B0. Các bạn lưu ảnh của mỗi trang sách theo tên trùng với cột Image\_name trong file Excel (vd: 001.png, 002.png, 003.png, ...) và lưu các ảnh đó trong thư mục Check\_label.

B1. Thay file Result.xlsx trong thư mục Check\_label thành file excel ngữ liệu GK của bạn.

Lưu ý:

* Cột ‘Image Box’ cần dùng dấu ngoặc vuông như hình. (có dạng [[x1,y1],[x2,y2],[x3,y3],[x4,y4]]) .
* Cột ‘Image\_name’ phải đặt theo đúng định dạng <số>.png, các box trong một hình sẽ đặt cùng một tên (xem hình ảnh ví dụ bên dưới).

B2. Copy thư mục hình ảnh cần check tọa độ vào folder check\_label, và đặt đúng tên image\_label

B3. Chạy file convert\_data\_to \_labelsPaddle\_v2.py để xuất ra Label.txt (gợi ý: có thể chạy dưới dạng command line trong terminal (python convert\_data\_to \_labelsPaddle\_v2.py))

B4: Mở tools PPOCRLabel trỏ tới thư mục chứa ảnh cần kiểm tra. Nhấn nút Auto Recognition (góc trái dưới ứng dụng để tạo định dạng labels tự động)

thư mục ảnh cần load

A screenshot of a computer

Description automatically generated

B5: Nhấn nút Auto Recogniton góc trái dưới

A screenshot of a computer

Description automatically generated

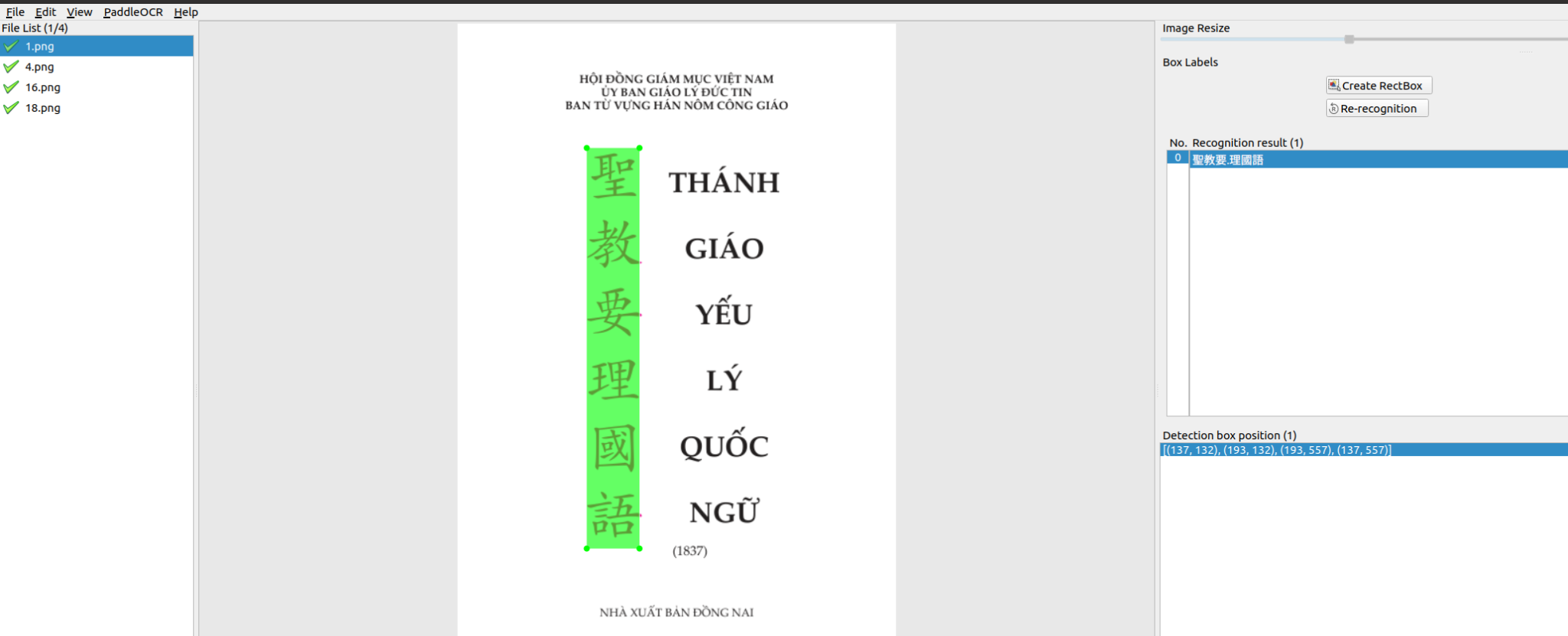
B6: Nhấn nút check để lưu labels tất cả các ảnh

A screenshot of a computer

Description automatically generated

Vào File->Export Label -> OK để lưu lại label.txt đã tự động recognition.

B7. Vào thư mục ảnh xóa file Cache.cache và Label.txt và paste file Label.txt đã tạo ở B3 vào. Sau đó mở lại tools PPOCRLabel load vào lại thư mục hình để xem kết quả. Như hình bên dưới, các box và text từ file result.txt đã được hiển thị.



B8. Vào File → Export Reconition Result → OK để xuất label cho giai đoạn nhận dạng văn bản. Kết quả như hình dưới

