P-ISSN 1859-3585 | E-ISSN 2615-9619

XÂY DỰNG HỆ THỐNG TRÍCH XUẤT THÔNG TIN GIẤY TỜ TUỲ THÂN TỪ HÌNH ẢNH CHO HỆ THỐNG ĐỊNH DANH KHÁCH HÀNG ĐIỆN TỬ

BUILDING A SYSTEM TO EXTRACT PERSONAL DOCUMENT INFOMATION FROM IMAGE FOR eKYC SYSTEM

Viên Thanh Nhã[1,*], Tiếp Sỹ Minh Phụng[1], Nguyễn Hoàng Tú[2], Đỗ Thị Kim Dung[3], Lê Đình Phú Cường[4]

TÓM TẮT

Trong thời buổi công nghệ thông tin hiện đại, tự động trích xuất thông tin từ giấy tờ tuỳ thân được áp dụng trong lĩnh vực ngân hàng, viễn thông, y tế, khách sạn,... Nghiên cứu này ứng dụng các kỹ thuật thị giác máy tính để xây dựng một hệ thống trích xuất thông tin từ giấy tờ tuỳ thân cho hệ thống định danh khách hàng điện tử. Ba bước chính: Nhận dạng giấy tờ, định vị thông tin và trích xuất thông tin. Đặc biệt, xác định thẻ thụt lề và thông tin định vị là được thực hiện bằng các kỹ thuật xử lý ảnh cơ bản để tăng tốc độ của toàn bộ hệ thống. Tại bước trích xuất thông tin, những thành tựu trong nhận dạng văn bản trên nền phức tạp sẽ được sử dụng để tăng độ chính xác. Sau khi phát triển và thử nghiệm, hệ thống đã đạt được độ chính xác tương đối cao. Độ chính xác của toàn hệ thống là trên 90%, kết quả này có thể được ứng dụng vào trong các hệ thống định danh điện tử của các ngân hàng, tổ chức tín dụng hiện nay.

Từ khóa: *Thị giác máy tính; YOLOv4; trích xuất thông tin từ hình ảnh; nhận dạng ký tự quang học.*

ABSTRACT

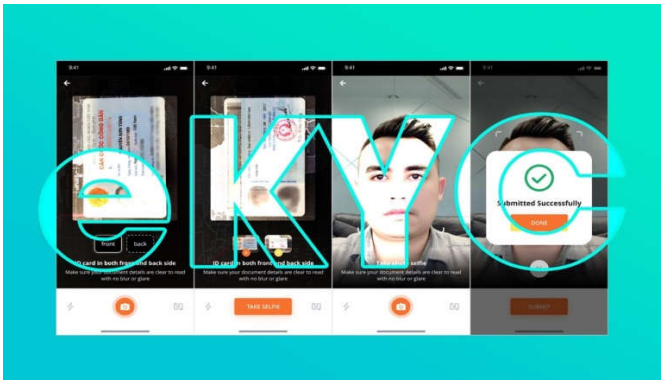
In this modern information technology era, automatically extracting information from identity documents is applied in bank, telecommunications, healthcare, hotels, and more. This paper uses of computer vision techniques to build a system which extracts information from identity documents for eKYC system. There are three main steps: Identifying identity documents, locating information and extracting information. In particular, identifying indentity cards and locating information are accomplished by basic image processing techniques to increase the speed of the entire system. At the information extraction step, the achievements in text recognition on complex background are used to increase accuracy. After development and experimentation, the system has achieved a relatively high degree of accuracy. The accuracy of the whole system is over 90%, this result can be applied to the eKYC system of banks and credit institutions.

Keywords: *Computer Vision; YOLOv4; extract information from images; Optical Character Recognition.*

[1]Trường Đại học Thuỷ Lợi - Phân hiệu Miền Nam
[2]Trung tâm Công nghệ thông tin, Trường Đại học Công nghiệp Hà Nội
[3]Trường Đại học Phan Thiết
[4]Trường Đại học Yersin Đà Lạt
[*]Email: vienthanhnha@tlu.edu.vn
Ngày nhận bài: 13/01/2022
Ngày nhận bài sửa sau phản biện: 05/3/2022
Ngày chấp nhận đăng: 25/4/2022

1. GIỚI THIỆU

Trong những năm gần đây, nhu cầu định danh khách hàng điện tử tăng trưởng nhanh chóng. Các doanh nghiệp về dịch vụ thi đua với nhau xây dựng các hệ thống định danh khách hàng điển tử nhằm tạo sự thuận lợi cho khách hàng và giúp doanh nghiệp giảm chi phí trong các khâu quản lí. Trước khi đến với thuật ngữ định danh khách hàng điện tử (eKYC) thì KYC (Know Your Customer) là nhận biết khách hàng của bạn. Đây là một quy trình nhằm xác minh danh tính của khách hàng khi tham gia vào các dịch vụ của ngân hàng, bảo hiểm, tài chính,... KYC là bước đầu tiên trong tất cả các hoạt động trước khi khách hàng sử dụng sản phẩm/dịch vụ của doanh nghiệp đó. Hiểu đơn giản KYC giúp các doanh nghiệp đảm bảo giao dịch đó của khách hàng là chính chủ, giúp các doanh nghiệp đánh giá và giảm rủi ro, ngăn ngừa các hành vi gian lận bất hợp pháp.



Hình 1. eKYC (Nguồn: https://fpt.ai/vi/ek)

eKYC giúp quy trình xác minh danh tính của khách hàng được thực hiện ngay lập tức, mọi lúc mọi nơi, làm gia tăng tỉ lệ chuyển đổi khách hàng thành công. Khách hàng chỉ mất khoảng 3 - 5 phút để hoàn tất xác minh để đăng ký tài khoản. Quá trình này thường có 3 bước:

- Bước 1 xác minh tài liệu (chụp giấy tờ tuỳ thân).
- Bước 2 trích xuất tự động thông tin giấy tờ tuỳ thân.
- Bước 3 đối chiếu hình ảnh người đăng ký bằng sách so khớp video hoặc hình ảnh chân dung với hình ảnh trong giấy tờ tuỳ thân.

Ở phạm vi của nghiên cứu này, chúng tôi nghiên cứu sâu vào bước 2 (trích xuất tự động thông tin từ giấy tờ tuỳ thân). Giấy tờ tuỳ thân được chúng tôi sử dụng là chứng minh nhân dân (CMND), căn cước công dân không có chip và hộ chiếu. Ở bài nghiên cứu này tác giả sử dụng mô hình YOLOv4 để phát hiện các trường thông tin có trong giấy tờ tuỳ thân và sử dụng 2 thư viện OCR là Google Tesseract [1] và VietOCR [2] cho phần trích xuất các ký tự có trong hình ảnh sau khi đã phát hiện. Cuối cùng chúng tôi sẽ xây dựng hệ thống này theo mô hình Client-Server [3] theo kiến trúc Restful API dưới sự hỗ trợ của thư viện Flask.

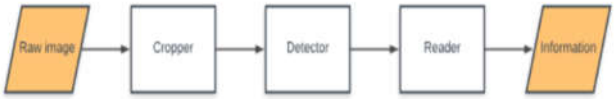
2. NHỮNG NGHIÊN CỨU LIÊN QUAN

Trong những năm gần đây, thị giác máy tính đã có một số thành tựu trong lĩnh vực nhận dạng chữ viết in. Trong đó, Tesseract OCRs được phát triển bởi Hewlett Packard (HP) và được tài trợ bởi Google, được dùng cho 116 ngôn ngữ khác nhau và đạt độ chính xác lên đến 99%. Tuy nhiên, độ chính xác của Tesseract phụ thuộc vào từng ngôn ngữ cần nhận dạng. Tesseract cũng nhằm mục đích giải quyết vấn đề với các tài liệu văn bản ở dạng quét hình ảnh với nền đơn giản. Có văn bản trên một nền tảng phức tạp, các nhà khoa học đang vẫn rất quan tâm đến và đã thực hiện những thành tựu nhất định, nhưng chủ yếu là với Dữ liệu ngôn ngữ tiếng Anh.

3. PHƯƠNG PHÁP ĐỀ XUẤT

3.1. Mô hình tổng quát

Kiến trúc tổng quát của hệ thống bao gồm ba thành phần cơ bản: Cropper, Detector và Reader. Ở phần Cropper và Detector, nhóm tác giả sử dụng mô hình YOLOv4 để huấn luyện và thu lại file trọng số của mô hình. Ở phần Reader, nhóm tác giả thực hiện một vài kỹ thuật xử lý ảnh cơ bản và sử dụng một trong hai thư viện là Google Tesseract hoặc VietOCR để trích xuất các ký tự. Mô hình tổng quát này được mô tả ở hình 2.



Hình 2. Mô hình tổng quát của hệ thống

3.2. Cropper và Detector

Để thực hiện phát hiện và nhận diện giấy tờ tuỳ thân, nhóm chúng tôi lựa chọn thuật toán YOLOv4 [4] là thuật toán YOLO mới nhất do Alexey B., Chien-Yao W., Hong-Yuan và Mark L. công bố vào ngày 23/04/2020. YOLO (You Only Look Once), tạm dịch là "Bạn chỉ nhìn một lần". Thuật toán YOLO xem bài toán phát hiện vật thể là một vấn đề hồi quy duy nhất trên toàn bộ bức ảnh, trực tiếp từ các pixel của ảnh thành các ô dự đoán (bounding box) cùng với xác suất phân loại vật thể thay vì phải thực hiện nhiều bài toán hồi quy cho từng vùng vật thể như các thuật toán R-CNN series. Như vậy việc chỉ sử dụng 1 bài toán hồi quy duy nhất cho toàn bộ ảnh, thuật toán YOLO giúp giảm số lượng phép toán, tăng tốc độ xử lý khi đó có thể đáp ứng bài toán thời gian thực tốt hơn so với các thuật toán R-CNN.

Yolo là một mô hình mạng CNN cho việc phát hiện, nhận dạng, phân loại đối tượng. Yolo được tạo ra từ việc kết hợp giữa các Convolutional layers và Connected layers. Trong đó các Convolutional layers sẽ trích xuất ra các feature của ảnh, còn full-connected layers sẽ dự đoán ra xác suất đó và tọa độ của đối tượng.

YOLOv4 là một loạt các cải tiến về tốc độ so với YOLOv3 và được cài đặt từ một bản fork của Darknet. Kiến trúc của YOLOv4 đã đưa bài toán object detection dễ tiếp cận hơn với những người không có tài nguyên tính toán mạnh. Chúng ta hoàn toàn có thể huấn luyện một mạng phát hiện vật với độ chính xác rất cao bằng YOLOv4 chỉ với GPU (1080ti hoặc 2080ti). Trong tương lai, việc tối ưu lại các mạng hiện tại để phù hợp với tài nguyên tính toán yếu hoặc tạo ra sự song song hóa cao ở các server chắc chắn phải được thực hiện để có thể đưa các ứng dụng thị giác máy tính vào thực tế.

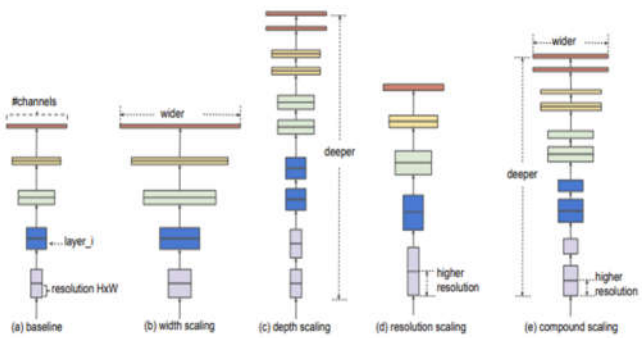
YOLOv4 được cấu tạo từ 4 phần:

- ❖ Backbone (xương sống) - trích xuất đặc trưng

Mạng xương sống cho nhận dạng vật thể thường được đào tạo trước (pre-train) thông qua bài toán phân loại ImageNet. Pre-train có nghĩa là trọng số của mạng đã được điều chỉnh để xác định các đặc điểm liên quan trong một hình ảnh, mặc dù chúng sẽ được tinh chỉnh trong nhiệm vụ mới là phát hiện đối tượng. Tác giả xem xét sử dụng các xương sống: CSPResNext50, CSPDarknet53 và EfficentNet-B3.

DenseNet được thiết kế để kết nối các lớp trong Mạng nơ-ron phức tạp nhằm mục đích để giảm bớt vấn đề gradient biến mất (khó có thể sao chép tín hiệu đã bị thất thoát trong một mạng sâu), để tăng cường lan truyền tính năng, khuyến khích mạng sử dụng lại các tính năng và giảm số lượng thông số mạng.

EfficientNet được thiết kế bởi Google Brain để chủ yếu nghiên cứu vấn đề mở rộng quy mô của Mạng nơ-ron tích chập. Có rất nhiều quyết định mà bạn có thể đưa ra khi mở rộng ConvNet của mình bao gồm kích thước đầu vào, tỷ lệ chiều rộng, tỷ lệ chiều sâu và mở rộng tất cả những điều trên. EfficientNet cho rằng có một điểm hoàn hảo, có thể tối ưu cho tất cả các thông số đó và thông qua tìm kiếm, họ đã tìm thấy điểm đó.



Hình 3. Mô hình EfficientNet (Nguồn: https://devai.info/)

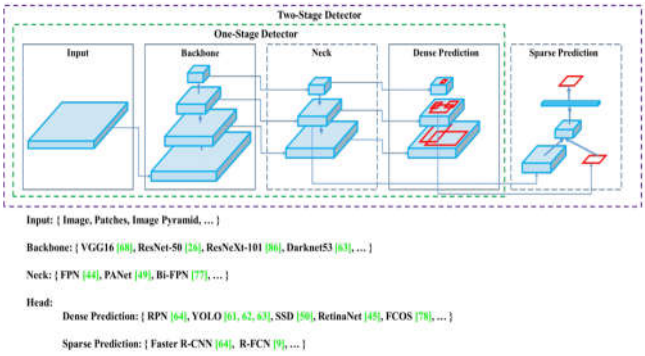
EfficientNet vượt trội hơn các mạng khác có cùng kích thước về phân loại hình ảnh. Tuy nhiên, tác giả của YOLOv4 cho rằng các mạng khác có khả năng hoạt động tốt hơn trong cài đặt để phát hiện đối tượng nên quyết định thử nghiệm với tất cả 3 mạng CNN ở trên. Cuối cùng thì nhóm tác giả chọn mạng CSPDarknet53 là backbone cho model.

❖ Neck (cổ)

Neck có nhiệm vụ trộn và kết hợp các bản đồ đặc trưng (Features map) đã học được thông qua quá trình trích xuất đặc trưng (backbone) và quá trình nhận dạng (YOLOv4 gọi là Dense prediction).

+ Dense prediction (dự đoán dày đặc): sử dụng các one-stage-detection như YOLO hoặc SSD.

+ Sparse Prediction (dự đoán thưa thớt): sử dụng các two-stage-detection như RCNN.



Hình 4. Cấu trúc mô hình YOLOv4

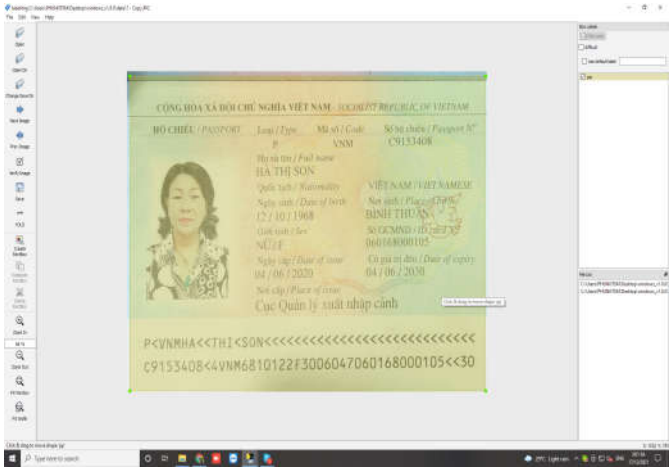
Để huấn luyện mô hình nhận diện giấy tờ tuỳ thân và trường bên trong giấy tờ sử dụng thư viện Darknet [5] trên server chạy hệ điều hành Ubuntu 18.04 đã được cài đặt môi trường CUDA 10.2 và CUDNN 7.6.5 với GPU RTX 2070 Super 8Gb giúp cho quá trình huấn huyện mô hình được nhanh hơn.

Trước khi huấn luyện mô hình cần tạo tập dữ liệu cho mô hình, trong nghiên cứu này, nhóm tác giả thu thập hình ảnh các loại giấy tờ tuỳ thân dựa trên mạng Internet và sau đó tiến hành gán nhãn cho các hình ảnh đó bằng phần mềm LabelImg [6].

Ở model Cropper ta huấn luyện mô hình nhằm mục đích tách hình ảnh giấy tờ tuỳ thân ra khỏi nền giống như hình 5.

Hình 5. Minh hoạ mô hình Cropper

Để được kết quả như hình 5 ta sẽ dùng phần mềm LabelImg để gán nhãn bên ngoài ảnh của giấy tờ tuỳ thân, ví dụ như hình 6.



Hình 6. Minh hoạ gán nhãn cho mô hình Cropper

Với mô hình Detector nhằm mục đích nhận diện các trường chứa thông tin ta sẽ huấn luyện mô hình để phát hiện được các trường thông tin giống như hình 7. Mô hình này sẽ được huấn luyện để có thể phát hiện được các trường thông tin như sau: ở hộ chiếu sẽ phát hiện trường chứa họ tên, quê quán và dãy ký MRZ (Machine Readable Zone), ở căn cước công dân không có chip và chứng minh nhân dân sẽ phát hiện các trường chứa họ tên, ngày sinh, số chứng minh nhân dân, số căn cước công dân, quê quán, địa chỉ, giới tính, quốc tịch,…

Hình 7. Minh hoạ mô hình Detection

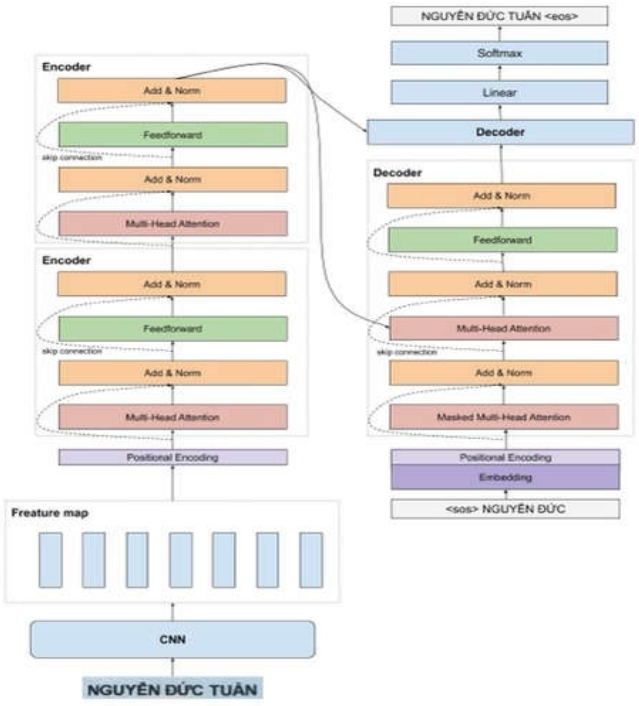
3.3. Mô hình Reader

Sau khi nhận diện được các trường chứa thông tin ta tiến hành ráp với mô hình đọc các thông tin từ trường đó ra. Ở hệ thống này, nhóm nghiên cứu sử dụng Thư viện VietOCR cho hai loại giấy tờ là căn cước công dân và chứng minh nhân dân, đối với hộ chiếu tác giả sử dụng cả hai thư viện VietOCR và Google Tesseract. Hai thư viện này mang lại độ chính xác khi đọc các chữ trong ảnh lên tới hơn 90%.

Thư viện VietOCR được tác giả Phạm Bá Cường Quốc xây dựng dựa trên mô hình Transformer OCR.

Thư viện Google Tesseract được sử dụng với bộ dữ liệu đã được huấn luyện tốt nhất (tessdata_best). Đối với font chữ MRZ trên Passport sử dụng mô hình mrz.trainneddata

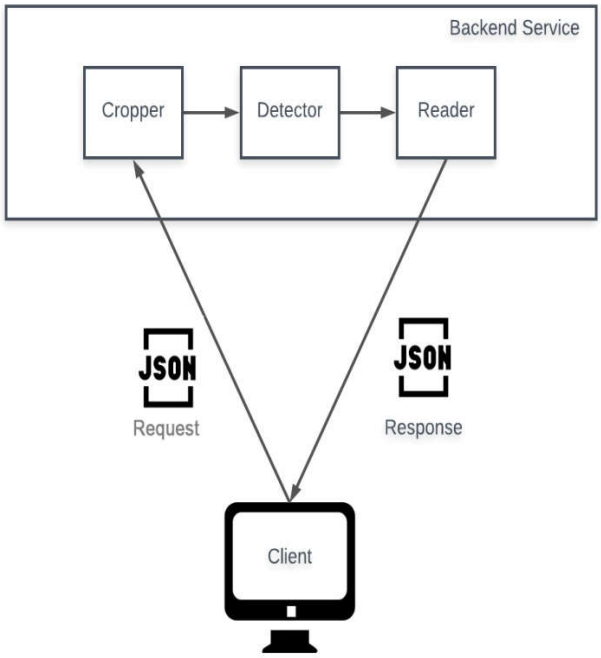
mô hình này do cộng đồng sử dụng Google Tesseract huấn luyện dựa trên font chữ MRZ.



Hình 8. Mô hình kiến trúc của thư viện VietOCR

3.4. Mô hình Backend Service

Sau khi đã huấn luyện xong mô hình Cropper, Dectector và Reader, nhóm nghiên cứu sử dụng mô hình Client-Server để triển khai hệ thống này vào sử dụng thực tế. Dưới sự hỗ trợ của thư viện Flask đã xây dựng mô hình này dưới dạng kiến trúc Restful API, được mô tả ở hình 9.

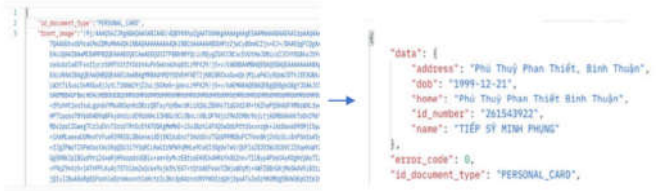


Hình 9. Kiến trúc hệ thống trích xuất

Client sẽ gửi cho Server yêu cầu có nội dung là một chuỗi JSON, trong chuỗi JSON này có chứa hình ảnh của giấy tờ tuỳ thân đã được mã hoá dưới dạng base64. Server sẽ tiến hành giải mã hình ảnh và trích xuất các thông tin trong hình ảnh sau đó sẽ gửi lại phản hồi về cho Client là một chuỗi JSON có chứa các trường thông tin của ảnh đó.

4. THỰC NGHIỆM VÀ KẾT QUẢ

Sau khi hoàn thiện xong hệ thống, nhóm nghiên cứu sử dụng phần mềm Postman để giả lập Client gửi yêu cầu lên Server. Hệ thống đang chạy trên server có chip xử lý Core I9-10900K cho thời gian phản hồi khá nhanh, khoảng 3 giây cho 1 yêu cầu từ máy Client và đạt được độ chính xác 99%.



Hình 10. Kết quả gửi yêu cầu và phản hồi của hệ thống

5. KẾT LUẬN

Mô hình phát hiện đối tượng YOLOv4 mang lại độ chính xác cao trong việc phát hiện các trường đối tượng và các thư viện Google Tesseract và VietOCR mang lại độ chính xác cao khi tiến hành trích xuất. Điều này mang lại trải nghiệm tốt cho khách hàng khi sử dụng dịch vụ định danh điện tử. Từ đó có thể áp dụng mô hình này cho nhiều loại giấy tờ khác nhau nhằm phục vụ cho việc số hoá tài liệu cũng như góp phần giảm chi phí hoạt động cho các doanh nghiệp dịch vụ.

TÀI LIỆU THAM KHẢO

[1]. Hewlett Packard, https://github.com/tesseract-ocr/tesseract

[2]. Pham Ba Cuong Quoc, https://github.com/pbcquoc/vietocr

[3]. Vu Thi Mai Phuong, https://viblo.asia/p/so-luoc-ve-mo-hinh-client-server-va-giao-thuc-http-jvElaarNlkw

[4]. Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, 2020. *YOLOv4: Optimal Speed and Accuracy of Object Detection.*

[5]. Joseph Redmon, https://pjreddie.com/darknet/

[6]. Tzutalin, https://github.com/tzutalin/labelImg

AUTHORS INFORMATION

Vien Thanh Nha[1], Tiep Sy Minh Phung[1], Nguyen Hoang Tu[2], Do Thi Kim Dung[3], Le Dinh Phu Cuong[4]

[1]Thuyloi University - Southern Campus

[2]Center of Information Technology, Hanoi University of Industry

[3]University of Phan Thiet

[4]Yersin University