

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KHOA HỌC MÁY TÍNH



BÁO CÁO ĐỒ ÁN CUỐI KÌ
SỐ HÓA TỬ SÁCH

Giảng viên hướng dẫn:

TS. Lê Đình Duy

ThS. Phạm Nguyễn Trường An

Sinh viên thực hiện:

21520725 – Bùi Lê Trọng Đức

21521472 – Nguyễn Tiến Thịnh

21522614 – Nguyễn Chí Thi

TP. Hồ Chí Minh, tháng 01 năm 2024

Mục lục

Tóm tắt đồ án	4
Chương 0: Cập nhật sau vấn đáp	5
Chương 1: Giới thiệu đồ án.....	5
Chương 2: Xây dựng bộ dữ liệu.....	6
1. Thu thập dữ liệu.....	6
2. Xử lý dữ liệu	7
3. Mô tả bộ dữ liệu	8
Chương 3: Thực nghiệm.....	8
1. Tổng quan các bước thực hiện.....	8
2. Các bước thực hiện	9
Chương 4: Kết quả đánh giá.....	12
1. Task Text Detection:	12
2. Task Text Recognition:.....	14
3. End to End	15
Chương 5: Kết luận và hướng phát triển.....	15
Tài liệu tham khảo	17

Danh mục hình ảnh

Hình 1: Input	6
Hình 2: Output	6
Hình 3: Giao diện công cụ gán nhãn PPOCRLabel.....	7
Hình 4: Định dạng file gán nhãn dữ liệu	8
Hình 5: Tổng quan quy trình thực hiện.....	8
Hình 6: Input YOLOv8.....	10
Hình 7: Output YOLOv8.....	11
Hình 8: Minh họa kết quả dự đoán của mô hình VietOCR.....	12

Tóm tắt đề án

Đề án “Số hóa tủ sách” được thực hiện với mục đích trích xuất thông tin từ ảnh bìa sách. Đầu vào là ảnh bìa sách được chụp chính diện và đầu ra là các thông tin bao gồm: Tiêu đề, Tác giả, Nhà xuất bản và những nội dung khác. Đề án được thực hiện bao gồm 3 task nhỏ là Text Detection, Text Recognition và Information Synthesis. Tại task Text Detection, nhóm chúng em sử dụng model YOLOv8. Tại task Text Recognition, nhóm sử dụng model VietOCR. Tại task Information Synthesis, nhóm chúng em tổng hợp lại các thông tin được trích xuất từ 2 task trước thành 4 trường thông tin đầu ra là: Tiêu đề, Tác giả, Nhà xuất bản và những nội dung khác. Nhóm tự thu thập và xây dựng bộ dữ liệu bao gồm 964 tấm ảnh bìa sách khác nhau, kết quả trích xuất thông tin đạt được kết quả theo độ đo WER là 62,88%

Github Repository: https://github.com/builetrongduc/CS114_SoHoaTuSach.git

Chương 0: Cập nhật sau vấn đáp

Ở buổi vấn đáp, chúng em chỉ hoàn thành được việc xây dựng bộ dữ liệu và thực nghiệm các task Text Detection và Text Recognition. Chúng em chỉ nhận diện được các vùng chứa văn bản và trích xuất được văn bản có trong vùng được nhận diện.

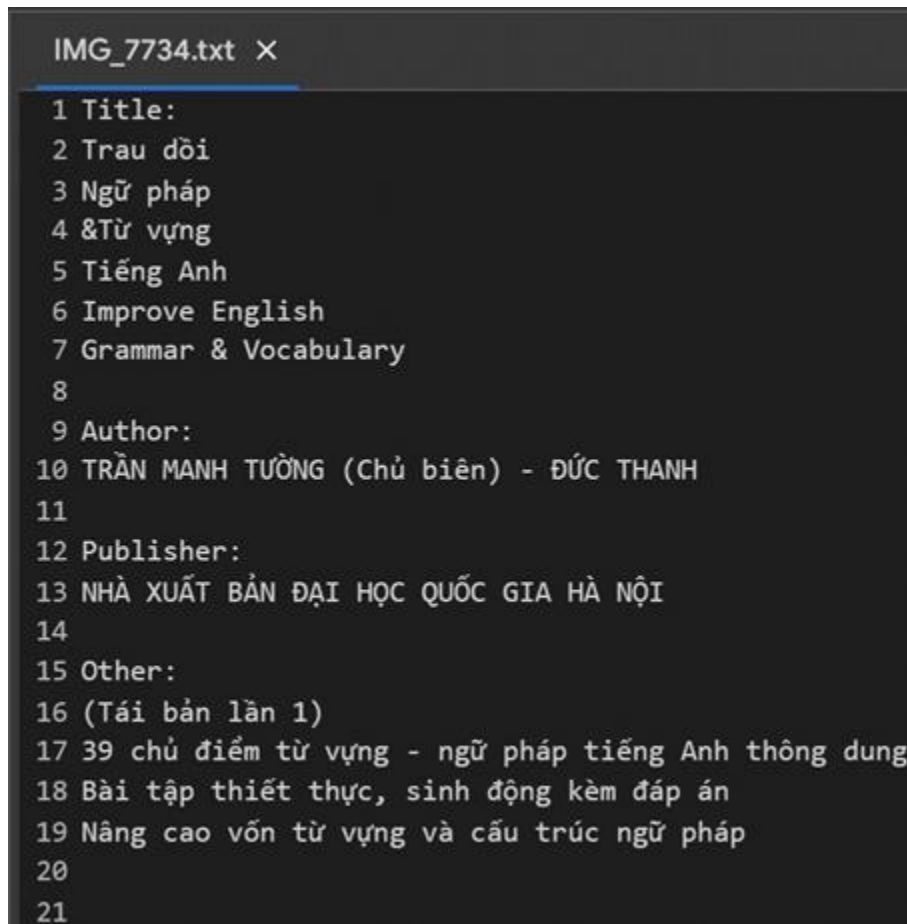
Sau buổi vấn đáp, chúng em đã hoàn thành thêm task Information Synthesis. Cụ thể là sắp xếp lại các phần tử theo tọa độ từ thấp đến cao của hoành độ và tung độ và tổng hợp lại theo các trường: Tiêu đề, Tác giả, Nhà xuất bản và những nội dung khác. Kết quả trả về các file .txt tương ứng với các ảnh, nội dung file text sẽ bao gồm thông tin trích xuất được từ ảnh bìa sách.

Chương 1: Giới thiệu đề án

Trong thời đại hiện nay, sự phát triển không ngừng của công nghệ thông tin đã mở ra nhiều cánh cửa mới, đầy tiềm năng và cơ hội cho sự đổi mới trong nhiều lĩnh vực khác nhau. Trong số đó, lĩnh vực quản lý thông tin đang trở thành một trong những lĩnh vực đầy hứa hẹn. Ví dụ điển hình có thể nhắc đến là quản lý tủ sách tại các thư viện hoặc nhà sách. Trước đây, khi nhập sách mới về các thư viện hoặc nhà sách, nhân viên phải thao tác thủ công như lọc thông tin tên sách, tác giả, nhà xuất bản, sau đó nhập vào máy tính để cho việc quản lý sách được hiệu quả. Tuy nhiên, quá trình thực hiện này tốn rất nhiều thời gian và công sức, mà độ hiệu quả cũng như chính xác không cao. Nhằm bắt được vấn đề này, nhóm chúng em xin trình bày đề án môn học là thực nghiệm phương pháp nhận diện văn bản Tiếng Việt từ ảnh các bìa sách tự chụp. Bài toán nhận đầu vào là một bức ảnh bìa sách, đầu ra là các thông tin trên bìa sách, bao gồm: Title, author, publisher và other. Thông qua đề án này, nhóm hi vọng đây sẽ là một trong những cách tiếp cận hiệu quả, nhằm tối ưu hóa thời gian và độ chính xác trong quá trình quản lý thông tin sách.



Hình 1: Input



Hình 2: Output

Chương 2: Xây dựng bộ dữ liệu

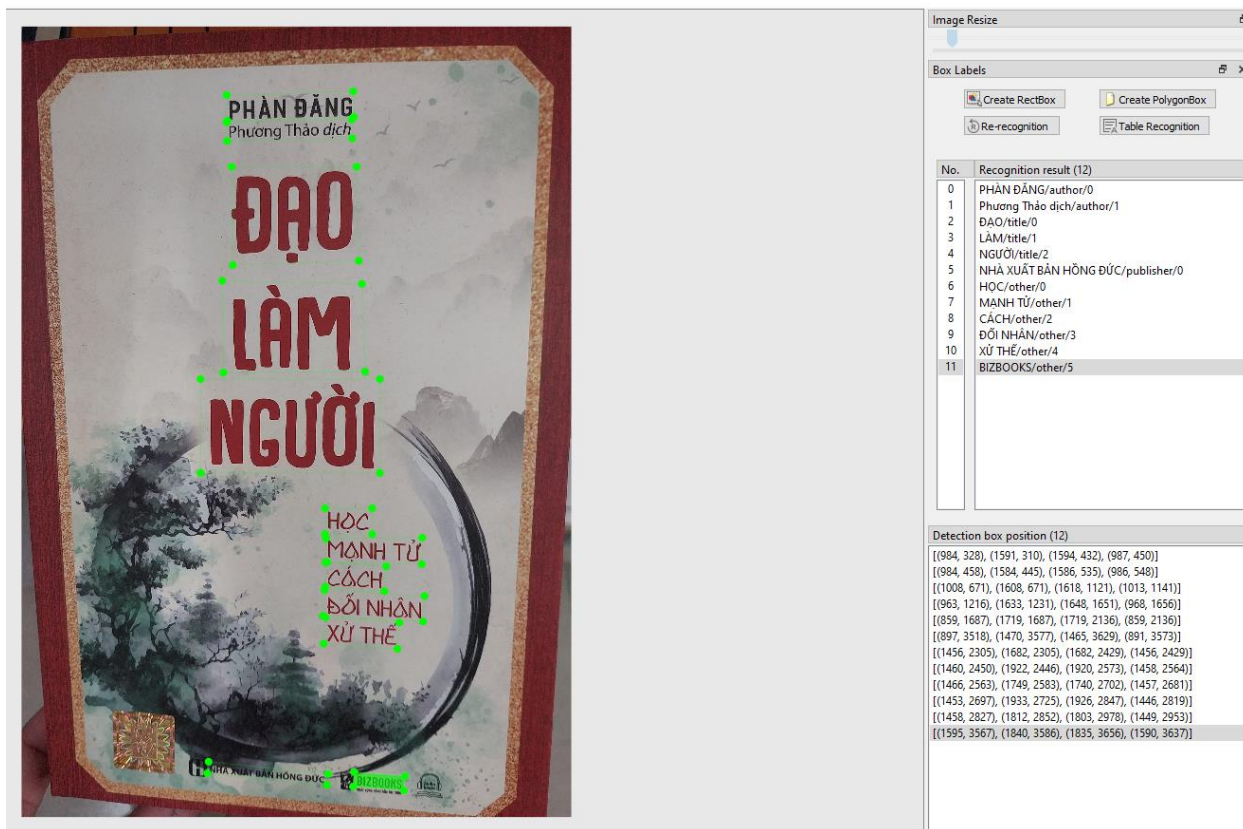
1. Thu thập dữ liệu

Để dữ liệu có chất lượng cao, nhóm chúng em quyết định chỉ thu thập bằng phương pháp thủ công, tức là tự đi chụp ảnh bìa sách tại nhà sách. Các ảnh được chụp bằng các thiết bị di động khác nhau, cụ thể ở đây là sử dụng Iphone, Samsung và Ipad, nhằm đảm bảo tính đa dạng về kích thước, chất lượng hình ảnh của bộ dữ liệu. Các hình bìa sách được yêu cầu chụp ở hướng chính diện, tuy nhiên trong số đó vẫn còn 1 vài tấm bị chói hoặc bị nghiêng nhưng không đáng kể. Ngoài ra, để mang tính đa dạng và phong phú về thể loại sách cho bộ dữ liệu, nhóm chúng em thu thập với nhiều thể loại khác nhau như sách giáo khoa, sách tham khảo, sách văn học, tâm lý, thiếu nhi,

Trong quá trình thực hiện đồ án, nhóm đã thu thập được khoảng 3000 ảnh. Tuy nhiên, vì thời gian thực hiện đồ án có hạn, cũng như loại bỏ những tấm hình không đạt yêu cầu, nhóm chỉ thu thập được bộ dữ liệu bao gồm 964 ảnh.

2. Xử lý dữ liệu

Để chuẩn bị dữ liệu cho việc huấn luyện cũng như kiểm tra mô hình, chúng em sử dụng công cụ hỗ trợ gán nhãn PPOCRLabel được phát triển bởi công ty Baidu (Trung Quốc). Lý do bọn em chọn công cụ này thay vì những công cụ khác, là vì công cụ này cho phép chúng ta nhận diện văn bản với bounding box là đa giác và có tính năng Auto-recognition hỗ trợ gán nhãn nhanh hơn các công cụ khác. Nhờ thực hiện gán nhãn vùng chứa văn bản bằng đa giác mà việc gán nhãn dữ liệu được hiệu quả hơn đặc biệt với vùng chứa văn bản tiếng Việt bởi vì các bounding box sẽ bao phủ được hết dấu của từ tiếng Việt, hạn chế vùng background thừa ra so với bounding box hình chữ nhật. Hình dưới đây minh họa gán nhãn bằng PPOCRLabel



Hình 3: Giao diện công cụ gán nhãn PPOCRLabel

Chúng em gán nhãn dữ liệu theo line-level, mỗi dòng được gán nhãn bằng một bounding box. Sau đó đánh dấu dòng đó thuộc Tiêu đề, Tác giả, Nhà xuất bản hay nội dung khác bằng các nhãn tương ứng: “title”, “author”, “publisher”, “other” kèm với số thứ tự của bounding box đó trong từng trường dữ liệu. Định dạng gán nhãn một vùng văn bản:

nội_dung/nhãn/stt. Ví dụ, tại khung có chứa văn bản là “ĐẠO”, nhóm sẽ thêm vào sau nội dung đó đuôi “/title/0”, với hàm ý rằng đây là kí tự thuộc label “title”, và đây là chữ đầu tiên của tiêu đề. Sau khi thực hiện gán nhãn, ta có được một file .txt bao gồm đường dẫn của hình ảnh, tọa độ 4 đỉnh của các bounding box và nội dung văn bản có trên bìa sách. File Label.txt có định dạng như hình dưới đây:

```

train_label.txt
C:\Users\DELL\Downloads> train_label.txt
1 train_images/IMG_7472.jpg [{"transcription": "BẢN ĐỒ VỀ Ý THỨC/title/0", "points": [[454, 414], [922, 424], [852, 3784], [384, 3775]], "difficult": false, {"transcript
2 train_images/IMG_7618.jpg [{"transcription": "Cảnh Điều/other/0", "points": [[350, 237], [638, 258], [630, 367], [341, 346]], "difficult": false, {"transcription": "BÙ
3 train_images/2023_12_29_13_29_IMG_5534.JPG [{"transcription": "Ngọc Phương/other/0", "points": [[820, 258], [1272, 245], [1274, 324], [822, 338]], "difficult": false, {
4 train_images/IMG_7869.jpg [{"transcription": "NGUYỄN ĐAI/author/0", "points": [[272, 92], [1345, 92], [1345, 270], [272, 270]], "difficult": false, {"transcription": "
5 train_images/VH_064.jpg [{"transcription": "SHARI LAPENA/author/0", "points": [[618, 195], [1565, 162], [1569, 255], [621, 288]], "difficult": false, {"transcription": "
6 train_images/2023_12_29_16_17_IMG_6130.JPG [{"transcription": "TRẦN LINH/author/0", "points": [[106, 189], [985, 142], [980, 276], [110, 238]], "difficult": false, {"tr
7 train_images/2023_12_29_16_17_IMG_6130.JPG [{"transcription": "QUẢN LÝ CÔNG VIỆC/title/0", "points": [[116, 151], [1129, 120], [1129, 305], [121, 338]], "difficult": fal
8 train_images/2023_12_29_15_51_IMG_6024.JPG [{"transcription": "SUNEEL GUPTA/author/0", "points": [[98, 130], [459, 120], [460, 158], [99, 168]], "difficult": false, {"t
9 train_images/2023_12_29_13_07_IMG_5461.JPG [{"transcription": "EM MUỐN BIẾT/title/0", "points": [[93, 102], [1080, 78], [1084, 359], [93, 326]], "difficult": false, {"t
10 train_images/IMG_7508.jpg [{"transcription": "PHẠM THẾ LONG (Chủ biên)/author/0", "points": [[221, 138], [967, 160], [965, 244], [215, 237]], "difficult": false, {"tra
11 train_images/20231229_130815.jpg [{"transcription": "GET SMART/other/0", "points": [[494, 244], [1915, 197], [1917, 397], [501, 431]], "difficult": false, {"transcrip
12 train_images/IMG_7592.jpg [{"transcription": "Cảnh Điều/other/0", "points": [[509, 292], [789, 313], [780, 422], [500, 401]], "difficult": false, {"transcription": "NG
13 train_images/IMG_7659.jpg [{"transcription": "NGUYỄN NAM ANH/author/0", "points": [[713, 485], [1216, 492], [1215, 567], [712, 560]], "difficult": false, {"transcripti
14 train_images/IMG_7506.jpg [{"transcription": "MAI LAN HUONG/author/0", "points": [[1555, 850], [2285, 850], [2285, 921], [1555, 921]], "difficult": false, {"transcript
15 train_images/20231229_124632.jpg [{"transcription": "OSHO/author/0", "points": [[191, 176], [2337, 157], [2341, 712], [195, 731]], "difficult": false, {"transcription
16 train_images/VH_119.jpg [{"transcription": "M.R.JAMES/author/0", "points": [[228, 307], [1142, 340], [1136, 510], [222, 476]], "difficult": false, {"transcription": "Lê
17 train_images/20231229_131724.jpg [{"transcription": "TRẦN HÙNG THIỆN/author/0", "points": [[762, 638], [1830, 481], [1846, 594], [778, 751]], "difficult": false, {"tr
18 train_images/IMG_7633.jpg [{"transcription": "NGUYỄN ĐỨC CHÍ/author/0", "points": [[1262, 406], [2072, 401], [2067, 504], [1250, 516]], "difficult": false, {"transcrip
19 train_images/IMG_062.jpg [{"transcription": "ANNA SEWELL/author/0", "points": [[736, 263], [1704, 229], [1707, 305], [738, 338]], "difficult": false, {"transcription": "N
20 train_images/IMG_7747.jpg [{"transcription": "NGUYỄN THỊ THU HUI/author/0", "points": [[996, 207], [1699, 167], [1704, 255], [1001, 295]], "difficult": false, {"transc
21 train_images/IMG_7791.jpg [{"transcription": "Biga/other/0", "points": [[1880, 166], [2073, 166], [2076, 249], [1876, 250]], "difficult": false, {"transcription": "BÙI
22 train_images/20231229_131105.jpg [{"transcription": "đời/title/0", "points": [[1083, 358], [1633, 358], [1633, 696], [1083, 696]], "difficult": false, {"transcription
23 train_images/2023_12_30_12_27_IMG_5704.JPG [{"transcription": "Hashtag/other/0", "points": [[355, 125], [616, 132], [614, 209], [353, 202]], "difficult": false, {"trans
24 train_images/IMG_7462.jpg [{"transcription": "LUẬT/title/0", "points": [[1028, 1288], [1690, 1280], [1694, 1598], [1016, 1594]], "difficult": false, {"transcription":
25 train_images/VH_023.jpg [{"transcription": "TRƯƠNG CHÍ HÙNG/author/0", "points": [[554, 195], [1854, 118], [1860, 218], [560, 295]], "difficult": false, {"transcription"
26 train_images/IMG_7744.jpg [{"transcription": "MAI LAN HUONG/author/0", "points": [[980, 600], [1770, 621], [1764, 710], [980, 705]], "difficult": false, {"transcriptio
27 train_images/20231229_124408.jpg [{"transcription": "Takashi Saito/author/0", "points": [[870, 139], [1968, 113], [1970, 224], [872, 250]], "difficult": false, {"tran

```

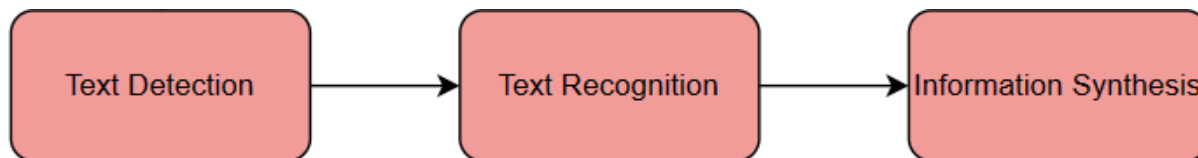
Hình 4: Định dạng file gán nhãn dữ liệu

3. Mô tả bộ dữ liệu

Bộ dữ liệu ảnh bìa sách được nhóm xây dựng bao gồm 964 ảnh, có định dạng là “.jpg” cùng với các annotation files. Bộ dữ liệu được chia thành 3 tập training set, validation set và testing set ngẫu nhiên với tỉ lệ của train/val/test = 7/1.5/1.5. Tương ứng, tập train gồm 700 ảnh, tập validation gồm 132 ảnh và tập test gồm 132 ảnh.

Chương 3: Thực nghiệm

1. Tổng quan các bước thực hiện



Hình 5: Tổng quan quy trình thực hiện

Chúng em chia công việc thành 3 task nhỏ là Text Detection, Text Recognition và Information Synthesis.

- **Task Text Detection (1):** nhiệm vụ của task này là chúng ta nhận diện được tọa độ bounding box chứa văn bản cùng với nhãn của bounding box đó.

- Input: Ảnh bìa sách
- Output: Tọa độ + Nhãn của bounding box
- Model: YOLOv8 (YOLOv8n)
- **Task Text Recognition (2):** nhiệm vụ của task này là trích xuất văn bản từ các bounding box đã nhận diện ở task (1)
 - Input: Thông tin tọa độ bounding box (task (1)) → Ảnh crop theo thông tin bounding box
 - Output: Văn bản trong ảnh đã được crop
 - Model: VietOCR
- **Task Information Synthesis (3):** nhiệm vụ của task này là tổng hợp lại các văn bản được trích xuất ở bước 2 thành các trường thông tin: Tiêu đề, Tác giả, Nhà xuất bản và những nội dung khác
 - Input: Văn bản trích xuất được từ task (2) (rời rạc)
 - Output: Thông tin hoàn chỉnh trên bìa sách theo các trường: Tiêu đề, Tác giả, Nhà xuất bản và những nội dung khác

2. Các bước thực hiện

2.1. Nhận diện đối tượng trên bìa sách (Text Detection)

- Ở bước này chúng em sử dụng mô hình YOLOv8 để thực nghiệm tác vụ phát hiện đối tượng trên bìa sách, cụ thể ở đây là tên sách, tác giả, nhà xuất bản và khác. Chúng em huấn luyện mô hình trên tập train và hiệu chỉnh trên tập validation.
- Chúng em đã tiến hành huấn luyện mô hình YOLOv8 khi gán nhãn được 700 ảnh trên train và 132 trên tập val. Các thông số được tinh chỉnh như sau: width = 640, height = 640, batch_size = 64, Max_batches = 6000, class = 4 (title, author, publisher, other)
- Sau khi dự đoán toàn bộ ảnh trên tập test, chúng em thực hiện sắp xếp lại các bounding box theo tọa độ trong tâm của box từ trên xuống dưới theo cách đọc của người Việt.



Hình 6: Input YOLOv8



Hình 7: Output YOLOv8

2.2. Nhận diện chữ trong các đối tượng (Text Recognition)

- Sau khi thực hiện task Text Detection bằng YOLOv8, chúng em sẽ cắt ảnh theo tọa độ bounding box được nhận diện và lưu thành các file ảnh crop trong thư mục Crop. Mỗi class có thể sẽ có nhiều bounding box. Sau đó, nhóm chúng em sử dụng VietOCR để thực hiện task Text Recognition
- Chúng em huấn luyện thêm model VietOCR dựa trên pretrained model của VietOCR bởi vì chúng em nhận thấy rằng văn bản xuất hiện trên bìa sách là vô cùng đa dạng về kích cỡ, kiểu chữ tuy nhiên lại lặp lại các thông tin như tác giả, nhà xuất bản nên chúng em quyết định train thêm model VietOCR với các tham số được cài đặt như sau: `iters=1600`; `batch_size=32`; `max_lr=0.001`; `image_height=32`; `image_max_width=512`; `image_min_width=32`



Hình 8: Minh họa kết quả dự đoán của mô hình VietOCR

2.3. Tổng hợp thông tin (Information Synthesis)

- Sau khi có được nội dung văn bản được trích xuất sau task (2), chúng em tổng hợp lại theo nhãn title, author, publisher, other đã được nhận diện từ task (1) theo thứ tự từ trái sang phải, từ trên xuống dưới để thành các trường thông tin hoàn chỉnh trên ảnh bìa sách: Tiêu đề, Tác giả, Nhà xuất bản và những nội dung khác.
- Thông tin cuối cùng sẽ được lưu vào file .txt có cùng tên với tên hình ảnh.

Chương 4: Kết quả đánh giá

Chúng em tiến hành đánh giá dựa trên từng task và đánh giá toàn bộ quá trình (end to end).

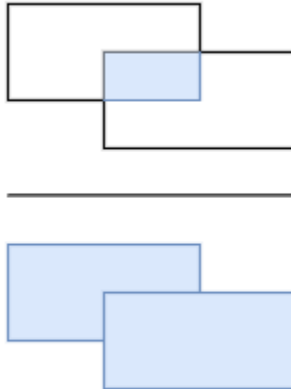
1. Task Text Detection:

Tại task này, mô hình nhóm sử dụng là mô hình YOLOv8. Không chỉ đối với mô hình YOLOv8 mà với bất kì mô hình nào, các số liệu này đều rất quan trọng, cụ thể:

- **Intersection over Union (IoU):** Đây là thước đo định lượng sự giao nhau giữa các bounding box dự đoán với những bounding box true. IoU được tính bằng tỉ lệ giữa diện tích bounding box thực sự và tổng diện tích của 2 bounding box đó. Công thức được minh họa bằng hình dưới đây:

$$\text{IoU} = \frac{\text{Intersection}}{\text{Union}}$$

(Intersection of Union)



Kết quả trả về của IoU sẽ nằm trong khoảng từ 0 đến 1, và đối với mỗi detection sẽ có một IoU cụ thể. Để xác định được rằng detection là true hay false thì chúng ta cần phải chọn một ngưỡng cụ thể. Nếu IoU của detection lớn hơn hoặc bằng ngưỡng mà chúng ta chọn thì đó là correct detection, còn ngược lại thì là wrong detection. Từ đó chúng ta sẽ có một số khái niệm sau:

- **True Position (TP):** đây là trường hợp correct detection, model dự đoán bounding box mà có phần trùng với bounding box true nằm trong ngưỡng quy định.
- **True Negative (TN):** đây là trường hợp mà model dự đoán rằng khu vực này không có bounding box mà thực tế rằng không có bounding box. Tuy nhiên, trong đề án này chúng ta không quan tâm đến nó vì chúng ta chỉ quan tâm đến những bounding box cần model dự đoán.
- **False Position (FP):** đây là trường hợp model dự đoán bounding box, tuy nhiên phần trùng với bounding box lại thấp hơn ngưỡng, hay IoU thấp hơn ngưỡng.
- **False Negative (FN):** đây là trường hợp có bounding box nhưng mode lại không dự đoán.

→ Từ những khái niệm trên, chúng ta có thể tính được **Precision, Recall và F1-Score** của model như sau:

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1\ score = \frac{2.Precision.Recall}{Precision+Recall}$$

- **Average Precision (AP):** Dựa vào các giá trị precision và recall, chúng ta có thể tính được AP. Giá trị AP là giá trị phía dưới đường biểu diễn mối quan hệ precision – recall.
- **Mean Average Precision (mAP):** mAP là sự mở rộng của AP bằng cách tính toán giá trị trung bình của AP qua nhiều class khác nhau. Điều này rất hữu ích để đánh giá toàn diện về hiệu suất của mô hình.

Để đánh giá hiệu suất của YOLOv8, chúng ta dùng hàm `metrics.box`. Sau đó chúng ta sẽ có được một số thông số như sau:

- **Class:** Đây là tên các class mà chúng ta quy định
- **Images:** Đây là số lượng hình ảnh trong nhóm xác thực có chứa lớp đối tượng.
- **Instances:** Số lần mà các ảnh thuộc lớp xuất hiện trong các hình ảnh thuộc validation.
- **P, R, mAP50, mAP50-95:** Số liệu cung cấp thông tin về hiệu suất của mô hình trong việc phát hiện đối tượng. Cụ thể với P (Precision) là độ chính xác của các đối tượng được phát hiện, cho biết có bao nhiêu phát hiện là chính xác. R (Recall) là khả năng mà mô hình xác định tất cả các trường hợp vừa các đối tượng trong hình ảnh. mAP50 là giá trị của mAP với ngưỡng IoU được chọn là 0.5. Còn mAP50-95 là giá trị của mAP với ngưỡng IoU được chọn nằm trong khoảng 0.5 đến 0.95. Với khoảng này thì chúng ta sẽ có cái nhìn hoàn thiện hơn về hiệu suất của mô hình thông qua các ngưỡng khác nhau.
- **Kết quả đánh giá model YOLOv8 được thể hiện qua bảng sau:**

Class	Images	Instances	P	R	F1-score	mAP50	mAP50-95
All	132	1442	0.666	0.636	0.6507	0.653	0.368
Other	132	564	0.651	0.47	0.54586	0.544	0.273
Author	132	232	0.637	0.578	0.60592	0.62	0.32
Title	132	469	0.8	0.868	0.83246	0.866	0.607
Publisher	132	177	0.577	0.627	0.60121	0.582	0.274

2. Task Text Recognition:

- Để tính được độ chính xác của từng ký tự (`Acc_per_char`), ta sẽ làm như sau: input là 1 phần tử chứa nhiều ký tự và một label, ta sẽ kiểm tra từng ký tự của phần tử input xem có giống với ký tự tương ứng bên label không. Sau đó ta sẽ tính được `Acc_per_char` bằng công thức:

$$Acc_per_char = \frac{sum(single_label_accuracy)}{label_nums}$$

- Tương tự với Acc_full_seq, input sẽ bao gồm các phần tử và các label tương ứng. Ta sẽ kiểm tra từng phần tử có giống với label tương ứng không. Sau đó ra sẽ tính được Acc_full_seq bằng công thức:

$$Acc_full_seq = \frac{single_label_accuracy}{labels_nums}$$

- Kết quả đánh giá model VietOCR được thể hiện qua bảng sau:

Model	Acc_full_seq (%)	Acc_per_char(%)
VietOCR	80.93	93.08

3. End to End

- Để đánh giá hiệu suất của toàn bộ mô hình, nhóm đã sử dụng độ đo WER (word error rate – tỉ lệ lỗi trên cấp độ từ) để tính độ lỗi. Công thức của WER được tính như sau:

$$WER = \frac{W + D + I}{N}$$

Trong đó: W: Số token bị sai chính tả.

D: Số token bị xóa mất hoặc bị thiếu.

I: Số token được thêm vào mà không có trong chuỗi đúng.

N: Tổng số token của chuỗi đúng

- Kết quả đánh giá toàn bộ quy trình trích xuất thông tin từ ảnh bìa sách trên độ đo WER (%):

Author	Title	Publisher	Other	All
76.05	49.22	36.61	89.65	62.88

Chương 5: Kết luận và hướng phát triển

Trong đồ án này, nhóm chúng em đã xây dựng được bộ dữ liệu ảnh chụp bìa sách gồm 964 ảnh với đa dạng thể loại để phục vụ cho bài toán nhận dạng, trích xuất thông tin từ bìa sách. Đồng thời, nhóm cũng đã thực hiện tiền xử lý dữ liệu để phục vụ cho mục đích huấn luyện và đánh giá các mô hình.

Chúng em đã tìm hiểu và thực hiện được một pipeline cho đề tài số hóa tủ sách. Nhóm đã thực hiện được các task Text Recognition, Text Detection và Information Synthesis và đánh giá được kết quả của từng task cũng như sử dụng được các độ đo để đánh giá kết quả của mô hình.

Trong tương lai, nhóm có định hướng sẽ hoàn thiện mô hình số hóa tủ sách, cũng như cập nhật thêm chức năng dựa theo thông tin đã nhận diện để phân loại thể loại của bìa sách. Bên cạnh đó, để tăng tính hiệu quả cho mô hình, nhóm sẽ tiếp tục tăng số lượng dữ liệu và đa dạng hơn các bộ dữ liệu để mô hình được học tốt hơn.

Tài liệu tham khảo

- [1] Tham Nguyen Thi et al, “*A Novel Two-stages Information Extraction Algorithm for Vietnamese Book Cover Images*”
- [2] Tham Nguyen Thi et al, “*Implementation of OCR system on extracting information from Vietnamese book cover images*”
- [3] VietOCR, <https://github.com/pbcquoc/vietocr/tree/master>
- [4] Ultralytics YOLOv8 Docs, <https://docs.ultralytics.com/>