

# BÁO CÁO ĐỒ ÁN CUỐI KỲ

Lớp: CS2225.CH1501

Môn: NHẬN DẠNG THỊ GIÁC VÀ ỨNG DỤNG

GV: PGS.TS Lê Đình Duy

Trường ĐH Công Nghệ Thông Tin, ĐHQG-HCM



# RÚT TRÍCH THÔNG TIN TRÊN HÓA ĐƠN THANH TOÁN

**Trần Anh Tuấn - CH2001018**

**Văn Thiên Luân - CH2001008**

**Nguyễn Ngọc Thừa - CH2001017**

**Link Github:** <https://github.com/tuananht/CS2225.CH1501.N013>

# Tóm tắt

- Tên đề tài: Rút trích thông tin từ hóa đơn Bách hóa xanh
- Tóm tắt về đề án và kết quả đạt được:

- Tóm tắt: Thực hiện rút trích một số thông tin từ hóa đơn bách hóa xanh như: thời gian bán hàng, tổng tiền phải trả, các mặt hàng đã được mua. Đề tài sử dụng phương pháp tesseract, yolov3 để phát hiện và nhận dạng thông tin.

- Kết quả đạt được: Xây dựng được một ứng dụng rút trích thông tin từ hóa đơn bách hóa xanh ở mức độ thử nghiệm đạt độ chính xác 73% trên tập dữ liệu tự thu thập (39 ảnh).

# Tóm tắt

- Ảnh của các thành viên của nhóm:



# Mô tả bài toán

- Mỗi ngày có rất nhiều giao dịch diễn ra, số lượng các hóa đơn giấy được in ra ngày càng nhiều. Phần lớn, các bộ phận nghiên cứu thị trường chưa tận dụng được lượng thông tin trên các hóa đơn giấy.
- Một số tổ chức đã cố gắng trong việc nhập liệu lại thông tin từ hóa đơn nhưng không khả thi, do phải nhập bằng tay, chi phí nhân lực cao.
- Bài toán được đặt ra làm thế nào để rút trích thông tin trên các hóa đơn nhanh chóng, thuận tiện, chi phí thấp, và đạt hiệu quả.

# Mô tả bài toán

- Bài toán thực hiện: Rút trích thông tin hóa đơn thanh toán của cửa hàng Bách hóa xanh.
- Đầu vào: ảnh chụp của một hóa đơn thanh toán.
- Đầu ra: Kiểm tra hóa đơn có từ cửa hàng Bách hóa xanh không, nếu có rút trích thông tin từ hóa đơn đó.
- Minh họa (slide kế tiếp)

- Minh họa



# Dữ liệu

- Thu thập và gán nhãn hóa đơn từ cửa hàng Bách hóa xanh(39 ảnh).
- Dùng 25 ảnh để train bộ phát hiện vùng thông tin.
- Chia 14 ảnh để test bộ phát hiện vùng thông tin.
- Dùng toàn bộ 39 ảnh để đánh giá ứng dụng.



BÁCH HÓA XANH		
www.bachhoaxanh.com		
112 Hoàng Hoa Thám, Phường 12, Quận Tân Bình, TP Hồ Chí Minh		
PHIẾU THANH TOÁN		
Số CT:	20110044050303618	
Ngày CT:	08/11/2020 17:22	
Nhân viên:	Lê Văn Thanh	
SL	Giá bán	T.Tiền
NẠC HEO XAY NHẬP KHẨU 0.306	119,000	36,414
KHỔ CHUA 0.312	30,000	9,360
THỊT ĐÙI HEO 0.362	145,000	52,490
BÍ NGÒI XANH 0.374	45,000	16,830
XƯƠNG HEO CÓ THỊT 0.45	119,000	53,550
CÀ CHUA (KG) 0.376	35,000	13,160
KHOAI TÂY 0.106	38,000	4,028
CÀ RỐT ĐL (KG) 0.17	40,000	6,800
KHOAI LANG NHẬT 0.168	27,000	4,536



# Phương pháp được sử dụng

**1. YOLOv3 (Object detection):** Dùng để phát hiện các vùng dữ liệu trên hóa đơn như logo, ngày bán, tổng tiền.

- Train lại trên tập 25 ảnh đã được gán nhãn tay.

- Kết quả thu được là tọa độ các vùng thông tin,

dựa vào các vùng này để xác định đây có

phải hóa đơn bách hóa xanh hay không và

matching kết quả nhận dạng cho phù hợp.



# Các phương pháp được sử dụng

## 2. Tesseract (Text Recognition)

- Là phương pháp nhận dạng chữ viết được sử dụng phổ biến.
- Nhóm sử dụng pretrain cho tiếng việt do tác giả cung cấp.
- Dùng để nhận dạng các ký tự trong hóa đơn sau khi đã qua tiền xử lý.



BÁCH HÓA XANH  
www.bachhoaxanh.com  
8 đường Bình Lợi, Phường 13, Quận Bình  
Thạnh, Thành phố Hồ Chí Minh, Việt Nam  
PHIẾU THANH TOÁN  
Số CT: 20080047450225488  
Ngày CT: 10/08/2020 18:36  
Nhân viên: Phạm Thị Cẩm Tú  
SL Giá bán T.Tiền  
KHOAI SỎ  
0.248 47,000 11,656  
KHOAI TÂY  
0.26 35,000 9,100  
HÀNH LÁ (KG)  
0.018 35,000 630  
SƯỜN NON HEO NHẬP KHẨU  
0.238 167,000 39,746  
BA RỢI HEO CỎ DA NHẬP KHẨU  
0.322 155,000 49,910  
Tổng tiền: 111,042  
Thanh toán: 111,000  
(Đã làm tròn)  
Tiền mặt: 200,000  
Tiền thối lại: 89,000  
(Giá trên đã bao gồm thuế GTGT)  
Lưu ý: Bách Hóa Xanh chỉ xuất hóa đơn trong  
ngày, Quý khách vui lòng liên hệ thu ngân để  
được hỗ trợ. Quý khách có thể in bản sao hóa  
đơn VAT tại quầy.

# Phương pháp đánh giá

**Khoảng cách Levenshtein:** độ đo tính theo dựa trên số lượng ký tự sai lệch giữa 2 chuỗi.

So sánh kết quả rút trích của từng vùng dữ liệu với nhãn. Sau đó tính trung bình kết quả của các trường ta được độ chính xác.

Ví dụ (cho trường ngày bán):

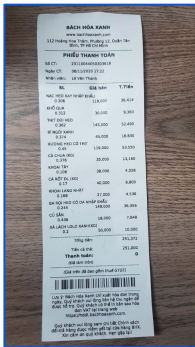
Nhãn: 09/01/2019 22:35

Kết quả: 09/11/2019 22:85

=> Sai lệch 2 kí tự => Độ chính xác =  $2/16 = 87.5\%$

# Các bước xử lý của ứng dụng

Ảnh chụp  
hóa đơn



Kiểm tra có phải là  
hóa đơn của hàng  
Bách hóa xanh  
không(yolov3)

Sai

Kết thúc  
xử lý.

Đúng

Sử dụng Tesseract với  
ngôn ngữ Tiếng Việt để rút  
trích thông tin

Sử dụng phương  
pháp lọc, khử nhiễu  
hình(opencv).

Rút trích thông  
tin.

Số HD	20110044050303618
Ngày lập HD	08/11/2020 17:22
Nhân viên	Lê Văn Thanh
Số sản phẩm	12
Thành tiền	N/a

# Một số hình ảnh của ứng dụng

Object Boxed

**BÁCH HOÀ XANH**  
www.bachhoaxanh.com

112 Hoàng Hoa Thám, Phường 12, Quận Tân Bình, TP Hồ Chí Minh

**PHIẾU THANH TOÁN**

Số CT: 20110044050303618  
Ngày CT: 08/11/2020 17:22  
Nhân viên: Lê Văn Thanh

SL	Giá bán	T.Tiền
<b>NẠC HEO XAY NHẬP KHẨU</b>		
0.306	119,000	36,414
<b>KHỔ QUA</b>		
0.312	30,000	9,360
<b>THỊT ĐÚI HEO</b>		
0.362	145,000	52,490
<b>BÍ NGŨI XANH</b>		
0.374	45,000	16,830
<b>XƯƠNG HEO CỎ THỊT</b>		
0.45	119,000	53,550
<b>CÁ CHUA (KG)</b>		
0.376	35,000	13,160
<b>KHOAI TÂY</b>		
0.106	38,000	4,028
<b>CÁ RỐT ĐL (KG)</b>		
0.17	40,000	6,800
<b>KHOAI LANG NHẬT</b>		
0.168	27,000	4,536
<b>BA RƠI HEO CỎ DA NHẬP KHẨU</b>		
0.244	149,000	36,356
<b>CỦ SẦN</b>		
0.435	18,000	7,740

✓ Preprocessing  
Tiền xử lý hình ảnh thành công

✓ Nhận dạng loại hoá đơn  
Cửa hàng: Bách Hoá Xanh (74.49%)

✓ Trích xuất thông tin hoá đơn  
Thành công

**Danh sách sản phẩm**

Website	www.bachhoaxanh.com
Địa chỉ	112 Hoàng Hoa Thám, Phường 12, Quận Tân Bình, TP Hồ Chí Minh
Số HD	20110044050303618
Ngày lập HD	08/11/2020 17:22
Nhân viên	Lê Văn Thanh
Số sản phẩm	12
Thành tiền	N/a

**Danh sách các sản phẩm**

#	Sản phẩm	Số lượng	Giá bán	Thành tiền
1	NẠC HEO XAY NHẬP KHẨU	0.306	119000	36414
2	KHỔ QUA	0.312	30000	9360

# Một số hình ảnh của ứng dụng

KHOAI TÂY	0.106	38,000	4,028
CÀ RỐT ĐL (KG)	0.17	40,000	6,800
KHOAI LANG NHẬT	0.168	27,000	4,536
BA RỢI HEO CÓ DA NHẬP KHẨU	0.244	149,000	36,356
CỦ SẴN	0.436	18,000	7,848
XÀ LÁCH LOLO XANH(KG)	0.2	50,000	10,000
Tổng tiền:		251,372	
Tiền cà thẻ:		251,000	
Thanh toán:		0	
(Đã làm tròn)			
(Giá trên đã bao gồm thuế GTGT)			
			
Lưu ý: Bách Hóa Xanh chỉ xuất hóa đơn trong ngày, Quý khách vui lòng liên hệ thu ngân để được hỗ trợ. Quý khách có thể in bản sao hóa đơn VAT tại trang web <a href="https://hddt.bachhoaxanh.com">https://hddt.bachhoaxanh.com</a> .			
Quý khách vui lòng xem chi tiết Chính sách đổi-trả hàng được niêm yết tại cửa hàng BHX. Xin cảm ơn quý khách. Hẹn gặp lại!			

Thành tiền N/a

## Danh sách các sản phẩm

#	Sản phẩm	Số lượng	Giá bán	Thành tiền
1	NẠC HEO XAY NHẬP KHẨU	0.306	119000	36414
2	KHỔ QUA	0:312	30000	9360
3	THỊT ĐÙI HEO	0.362	145000	52490
4	BÍ NGÒI XANH	0.374	45000	16830
5	XƯƠNG HEO CÓ THỊT	0.45	118000	53550
6	CÀ CHUA (KG)	0.376	35000	13160
7	KHOAI TÂY	0:108	38000	4028
8	CÀ RỐT ĐL (KG)	0.17	40000	6800
9	KHOAI LANG NHẬT	0,168	27000	4536
10	BA RỢI HEO CÓ DA NHẬP KHẨU	0.244	148000	36356
11	CỦ SẴN	0.436	18000	7848
12	XÀ LÁCH LOLO XANH(KG)	0.2	50000	10000

# Định hướng phát triển.

- Tìm cách tăng độ chính xác của ứng dụng.
- Tổ chức lại luồng xử lý để giảm thời gian xử lý.
- Tìm hiểu cách để cắt ảnh hóa đơn ra khỏi background.
- Mở rộng dữ liệu để có thể xử lý một số hóa đơn của cửa hàng khác.