



# Nhận dạng văn bản cho chứng minh nhân dân Việt Nam dựa trên mạng tính năng sâu

Duc Phan Van Hoa<sup>1</sup> · Huu-Thanh Duong<sup>1</sup> · Vinh Truong Hoang<sup>1</sup>



Đã nhận: 8 tháng 4 năm 2020 / Sửa đổi: 21 tháng 1 năm 2021 / Đã chấp nhận: 26 tháng 1 năm 2021 © Tác giả, theo giấy phép độc quyền cho Springer-Verlag GmbH, một phần DE của Springer Nature 2021

trừu tượng

Nhận dạng ký tự quang học (OCR) là công nghệ chuyển đổi văn bản tự động trên hình ảnh thành chuỗi dữ liệu để chỉnh sửa, lập chỉ mục và tìm kiếm. Các chuỗi có thể được áp dụng cho nhiều tác vụ như số hóa các tài liệu cũ, dịch sang các ngôn ngữ khác hoặc để kiểm tra và xác minh các vị trí văn bản. Gần đây, Biết khách hàng của bạn (KYC) đã trở thành một tiêu chuẩn ngành để đảm bảo rằng mọi người đúng như họ nói. Mặc dù phạm vi Biết khách hàng của bạn không ngừng mở rộng, xác minh ID vẫn là bước đầu tiên quan trọng trong quy trình KYC. OCR di động là một trong những giải pháp công nghệ giúp khách hàng tuân thủ phần này của KYC dễ dàng hơn bao giờ hết. Quy trình KYC yêu cầu các công ty dịch vụ tài chính xác minh danh tính của khách hàng OCR để trích xuất dữ liệu bằng cách đọc ID, thẻ ngân hàng và tài liệu. Trong bài báo này, chúng tôi nghiên cứu để phát triển một phương pháp nhận dạng chứng minh nhân dân Việt Nam dựa trên mạng lưu ý tính năng chuyên sâu. Trên một số dữ liệu chính của chứng minh thư, nó đạt được độ chính xác tương ứng là hơn 96,7% và 89,7% ở cấp độ ký tự và cấp độ từ.

Từ khóa OCR · Nhận dạng văn bản · Quy trình KYC · Xác minh ID · Mạng tính năng chuyên sâu · Thẻ ID Việt Nam

## 1. Giới thiệu

Nhận dạng ký tự quang học (OCR) là một trong những chủ đề nghiên cứu chính trong thị giác máy, biến hình ảnh được quét của văn bản viết tay, đánh máy thành tài liệu hoàn toàn có thể tìm kiếm và có thể chỉnh sửa với nội dung văn bản. Nó có các ứng dụng công nghiệp tiềm năng khác nhau như dịch máy, chuyển văn bản thành giọng nói, khai thác văn bản, xác minh bảo mật, hỗ trợ người khuyết tật và trình đọc tự động. Vì vậy, công nghệ OCR đã được áp dụng để cải cách quy trình quản lý tài liệu.

Gần đây, công nghệ tài chính được cách mạng hóa thành một mô hình mới với nhiều ứng dụng có thể xảy ra cho các công ty dịch vụ tài chính và ngân hàng. Gần đây, Biết khách hàng của bạn (KYC) trở thành một tiêu chuẩn để xác minh xem họ nói họ là ai [5]. Trong khi phạm vi của KYC liên tục được mở rộng, xác minh tài liệu nhận dạng (ID) vẫn là giai đoạn đầu tiên cần thiết trong các quy trình KYC. OCR di động là một trong những giải pháp công nghệ giúp khách hàng tuân thủ phần này của KYC dễ dàng hơn bao giờ hết.

Rất nhiều phương pháp tiếp cận nhận dạng văn bản đã được đưa ra chuyên nghiệp, và những cách tiếp cận này có thể được nhóm thành 3 loại như phương pháp tiếp cận dựa trên ký tự, phân loại toàn bộ từ và phương pháp tiếp cận nhận dạng dựa trên trình tự. Fernandez và cộng sự. [15] đã giới thiệu một phương pháp cho hệ thống OCR thư ứng mại dựa trên việc tăng cường màu sắc và mã hóa ký tự văn bản. Clausner và cộng sự. [8] cải thiện công cụ Tesseract OCR bằng cách sử dụng phân tích và nhận dạng tài liệu Aletheia. Ba thành phần giai đoạn chính được yêu cầu để tích hợp vào Aletheia như quy trình đào tạo của công cụ OCR, nhận dạng văn bản và đánh giá định lượng. Pramanik và Bag [36] trình bày một phương pháp nhận dạng văn bản Bangla giúp phân tách vùng phân đoạn với hai hình dạng cơ bản.

Sau đó, một tính năng biểu đồ mã dựa trên tính năng perceptron nhiều lớp được sử dụng để phân loại. Attivissimo và cộng sự. [2] trình bày một hệ thống nhận dạng văn bản dựa trên hình ảnh của các tài liệu nhận dạng của Ý. Thử nghiệm được thực hiện trên tập dữ liệu tổng hợp do các vấn đề về quyền riêng tư. Phương pháp này chủ yếu dựa trên khung CNN bằng cách sử dụng mô hình phụ đề văn bản một dòng. Holanda và cộng sự. [20] đề xuất một hệ thống OCR để đọc văn bản dựa trên tính năng Central Moments dành cho người khiếm thị. Văn bản được tạo từ hình ảnh sẽ được dịch sang chữ nổi trên thiết bị di động. Du và cộng sự. [13] đề xuất một phương pháp hồi quy tuyến tính để cải thiện trình phân loại dựa trên nguyên mẫu trong ngôn ngữ Trung Quốc. Nghiên cứu này

B Vinh Truong Hoang  
vinh.th@ou.edu.vn

<sup>1</sup> Khoa Công nghệ Thông tin, Thành phố Hồ Chí Minh Mở rộng  
Đại học, 97 Võ Văn Tần, Quận 3, Thành phố Hồ Chí Minh,  
Việt Nam



Hình 1 Một mẫu CMND Việt Nam mặt trước và mặt sau. Có một số thông tin cần thiết như ID quốc gia, Tên & DOB, Tên, địa chỉ & DOB, và Sinh trắc học. Thẻ ID Việt Nam là một nhiệm vụ rất khó để nhận dạng vì nó có nhiều phông chữ và biến dạng văn bản khác nhau

cũng kiểm tra tác động của sự thích ứng được giám sát đối với kiểu phông chữ.

Ngoài ra, xử lý sau OCR là một quá trình để xác minh và sửa văn bản được tạo OCR từ hình ảnh được quét.

Elagouni và cộng sự. [14] trích hết áp dụng phân đoạn bằng cách quét ảnh đa tỷ lệ để phát hiện văn bản. Mô hình biểu đồ dựa trên cửa sổ trích xuất nhiều tỷ lệ được áp dụng cho các ký tự nhận lại ký tự ở mỗi vị trí và tỷ lệ. Một số mô hình ngôn ngữ cũng được tích hợp để sửa lỗi nhận dạng văn bản.

Mei và cộng sự. [33] tích hợp kho ngữ liệu quy mô web gồm một tập hợp các tính năng ngôn ngữ để sửa lỗi và đề xuất cách sửa lỗi cho OCR của sách sinh học lịch sử. Doush và cộng sự. [11] đề xuất thuật toán Căn chỉnh văn bản bám tự động nhanh cho tài liệu (FAHTA) để đánh giá độ chính xác OCR của tài liệu được quét. Đây là một phương pháp thống kê cho

Xử lý hậu OCR kết hợp hệ thống dựa trên quy tắc và hệ thống kết hợp.

OCR đã được nghiên cứu thành công bằng các ngôn ngữ khác nhau. Ví dụ, Oni et al. [34] trình bày một phương pháp để nhận dạng ký tự Yorùbá dựa trên mô hình Bộ nhớ Ngắn hạn Dài hạn (LSTM) kết hợp trình sửa lỗi kiểm tra chính tả. Piroonsup và cộng sự. [35] kết hợp bộ phân loại bán giám sát và phương pháp giảm nhiễu để nhận dạng ảnh tài liệu Thái Lan. Arafat và cộng sự. [1] giới thiệu FasterRCNN tùy chỉnh kết hợp với một số CNN như Squeezenet, Googlenet, Resnet18 và Resnet50 để phát hiện và nhận dạng văn bản tiếng Urdu trong tự nhiên. Darwish và cộng sự. [9] trình bày một phương pháp mới dựa trên Thuật toán di truyền (GA) và bộ phân loại Mờ K-Nearest Neighbor cho ứng dụng OCR tiếng Ả Rập. Trên thực tế, việc ứng dụng OCR vào nhận dạng văn bản tiếng Anh đã đạt được những kết quả tốt [4].

Tuy nhiên, các phương pháp này không thể áp dụng trực tiếp cho văn bản tên tiếng Việt do có sự khác biệt đáng kể giữa tiếng Anh và tiếng Việt với sáu thanh điệu riêng biệt. Có tất cả 29 chữ cái trong bảng chữ cái tiếng Việt. Hình 1 minh họa một ví dụ về thẻ căn cước Việt Nam có mặt trước và mặt sau.

Lưu ý rằng, tất cả các thẻ ID đều được bao phủ bởi một lớp nhựa nên đôi khi rất khó chụp ảnh trong một số điều kiện ánh sáng trực tiếp.

Trong bài báo này, chúng tôi tập trung vào nhận dạng văn bản trên ảnh thẻ chứng minh nhân dân Việt Nam hơn là ảnh cảnh thiên nhiên. Do đó, CMND Việt Nam được sử dụng trong nhiều bối cảnh khác nhau như giao dịch ngân hàng, kiểm soát an ninh, đặt phòng trực tuyến, đăng ký hoặc quy trình nhận dạng. Hiện nay, ở Việt Nam, thông tin từ CMND được trích xuất thủ công, rất mất thời gian. Cần có hệ thống OCR cho nhiệm vụ này để thay thế con người. Tuy nhiên, OCR trên thẻ ID là một nhiệm vụ đầy thách thức vì nó đòi hỏi độ chính xác cao để áp dụng vào thực tế.

Do màu nền và hình in trên thẻ căn cước gần giống nhau nên ảnh thu được có độ nhiễu cao.

Ngoài ra, nó yêu cầu một lượng lớn dữ liệu do con người gắn nhãn để xây dựng các mô hình và quá trình xử lý phải nhanh chóng đối với một ứng dụng KYC.

Do đó, học sâu được coi là một phương pháp hiệu quả để xử lý dữ liệu lớn và cải thiện hiệu suất nhận dạng. Gần đây, nó đã được áp dụng thành công trong các ứng dụng khác nhau [17,18,21]. Pratama và cộng sự. [37] áp dụng CNN để nhận dạng thẻ ID Indonesia và cho thấy rằng cách tiếp cận được đề xuất vượt trội hơn các phương pháp thị giác máy tính truyền thống. Bulatov và cộng sự. [3] đã giới thiệu tập dữ liệu mới có tên là MIDV-2019, chứa các video clip về 50 loại tài liệu riêng biệt (thẻ ID, hộ chiếu và giấy phép lái xe) được chụp bằng điện thoại di động trong điều kiện ánh sáng yếu. Sau đó, thử nghiệm nhận dạng văn bản được đưa ra bởi khung Tesseract. Yin và cộng sự. [43] áp dụng ba mạng nơ-ron dựa trên Mạng nơ-ron Convolution (CNN) bao gồm Nhóm Hình học Trục quan (VGG), Mạng Capsule (CapsNet) và Mạng Dư (ResNet) để nhận dạng văn bản Trung Quốc. Lee và cộng sự. [26] áp dụng phương pháp lựa chọn tính năng để loại bỏ các tính năng không liên quan của OCR để nhanh chóng xác định thông tin cá nhân. Liem và cộng sự. giới thiệu ứng dụng dò tìm văn bản và nhận dạng CMND đầu tiên của Việt Nam [29]. Giai đoạn phát hiện văn bản được thực hiện bởi RetinaNet và mô hình Inception v3 được áp dụng để nhận dạng văn bản. Thử nghiệm được đánh giá trên tập dữ liệu riêng với 1.000 hình ảnh của thẻ căn cước.

Do những tiến bộ và hiệu quả của học sâu trong những năm gần đây, có một số phương pháp tiếp cận thành công đã được đề xuất để phát hiện và nhận dạng văn bản trong tự nhiên:

- Để phát hiện văn bản: Trình phát hiện văn bản theo cảnh hiệu quả và chính xác (EAST) [44] và Mạng đề xuất văn bản kết nối (CTPN) [42] là những cách tiếp cận phổ biến nhất. EAST là một phương pháp phân đoạn văn bản dựa trên mô hình UNet [39]. EAST sử dụng bản đồ đối tượng địa lý cuối cùng để hồi quy điểm số và hình dạng của hộp giới hạn tứ giác của dòng văn bản. EAST có thể phát hiện văn bản trong hình ảnh quadrilateral tùy ý. CTPN là một phương pháp cơ sở neo được truyền cảm hứng bởi

Máy dò MultiBox Single Shot (SSD) [31] để xuất sử dụng lớp Bộ nhớ ngắn hạn dài (LSTM) [19] để nắm bắt thông tin tuần tự của văn bản. CTPN trước hết được thiết kế để chỉ chụp các dòng văn bản ngang. Gần đây, Deng et al. [10] giới thiệu một phương pháp mới để phát hiện văn bản cảnh, cụ thể là công việc Mạng tinh chỉnh tính năng định kỳ (RFRN). Đánh giá toàn diện về phát hiện văn bản có thể được tìm thấy trong [30].

- Đối với nhận dạng văn bản, mô hình được sử dụng rộng rãi nhất là Mạng nơ-ron tái diễn Con vol Hiên (CRNN) [38,40], bao gồm ba giai đoạn: phần phức hợp để trích xuất các đặc điểm hình ảnh, phần lặp lại để thu thập thông tin tuần tự từ các đặc điểm hình ảnh và dự đoán bộ kích hoạt ký tự ở mỗi vị trí, phần phiên mã để chuyển đổi các dự đoán và tạo văn bản.
- Để phát hiện và nhận dạng văn bản: Một số cách tiếp cận đã đề xuất các mô hình end-to-end để phát hiện và nhận dạng văn bản [28,32] đồng thời. Các ches khoảng này đã cho thấy làm giảm đáng kể thời gian tính toán bằng cách chia sẻ bước trích xuất đối tượng địa lý. Tuy nhiên, những mô hình này phức tạp hơn để được đào tạo.

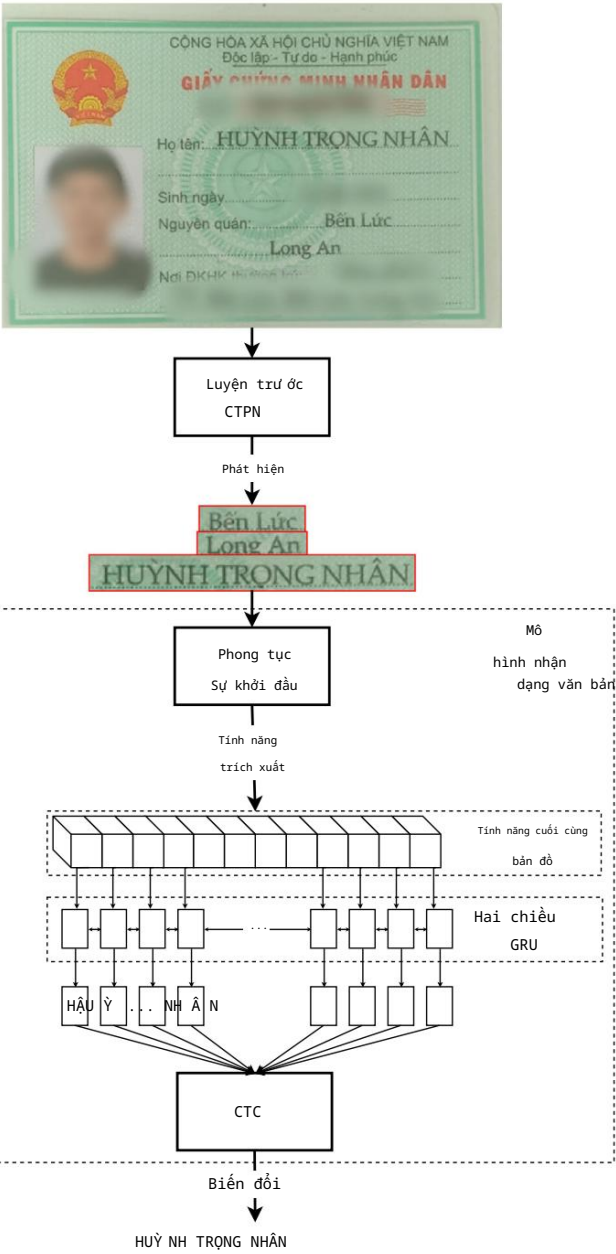
Để giải quyết các vấn đề về OCR cho CMND Việt Nam, chúng tôi đã đề xuất một khung bao gồm hai giai đoạn: phát hiện văn bản và nhận dạng văn bản. Mô hình CTPN đa ngôn ngữ được đào tạo trước được sử dụng làm mô hình phát hiện văn bản. Đối với nhận dạng văn bản, mô hình CRNN tùy chỉnh được áp dụng, mô hình này cũng bao gồm mô hình Xception tùy chỉnh [7], lớp Đơn vị lặp lại định hướng hai chiều (GRU) [6] và Phân loại thời gian kết nối (CTC) [16] để chuyển đổi out đư a vào một kết quả dòng văn bản. Hình 2 minh họa chi tiết cách tiếp cận được đề xuất đối với thẻ căn cước Việt Nam. Ví dụ về chứng minh thư Việt Nam bị làm mờ (thông tin cần thiết) để bảo mật. Từ phần trên cùng của hình này, một CTPN được đào tạo trước được áp dụng để phát hiện văn bản và cắt thành các bản vá. Những hình ảnh này sau đó được chuyển sang mô hình nhận dạng văn bản trên cus tom Xception để tạo ra văn bản.

Phần còn lại của bài báo này được tổ chức như sau. Phần 2 giới thiệu sơ lược về các tác phẩm liên quan. Tiếp theo, các tính năng tạo tư thế chuyên nghiệp và phương pháp lựa chọn tính năng được giới thiệu trong Sect. 3. Sau đó, Sect. 4 trình bày kết quả thực nghiệm của các phương pháp tiếp cận được đề xuất. Cuối cùng, kết luận được thảo luận trong Sect. 5.

2 tác phẩm liên quan

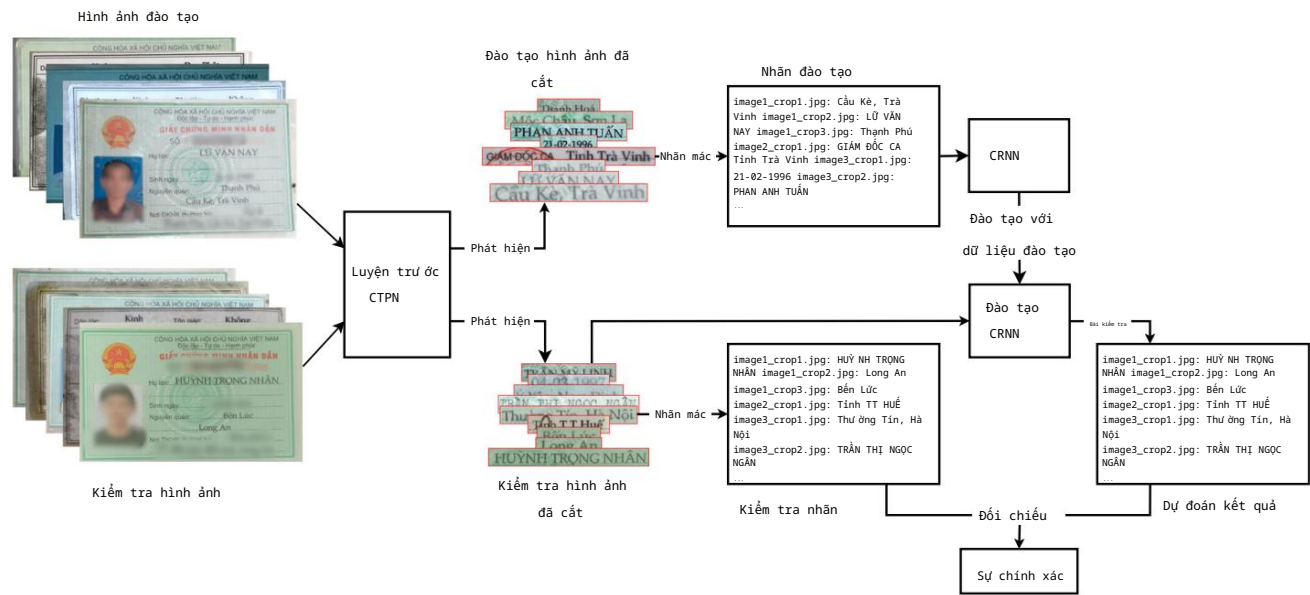
Trong phần này, chúng tôi xem xét ngắn gọn công việc Mạng nơron kết hợp (CNN), Mạng nơron định kỳ (RNN), VGG và CTPN.

CNN là một loại mạng nơ-ron nhân tạo, nó chủ yếu được ứng dụng cho các nhiệm vụ thị giác máy tính như phân loại đối tượng, phát hiện đối tượng và phân đoạn đối tượng. CNN là



Hình 2 Minh họa phương pháp đề xuất cho thẻ căn cước Việt Nam. Mô hình CTPN đa ngôn ngữ được đào tạo trước được sử dụng để phát hiện các vùng văn bản từ hình ảnh. Sau đó, mỗi vùng được cắt được thay đổi kích thước chiều cao thành 64 pixel theo tỷ lệ khung hình và được đưa vào mô hình nhận dạng văn bản. Mô hình Xception tùy chỉnh được sử dụng để trích xuất đối tượng địa lý từ vùng đã cắt với kích thước 64 × w × 3. Bản đồ đối tượng cuối cùng có hình dạng 1 × × 256 sau đó được đưa vào lớp GRU để trích xuất đối tượng cuối cùng. Cuối cùng, CTC được sử dụng để chuyển đổi thành kết quả văn bản cuối cùng.

lần đầu tiên được giới thiệu vào năm 1989 [25]. Tuy nhiên, nó chỉ trở nên phổ biến sau chiến thắng trên ImageNet Large Scale Visual Recognition Challenge vào năm 2012 [23]. Hai yếu tố chính góp phần tạo nên thành công của CNN là sự xuất hiện của các



Hình 3 Minh họa khung triển khai hệ thống OCR cho thẻ căn cước Việt Nam. Thứ nhất, đa ngôn ngữ được đào tạo trước của mô hình CTPN được sử dụng để cắt các vùng văn bản trong cả hình ảnh đào tạo và thử nghiệm.

Tiếp theo, một công cụ gắn nhãn OCR được sử dụng để gắn nhãn cho tất cả các vùng văn bản đã cắt. Sau đó, tất cả các vùng được cắt từ hình ảnh đào tạo và nhân của nó được sử dụng

để đào tạo mô hình CRNN được đề xuất. Sau đó, tất cả các vùng được cắt từ ảnh thử nghiệm được đưa vào mô hình CRNN đã được đào tạo để dự đoán kết quả. Cuối cùng, các nhãn thử nghiệm và kết quả dự đoán được so sánh để tính độ chính xác

quy mô tập dữ liệu và cải thiện sức mạnh tính toán của Bộ xử lý đồ họa (GPU).

RNN là một loại mạng nơ-ron nhân tạo được thiết kế để xử lý dữ liệu chuỗi thời gian, chẳng hạn như văn bản, lời nói, chứng khoán và thời tiết. RNN được phân biệt với các loại mạng nơ-ron nhân tạo khác bởi kiến trúc của nó, đầu ra của bộ nhớ thời gian trước được xử lý và sử dụng làm đầu vào của bộ nhớ thời gian tiếp theo. Do đó, khi RNN đưa ra quyết định, nó không chỉ dựa trên dữ liệu đầu vào hiện tại mà còn cả dữ liệu ưu việt. Hai trong số các kiến trúc RNN phổ biến nhất là Đơn vị dòng lặp Gated (GRU) và Bộ nhớ ngắn hạn dài (LSTM). Cả hai kiến trúc đều có các đơn vị bộ nhớ để lưu giữ thông tin từ các bước thời gian trước đó. GRU có hai cổng để kiểm soát dữ liệu đầu vào từ bộ nhớ thời gian trước và đầu vào hiện tại. LSTM dựa trên GRU và nó có thêm hai cổng.

VGG là một kiến trúc dựa trên AlexNet [24] được Simonyan và Zisserman đề xuất vào năm 2015 [41]. Mạng này đạt được độ chính xác top 5 là 92,3% trên ImageNet Challenge. Để thiết kế VGG trở thành một cấu trúc lưu trữ sâu hơn, họ chỉ sử dụng các lớp chập 3 × 3. Bằng cách xếp chồng một số lớp chập 3 × 3, chúng có được cùng một trường tiếp nhận hiệu quả như một lớp tích chập kích thước hạt nhân lớn hơn. Bằng cách đó, mô hình sâu hơn và có nhiều mối quan hệ phi tuyến hơn. Có một số biến thể VGG, trong số các phần mở rộng này, VGG16 và VGG19 được sử dụng rộng rãi nhất.

Một trong những mô hình phổ biến nhất để phát hiện văn bản là CTPN, tương tự như mô hình phát hiện đối tượng

Ổ cứng SSD [31]. CTPN sử dụng mô hình VGG16 để trích xuất các tính năng từ một hình ảnh, sau đó một LSTM hai hướng được sử dụng trên mỗi hàng của bản đồ đặc trưng để phát hiện văn bản ngang trong hình ảnh. Tại mỗi vị trí, CTPN dự đoán điểm số văn bản / không phải văn bản, tọa độ trục y và hiệu số chỉnh bên của neo liên kết. Cuối cùng, họ phát triển một thuật toán để kết nối văn bản gần đó thành một dòng văn bản.

### 3 Cách tiếp cận được đề xuất

Trong phần này, chúng tôi mô tả khuôn khổ để triển khai hệ thống OCR cho thẻ ID. Khung này có thể được sử dụng để phát triển một hệ thống OCR tương tự cho các loại tài liệu in khác bằng bất kỳ ngôn ngữ nào như hệ chiếu, giấy phép lái xe và biên lai. Hình 3 là một minh họa của phương pháp được đề xuất.

Đa ngôn ngữ được đào tạo trước của CTPN model1 được sử dụng để cắt tất cả các vùng văn bản trong tất cả các hình ảnh trong tập huấn luyện và thử nghiệm. Phần trên và dưới của vùng cắt được mở rộng bằng một phần năm (20%) chiều cao của vùng do các từ tiếng Việt có dấu thanh, điều này thường bị bỏ qua bởi mô hình CTPN. Tất cả các vùng đã cắt từ tập huấn luyện và tập thử nghiệm được lưu trữ để dán nhãn.

Để đơn giản hóa chú thích văn bản, chúng tôi phát triển một công cụ gắn nhãn đơn giản trong Máy tính xách tay Jupyter để gắn nhãn đã cắt

<sup>1</sup> <https://github.com/eragonxuan/text-detection-ctpn>.



Hình 4 Công cụ chú thích hình ảnh. Nút “Trở về” được sử dụng để điều hướng đến hình ảnh đã cắt trước đó. Nút “Tiếp theo” được sử dụng để điều hướng đến hình ảnh được cắt tiếp theo. Hộp văn bản bên dưới hai nút được sử dụng để nhập văn bản nhập văn bản chân thực của hình ảnh đã cắt tương ứng. Người dùng có thể nhấn “Enter” để nhập vào hộp văn bản để lưu văn bản sự thật cơ bản và chuyển tiếp đến hình ảnh tiếp theo. Dòng văn bản bên dưới hộp văn bản cho biết tên tệp của hình ảnh đã cắt hiện tại

vùng. Hình 4 cho thấy giao diện công cụ ghi nhãn. Tất cả các thay đổi được lưu vào tệp định dạng JSON bằng cách nhấn nút “Tiếp theo” hoặc nhấn “Enter” để nhập vào hộp văn bản.

Sau khi gắn nhãn tất cả các hình ảnh đã cắt từ tập huấn luyện. 80% hình ảnh đã cắt được sử dụng để đào tạo mô hình CRNN. Tất cả các vùng đã cắt được thay đổi kích thước chiều cao thành 64 pixel theo tỷ lệ khung hình. Mô hình CRNN được đề xuất bao gồm mô hình Xception tùy chỉnh để trích xuất tính năng, lớp GRU hai hướng để dự đoán ký tự tại mỗi vị trí và lớp CTC để tính toán tổn thất. Xception tùy chỉnh bao gồm một số lớp phức hợp, mô hình bao gồm ba lớp tổng hợp tối đa với bước đi (2, 2) và ba lớp tổng hợp tối đa với bước đi (2, 1). Do đó, nếu hình ảnh đầu vào có hình dạng 64 × w × 3, thì bản đồ tính năng cuối cùng có hình dạng 1 × × 256, được sử dụng làm đặc trưng chuỗi thời gian với các bước thời gian, mỗi bước thời gian có 256 đặc trưng. Bản đồ tính năng này được xử lý bởi lớp softmax để dự đoán nhân vật tại mỗi bước thời gian.

Thuật toán CTC được sử dụng để tính toán sự mất mát của các kết quả được làm mờ trước. Thuật toán giải mã được dẫn tốt nhất được sử dụng để giải mã các kết quả dự đoán trong giai đoạn thử nghiệm. Cuối cùng, để kiểm tra mô hình CRNN đã được đào tạo, tất cả các hình ảnh đã cắt từ bộ kiểm tra được đưa vào mô hình đã được đào tạo để tạo ra kết quả dự đoán. Các kết quả dự đoán sau đó được so sánh với nhãn sự thật trên mặt đất để có được độ chính xác. Để có được độ chính xác của từng trường dữ liệu cụ thể, công cụ OCR được sử dụng để gắn nhãn tất cả các hình ảnh đã cắt từ bộ thử nghiệm. Mỗi trường dữ liệu được gán cho một số xác định, ví dụ: trường số ID được gán nhãn là 1, trường tên được gán nhãn là 2, v.v.

## 4 thử nghiệm

Trong phần này, chuẩn bị dữ liệu, thiết lập thử nghiệm và kết quả được trình bày.

### 4.1 Mô tả tập dữ liệu

Cần lưu ý rằng dữ liệu là một yêu cầu thiết yếu để đào tạo các thuật toán OCR. Do đó, cần một số lượng hình ảnh đủ lớn để có được một mô hình tốt. Chúng tôi sở hữu bộ dữ liệu riêng gồm 2.500 ảnh thẻ ID. Những hình ảnh này được chụp bởi máy ảnh di động trong nhiều điều kiện khác nhau như bị mờ màu, nhòe hoặc nhãn với các ánh sáng khác nhau (ánh sáng yếu, đèn pin, ánh sáng mặt trời).

Theo cách tiếp cận của chúng tôi, tất cả hình ảnh được cắt thành hình chữ nhật với một sự xoay nhẹ do thực tế là các hệ thống OCR thường có thể hoạt động trong điều kiện cố định hướng. Cần có một thuật toán biệt danh để phát hiện thẻ ID nếu hệ thống OCR không hoạt động trong điều kiện cố định vị trí. Tập dữ liệu sau đó được chia thành tập huấn luyện và thử nghiệm với tỷ lệ 4: 1 (2.000 hình ảnh được sử dụng để huấn luyện và các hình ảnh còn lại được sử dụng để thử nghiệm. Chúng tôi chọn ngẫu nhiên 750 hình ảnh đào tạo và gắn nhãn cho tất cả các hình ảnh đã cắt được trích xuất từ những hình ảnh này. Đối với các hình ảnh khác được cắt từ 1.250 tập huấn luyện và 500 thử nghiệm, chúng tôi chỉ gắn nhãn cho các trường văn bản mong muốn như số ID, tên, ngày sinh, địa chỉ, ngày cấp, nơi cấp. Lý do là mất khoảng 2 tuần để dán nhãn cho 500 hình ảnh bằng một người.

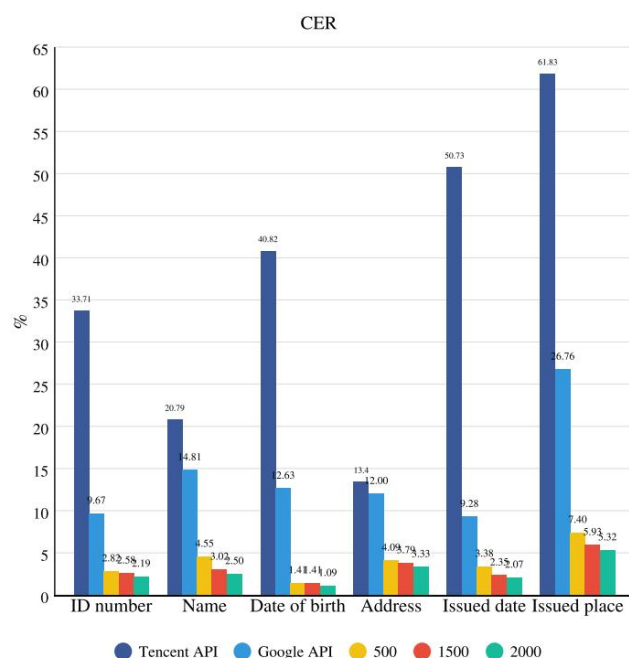
### 4.2 Thiết lập thử nghiệm

Tất cả các thử nghiệm đều được tiến hành trên phiên bản Linux của Google Cloud Computing Engine với cấu hình 8 CPU 2,5 GHz, 52 GB RAM, 1 GPU NVIDIA Tesla P100 16 GB RAM. Mô hình CRNN được triển khai bằng Keras phiên bản 2.2.4 với phiên bản phụ trợ Tensorflow 1.13.1. Trình tối ưu hóa Adam [22] được sử dụng để đào tạo mô hình CRNN. Tốc độ học ban đầu được đặt ở 10<sup>-4</sup>, sau đó tốc độ học tập giảm 75% khi không thể giảm mất bộ xác nhận trong 6 kỷ nguyên.

Ba thử nghiệm để huấn luyện CRNN với các lượng dữ liệu huấn luyện khác nhau đã được thực hiện. Sau khi đào tạo về dữ liệu thực trong 50 kỷ nguyên, điểm kiểm tra có mức tổn thất thấp nhất được tải và đánh giá bằng cách sử dụng hai số liệu: Tỷ lệ lỗi ký tự (CER) và Tỷ lệ lỗi từ (WER) [12]. Khoảng cách Levenshtein (LD) [27] được sử dụng để tính CER, WER. LD là một thuật toán để tính toán chi phí tối thiểu để biến đổi một chuỗi thành một chuỗi khác. Có ba loại lỗi cần được xem xét: thêm một ký tự mới, xóa một bộ kích hoạt ký tự hiện có hoặc thay thế một ký tự hiện có bằng một ký tự khác. Chi phí của mỗi hoạt động có thể được sửa đổi để phù hợp với yêu cầu của vấn đề. Trong các thử nghiệm của chúng tôi, chi phí của mỗi thao tác được đặt thành 1, do đó chi phí LD là số ký tự sai. Do đó, CER được tính là:

$$CER = \frac{\text{Chi phí LD}}{\text{Độ dài của chuỗi nhãn}} \times 100 \tag{1}$$





Hình 5 CER trên sáu trường văn bản mong muốn của mô hình CRNN sử dụng hình ảnh 500, 1500, 2000 làm dữ liệu đào tạo

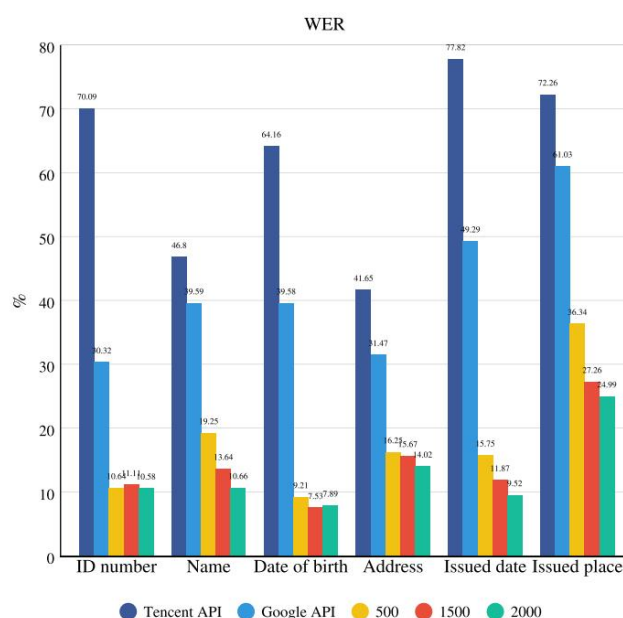
Để tính WER, mỗi từ của hai chuỗi được gán là một số nguyên duy nhất, sau đó thuật toán LD được sử dụng để tính toán chi phí chuyển đổi chuỗi số nguyên thứ nhất sang thứ hai. Cuối cùng, WER được tính là:

$$WER = \frac{\text{Chi phí LD}}{\text{Số từ trong chuỗi nhân}} \times 100. \quad (2)$$

Chúng tôi sử dụng mã nguồn từ <https://github.com/aflc/editdistance> để tính toán chi phí LD. Hơn nữa, mã nguồn của bài báo này có sẵn theo yêu cầu của các tác giả.

#### 4.3 Kết quả

Hình 5 cho thấy kết quả nhận dạng 500 thẻ căn cước Việt Nam theo phương pháp đề xuất bằng cách sử dụng các bộ dữ liệu đào tạo khác nhau. Kết quả này cũng được so sánh với Google API OCR Vision2 và Tencent API OCR.3 Các API này có thể phát hiện nhiều loại ngôn ngữ (bao gồm cả tiếng Việt) và có thể phát hiện nhiều ngôn ngữ trong một hình ảnh duy nhất. Để so sánh công bằng, tất cả hình ảnh được cắt từ giai đoạn CTPN đều được gửi đến API đó để nhận dạng văn bản. Kết quả tốt nhất nhận được bằng cách chọn ngôn ngữ tiếng Việt làm tham số tùy chọn. Từ hình này, chúng tôi quan sát thấy CER nhỏ hơn 8% trên tất cả sáu trường văn bản và tỷ lệ lỗi giảm dần khi sử dụng nhiều dữ liệu đào tạo hơn. Mô hình CTPN luôn phát hiện dòng văn bản, do đó nó thường cắt



Hình 6 WER trên sáu trường văn bản mong muốn của mô hình CRNN sử dụng hình ảnh 500, 1500, 2000 làm dữ liệu đào tạo

vùng văn bản ở phía trước vùng dữ liệu quan trọng. Tuy nhiên, độ chính xác của trường văn bản số CMND, ngày sinh, ngày cấp đạt hơn 96,5% do nó có một định dạng cố định và chủ yếu chứa số. Mặt khác, trường văn bản địa điểm đã phát hành đạt tỷ lệ lỗi cao nhất là hơn 5,32% do vùng văn bản bị nhiễu ở phía trước vùng dữ liệu mong muốn. Bằng cách sử dụng nhiều dữ liệu đào tạo hơn, CER của số ID, ngày sinh và trường dữ liệu địa chỉ sẽ giảm nhẹ. Do đó, để triển khai một hệ thống OCR cho các trường dữ liệu này, chúng ta chỉ cần một lượng nhỏ dữ liệu huấn luyện. Ngược lại, kết quả thu được bằng cách sử dụng Google API OCR Vision hoặc Tencent API OCR yếu hơn so với cách tiếp cận được đề xuất. Tỷ lệ lỗi của các trường quan trọng (số ID và tên) tương ứng là hơn 9% và 14% bởi Google API OCR. Vì vậy, nó đòi hỏi phải đào tạo và xây dựng mô hình cho một ứng dụng OCR cụ thể.

Ngoài ra, Hình 6 cho thấy rằng ngay cả CER rất thấp, WER cũng bị ảnh hưởng nghiêm trọng. CER của số ID, ngày sinh và trường văn bản ngày cấp dưới 3,5%. Đã bao giờ, WER của những trường này là hơn 7,5%. WER của trường dữ liệu địa điểm đã cấp trên 25%. Bằng cách sử dụng 2.000 hình ảnh thay vì 500 hình ảnh cho dữ liệu đào tạo, WER giảm đáng kể lần lượt là 8,6%, 6,2%, 11,4% về tên, ngày cấp, trường dữ liệu địa điểm. Tương tự, phương pháp tiếp cận tự thể chuyên nghiệp của chúng tôi vượt trội hơn đáng kể so với Google API OCR Vision và Tencent API OCR. WER và CER là các thước đo đánh giá phổ biến nhất để đo hiệu suất của hệ thống OCR. Hình 7 minh họa kết quả đầu ra mẫu của Google OCR API, Tencent OCR và cách tiếp cận của chúng tôi về hai hình ảnh được cắt xén về nơi cấp và ngày sinh.

<sup>2</sup> <https://cloud.google.com/vision/docs/ocr>.

<sup>3</sup> <https://intl.cloud.tencent.com/product/ocr>.

Nhận dạng văn bản cho chứng minh nhân dân Việt Nam ...

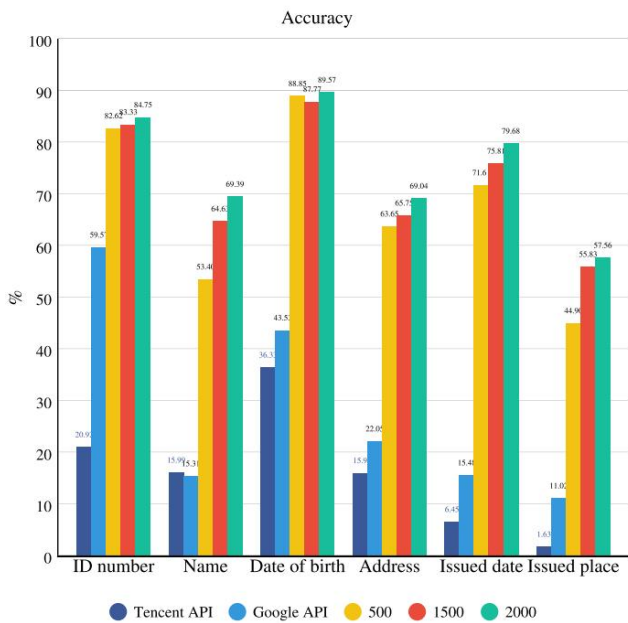
GIẤM ĐỐC CA ĐỒNG THÁP

	Google OCR	Tencent OCR	Our approach
Label	GIẤM ĐỐC CA ĐỒNG THÁP		
Prediction	GIẤM ĐỐC CA ĐỒNG THÁP	G ĐỒNG THÁP	GIẤM ĐỐC CA ĐỒNG THÁP
CER	0.09523809523809523	0.5238095238095238	0.0
WER	0.4	0.8	0.0

Sinh ngày 1978

	Google OCR	Tencent OCR	Our approach
Label	Sinh ngày 1978		
Prediction	Sinh ngày. 1978	Kebu quis 8Z6L	Sinh ngày 1978
CER	0.07142857142857142	0.8571428571428571	0.0
WER	0.3333333333333333	1.0	0.0

Hình 7 Minh họa nhận dạng văn bản từ ảnh cắt giấy CMND của người Việt Nam trên hai ô: Nơi cấp (trên) và Ngày tháng năm sinh (dưới)



Hình 8 Độ chính xác trên sáu trường văn bản mong muốn của mô hình CRNN sử dụng hình ảnh 500, 1500, 2000 làm dữ liệu đào tạo

Mặc dù các API này cho phép chọn ngôn ngữ tiếng Việt trước giai đoạn nhận dạng. Họ không thể nhận dạng chính xác văn bản tiếng Việt có dấu.

Để so sánh kết quả của chúng tôi với các phương pháp trước đó trong tài liệu, chúng tôi cũng chứng minh tỷ lệ nhận dạng theo độ chính xác như được minh họa trong Hình 8. Độ chính xác tốt nhất thu được là 68,5% độ chính xác cho các trường Số ID, Tên và Ngày sinh trong [ 29]. Thật không công bằng khi so sánh các kết quả thu được vì chúng tôi đã thử nghiệm trên các bộ dữ liệu khác nhau. Tuy nhiên, phương pháp đề xuất của bài báo này đạt được kết quả trung bình của 3 trường quan trọng (Số CMND, Tên và Ngày sinh) là 78,0% về độ chính xác.

Bằng cách so sánh ba kịch bản dữ liệu huấn luyện được sử dụng để xây dựng mô hình CRNN, chúng tôi nhận thấy rằng càng sử dụng nhiều dữ liệu, tỷ lệ lỗi của cả CER và WER càng giảm. Thật thú vị khi cải thiện hiệu suất bằng các kỹ thuật tăng dữ liệu hoặc thu thập dữ liệu.

5. Kết luận

Trong bài báo này, một khuôn khổ để triển khai hệ thống OCR cho thẻ ID được mô tả chi tiết. Hệ thống OCR này được triển khai dưới dạng API với hình ảnh thẻ ID làm dữ liệu đầu vào nhằm mục đích tích hợp vào các trang web hoặc ứng dụng hiện có. Chúng tôi cũng đã chuẩn bị phương pháp tiếp cận được đề xuất với Google API OCR Vision và Tencent OCR. Thử nghiệm cho thấy hệ thống này chỉ cần 500 ảnh CMND làm dữ liệu đầu vào để đạt CER dưới 3,4% vàWER dưới 15,8% trên một số trường văn bản chính của CMND như số CMND, tên, trường dữ liệu nơi cấp. Bằng cách sử dụng 2.000 ảnh thẻ ID làm dữ liệu đào tạo, CER của trường dữ liệu địa điểm đã cấp là 5,3% và CER nhỏ hơn 3,3% trên các trường dữ liệu khác, WER của trường dữ liệu địa điểm đã cấp là 25% và WER nhỏ hơn 10,7 % trên dữ liệu khác

lĩnh vực. Mô hình CTPN đào tạo trước thư vùng cắt vùng văn bản thừa trước vùng dữ liệu quan trọng, hơn nữa, việc phát hiện dấu thanh của văn bản tiếng Việt cũng kém hơn.

Phương pháp này có thể được sử dụng để triển khai hệ thống OCR như vậy bằng bất kỳ ngôn ngữ nào cho nhiều loại tài liệu in như hộ chiếu, giấy phép lái xe và biên lai. Do đó, tư duy lại của công việc này là đào tạo một mô hình phát hiện các dòng văn bản cho một vấn đề cụ thể. Quan điểm thứ hai là áp dụng và so sánh các kỹ thuật gia tăng dữ liệu để nâng cao chất lượng mỗi hình thức. Hơn nữa, tỷ lệ nhận dạng của hai trường văn bản (Địa chỉ và Nơi cấp) có thể được cải thiện bằng cách sử dụng các kỹ thuật xử lý hậu kỳ dựa trên một từ điển cụ thể.

Tuân thủ tiêu chuẩn đạo đức

Xung đột lợi ích Công việc này được hỗ trợ bởi Đại học Mở Thành phố Hồ Chí Minh, Việt Nam Xung đột lợi ích: Chúng tôi không có xung đột lợi ích để tuyên bố.

Người giới thiệu

1. Arafat, SY, Iqbal, MJ: Phát hiện và nhận dạng văn bản Urdu trong hình ảnh cảnh tự nhiên bằng cách sử dụng Deep Learning. IEEE Access 8, 96787-96803 (2020)

2. Attivissimo, F., Giaquinto, N., Scarpetta, M., Spadavecchia, M.: Trình đọc tự động tài liệu nhận dạng. Vào: Hội nghị liên quốc gia IEEE 2019 về Hệ thống. Con người và Điều khiển học (SMC), pp. 3525-3530. IEEE, Bari, Ý (2019)

3. Bulatov, K., Matalov, D., Arlazarov, VV: MIDV-2019: những thách thức của tài liệu OCR dựa trên thiết bị di động hiện đại. Trong: W. Osten, DP Nikolaev (eds.) Hội nghị quốc tế lần thứ mười hai về máy

- Tầm nhìn (ICMV 2019), tập 11433, trang 717 - 722. SPIE (2020), dự phòng Nhà xuất bản: Hiệp hội Quang học và Quang tử Quốc tế 4. Chaudhuri, A., Mandaviya, K., Badelia, P., K Ghosh, S.: Hệ thống nhận dạng ký tự quang học cho Các ngôn ngữ khác nhau với Máy tính mềm, Nghiên cứu về Tính mờ ảo và Máy tính mềm, tập. 352. Springer International Publishing, Cham (2017)
5. Chen, TH: Bạn có biết khách hàng của mình không? Đánh giá rủi ro ngân hàng dựa trên máy học. Máy tính mềm ứng dụng 86, 105779 (2020)
  6. Cho, K., van Merriënboer, B., Bahdanau, D., Bengio, Y.: Về đặc tính của dịch máy thần kinh: Phương pháp tiếp cận bộ mã hóa-giải mã. Trong: Kỷ yếu của SSST-8, Hội thảo lần thứ tám về cú pháp, ngữ nghĩa và cấu trúc trong bản dịch thống kê, pp. 103-111. Hiệp hội ngôn ngữ học tính toán, Doha, Qatar (2014)
  7. Chollet, F.: Xception: Học sâu với các bài tập có thể phân tách theo chiều sâu. Hội nghị IEEE 2017 về Thị giác Máy tính và Nhận dạng Mẫu (CVPR) trang 1800-1807 (2017)
  8. Clausner, C., Antonacopoulos, A., Pletschacher, S.: Đào tạo động cơ OCR hiệu quả và hiệu quả. Tạp chí Quốc tế về Phân tích và Ghi nhận Tài liệu (IJ DAR) (2019)
  9. Darwish, SM, Elzoghaly, KO: Mô hình Ara bic OCR được in ngoại tuyến nâng cao dựa trên Bộ phân loại mở lấy cảm hứng từ sinh học. IEEE Access 8, 117770-117781 (2020)
  10. Deng, G., Ming, Y., Xue, JH: RFRN: Mạng tính chính tính năng lặp lại để phát hiện văn bản cảnh chính xác và hiệu quả. Máy tính thần kinh p. S0925231220317124 (2020)
  11. Doush, IA, Alkhateeb, F., Gharaibeh, AH: Xử lý hậu kỳ OCR tiếng Ả Rập mới lạ sử dụng các kỹ thuật ngữ cảnh từ và dựa trên quy tắc. Tạp chí Quốc tế về Phân tích và Ghi nhận Tài liệu (IJ DAR) 21 (1-2), 77-89 (2018)
  12. Dreuw, P., Heigold, G., Ney, H.: Đào tạo về phân biệt MMI / MPE dựa trên độ tin cậy và dựa trên lễ cho khả năng nhận lại chữ viết tay không đúng dòng. Tạp chí Quốc tế về Phân tích và Nhận biết Tài liệu (IJ DAR) 14 (3), 273-288 (2011)
  13. Du, J., Huo, Q.: Một phương pháp hồi quy tuyến tính có phân biệt đối với sự thích ứng của các bộ phân loại dựa trên đa nguyên mẫu và các ứng dụng của nó cho OCR của Trung Quốc. Nhận dạng mẫu 46 (8), 2313-2322 (2013)
  14. Elagouni, K., Garcia, C., Mamalet, F., Sébillot, P.: Nhận dạng văn bản trong tài liệu đa phương tiện: một nghiên cứu về hai OCR dựa trên thần kinh sử dụng và tránh phân đoạn ký tự. Tạp chí Quốc tế về Phân tích và Công nhận Tài liệu (IJ DAR) 17 (1), 19-31 (2014)
  15. Fernández-Caballero, A., López, MT, Castillo, JC: Hiện thị phân đoạn văn bản sau khi học các tham số binarization OCR phù hợp nhất. Hệ thống Chuyên gia với Ứng dụng 39 (4), 4032-4043 (2012)
  16. Graves, A., Fernández, S., Gomez, F., Schmidhuber, J.. trang 369-376 (2006)
  17. Hoài, DPV, Hoàng, VT: Cung cấp tác phẩm Mạng Nơ-ron Convolutions bằng các tính năng thủ công dựa trên Hình ảnh Khác biệt Vùng lân cận Trung tâm Nâng cao để phân loại kết cấu màu sắc. Trong: Hội nghị quốc tế về phân tích đa phương tiện và nhận dạng mẫu (MAPR) năm 2019, trang 1-6 (2019)
  18. Hoai, DPV, Surinwarangkoon, T., Hoang, VT, Duong, HT, Meethongjan, K.: Một nghiên cứu so sánh về phân loại giống lúa dựa trên các tính năng học sâu và thủ công tr. 10 (2020)
  19. Hochreiter, S., Schmidhuber, J.: Trí nhớ ngắn hạn dài. Thần kinh Tính toán 9, 1735-1780 (1997)
  20. Holanda, GB, Souza, JWM, Lima, DA, Marinho, LB, Girão, AM, Bezerra Frota, JB, Rebouças Filho, PP: Phát triển hệ thống OCR trên nền tảng Android để hỗ trợ đọc với màn hình chữ nổi có thể làm mới trong thời gian thực . Phép đo 120, 150-168 (2018)
  21. Khosravi, P., Kazemi, E., Zhan, Q., Malmsten, JE, Toschi, M., Zisimopoulos, P., Sigaras, A., Lavery, S., Cooper, LAD, Hickman, C., Mesquer, M., Rosenwaks, Z., Elemento, O., Zaninovic, N., Hajirasouliha, I.: Học sâu cho phép đánh giá mạnh mẽ và tuyển chọn phối nang ngữ ời sau khi thụ tinh trong ống nghiệm. npj Digital Medicine 2 (1), 21 (2019)
  22. Kingma, D., Ba, J.: Adam: Một phương pháp tối ưu hóa ngẫu nhiên. Hội nghị quốc tế về đại diện học tập (2014)
  23. Krizhevsky, A., Sutskever, I., Hinton, GE: Phân loại ImageNet với mạng nơ-ron phức hợp sâu. Thông tin liên lạc của ACM 60 (6), 84-90 (2017)
  24. Krizhevsky, A., Sutskever, I., Hinton, GE: Phân loại mạng hình ảnh với mạng nơ-ron phức hợp sâu. Commun ACM. 60 (6), 84-90 (2017)
  25. LeCun, Y., Boser, B., Denker, JS, Henderson, D., Howard, RE, Hubbard, W., Jackel, LD: Backpropagation Được áp dụng cho Nhận dạng mã Zip viết tay. Tính toán thần kinh 1 (4), 541-551 (1989)
  26. Lee, Y., Song, J., Won, Y.: Cải thiện khả năng phát hiện thông tin cá nhân bằng cách sử dụng tỷ lệ nhận dạng tính năng OCR. Tạp chí Siêu máy tính 75 (4), 1941-1952 (2019)
  27. Levenshtein, VI: Mã nhị phân có khả năng sửa lỗi. Chèn và Đảo ngược. Vật lý Liên Xô Doklady 10, 707 (1966)
  28. Liao, M., Shi, B., Bai, X.: Textboxes ++: Trình dò tìm văn bản theo cảnh định hướng một lần chụp. Giao dịch IEEE về Xử lý hình ảnh 27 (8), 3676-3690 (2018)
  29. Liem, HD, Minh, ND, Trung, NB, Đức, HT, Hiệp, PH, Dũng, DV, Vũ, DH: FVI: An End-to-end Nhận dạng và Nhận dạng Thê Căn cứ Việt Nam trong Hình ảnh. Trong: Hội nghị NAFOS TED lần thứ 5 năm 2018 về Khoa học Máy tính và Thông tin (NICS), trang 338-340. IEEE, Thành phố Hồ Chí Minh (2018)
  30. Lin, H., Yang, P., Zhang, F.: Đánh giá về Nhận dạng và Phát hiện Văn bản Cảnh. Lưu trữ các phương pháp tính toán trong kỹ thuật 27 (2), 433-454 (2020)
  31. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, CY, Berg, AC: Ssd: Máy dò multibox bắn một lần. Trong: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision - ECCV 2016, pp. 21-37. Springer International Publishing, Cham (2016)
  32. Liu, X., Liang, D., Yan, S., Chen, D., Qiao, Y., Yan, J.: Fots: Phát hiện văn bản có định hướng nhanh với một mạng thống nhất. Hội nghị IEEE / CVF 2018 về Thị giác Máy tính và Nhận dạng Mẫu trang 5676- 5685 (2018)
  33. Mei, J., Islam, A., Moh'd, A., Wu, Y., Milios, E.: Học thống kê để sửa lỗi OCR. Xử lý & Quản lý Thông tin 54 (6), 874-887 (2018)
  34. Oni, OJ, Asahiah, FO: Mô hình tính toán của hệ thống nhận dạng ký tự quang học cho hình ảnh văn bản in Yorubá. Khoa học viễn tư ở Châu Phi 9, e00415 (2020)
  35. Piroonsup, N., Sinthupinyo, S.: Nhóm và nhân bản giám sát với tính năng tái phân nhóm dựa trên tính năng để giảm nhiễu trong ảnh tài liệu tiếng Thái. Hệ thống dựa trên tri thức 90, 58-69 (2015)
  36. Pramanik, R., Bag, S.: Nhận dạng ký tự ghép viết tay dựa trên sự phân hủy hình dạng cho Bangla OCR. Tạp chí Truyền thông Hình ảnh và Biểu diễn Hình ảnh 50, 123-134 (2018)
  37. Pratama, MO, Satyawan, W., Fajar, B., Fikri, R., Hamzah, H.: Nhận dạng thể căn cứ Indonesia sử dụng Mạng Neural Hợp lệ. Tại: Hội nghị Quốc tế về Kỹ thuật Điện lần thứ 5 năm 2018. Khoa học Máy tính và Tin học (EECSI), trang 178-181. IEEE, Malang, Indonesia (2018)
  38. Puigcerver, J.: Các lớp lặp lại đa chiều có thực sự cần thiết cho việc nhận dạng văn bản viết tay không? Trong: Hội nghị Quốc tế IAPR lần thứ 14 về Phân tích và Ghi nhận Tài liệu (ICDAR), vol. 01, trang 67-72 (2017)
  39. Ronneberger, O., Fischer, P., Brox, T.: U-net: Mạng phù hợp hoạt động để phân đoạn hình ảnh y sinh. Trong: Navab, N., Hornegger, J., Wells, WM, Frangi, AF (eds.) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015, pp. 234-241. Springer International Publishing, Cham (2015)
  40. Shi, B., Bai, X., Yao, C.: Một mạng nơ-ron có thể huấn luyện đầu cuối để nhận dạng trình tự dựa trên hình ảnh và ứng dụng của nó cho văn bản cảnh



sự công nhận. Giao dịch IEEE trên Phân tích mẫu và Trí tuệ máy 39 (11), 2298-2304 (2017)

41. Simonyan, K., Zisserman, A.: Mạng tích tụ rất sâu để nhận dạng hình ảnh quy mô lớn. Trong: Hội nghị quốc tế về đại diện học tập (2015)

42. Tian, Z., Huang, W., He, T., He, P., Qiao, Y.: Phát hiện văn bản trong hình ảnh tự nhiên với mạng đề xuất văn bản liên kết. Trong: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision - ECCV 2016, trang 56-72. Springer International Publishing, Cham (2016)

43. Yin, Y., Zhang, W., Hong, S., Yang, J., Xiong, J., Gui, G.: Kỹ thuật OCR có hỗ trợ học tập sâu cho các bộ phân tích ký tự chữ hoa Trung Quốc trong ứng dụng Internet of Things . IEEE Access 7, 47043-47049 (2019)

44. Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., Liang, J.: East: Trình phát hiện văn bản cảnh hiệu quả và chính xác. Hội nghị IEEE 2017 về Thị giác Máy tính và Nhận dạng Mẫu (CVPR) trang 2642-2651 (2017)

Ghi chú của nhà xuất bản Springer Nature vẫn giữ thái độ trung lập đối với các tuyên bố theo hướng pháp lý trong các bản đồ đã xuất bản và các tổ chức liên kết.