


## MẪU BÁO CÁO CỦA MỖI HV

<b>Họ và tên (IN HOA)</b>	NGUYỄN NGỌC THỪA - CH2001017
<b>Ảnh</b>	
<b>Số buổi vắng</b>	3
<b>Bonus</b>	4
<b>Tên đề tài (VN)</b>	XÂY DỰNG HỆ THỐNG RÚT TRÍCH THÔNG TIN HÓA ĐƠN TIẾNG VIỆT TỪ ẢNH
<b>Tên đề tài (EN)</b>	AN END TO END RECOGNITION FRAMEWORK FOR VIETNAMESE RECEIPTS
<b>Giới thiệu</b>	<p>Thông tin từ dữ liệu văn bản (text) trong ảnh là một trong những thông tin có nhiều giá trị khi được trích xuất và khai thác hợp lý. Hiện nay có khá nhiều ứng dụng thực tiễn được triển khai nhờ việc trích xuất thông tin văn bản từ ảnh như chứng minh thư, căn cước công dân, bằng lái ... Trong đó, thông tin từ hóa đơn chứa những nội dung quan trọng phục vụ cho một số nhu cầu tự động hóa hỗ trợ các công việc liên quan đến lưu trữ, thanh toán, kiểm kê của doanh nghiệp. Theo đó, trích xuất thông tin từ hóa đơn được chọn là bài toán nghiên cứu chính trong đề tài này.</p>

Bài toán được mô tả như sau: đầu vào hệ thống là ảnh chụp hóa đơn, đầu ra của hệ thống là các thông tin văn bản cần trích xuất trong nội dung hóa đơn. Trong đó trong đề tài này hóa đơn được chụp từ ảnh điện thoại, các thông tin được trích xuất bao gồm: tên cửa hàng hay nơi xuất hóa đơn (Seller), địa chỉ nơi xuất hóa đơn (address), ngày giờ xuất hóa đơn (date) và tổng số tiền thanh toán (total) (Hình 1: mô tả đầu ra đầu vào của bài toán).



Đầu vào

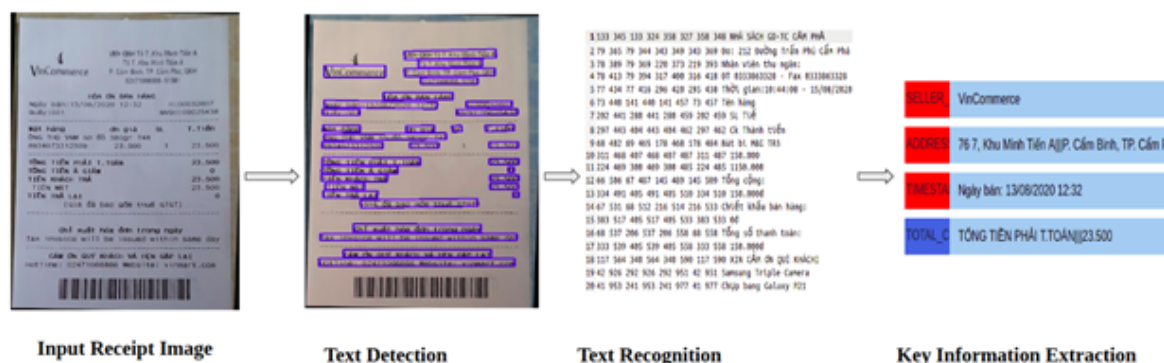


SELLER:	THE COFFEE HOUSE
ADDRESS	403 Phan Huy Ích, P.14, Q.Gò Vấp
DATE:	Thời gian: 02:10:2020 11:49
TOTAL:	Thanh Toán   49 000

Đầu ra

Hình 1: Mô tả đầu ra đầu vào của bài toán.

Về mặt tổng quát, bài toán trích xuất thông tin trong hóa đơn là bài toán phức tạp bao gồm nhiều bài toán con. Vì hóa đơn được chụp từ điện thoại với nhiều góc chụp và cách thức khác nhau, do đó ở bước đầu tiên hệ thống cần xác định vị trí hóa đơn trong bức ảnh và căn chỉnh hóa đơn (crop và align) theo chiều phù hợp làm đầu vào cho bước sau. Ở bước tiếp theo hệ thống sẽ thực hiện phát hiện vị trí văn bản trong ảnh (text detection), sau đó các bộ nhận diện (text recognition) được áp dụng để xác định nội dung văn bản. Ở bước cuối cùng các thông tin cần rút trích từ hóa đơn được rút ra thông qua bộ trích xuất (information extraction) (Hình 2 minh họa các bài toán con trong hệ thống). Như vậy, bên cạnh thách thức về mặt dữ liệu khi bị ảnh hưởng bởi yếu tố môi trường như mờ, che khuất, chói sáng .. thì việc lựa chọn các phương pháp phù hợp nhất trong các bài toán con ở các giai đoạn trong hệ thống cũng là thách thức.



Hình 2. Một số bài toán con trong hệ thống rút trích hóa đơn .

Rút trích thông tin văn bản từ ảnh là bài toán được quan tâm trong cộng đồng nghiên cứu về thị giác máy và xử lý ngôn ngữ tự nhiên. Trong đó, đã có khá nhiều kết quả có độ chính xác cao tương ứng với các bài toán con có ảnh hưởng lớn đến độ chính xác của hệ thống như Text Detection và Text Recognition. Đối với bài toán text detection, hầu hết các phương pháp tiên tiến hiện nay đều tiếp cận theo hướng áp dụng Deep Learning (học sâu) như ABCNet [1], Craft [2], Paddle [3], ... các detector này được đánh giá trên một số tập dữ liệu chuẩn như COCO-Text [4], ICDAR 2013, ICDAR 2015 [5]. Với text recognition, các phương pháp hiện nay phần lớn dựa trên Recurrent Neural Network (RNN, tạm dịch: mạng nơ-ron hồi quy), được chứng minh là phương pháp deep learning phù hợp với dữ liệu dạng chuỗi và time-series trong nhiều bài toán liên quan khác. Ảnh đầu vào được resize với kích thước cố định (thường cố định chiều cao) và chia thành các chuỗi (sequence) hay cột (column). Sau đó, một mạng nơ-ron hồi quy (RNN) được dùng để dự đoán các kí tự xuất hiện trong các chuỗi hay cột này. Một số phương pháp tiêu biểu dựa trên RNN như: phương pháp được giới thiệu trong [6],[7] kết hợp RNN với CNN, hoặc các phương pháp sử dụng attention model [8-9]. Tuy nhiên các nghiên cứu này đang được đánh giá trên các bài toán với dữ liệu dạng scene text (ảnh được chụp tự nhiên), đối với bài toán có lĩnh vực cụ thể như hóa đơn thì chưa có nhiều nghiên cứu. Năm 2019 hội nghị hàng đầu trong lĩnh vực nhận dạng chữ ICDAR đã tổ chức cuộc thi về nhận dạng thông tin hóa đơn trên dữ liệu hóa đơn scan trong đó ngôn ngữ chính của dữ liệu là tiếng anh. Theo đó, các đánh giá trên dữ liệu được chụp từ ảnh điện thoại và áp dụng cho dữ liệu Tiếng Việt chưa có nhiều đánh giá.

## Mục tiêu

- Tìm hiểu tổng quan về bài toán phát trích xuất thông tin văn bản trong ảnh hóa đơn.

	<ul style="list-style-type: none"> <li>- Phân tích, đánh giá một số phương pháp tiên tiến hiện nay cho bài toán trên tập dữ liệu chuẩn.</li> <li>- Đề xuất mô hình và xây dựng hệ thống minh họa cho bài toán rút trích thông tin hóa đơn tiếng Việt từ ảnh.</li> </ul>
<b>Nội dung và phương pháp thực hiện</b>	<p><b>Nội dung 1: Tìm hiểu tổng quan về bài toán rút trích thông tin hóa đơn từ ảnh.</b></p> <p><b>*Mục tiêu nội dung:</b></p> <ul style="list-style-type: none"> <li>- Khảo sát, tìm hiểu tổng quan bài toán rút trích thông tin hóa đơn từ ảnh.</li> <li>- Phân tích các phương pháp tiên tiến nhất hiện nay được áp dụng cho bài toán này thông qua các nghiên cứu mới nhất.</li> <li>- Tìm hiểu một số tập dữ liệu phổ biến cho bài toán.</li> </ul> <p><b>*Sản phẩm khoa học dự kiến và chỉ tiêu đánh giá:</b></p> <p>Tài liệu tổng quan về bài toán rút trích thông tin hóa đơn từ ảnh.</p> <p>Tài liệu về tập dữ liệu sử dụng cho bài toán.</p> <p><b>*Phương pháp:</b></p> <p>Khảo sát, tìm hiểu tổng quan bài toán thông qua các nghiên cứu mới nhất công bố trên các hội nghị, tạp chí có uy tín.</p> <p><b>Nội dung 2: Phân tích, đánh giá một số phương pháp tiên tiến hiện nay cho bài toán rút trích thông tin hóa đơn trong ảnh.</b></p> <p><b>*Mục tiêu nội dung :</b></p> <ul style="list-style-type: none"> <li>- Phân tích, đánh giá một số phương pháp tiên tiến hiện nay trên tập dữ liệu chuẩn. Từ đó lựa chọn các phương pháp phù hợp để xây dựng hệ thống minh họa.</li> </ul>

**\*Sản phẩm khoa học dự kiến và chỉ tiêu đánh giá:**

- Tài liệu phân tích, đánh giá kết quả một số phương pháp tiên tiến hiện nay trên tập dữ liệu chuẩn.

**\*Phương pháp:**

Khảo sát, tìm hiểu một số hướng tiếp cận tiên tiến hiện nay cho bài toán rút trích thông tin từ ảnh. Thử nghiệm sẽ được đánh giá trên tập dữ liệu chuẩn, dựa trên các phân tích trên kết quả thực nghiệm phương pháp phù hợp nhất ở mỗi giai đoạn trong toàn hệ thống sẽ được chọn ra để xây dựng hệ thống minh họa.

**\*Nội dung 3: Đề xuất mô hình rút trích thông tin hóa đơn Tiếng Việt từ ảnh.**

**\*Mục tiêu nội dung :**

- Thử nghiệm, đánh giá một số phương pháp rút trích hóa đơn trên bộ dữ liệu tiếng Việt.

- Đề xuất mô hình cho bài toán trên dữ liệu hóa đơn tiếng Việt.

- Công bố 1 bài báo.

**\*Sản phẩm khoa học dự kiến và chỉ tiêu đánh giá:**

- Bài báo khoa học.

**\*Phương pháp:**

Thử nghiệm huấn luyện các mô hình cho bài toán trên dữ liệu tiếng Việt. Dựa trên các đánh giá và phân tích trên các kết quả thử nghiệm đề xuất các phương pháp trong toàn bộ quy trình để rút trích thông tin hóa đơn Tiếng Việt từ ảnh.

**Nội dung 4: Xây dựng hệ thống minh họa.**

**\*Mục tiêu nội dung :**

	<ul style="list-style-type: none"> <li>- Đề xuất kiến trúc hệ thống dựa trên các đánh giá và thực nghiệm</li> <li>- Xây dựng ứng dụng minh họa.</li> </ul> <p>Sản phẩm khoa học dự kiến và chỉ tiêu đánh giá:</p> <ul style="list-style-type: none"> <li>- Hệ thống minh họa</li> </ul> <p><b>*Phương pháp:</b></p> <p>Phân tích, đánh giá kết quả thực nghiệm các phương pháp trên tập dữ liệu thực. Từ đó đề xuất kiến trúc hệ thống tốt nhất cho quá trình xây dựng ứng dụng minh họa.</p>
<b>Kết quả dự kiến</b>	<p><b>Các kết quả nghiên cứu của đề tài này là:</b></p> <ul style="list-style-type: none"> <li>-Tài liệu báo cáo tổng hợp về bài toán rút trích thông tin hóa đơn.</li> <li>-Ứng dụng minh họa cho kiến trúc đã đề xuất.</li> </ul>
<b>Tài liệu tham khảo</b>	<p>[1] Liu, Yuliang, et al. "ABCNet: Real-time scene text spotting with adaptive bezier-curve network." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.</p> <p>[2] Baek, Youngmin, et al. "Character region awareness for text detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019</p> <p>[3] Shang, Mingyu, Jie Gao, and Jun Sun. "Character Region Awareness Network For Scene Text Recognition." 2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2020</p> <p>[4] Raul Gomez, Baoguang Shi, Lluís Gomez, Lukas Numann, Andreas Veit, Jiri Matas, Serge Belongie and Dismosthenis Karatzas, "ICDAR2017 Robust Reading Challenge on COCO-Text", 14th IAPR International Conference on Document Analysis and Recognition, 2017</p>

- |  |   |
|--|---|
|  | <p>[5] ICDAR Robust Reading Competition (RRC), <a href="https://rrc.cvc.uab.es/">https://rrc.cvc.uab.es/</a>.</p> <p>[6] Y. Gao, Y. Chen, J. Wang, and H. Lu. Reading scene text with attention convolutional sequence modeling. arXiv preprint arXiv:1709.04303, 2017.</p> <p>[7] B. Shi, X. Bai, and C. Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. TPAMI, 2017.</p> <p>[8] Z. Cheng, F. Bai, Y. Xu, G. Zheng, S. Pu, and S. Zhou. Focusing attention: Towards accurate text recognition in natural images. ICCV, 2017.</p> <p>[9] Z. Cheng, X. Liu, F. Bai, Yi Niu, S. Pu, and S. Zhou. Arbitrarily-oriented text recognition. CVPR, 2018.</p> |
|--|---|