Northeastern Experiential Network

Experiential Project – ReMo

Team Members	Yiu Kwan Leung, Zirui Qiao
Team Lead	Ka Lai Yeung
Project Description	The objective of the project is to incorporate a bulk import feature into the ReMo web-based application, enabling educators to import photos of library bookshelves. These photos will undergo automatic processing utilizing Optical Character Recognition (OCR) and Artificial Intelligence (AI) algorithms to precisely identify and extract book titles. The extracted data will be converted into a comprehensive book list, seamlessly imported into the ReMo system. Throughout the project duration, the team will explore innovative approaches to improve the cataloging process, ensuring both efficiency and accuracy when handling a substantial number of books.

Deliverable

Status Summary

We have received 4 samples of photos taken at the bookshelves from Remo and public library. These photos underwent analysis using both Google OCR and Amazon OCR to evaluate the efficiency of their services. Our algorithm then compares the results obtained from human assessment with those from OCR to gauge the accuracy of the OCR services. We anticipate an enhancement in accuracy following the ongoing development of our algorithm.

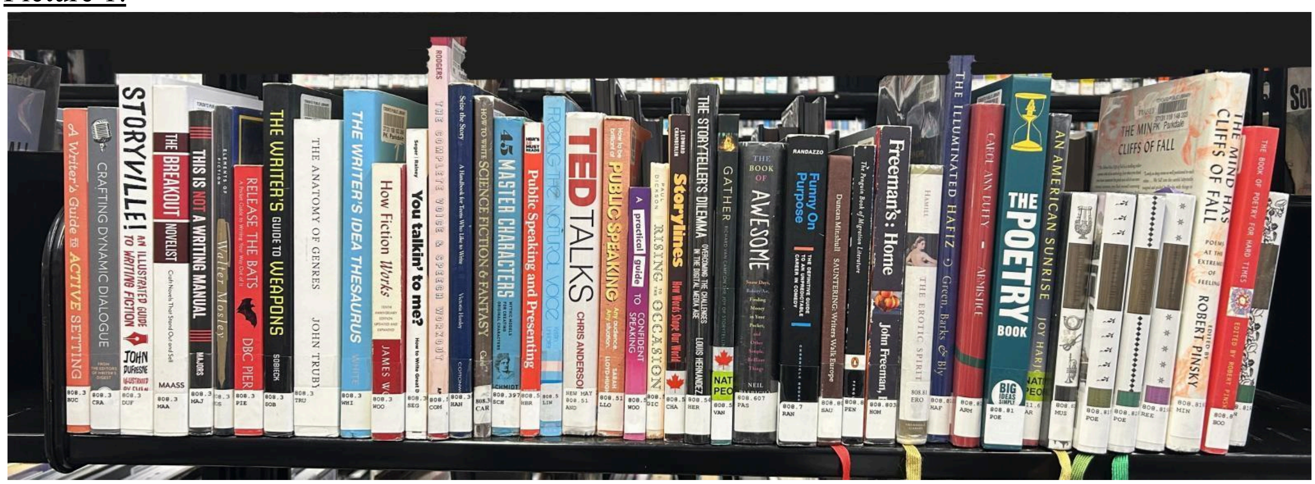
Methodology

1. Manually identify the titles of all books on the bookshelf by human assessment.
2. Utilize the API and algorithm to automatically extract the book titles.
3. Compare the manually identified titles with the automatically detected titles to create a confusion matrix.
4. OCR Confusion Matrix: OCR detects only words that appears in the image, but not book names.
5. Text Group Confusion Matrix: Algorithm detects and returns group book names without sorting.

Results

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

Picture 1:



OCR Confusion Matrix

	Google Vision API	Amazon Rekognition
Human Readable Word Count	275	275
OCR Readable Word Count	424	100
1. Matched by both	258	66
2. Only by OCR	166	34
3. Only by Human	17	209
Matched Precision (with respect to Human Readable Word Count)	93.8%	24%
Matched Precision (with respect to OCR Readable Word Count)	60.8%	66%

Text Group Confusion Matrix

	Google Vision API	Amazon Rekognition
Human Readable Book Name Count	37	37
OCR Readable Book Name Count	182	47
1. Matched by both	0	0
2. Only by OCR	182	47
3. Only by Human	37	37
Matched Precision (with respect to Human Readable Book Name Count)	0%	0%

Matched Precision (with respect to OCR Readable Book Name Count)	0%	0%

Picture 2:

OCR Confusion Matrix

	Google Vision API	Amazon Rekognition
Human Readable Word Count	292	292
OCR Readable Word Count	510	200
1. Matched by both	272	111
2. Only by OCR	238	89
3. Only by Human	20	181
Matched Precision (with respect to Human Readable Word Count)	93.2%	38%
Matched Precision (with respect to OCR Readable Word Count)	53.3%	55.5%

Text Group Confusion Matrix

	Google Vision API	Amazon Rekognition
Human Readable Book Name Count	58	58
OCR Readable Book Name Count	191	106
1. Matched by both	0	0
2. Only by OCR	191	106

3. Only by Human	58	58
Matched Precision (with respect to Human Readable Book Name Count)	0%	0%
Matched Precision (with respect to OCR Readable Book Name Count)	0%	0%

Picture 3:

OCR Confusion Matrix

	Google Vision API	Amazon Rekognition
Human Readable Word Count	438	438
OCR Readable Word Count	439	429
1. Matched by both	341	241
2. Only by OCR	98	188
3. Only by Human	97	197

Matched Precision (with respect to Human Readable Word Count)	77.9%	55%
Matched Precision (with respect to OCR Readable Word Count)	77.7%	56.2%

Text Group Confusion Matrix

	Google Vision API	Amazon Rekognition
Human Readable Book Name Count	95	95
OCR Readable Book Name Count	198	251
1. Matched by both	0	0
2. Only by OCR	198	251
3. Only by Human	95	95
Matched Precision (with respect to Human Readable Book Name Count)	0%	0%
Matched Precision (with respect to OCR Readable Book Name Count)	0%	0%

Picture 4:

OCR Confusion Matrix

	Google Vision API	Amazon Rekognition
Human Readable Word Count	539	539
OCR Readable Word Count	634	300
1. Matched by both	442	176
2. Only by OCR	192	124
3. Only by Human	97	363
Matched Precision (with respect to Human Readable Word Count)	82%	32.7%
Matched Precision (with respect to OCR Readable Word Count)	69.7%	58.7%

Text Group Confusion Matrix

	Google Vision API	Amazon Rekognition
Human Readable Book Name Count	155	155
OCR Readable Book Name Count	255	129
1. Matched by both	0	0
2. Only by OCR	255	129
3. Only by Human	155	155
Matched Precision (with respect to Human Readable Book Name Count)	0%	0%
Matched Precision (with respect to OCR Readable Book Name Count)	0%	0%

Discussion

<u>OCR Word Detection</u>

Figure 1.0: Sample Amazon Rekognition Detection

The evaluation indicates that Google OCR Text Detection achieves a high accuracy rate of 93.8% with neatly arranged books and a satisfactory accuracy rate of 79% when books are disorganized. In contrast, Amazon OCR Text Detection often overlooks sections of the image, leading to a significantly lower accuracy of 24% for neatly placed books and 71% for disorganized books. Additionally, neither Google's nor Amazon's line detection methods succeeded in accurately identifying book titles, with a 0% success rate across all images. Object Detection was also largely ineffective, failing to recognize any of the 22 books in Google's case and only identifying 5 with Amazon. As a result, Google OCR Text Detection will be the sole method used for subsequent analysis.

Text Group Detection Algorithm

Given the challenge of accurately determining the positions of books, largely due to a limited dataset for training (which requires: 1. images of bookshelves; 2. accurate book names or the precise positions of book objects), the most effective approach is to create an algorithm designed to calculate the positions of text.

I. 4 Algorithms have been created and tested for efficiency.

1. Vertical Line Algorithm

For each word, the process involves calculating its slope and centre point. Following this, the slope and vertices of the first word are utilized to form two line equations. The method then checks whether the centre point of the second word lies within the boundaries defined by these two lines. If the centre point of the second word is indeed between these two lines, then the first and second words are grouped together. If it does not, the analysis continues with the next pair of words.

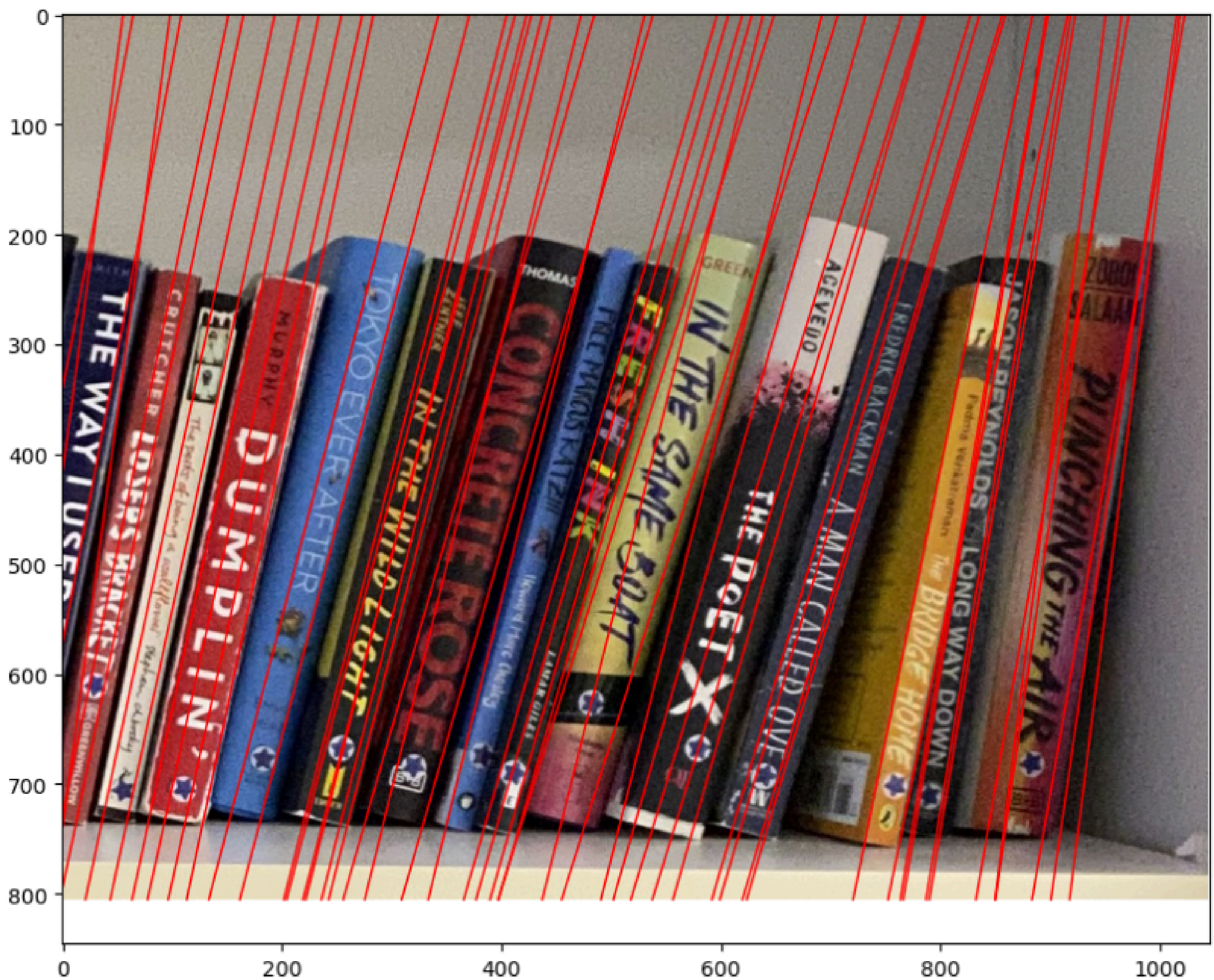


Figure 2.0: Schematic of vertical line algorithm

After testing, this algorithm provides a poor accuracy:

	Picture 1	Picture 2	Picture 3	Picture 4
1. Matched by both	0	0	1	0
2. Only by OCR	87	69	119	153
3. Only by Human	37	58	94	155
Matched Precision (with respect to Human Readable Book Name Count)	0%	0%	0.8%	0%
Matched Precision (with respect to OCR Readable Book Name Count)	0%	0%	1%	0%

(* When evaluating a book name, only cases where at least 80% of the words match are considered a valid match.)

2. Same Weight Ellipse Group Algorithm

For each word, the procedure entails calculating its centre point and creating an ellipse shape around it, with the ellipse's height and width being proportional to those of the word, adjusted by a constant factor (**dt**). Subsequently, it is determined whether the centre point of another word (word 2) falls within this elliptical area. If the centre point of word 2 is indeed within the ellipse, then word 1 and word 2 are grouped together. Otherwise, the analysis moves on to the next pair of words to evaluate their potential grouping.

Figure 3.0: Schematic of Same Weight Ellipse Group algorithm (dt=1, randomly selected)

After testing, this algorithm provide a much better accuracy:

	Picture 1	Picture 2	Picture 3	Picture 4
1. Matched by both	8	9	4	5
2. Only by OCR	160	237	156	105
3. Only by Human	29	49	91	150
Matched Precision (with respect to Human Readable Book Name Count)	21.6%	15.5%	4.2%	3.2%
Matched Precision (with respect to OCR Readable Book Name Count)	4.8%	3.7%	2.5%	4.5%

(* When evaluating a book name, only cases where at least 80% of the words match are considered a valid match.)

3. Different Weight Ellipse Group Algorithm

This algorithm bears resemblance to the Same Weight Ellipse Group Algorithm, with a key distinction: the height and width of the ellipse are determined in relation to the height and width of each word, adjusted by separate **constants ydt** and **xdt**, respectively.

Figure 4.0: Schematic of Different Weight Ellipse Group algorithm (ydt=5, xdt=1, randomly selected)

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

	Picture 1	Picture 2	Picture 3	Picture 4
1. Matched by both	7	8	13	6
2. Only by OCR	43	64	62	60
3. Only by Human	30	50	82	149
Matched Precision (with respect to Human Readable Book Name Count)	18.9%	13.8%	13.7%	3.8%
Matched Precision (with respect to OCR Readable Book Name Count)	14%	11.1%	17.3%	9%

(* When evaluating a book name, only cases where at least 80% of the words match are considered a valid match.)

4. Unsupervised Learning Clustering

KMeans is a popular unsupervised machine learning algorithm designed for clustering tasks. It organizes observations into a predetermined number of clusters (denoted as K) by aiming to minimize the within-cluster variance. The process starts by randomly selecting K centroids, then assigns each data point to the nearest centroid, effectively forming clusters. After assignment, the centroids are recalculated to be the mean of all points in their respective clusters. This cycle of assignment and recalibration repeats iteratively until the centroids reach a point of stability, where their positions do not undergo significant changes. The simplicity and efficiency of KMeans make it a favored choice for identifying groups within data based on similarities. However, challenges arise in selecting an appropriate value for K and managing outliers. For this task, K=40 is selected as the number of clusters and proceeded to test the algorithm on the images.

	Picture 1	Picture 2	Picture 3	Picture 4
1. Matched by both	0	3	5	0
2. Only by OCR	40	77	155	120
3. Only by Human	37	55	90	155
Matched Precision (with respect to Human Readable Book Name Count)	0%	5.2%	5.3%	0%
Matched Precision (with respect to OCR Readable Book Name Count)	0%	3.8%	3.1%	0%

(* When evaluating a book name, only cases where at least 80% of the words match are considered a valid match.)

II. Considering that both the **Same Weight Ellipse Group Algorithm** and the **Different Weight Ellipse Group Algorithm** exhibit better accuracy, these methodologies will be employed for additional testing phases.

1. Hyperparameter Tuning

Since the values for xdt, ydt, and dt are chosen randomly, various combinations were explored to determine the optimal values for xdt, ydt, and dt.

Same Weight Ellipse Group Algorithm Tuning

dt = 1.1

Picture	Max TP Number	dt
Picture 1	14	1.1
Picture 2	10	1.0
Picture 3	10	1.1
Picture 4	12	1.1
Picture 5	9	1.2

	Picture 1	Picture 2	Picture 3	Picture 4
1. Matched by both	11	10	10	7
2. Only by OCR	128	173	133	90
3. Only by Human	26	48	85	148
Matched Precision (with respect to Human Readable Book Name Count)	29.7%	17.2%	10.5%	4.5%
Matched Precision (with respect to OCR Readable Book Name Count)	7.9%	5.5%	7%	7.2%

(* When evaluating a book name, only cases where at least 80% of the words match are considered a valid match.)

For all pictures, the algorithm does not perform well.

2. Different Weight Ellipse Group Algorithm Tuning

Given that the optimal (xdt, ydt) pairings differ across various images, and considering the variability in the (xdt, ydt) combinations, the most frequently occurring values, 0.7 for xdt and 1.5 for ydt were chosen as the standard parameters.

Picture	Max TP Number	xdt	ydt
Picture 1	17	0.7	1.5
Picture 1	17	0.8	1.5
Picture 2	15	0.4	4.0
Picture 2	15	0.7	1.5
Picture 2	15	0.7	3.5
Picture 2	15	0.7	4.0
Picture 3	18	0.3	3.5
Picture 3	18	0.4	2.5
Picture 4	14	0.4	2.0

	Picture 1	Picture 2	Picture 3	Picture 4
1. Matched by both	17	15	15	10
2. Only by OCR	120	157	116	89
3. Only by Human	20	43	80	145
Matched Precision (with respect to Human Readable Book Name Count)	45.9%	25.9%	15.8%	6.5%
Matched Precision (with respect to OCR Readable Book Name Count)	12.4%	8.7%	11.5%	10.1%

(* When evaluating a book name, only cases where at least 80% of the words match are considered a valid match.)

For picture 1 alone, the algorithm maintains an acceptable level of precision, but its performance deteriorates for the remainder of the images.

Conclusion & Future Action

The project aims to automate the identification of book titles using OCR technology on images captured from bookshelves, subsequently searching for these titles in a database to retrieve and record their ISBN codes. Currently, the focus is on evaluating the effectiveness of OCR for this purpose. However, several challenges have been encountered, including unclear photographs, obstructions in front of the bookshelf, stickers obscuring book titles, incomplete titles on book spines, small text size, and insufficient spine information to ascertain ISBN numbers. Thus, relying solely on images of book spines proves inadequate for extracting comprehensive book information.

A proposed next step involves taking photographs of book covers or backs, which may display more detailed information, albeit less efficiently than capturing images of book spines. Another strategy could involve photographing open pages that display the ISBN, potentially capturing up to 10 books in a single image. This method could expedite data entry more effectively than scanning individual codes.

Despite the challenges, the project's concept remains promising, especially for streamlining the process of entering book data into databases for teachers. Optimizing OCR's application is crucial for enhancing its utility and achieving the project's goals.