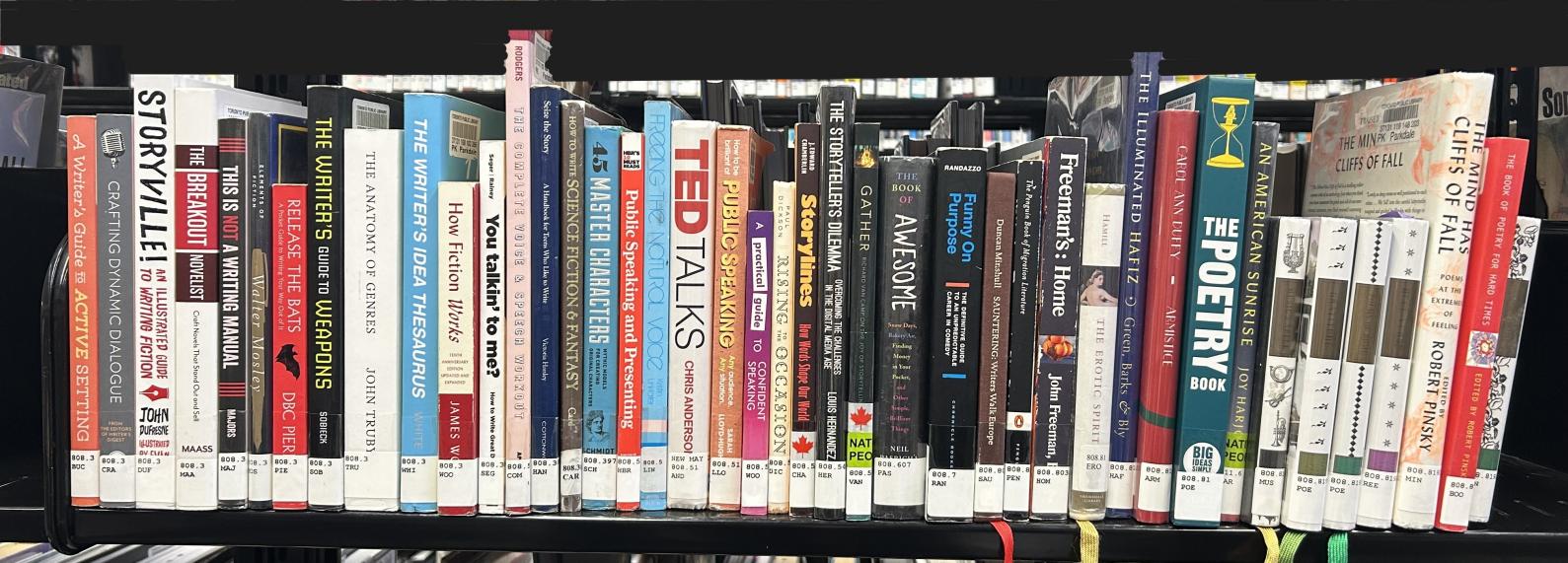
Tested four pictures.

IMG\_7940.jpeg:



Bookshelves\_1.jpg:



Bookshelves\_4.jpg:



Bookshelves\_5.jpg:



# Metric

For a picture of a bookshelf, the first step in trying to determine the accuracy of the api and algorithm is to find the names of all the books by the human eye(***Human Eye Book Name List***), Then, applied API and Algorithm to get all names of books by Computing(***Detected Book Name List***). Finally, compare the two list of book names to get a ***confusion matrix***.

## OCR Confusion Matrix

Expect OCR to return only all the text that appears in the image, not all the book names

1. Matched By Both: Text exists in both ***Human Eye Book Name List*** and ***Detected Book Name List***
2. Only by OCR: Text exists in only ***Detected Book Name List*** but not ***Human Eye Book Name List***
3. OCR Book Name Total Amount: Total number of Text in ***Detected Book Name List***
4. Only by Human: Text exists in only ***Human Eye Book Name List*** but not ***Detected Book Name List***
5. Human Eye Book Name Total Amount: Total number of Text in ***Human Eye Book Name List***

## Text Group Confusion Matrix

Expect algorithm to only group book names without sorting

1. Matched By Both: Book Names exists in both ***Human Eye Book Name List*** and ***Detected Book Name List***
2. Only by Algorithm: Book Names exists in only ***Detected Book Name List*** but not ***Human Eye Book Name List***
3. Algorithm Book Name Total Amount: Total number of Book Names in ***Detected Book Name List***
4. Only by Human: Book Names exists in only ***Human Eye Book Name List*** but not ***Detected Book Name List***
5. Human Eye Book Name Total Amount: Total number of Book Names in ***Human Eye Book Name List***

Recall = ***Matched By Both*** / number of ***Detected Book Name List***

Precision = ***Matched By Both*** / number of ***Human Eye Book Name List***

# OCR API Accuracy

The followings are Text and Book Names accuracy for Google Vision api and Amazon Rekognition:

Google IMG\_7940.jpeg:

=========================================================

GOOGLE OCR:

./pics/IMG\_7940.jpeg GOOGLE Text Confusion Matrix:

Matched by both 258

Only by OCR 166

OCR Text Total Amount 424

Only by Human 17

Human Eye Book Name Total Amount 275

OCR Text Recall 0.6084905660377359

OCR Text Precision 0.9381818181818182

./pics/IMG\_7940.jpeg GOOGLE Book Names Confusion Matrix:

Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 182

OCR Book Name Total Amount 182

Only by Human 37

Human Eye Book Name Total Amount 37

OCR Book Name Recall 0.0

OCR Book Name Precision 0.0

=========================================================

Google Bookshelves\_1.jpg:

=========================================================

./pics/Bookshelves\_1.jpg GOOGLE Text Confusion Matrix:

Matched by both 272

Only by OCR 238

OCR Text Total Amount 510

Only by Human 20

Human Eye Book Name Total Amount 292

OCR Text Recall 0.5333333333333333

OCR Text Precision 0.9315068493150684

./pics/Bookshelves\_1.jpg GOOGLE Book Names Confusion Matrix:

Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 191

OCR Book Name Total Amount 191

Only by Human 58

Human Eye Book Name Total Amount 58

OCR Book Name Recall 0.0

OCR Book Name Precision 0.0

=========================================================

Google Bookshelves\_4.jpg:

=========================================================

./pics/Bookshelves\_4.jpg GOOGLE Text Confusion Matrix:

Matched by both 341

Only by OCR 98

OCR Text Total Amount 439

Only by Human 97

Human Eye Book Name Total Amount 438

OCR Text Recall 0.7767653758542141

OCR Text Precision 0.7785388127853882

./pics/Bookshelves\_4.jpg GOOGLE Book Names Confusion Matrix:

Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 198

OCR Book Name Total Amount 198

Only by Human 95

Human Eye Book Name Total Amount 95

OCR Book Name Recall 0.0

OCR Book Name Precision 0.0

=========================================================

Google Bookshelves\_5.jpg:

./pics/Bookshelves\_5.jpg GOOGLE Text Confusion Matrix:

Matched by both 442

Only by OCR 192

OCR Text Total Amount 634

Only by Human 97

Human Eye Book Name Total Amount 539

OCR Text Recall 0.6971608832807571

OCR Text Precision 0.8200371057513914

./pics/Bookshelves\_5.jpg GOOGLE Book Names Confusion Matrix:

Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 255

OCR Book Name Total Amount 255

Only by Human 155

Human Eye Book Name Total Amount 155

OCR Book Name Recall 0.0

OCR Book Name Precision 0.0

=========================================================

Amazon IMG\_7940.jpeg:

=========================================================

GOOGLE OCR:

./pics/IMG\_7940.jpeg GOOGLE Text Confusion Matrix:

Matched by both 66

Only by OCR 34

OCR Text Total Amount 100

Only by Human 209

Human Eye Book Name Total Amount 275

OCR Text Recall 0.66

OCR Text Precision 0.24

./pics/IMG\_7940.jpeg GOOGLE Book Names Confusion Matrix:

Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 47

OCR Book Name Total Amount 47

Only by Human 37

Human Eye Book Name Total Amount 37

OCR Book Name Recall 0.0

OCR Book Name Precision 0.0

=========================================================

Amazon Bookshelves\_1.jpg:

=========================================================

./pics/Bookshelves\_1.jpg GOOGLE Text Confusion Matrix:

Matched by both 111

Only by OCR 89

OCR Text Total Amount 200

Only by Human 181

Human Eye Book Name Total Amount 292

OCR Text Recall 0.555

OCR Text Precision 0.3801369863013699

./pics/Bookshelves\_1.jpg GOOGLE Book Names Confusion Matrix:

Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 106

OCR Book Name Total Amount 106

Only by Human 58

Human Eye Book Name Total Amount 58

OCR Book Name Recall 0.0

OCR Book Name Precision 0.0

=========================================================

Amazon Bookshelves\_4.jpg:

=========================================================

./pics/Bookshelves\_4.jpg GOOGLE Text Confusion Matrix:

Matched by both 241

Only by OCR 188

OCR Text Total Amount 429

Only by Human 197

Human Eye Book Name Total Amount 438

OCR Text Recall 0.5617715617715617

OCR Text Precision 0.5502283105022832

./pics/Bookshelves\_4.jpg GOOGLE Book Names Confusion Matrix:

Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 251

OCR Book Name Total Amount 251

Only by Human 95

Human Eye Book Name Total Amount 95

OCR Book Name Recall 0.0

OCR Book Name Precision 0.0

=========================================================

Amazon Bookshelves\_5.jpg:

./pics/Bookshelves\_5.jpg GOOGLE Text Confusion Matrix:

Matched by both 176

Only by OCR 124

OCR Text Total Amount 300

Only by Human 363

Human Eye Book Name Total Amount 539

OCR Text Recall 0.5866666666666667

OCR Text Precision 0.32653061224489793

./pics/Bookshelves\_5.jpg GOOGLE Book Names Confusion Matrix:

Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 129

OCR Book Name Total Amount 129

Only by Human 155

Human Eye Book Name Total Amount 155

OCR Book Name Recall 0.0

OCR Book Name Precision 0.0

=========================================================

## Conclusion

The test shows that Google OCR Text Detection has a relatively high accuracy (93.8%) when the books are placed neatly, and an acceptable accuracy (79%) when books are in mess.



Figure 1.0: Sample Amazon Rekognition Detection

Amazon OCR Text Detection sometimes ignore a part of the picture therefore result in a low accuracy, neat (24%), mess (71%).

Neither lines detection for Google nor Amazon can extract correct book names (0% for all pictures)  
Object Detection also not work in most case (Of the 22 books, 0 were detected by Google and 5 by Amazon)

Therefore, we only use Google OCR Text Detection for the later analysis.

# Text Group Algorithm

Since the position of books cannot be retrieved easily, due to limited number of training data, (Training Data require: 1. picture of bookshelf; 2. True Book Names or True Book Objects Position;) the best way to group them is to develop an algorithm to calculate text positions.

## Vertical Line Algorithm

For each word, calculate its **slope** and **center point**, use the **slope and vertices of word1** to get two line equation and check if the **center point of word 2** is between the two line. If so, group word1 and word2 together, else, continue for next pair of words.

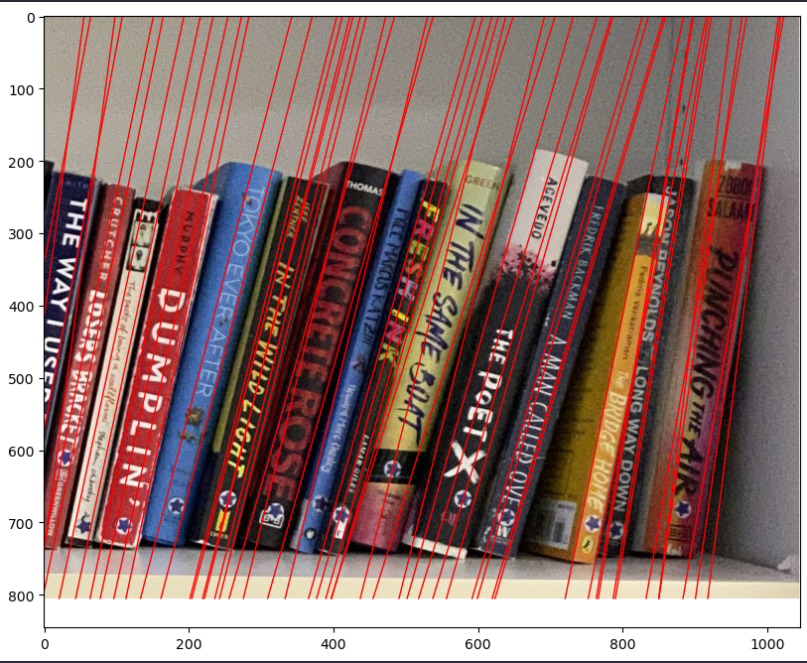


Figure 2.0: Schematic of vertical line algorithm

After testing, this algorithm provide a poor accuracy:

./pics/IMG\_7940.jpeg Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 87

Only by Human 37

Recall 0.0

Precision 0.0

=========================================================

./pics/Bookshelves\_1.jpg Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 69

Only by Human 58

Recall 0.0

Precision 0.0

=========================================================

./pics/Bookshelves\_4.jpg Book Name Group Larger than 80%:

Matched by both 1

Only by OCR 119

Only by Human 94

Recall 0.010526315789473684

Precision 0.008333333333333333

=========================================================

./pics/Bookshelves\_5.jpg Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 153

Only by Human 155

Recall 0.0

Precision 0.0

=========================================================

## Same Weight Ellipse Group Algorithm

For each word, calculate its **center point** and an **ellipse** shape around it, the height and width of ellipse is related to the height and width with a constant (dt) of each word, check if **center point of word 2** is in the ellipse area, If so, group word1 and word2 together, else, continue for next pair of words.

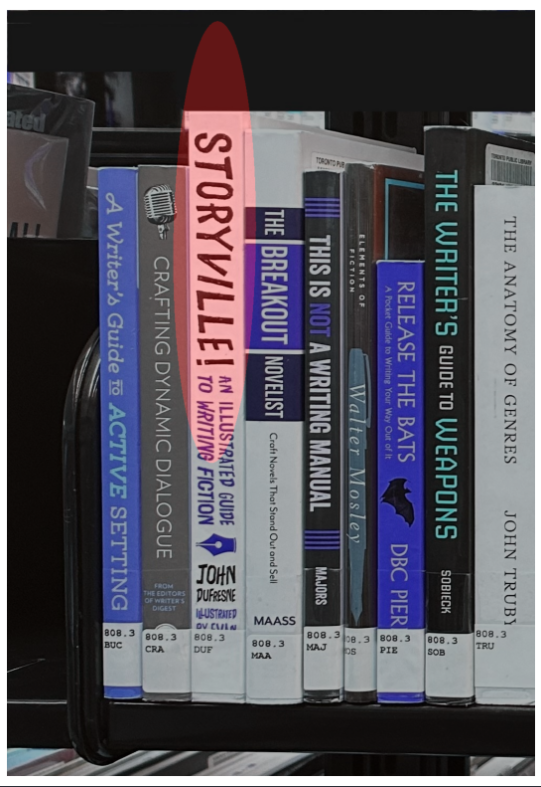


Figure 2.0: Schematic of Same Weight Ellipse Group algorithm (dt=1, randomly selected)

After testing, this algorithm provide a much better accuracy:

./pics/IMG\_7940.jpeg Book Name Group Larger than 80%:

Matched by both 8

Only by OCR 160

Only by Human 29

Recall 0.047619047619047616

Precision 0.21621621621621623

=========================================================

./pics/Bookshelves\_1.jpg Book Name Group Larger than 80%:

Matched by both 9

Only by OCR 237

Only by Human 49

Recall 0.036585365853658534

Precision 0.15517241379310345

=========================================================

./pics/Bookshelves\_4.jpg Book Name Group Larger than 80%:

Matched by both 4

Only by OCR 156

Only by Human 91

Recall 0.025

Precision 0.042105263157894736

=========================================================

./pics/Bookshelves\_5.jpg Book Name Group Larger than 80%:

Matched by both 5

Only by OCR 105

Only by Human 150

Recall 0.045454545454545456

Precision 0.03225806451612903

=========================================================

## Different Weight Ellipse Group Algorithm

This algorithm is similar to Same Weight Ellipse Group Algorithm, but the height and width of ellipse is related to the height and width with **constant ydt and constant xdt** respectively of each word.



Figure 2.0: Schematic of Different Weight Ellipse Group algorithm (ydt=5, xdt=1, randomly selected)

Until now, this algorithm provide a similar accuracy as the previous algorithm:

./pics/IMG\_7940.jpeg Book Name Group Larger than 80%:

Matched by both 7

Only by OCR 43

Only by Human 30

Recall 0.14

Precision 0.1891891891891892

=========================================================

./pics/Bookshelves\_1.jpg Book Name Group Larger than 80%:

Matched by both 8

Only by OCR 64

Only by Human 50

Recall 0.1111111111111111

Precision 0.13793103448275862

=========================================================

./pics/Bookshelves\_4.jpg Book Name Group Larger than 80%:

Matched by both 13

Only by OCR 62

Only by Human 82

Recall 0.17333333333333334

Precision 0.1368421052631579

=========================================================

./pics/Bookshelves\_5.jpg Book Name Group Larger than 80%:

Matched by both 6

Only by OCR 60

Only by Human 149

Recall 0.09090909090909091

Precision 0.03870967741935484

=========================================================

## KMeans

KMeans is an unsupervised machine learning algorithm used for clustering. It partitions observations into a specified number of clusters (K) by minimizing the variance within each cluster. The algorithm initializes with K random centroids and assigns each data point to the nearest centroid, forming clusters. Centroids are then recalculated as the mean of all points in the cluster. This process repeats iteratively until the centroids stabilize, meaning their positions no longer change significantly. KMeans is widely used for its simplicity and efficiency in grouping data into clusters based on similarity, but choosing the right K and handling outliers can be challenging. I choose K=40 and test the pictures.

./pics/IMG\_7940.jpeg Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 40

Only by Human 37

Recall 0.0

Precision 0.0

=========================================================

./pics/Bookshelves\_1.jpg Book Name Group Larger than 80%:

Matched by both 3

Only by OCR 77

Only by Human 55

Recall 0.0375

Precision 0.05172413793103448

========================================================

./pics/Bookshelves\_4.jpg Book Name Group Larger than 80%:

Matched by both 5

Only by OCR 155

Only by Human 90

Recall 0.03125

Precision 0.05263157894736842

=========================================================

./pics/Bookshelves\_5.jpg Book Name Group Larger than 80%:

Matched by both 0

Only by OCR 120

Only by Human 155

Recall 0.0

Precision 0.0

=========================================================

## Conclusion

Test Same Weight Ellipse Group Algorithm and Different Weight Ellipse Group Algorithm now.

# Hyperparameter Tuning

Since xdt, ydt and dt are randomly selected, I different combination to find the best xdt, ydt and dt.

## Same Weight Ellipse Group Algorithm Tuning

|  |  |  |
| --- | --- | --- |
| Picture | Max TP Number | dt |
| IMG\_7940.jpeg | 14 | 1.1 |
| Bookshelves\_1.jpg | 10 | 1.0 |
| Bookshelves\_1.jpg | 10 | 1.1 |
| Bookshelves\_4.jpg | 12 | 1.1 |
| Bookshelves\_5.jpg | 9 | 1.2 |

Choose 1.1 as dt

./pics/IMG\_7940.jpeg Book Name Group Larger than 80%:

Matched by both 11

Only by OCR 128

Only by Human 26

Recall 0.07913669064748201

Precision 0.2972972972972973

=========================================================

./pics/Bookshelves\_1.jpg Book Name Group Larger than 80%:

Matched by both 10

Only by OCR 173

Only by Human 48

Recall 0.0546448087431694

Precision 0.1724137931034483

=========================================================

./pics/Bookshelves\_4.jpg Book Name Group Larger than 80%:

Matched by both 10

Only by OCR 133

Only by Human 85

Recall 0.06993006993006994

Precision 0.10526315789473684

=========================================================

./pics/Bookshelves\_5.jpg Book Name Group Larger than 80%:

Matched by both 7

Only by OCR 90

Only by Human 148

Recall 0.07216494845360824

Precision 0.04516129032258064

=========================================================

For all pictures, the algorithm performs poor.

## Different Weight Ellipse Group Algorithm Tuning

For different picture, the best (xdt, ydt) are different,

|  |  |  |  |
| --- | --- | --- | --- |
| Picture | Max TP Number | xdt | ydt |
| IMG\_7940.jpeg | 17 | 0.7 | 1.5 |
| IMG\_7940.jpeg | 17 | 0.8 | 1.5 |
| Bookshelves\_1.jpg | 15 | 0.4 | 4.0 |
| Bookshelves\_1.jpg | 15 | 0.7 | 1.5 |
| Bookshelves\_1.jpg | 15 | 0.7 | 3.5 |
| Bookshelves\_1.jpg | 15 | 0.7 | 4.0 |
| Bookshelves\_4.jpg | 18 | 0.3 | 3.5 |
| Bookshelves\_4.jpg | 18 | 0.4 | 2.5 |
| Bookshelves\_5.jpg | 14 | 0.4 | 2.0 |

Since (xdt, ydt) combination varies, I choose the most often appeared 0.7 as xdt and 1.5 as ydt.

./pics/IMG\_7940.jpeg Book Name Group Larger than 80%:

Matched by both 17

Only by OCR 120

Only by Human 20

Recall 0.12408759124087591

Precision 0.4594594594594595

=========================================================

./pics/Bookshelves\_1.jpg Book Name Group Larger than 80%:

Matched by both 15

Only by OCR 157

Only by Human 43

Recall 0.0872093023255814

Precision 0.25862068965517243

=========================================================

./pics/Bookshelves\_4.jpg Book Name Group Larger than 80%:

Matched by both 15

Only by OCR 116

Only by Human 80

Recall 0.11450381679389313

Precision 0.15789473684210525

=========================================================

./pics/Bookshelves\_5.jpg Book Name Group Larger than 80%:

Matched by both 10

Only by OCR 89

Only by Human 145

Recall 0.10101010101010101

Precision 0.06451612903225806

=========================================================

For IMG\_7940.jpeg only, the algorithm remain an acceptable precision, it performs bad for the rest.