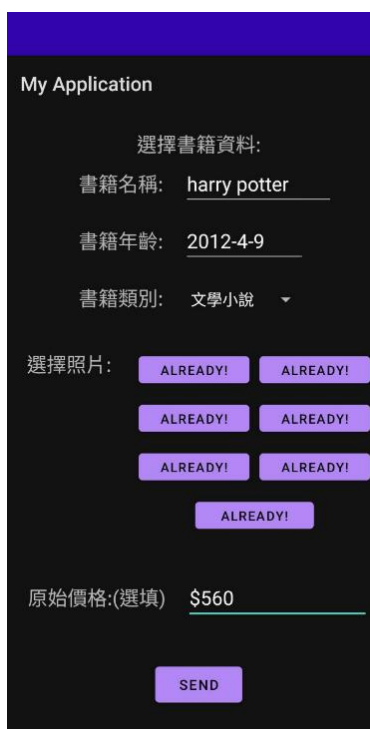
4. 主要成果與評估

這裡簡要介紹了本專題之研製成

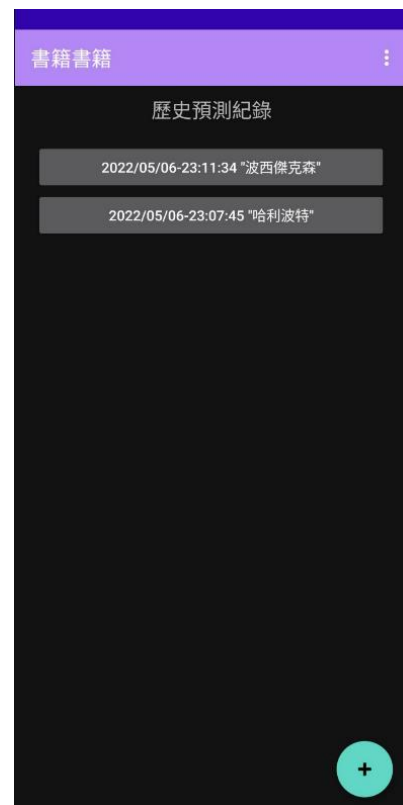
果, 包括完成之系統規模、測試上線情形、與其它具體成效說明。簡要評估本專題之研製過程, 包括未來可能之擴展方向、與其他客觀之評估等

4.1 結果呈現

我們利用 android studio 實作 client-server 的方式收集使用者書況資料並分析, 加上本地儲存系統可供使用者查詢使用紀錄。以手機 APP 的方式呈現, 兼具方便性和實用性。能夠讓使用者可以選擇書籍照片並填寫一些書籍資訊就能得知預測書評。

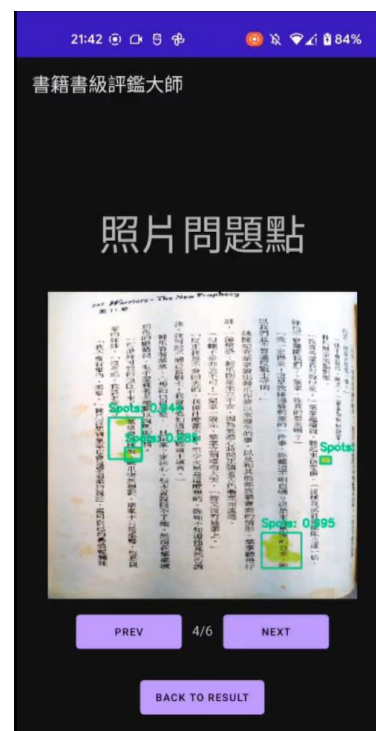


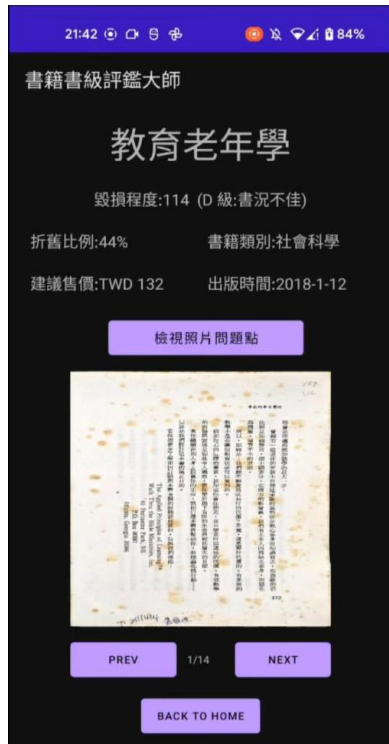
填入書籍名稱、年齡、類別、照片及價格



透過歷史紀錄功能來查詢之前輸入的資訊

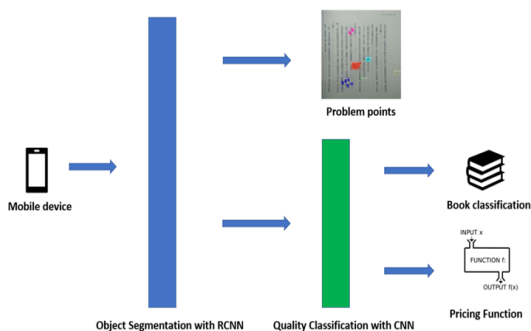
且能在 Android 平台上標記出污漬、裂痕、字跡等位置, 為使用者提供一些實體書況不良的證明。





顯示折舊比例、建議售價、書況等級等重要資訊

4.2 系統規模架構



架構圖如上，先運用 Mask-RCNN 的技術先行將書籍上的污漬、裂痕、筆跡等問體點給標示出來，給使用者更加清楚的資訊來瞭解書況的問體點。另外，透過權重訓練完成後的 RCNN 中的 RoI 層來取得各圖片資訊濃縮後的數值，該數值一方面能放進後續的 CNN 模型藉此增加準確性，試想如果將判斷書況與問體點拆開來訓練並預測，必定會重疊進而導致效率降低且

會產生判斷上的問題。

接下來會使用 CNN 將書況結果預測出來，而輸入資料會使用書籍中的各角度原圖像，如封面、書背、數張書頁等資訊，並將上一步驟 RCNN 的濃縮資訊與 CNN 中途的 tensor 做合併。並選定 EfficientNet 來作為 CNN 的基底模型，兼顧預測準確性與速度。

而最後 CNN 將把書況以 0~1 的數值進行評級，並透過簡單的 threshold 將書況平分成 5 個等級給使用者參考。且透過自定義的函式結合使用者輸入的出版年月日、書籍類別、書籍原價與爬蟲取得的書籍討論度等資訊與 CNN 模型的結果來產生出一個建議的二手書售價。

5. 結語與展望

此計畫開發出一個簡易且方便的 APP，希望能套用於各大二手書店的網站上方便買賣雙方使用，並期待有機會能順利申請專利。此 app 達成讓使用者不需額外的平台即可自動為自己手上的二手書估價，十分的方便又快速。而我們在做專題的途中深刻感受到了收集資料的困難及辛苦，未來我們期望對資料及模型進行改良，想要利用 CycleGan 技術增加模型的準確率。

6. 銘謝

非常感謝吳信龍教授這兩年來的指導和鼓勵，為我們的專題帶來源源不絕的動力，還有一起奮鬥的組員們，我們都不會忘記這段雖然很辛苦但非常重要的專題時光。

7. 参考文献

- [1] Mingxing Tan, Quoc V. Le,
“EfficientNet: Rethinking Model
Scaling for Convolutional Neural
Networks” ,2019.
- [2] Kaiming He, Georgia Gkioxari,
Piotr Dollár, Ross Girshick,
“Mask R-CNN” ,2017.
- [3] Junho Jo, Hyung Il Koo,
Jae Woong Soh, Nam Ik Cho,
“Handwritten Text Segmentation via
End-to-End Learning of Convolutional
Neural Network” ,2019