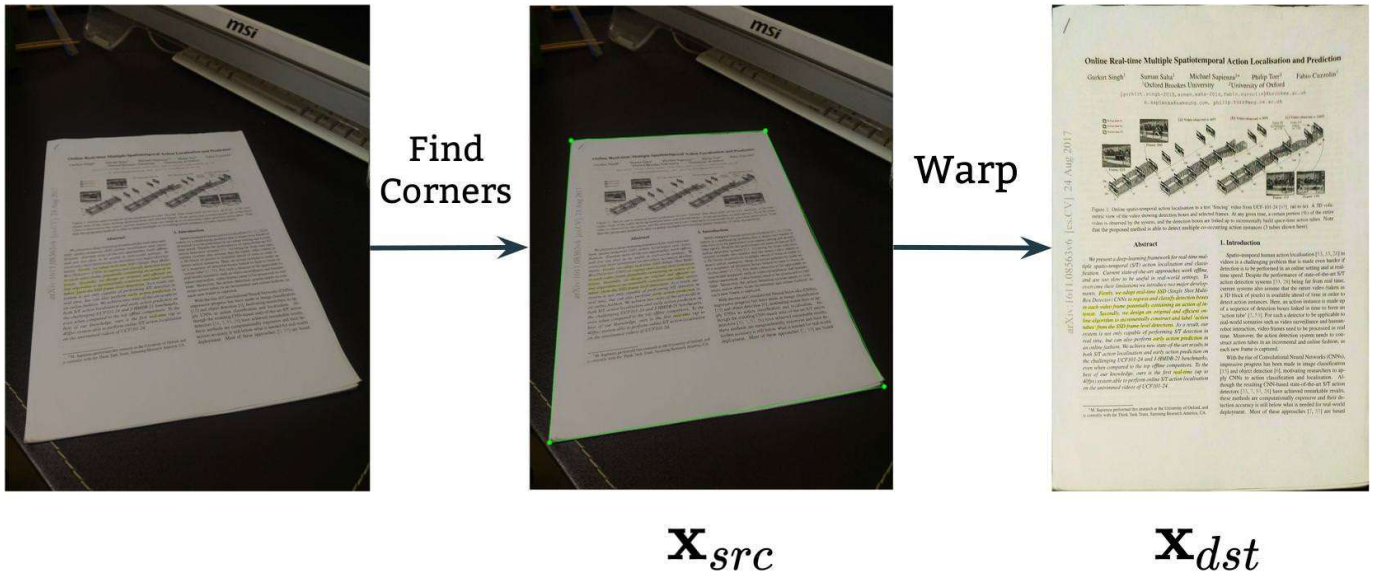
APT-Scan

403410034 資工四 黃鈺程
403410071 資工四 李晨維
404410030 資工三 鄭光宇

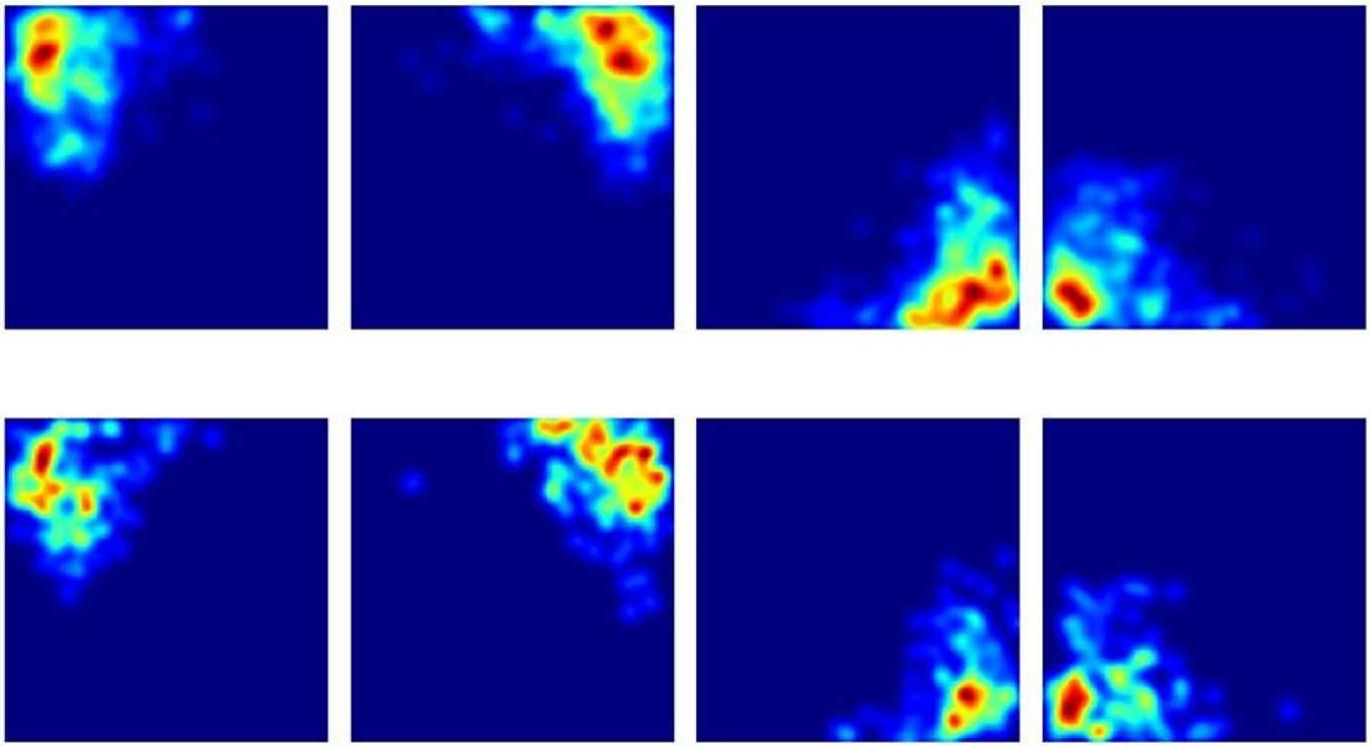
Abstract

我們打造了一個功能相同於 Office Lens 的系統。這個系統的輸入為一張關於紙張、書本或海報的照片，系統會自動偵測照片中的目標，將之轉正然後輸出。我們嘗試了幾種不同的 CNN，其中最好的在我們自己收集的 Dataset 上做出了 92% 的 Accuracy 且預測的 corner 與實際 corner 的距離平均只有 0.003 倍的圖片長。



Dataset

我們自己收集了 649 張圖片，並以 labelme 標記，最後以 5:1 將之切割成 train 與 test。

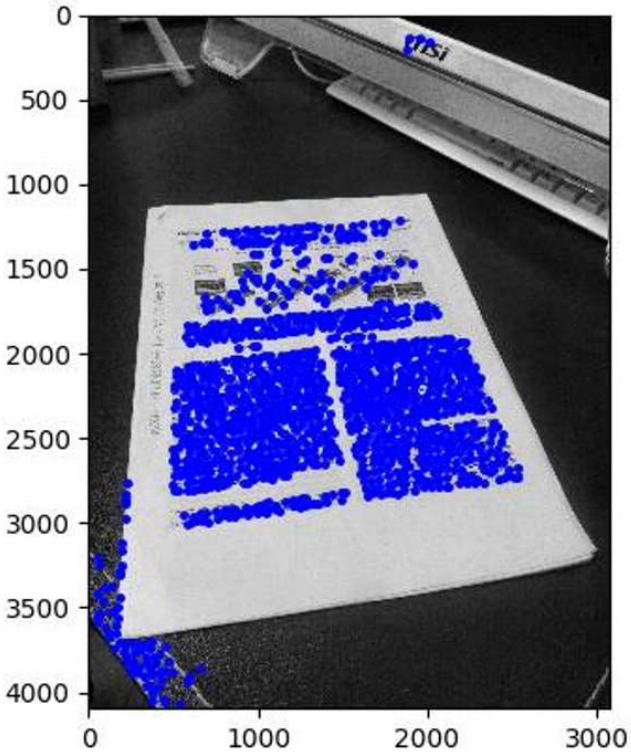


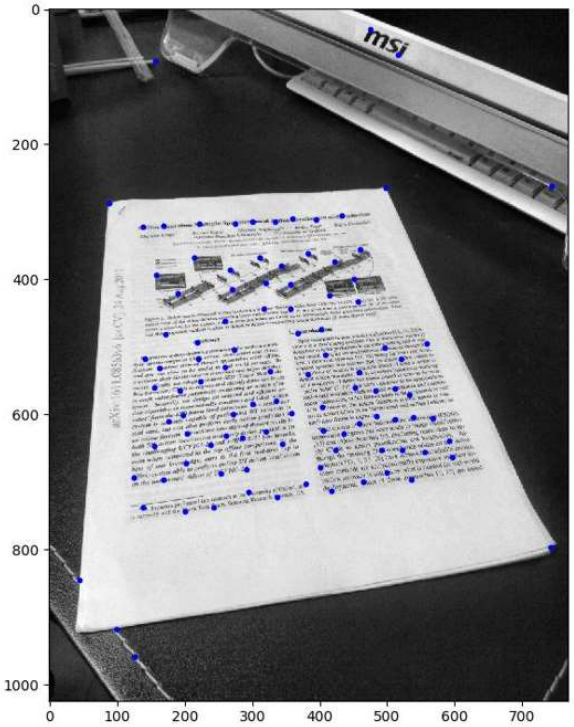
上圖是將 corners 順時針排序後，4 個 corners 各別的分佈圖。上半部是 train，下半圖是 test，以 jet colormap 顯示。

Corners Detection

Harris Corner

我們原先想嘗試使用 Harris Corner 來偵測出所有候選的 corners，然後透過一些幾何方法，找出目標的 4 個 corner。但最終發現，Harris Corner Detection 非常的不穩定且雜訊非常多。甚至許多圖片，他生出了非常多的候選點，其中卻沒有目標的 4 個點，這是不可接受的，因此我們放棄了使用這個方法做 Corners Detection。





Semantic Segmantation

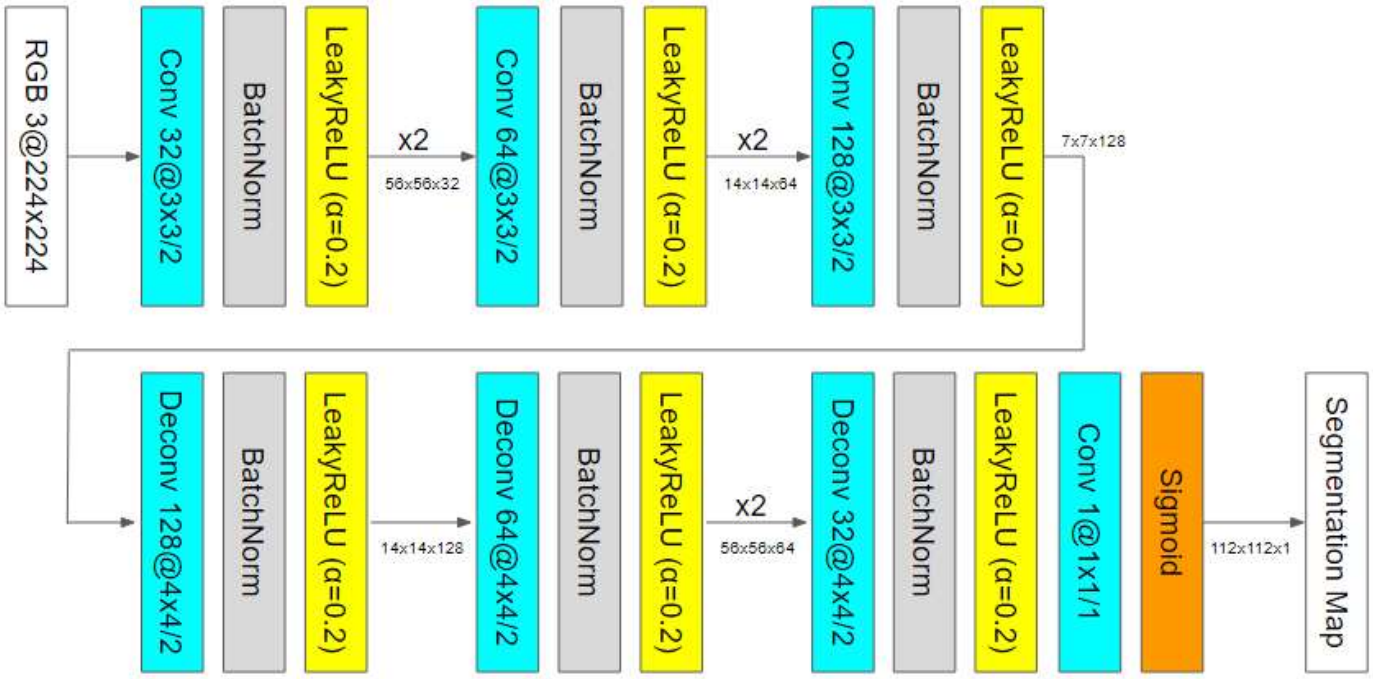
我們將角點偵測轉換為 Semantic Segmentation 的問題。

我們將輸入圖片縮放至 224x224，作為模型的輸入資料。之後將四個角點座標對應到 112x112 尺寸的語意分割圖上，以角點座標為圓心、固定半徑 3.2 pixel 畫出 4 個實心圓，作為之後 Semantic Segmentaion 的 ground truth 。

我們的模型輸入 224x224 的原始圖片，輸出 112x112 的語意分割圖。

Model

我們設計了一個小型的 FCN (Fully Convolutional Network) 模型，用來執行語意分割任務，模型架構如下圖所示:



Loss

模型訓練使用的 Loss 結合 Binary Cross-Entropy (BCE) 與 Dice Coefficient (Dice)。

Dice Coefficient:

$$s = \frac{2|X \cap Y|}{|X| + |Y|}$$

Binary Cross-Entropy:

$$-\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log \hat{y}_n + (1-y_n)\log(1-\hat{y}_n)\right]$$

我們的 Loss 為

$$Loss = 0.5 \times BCE - Dice + 1$$

之後用 Adam 以 0.001 的 learning rate，訓練 178 個 epoch。

訓練時，使用一些 data augmentation 來增加資料多樣性。如果 data augmentation 某些操作會改變角點座標（例如：旋轉、裁切），我們會重新計算角點座標，照前述方式重新產生語意分割圖的 ground truth。

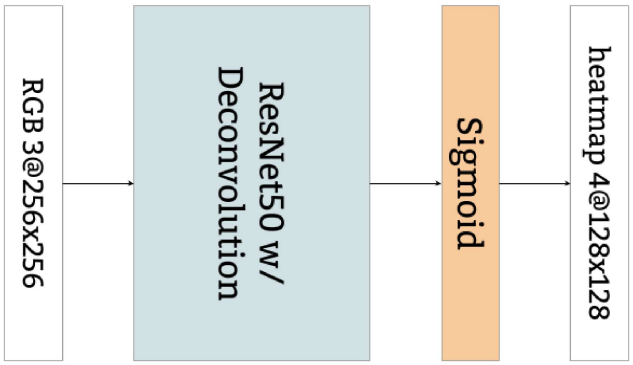
Post Processing

1. 將模型輸出結果以 threshold=0.5 二值化，得到語意分割圖
2. 標記分割圖中的每個連通塊
3. 找到面積前 4 大的連通塊
4. 4 個連通塊的重心視為角點

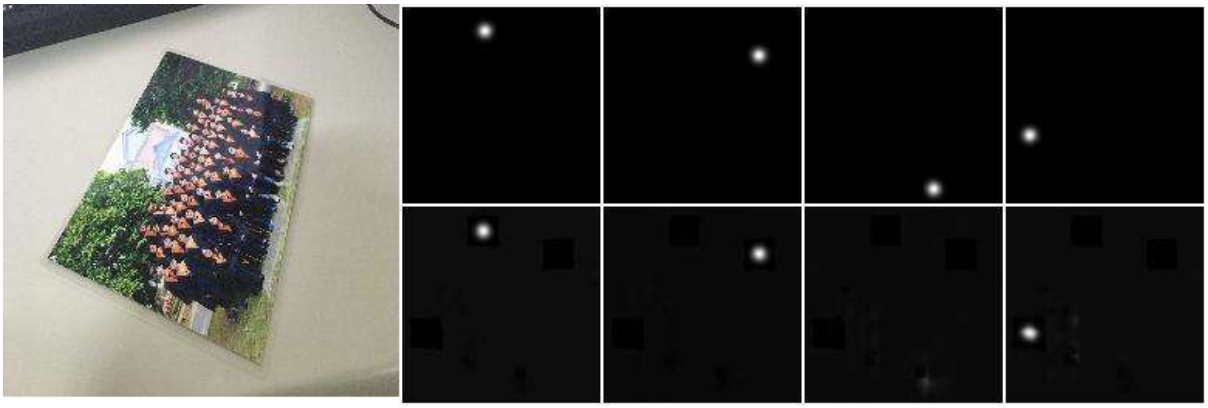
圖片說明:
最左測是原始圖片
最右側是 warp 後的結果
中間左上是圖片縮放後，與 groud truth 疊圖結果
中間右上是 groud truth
中間左下是模型輸出結果
中間右下是模型輸出二值化的結果

Pose Estimation

我們也嘗試將 Corners Detection 視為 Pose Estimation，並使用 Pose Estimation 的方法來解決：維度為 (3, 256, 256) 的圖片通過 model 後，生出 4 個大小為 (128, 128) 的 heatmap，分別對應 4 個 corners。Model 我們使用 pretrained 在 ImageNet 上 ResNet50，並把 ResNet50 的 FC 換成了 2 個 Deconvolution。Heatmaps 的 ground truth 為高斯分佈，即在對應的 corner 的正確位置會有一個 sigma=[3, 3] 的 2D Gaussian Distribution。訓練時我們使用 Adam(lr=0.001) 與一些的 data augmentation。而 Loss 使用帶權的 MSE，corner 附近的矩形有著比較大的權重。



Model 架構圖



範例。左圖是 input，右邊上排四個是 ground truth heatmap，而下排四個是預測結果。

最後的座標使用 skimage.feature.peak_loca_max 從 heatmap 中抽出，若有多個 local max 時，選取抽取出來的第一個。

Experimental Results

為了比較不同 model 之間的 robustness，我們使用以下 metric 做為衡量標準，假設座標都已被縮放至 $[0, 1]$:

對於圖片 i, Corner c 我們使用 L2 來衡量預測座標 x_{pred} 與實際座標 x_{gt}

$$L2_{i,c} = (x_{gt} - x_{pred})^2$$

而對於 Corner c 所有圖片的平均為

$$MSE_c = \frac{\sum_i L2_{i,c}}{n(Samples)}$$

最終的衡量標準是所有 Corner 的平均

$$mMSE = \frac{\sum_c MSE_c}{4}$$

如果 model 針對某張圖片沒辦法輸出 4 個座標，則視為一個 Failure，在這種情況下，model 預測的座標視為原圖的四個頂點 $(0, 0), (1, 0), (1, 1), (0, 1)$ 。

我們另外衡量了各 model 的 Failure Rate:

$$R = \frac{n(Failure)}{n(samples)}$$

Result

	MSE_{TL}	MSE_{TR}	MSE_{BR}	MSE_{BL}	$mMSE$	R
Segment.	**0.0297**	**0.0353**	**0.0184**	**0.0567**	**0.0350**	**0.0800**
Pose Est.	0.1118	0.1830	0.2618	0.1922	0.1822	0.1473

TL, TR, BR, BL 分別代表左上、右上、右下、左下 corner

從表格中可以看到 Segmentation 的方法全面性的比 Pose Estimation 方法好，尤其是以 model 大小來看，前者只是後者的 1/10 倍。我們相信這是來自於 loss 設計不同造成的，Segmentation 的混合 loss 能有效的辨識出那些邊界的條件。

Warping

找出 corner 後，我們需要找出原圖片與轉正後圖片座標之間的關係。電腦視覺告訴我們，事實上他們座標只差一個 Homography Transform（Projection Transform, Perspective Transform）。而這個 Transform 在**齊次座標系**下是一個線性變換。因些針對原圖上的某一點座標 (src_x, src_y) 與轉正後圖片上的對應座標 (dst_x, dst_y)，以下式子恆成立：

$$\begin{bmatrix} dst_x \\ dst_y \\ 1 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & 1 \end{bmatrix} \begin{bmatrix} src_x \\ src_y \\ 1 \end{bmatrix}$$

我們的目標是求出 Homograph Matrix H，如果求出來後，我們就可以用這個式子去做 Warping。

Homograph Matrix H 總共只有 8 個變數，所以我們只需要 4 組點對即可將上式轉成聯立方程。剛好我們的 4 個 corners，我們只需要知道他們在轉正後圖片上的對應座標。

明顯的，這四個對應座標就會是轉正後圖片的四個頂點 $(0,0), (W,0), (W,H), (0,H)$，所以我們的問題變成 W, H 是多少呢？我們並沒有想到太特殊的方法來求出 W, H，我們目前的方法為用估計的：

假設我們已求出原圖上 4 個 corners，那我們也可以得到我們目標的 4 條邊的邊長，假設分別為 T, R, B, L，那我們設

$$W = (T + B)/2$$

$$H = (L + R)/2$$

至此，我們即可求出 Homograph Matrix H，最後利用這個座標之間的關係式，我們即可以使用 Warping 將原圖轉正。

因為 W, H 我們使用估計的，在一些特定圖片中，會造出轉正後的圖片比例怪怪的，但這各問題我們還沒有想到要怎麼解決。

Visualization

FamilyMart

電子發票證明聯

107年05-06月

DG-08161158

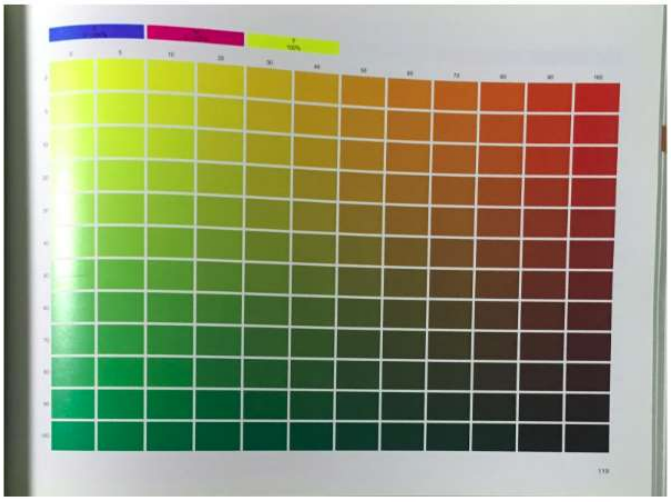
2018-05-24 18:12:37

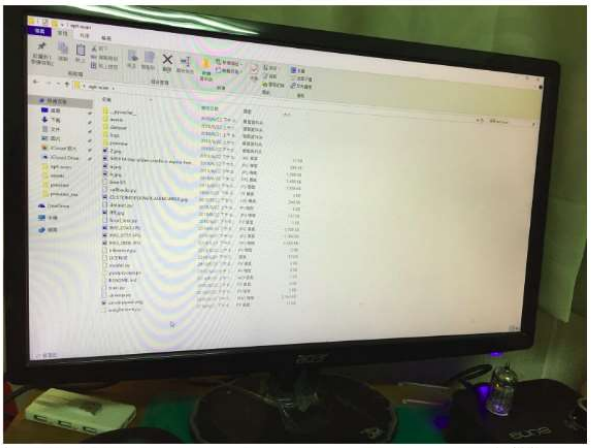
隨機碼:2324 總計:$8

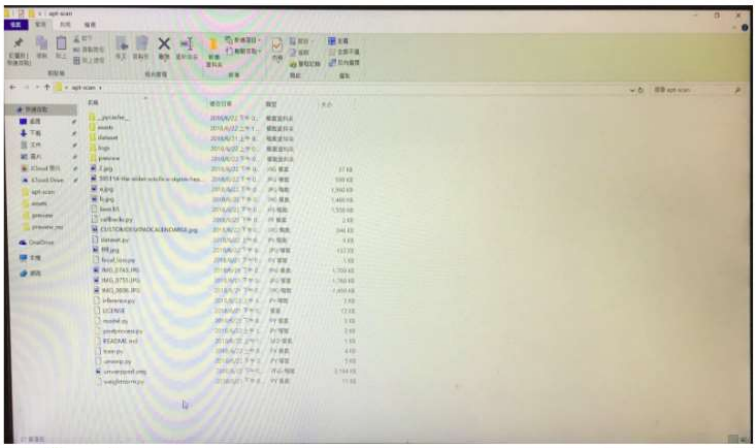
賣方：53665764

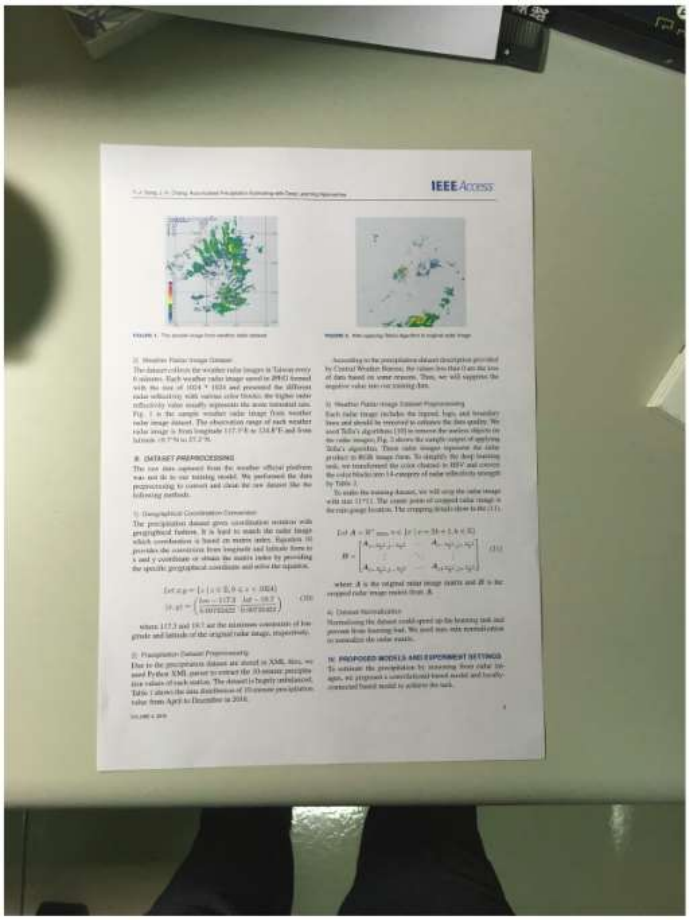
016781民雄神農店01序24151714

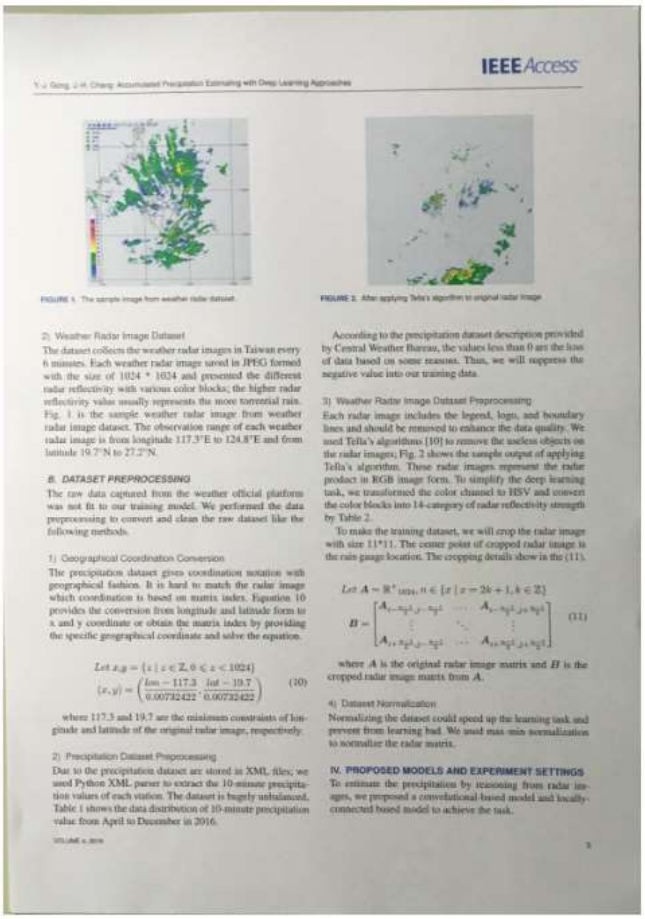
**退貨時請攜帶電子發票證明聯



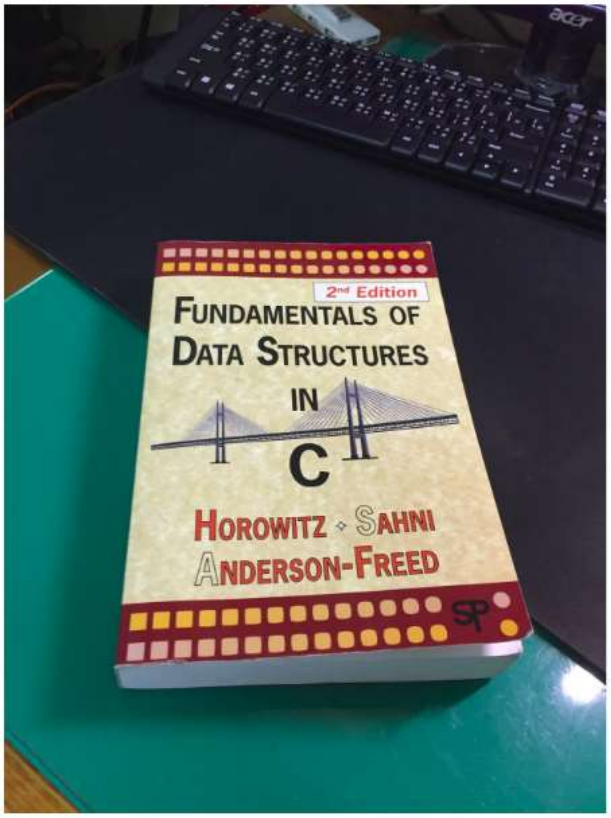


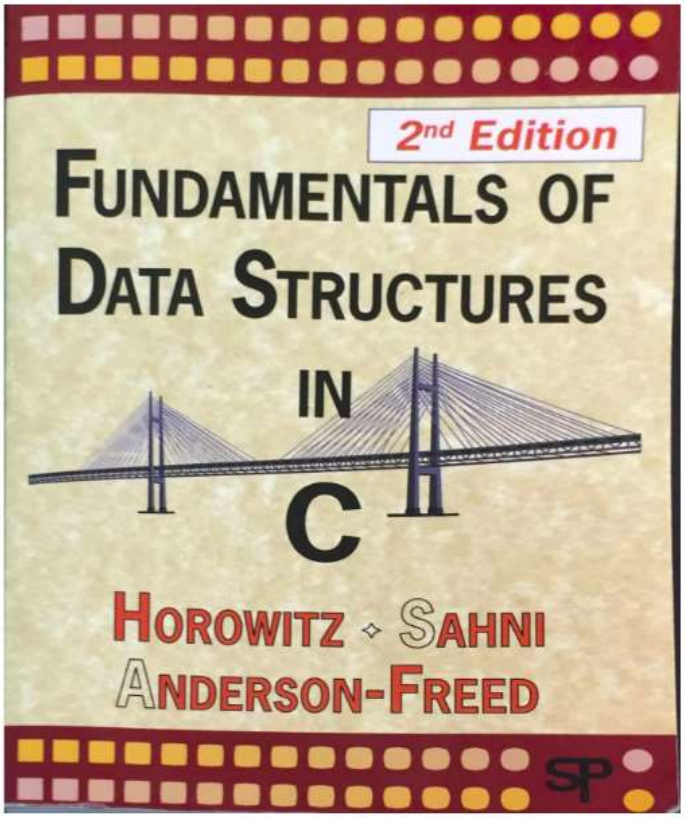






Failure Case





Incorrect Aspect Ratio

Corner Detection Failure