

Nhận diện biển số trong hệ thống giám sát ô tô từ các camera giao thông

Khanh Nguyen Quoc, Dan Pham Van

Faculty Of Information Technology, Le Quy Don Technical University
khanh29bk@mta.edu.vn, phamvandan.cntt2@gmail.com

Tóm tắt nội dung—Nhận diện biển số xe ô tô tự động là bài toán có nhiều ứng dụng trong thực tế và rất quan trọng trong hệ thống giao thông thông minh như trạm thu phí, theo dõi xe, quản lý bãi đỗ xe. Báo cáo này hướng tới nghiên cứu các kỹ thuật học sâu áp dụng cho nhận diện biển số xe ô tô Việt Nam ứng dụng trong hệ thống nhận diện biển số xe tự động cho các camera giao thông. Hệ thống chúng tôi đề xuất có ba phần chính: phát hiện ô tô, phát hiện biển số và nhận diện biển số. Phần phát hiện ô tô chúng tôi tận dụng kết quả của các nghiên cứu trước. Nhằm giải quyết vấn đề non-rectangular object detection, phần phát hiện biển số được lấy ý tưởng từ bài toán keypoints detection, mỗi góc của biển số là một keypoint. Chúng tôi thực nghiệm trên tập 15571 ảnh. Kết quả tốt nhất trên tập kiểm thử 1558 ảnh với $mIOU = 95\%$ và $Precision_{75} = 99,2\%$ dựa trên thay đổi từ kiến trúc segmantic segmentation DDRNet [ddrnet]. Phần nhận diện biển số với cách tiếp cận segmentation free dựa trên kiến trúc mạng encoder decoder. Thuận lợi chính của cách tiếp cận này là không cần character segmentation. Chúng tôi đã thực nghiệm phần nhận diện với hơn 16012 ảnh biển số xe Vietnam. Kết quả tốt nhất trên tập test 2472 ảnh với độ chính xác Full sequence là $Acc_{seq} = 99.4\%$, Mean char là $Acc_{char} = 99.8\%$ dựa trên kiến trúc [1] áp dụng trong bài toán dịch máy. Ngoài ra, chúng tôi đề xuất kiến trúc hệ thống nhận diện biển số xe ô tô tự động từ camera giao thông tích hợp thêm các thuật toán motion detection, object tracking, kiểm tra kết quả ocr tăng độ tin cậy và tính ứng dụng của hệ thống.

Key words: Car plate detection and recognition, tracking car, segmentation free, sequence modeling.

Từ khóa—Car plate detection and recognition, tracking car, segmentation free, sequence modeling.

I. GIỚI THIỆU

Ngày nay, trước sự gia tăng nhanh chóng của các phương tiện giao thông thì hệ thống nhận diện biển số xe tự động (ALPR) thông minh ngày càng thu hút được sự quan tâm. Hệ thống nhận diện biển số xe tự động đã được cài đặt ở nhiều quốc gia để phục vụ cho các nhiệm vụ giám sát giao thông. Bên cạnh đó, hệ thống ALPR cũng được sử dụng để kiểm soát ra vào các bãi đỗ xe, trạm thu phí tự động và kiểm soát an ninh tại các khu vực hạn chế ra vào. Hệ thống ALPR hỗ trợ tìm kiếm những phương tiện mất tích hoặc liên quan đến trộm cắp, mà nếu không sử dụng ALPR thì đây quả là một công việc rất tốn nhân lực và công sức, hơn nữa dễ dẫn đến sai lầm. Hơn nữa, với một chiếc xe đang di chuyển trên đường, con người rất khó đọc và nhớ được biển số của nó.

Hệ thống ALPR là một hệ thống lấy hình ảnh hoặc video làm đầu vào. Nếu đầu vào có chứa phương tiện thì hệ thống sẽ xuất ra được nội dung của biển số dưới dạng văn bản. Các hệ thống này bao gồm máy ảnh để chụp hình ảnh các phương

tiện. Các kỹ thuật để giải quyết bài toán được đưa ra như object detection [2], image processing [3], pattern recognition [4]. Hầu hết hệ thống ALPR được thiết kế để triển khai ngoài trời. Tuy nhiên lại gặp khó khăn trước tác động của môi trường và thời tiết, như trong điều kiện mưa hay sương mù. Thêm cả việc thay đổi ánh sáng giữa ngày và đêm. Thực tế cũng rất khó khi xử lý khi hình ảnh đầu vào có nhiều biển số xe, và chất lượng ảnh kém.

Hầu hết hệ thống ALPR được thiết kế để triển khai ngoài trời. Tuy nhiên lại gặp khó khăn trước tác động của môi trường và thời tiết, như trong điều kiện mưa hay sương mù. Thêm cả việc thay đổi ánh sáng giữa ngày và đêm. Thực tế cũng rất khó khi xử lý khi hình ảnh đầu vào có nhiều biển số xe, và chất lượng ảnh kém. Do góc nhìn của máy ảnh đối với phương tiện khác nhau, điều kiện ánh sáng và màu sắc bản chất của biển số nên biển số xe sẽ có kích thước, màu sắc, phong chữ và độ lệch khác nhau gây khó khăn cho việc giải quyết bài toán nhận diện biển số xe cho tất cả các kịch bản. Hơn nữa, giữa các quốc gia sẽ có chuẩn riêng cho biển số xe của quốc gia đó, nên việc áp dụng một hệ thống ở bất kỳ nơi nào trên toàn thế giới thực sự khó khăn. Chính vì thế, các kỹ thuật để giải quyết bài toán này rất phức tạp [5], [6].

Theo khảo sát [7], [8] có rất nhiều nghiên cứu đặt ra kịch bản hạn chế về điều kiện phong nền và giới hạn khoảng cách giữa máy ảnh và phương tiện. Tuy nhiên điều này khó để đọc được biển số của các phương tiện đang di chuyển. Để giải quyết được sự thay đổi ánh sáng thì có nhiều nghiên cứu đã tập trung vào việc phát triển các hệ thống có thể hoạt động cả ngày và đêm, và có đánh giá chung là hiệu suất hoạt động vào đêm thường kém hơn ban ngày [7]. Ngoài ra hệ thống ALPR còn phải đáp ứng những yêu cầu về khả năng vận hành và kết nối. Trong khi các phương pháp học sâu có thể mang lại hiệu suất cao về độ chính xác nhưng lại rất tốn kém về tính toán và tài nguyên để triển khai trên các thiết bị thực tế. Trong khi các phương pháp tiếp cận thị giác máy tính cổ điển có độ chính xác hợp lý trong các kịch bản cụ thể và tốn ít chi phí tính toán. Việc triển khai một hệ thống ALPR cần có sự cân đối toàn diện giữa các yếu tố lợi ích của các phương pháp.

Các công trình trước tập trung vào độ chính xác trên tập dữ liệu nghiên cứu còn ít tập trung vào vấn đề triển khai trong thực tế. Trong công trình này, chúng tôi tập trung xây dựng xây dựng một hệ thống ALPR áp dụng thực tế cho biển số xe ô tô Việt Nam, ứng dụng cho các bãi đỗ không dừng và các trạm theo dõi của đường cao tốc. Đóng góp của chúng tôi gồm có:

- Mô hình hệ thống hoàn chỉnh cho triển khai thực tế. Hệ thống được tích hợp thêm thuật toán tracking và thuật toán kiểm tra kết quả nhận dạng 2 có đúng theo quy định của biển số Việt Nam mà chính phủ ban hành. Nhờ đó nâng tính áp dụng thực tế của hệ thống tại các camera giao thông ở Việt Nam.
- Đối với nhận diện biển số theo cách tiếp cận segmentation-free, các nghiên cứu trước chủ yếu sử dụng các kiến trúc RNN kết hợp với CTC loss [9]. Tuy nhiên, cách tiếp cận này sẽ gặp phải một số nhược điểm nhất định (xem mục II-A2). Khắc phục những hạn chế đó, chúng tôi kiến trúc encoder-decoder sử dụng GRU [10], tích hợp thêm cơ chế attention lấy ý tưởng từ [1]. Phần trích xuất đặc trưng được thiết kế dựa trên VGG-19 [11] với một số tinh chỉnh.
- Chúng tôi đề xuất bộ dữ liệu VLP hơn 18803 ảnh biển số xe ô tô Vietnam ở các điều kiện khác nhau và các kiểu biển số khác nhau. Kết quả thực nghiệm trên bộ dữ liệu VLP đạt độ chính xác Full sequence là 98,36%, Mean char là 99.65% cho thấy tiềm năng của cách tiếp cận này.

II. CÁC CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN

ALPR thường có 2 cách tiếp cận chính: multi-stage và single stage license plate recognition

A. Multi-stage license plate recognition systems

Các hệ thống multi-stages bao gồm : giai đoạn đầu tiên là trích xuất biển số xe, ở đây có thể sử dụng kỹ thuật xử lý ảnh truyền thống hoặc sử dụng các phương pháp học sâu với tính tăng phát hiện đối tượng để xác định vị trí biển số xe trong một hình ảnh. Các kỹ thuật xử lý ảnh truyền thống dựa trên các tính năng của biển số xe như hình dạng [12], màu sắc [13], đối xứng [14], kết cấu [15]. Với sự phát triển của deep learning, kỹ thuật bắt đầu được chuyển sang hướng khác, ngày các nhiều nghiên cứu áp dụng CNN để phát hiện đối tượng biển số xe do có độ chính xác cao và đã được công nhận [16], [13].

1) *License plate detection*: Sự thành công của YOLO đã truyền cảm hứng cho nhiều nghiên cứu gần đây [17], các tác giả đã mở rộng độ rộng mức độ chi tiết của output để tăng số lượng detection của mạng và đạt xác suất cho hai lớp là biển số xe và nền. Nhưng những nghiên cứu thiếu đánh giá chi tiết về bounding box được trích xuất. Trong nghiên cứu [18], tác giả đã thiết lập hai mạng dựa trên YOLO với mục tiêu phát hiện biển số xe với các góc quay khác nhau. Mạng đầu tiên được sử dụng để tìm vùng chứa biển số xe và mạng thứ hai với đầu ra là bounding box chứa biển số xe. Tuy nhiên ở đây nghiên cứu chỉ xem xét biển số xe dưới góc độ các góc quay, chứ không phải các biến dạng phức tạp, ví dụ như góc nghiêng lớn. Việc sử dụng sliding window hoặc lọc candidates cùng với việc sử dụng CNN cũng có thể được tìm thấy trong nhiều tài liệu [19], tuy nhiên kém hiệu quả về mặt tốc độ tính toán.

Các phương pháp Scene Text Spotting (STS) chủ yếu tập trung vào các thông tin về phong chữ, từ vựng và ngữ nghĩa,

có thể nói đến một số phương pháp xử lý văn bản có thể được áp dụng để phát hiện biển số xe với các góc lệch khác nhau [20],[21]. Gần đây hơn nghiên cứu [22] đã trình bày cách tiếp cận để phát hiện văn bản gọi là Instance Transformation Network (ITN), cách tiếp cận này về lý thuyết có thể xử lý các phép quay của biển số xe, nhưng chỉ đưa ra được bounding box bao quanh văn bản chứ không phải là các góc giới hạn của văn bản.

2) *License plate recognition*: Giai đoạn thứ hai là sử dụng các kỹ thuật nhận diện ký tự quang học. Bài toán này có hai hướng tiếp cận chính là *segmentation* và *segmentation-free*.

Với segmentation approaches, các ký tự được trải qua bước segmentation rồi sau đó tiến hành phân lớp. Các phương pháp để thực hiện phân đoạn ký tự như projection profiles methods, hay sử dụng pixel connectivity, prior knowledge đều dựa trên các kịch bản nhất định và phụ thuộc rất lớn vào đặc điểm đầu vào, ngưỡng được áp dụng [23], [24], [25]. Ngày nay các tiếp cận sử dụng CNN để phân đoạn ký tự trở nên phổ biến [26], tuy nhiên tùy trên tập dữ liệu thì việc thực thi CNN tiêu tốn nhiều thời gian và tài nguyên hơn so với kỹ thuật xử lý ảnh truyền thống.

Các kỹ thuật phân lớp ký tự có thể được giới thiệu là: template and pattern matching so sánh trực tiếp tất cả các giá trị pixel của dữ liệu ảnh thô với các mẫu được xác định từ trước [2], character recognition using feature extractors sử dụng kỹ thuật xử lý ảnh và machine learning để trích xuất đặc trưng trước khi phân loại [4] hay sử dụng kỹ thuật học sâu để phân loại [23].

Với segmentation-free, bước phân đoạn ký tự bị bỏ qua để giảm chi phí tính toán, gán nhãn dữ liệu, lấy toàn bộ hình ảnh biển số xe làm đầu vào và đầu ra là văn bản. Để giải quyết vấn đề này, nhiều nghiên cứu đã đề xuất các giải pháp dựa trên CNN và RNN. Các mô hình trong nghiên cứu [27], [20] dựa trên variable length sequence decoding được điều khiển bằng connectionist temporal classification (CTC) loss [9]. Mô hình [28] tuân theo cách tiếp cận được mô tả trong [29] nhưng phương pháp sử dụng sliding window đã được thay thế bằng cách CNN output spatial splitting to the RNN input sequence.

Các nghiên cứu với cách tiếp cận segmentation free chủ yếu là sự kết hợp giữa CNN và RNN, tối ưu thông qua CTC loss. Tuy nhiên, CTC có xu hướng sắp xếp từng dự đoán nhãn với phần tương ứng của chuỗi đầu vào. Đây cũng là nhược điểm chính của nó trong việc chọn đặc trưng cố định của phần nào trong bức ảnh để làm đầu ra ở mỗi bước của CTC. Ngoài ra, khi sử dụng CTC số lượng đầu ra luôn nhỏ hơn hoặc bằng chuỗi đầu vào. Dẫn tới, nếu mô hình được huấn luyện cho các loại biển số mà kích cỡ chữ lớn có tối đa m chữ cái sẽ không áp dụng được cho các biển khác có cùng kích thước nhưng cỡ chữ nhỏ hơn do có tối đa n chữ cái, với $n > m$. Một đặc điểm khác biệt nữa của CTC là nó không mô hình hóa rõ ràng các phụ thuộc liên nhãn, nó giải quyết vấn đề này thông qua việc thêm ký tự blank để giải quyết vấn đề trùng lặp.

Trong công trình này, chúng tôi tiếp cận theo hướng segmentation-free với kiến trúc encoder-decoder sử dụng GRU [10] nhằm giải quyết vấn đề giới hạn số lượng đầu ra và phụ thuộc liên nhãn. Chúng tôi tích hợp thêm cơ chế attention lấy

ý tưởng từ [1] giúp mô hình trích chọn được những đặc trưng quan trọng trong từng phần của bức ảnh, cái ảnh hưởng tới đầu ra mong muốn.

B. Single-stage license plate recognition systems

Khi hầu hết các nghiên cứu đều tập trung vào quy trình multi-stage trong nhận dạng biển số xe thì gần đây đã có một số nghiên cứu thành công đối với các quy trình single-stage. Có thể được hiểu là sử dụng một mạng deep neural network duy nhất, được đào tạo cho end-to-end detection, localization and recognition of the license plate in a single forward pass. Điều này cho phép các mô hình chia sẻ các thông số và có ít tham số hơn so với mô hình multi-stages, do đó nhanh hơn và có hiệu quả cao hơn [28], [30]. Nghiên cứu [30] tiếp cận sử dụng VGG16 như một bộ feature extractor, và đã sửa đổi VGG16 và bỏ đi ba lớp và đầu ra được đưa vào một Region Proposal Network (RPN) [28]. Họ sử dụng hai separate sub-networks để phát hiện biển số xe và nhận dạng biển số xe. Mạng con phát hiện biển số xe đưa ra xác suất và tọa độ bounding box, để tránh phân đoạn ký tự họ đã mô hình hóa như một sequence labelling problem sử dụng Bidirectional RNNs.

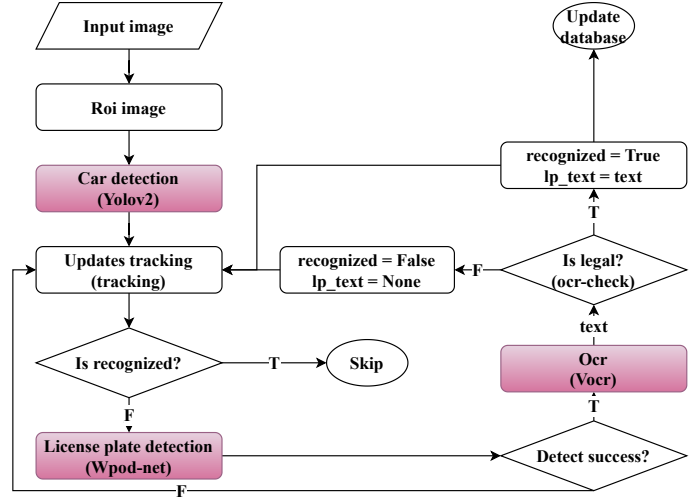
Một cách tiếp cận tương tự đã được đề xuất ở nghiên cứu [28] thay thế VGG16, họ sử dụng mạng CNN với 10 lớp, sử dụng các bộ phân loại riêng lẻ và trích xuất tính năng đơn giản hơn cho phép họ xử lý nhanh hơn các phương pháp tiếp cận trước đó và được đánh giá với độ chính xác rất cao

III. PHƯƠNG PHÁP ĐỀ XUẤT

Nhằm đáp ứng được nhu cầu triển khai thực tế hệ thống nhận diện biển số xe tự động ở các highway surveillance camera, khi mà các phương tiện không ngừng chuyển động. Hệ thống chúng tôi đề xuất gồm có ba phần chính: phát hiện ô tô, phát hiện biển số và nhận diện biển số, như minh họa trong hình 1. Đầu vào là một khung hình thu được từ camera giám sát trên đường quốc lộ, module car detection (Yolov2) phát hiện và theo dõi các ô tô có trong vùng quan tâm (ROI). Mỗi vùng phát hiện được sẽ được cập nhật thuật toán tracking. Thuật toán tracking cho biết đối tượng đã được xử lý ở khung hình trước hay chưa, nếu chưa, module License plate detection (Wpod-net) tiến hành phát hiện biển số. Module ocr (Vocr) tiến hành nhận diện ký tự quang học và đi qua bước kiểm tra (ocr-check) về tính hợp lệ của kết quả thu được để xác nhận kết quả dự đoán hợp lệ hay không. Nếu hợp lệ, đối tượng được xác định hoàn tất quá trình nhận dạng và kết quả được lưu trữ vào cơ sở dữ liệu. Ngược lại, đối tượng chưa nhận dạng xong, và tiếp tục được xử lý ở các khung hình sau.

A. Car detection

Ô tô là một trong những đối tượng cơ bản có mặt trong nhiều bộ dữ liệu lớn, chẳng hạn như PASCAL-VOC [31], ImageNet [32] và COCO [33], do đó, chúng tôi quyết định không huấn luyện bộ phát hiện từ đầu, thay vào đó chọn một mô hình đã biết để thực hiện phát hiện ô tô dựa trên các tiêu chí tỉ lệ recall cao, vì việc từ chối nhầm ô tô có thể dẫn tới lỗi sai tổng thể, đồng thời, tỉ lệ precision cũng cần được quan tâm vì việc phát hiện nhầm các vật thể không phải là ô tô làm tốn thời gian



Hình 1: Sự minh họa hệ thống được đề xuất.

xử lý vì phải đi qua các bước sau. Dựa trên những tiêu chí này, chúng tôi đã quyết định sử dụng mạng YOLOv2 do tốc độ thực thi nhanh khoảng 70 FPS và độ chính xác đạt 76,8% mAP so với bộ dữ liệu PASCAL-VOC. Chúng tôi đã không thực hiện bất kỳ thay đổi hoặc cải tiến nào đối với YOLOv2.

Đối với mỗi vùng phát hiện được, các đối tượng sẽ cập nhật thuật toán tracking mà chúng tôi đề xuất. Thuật toán tracking này được gọi là centroid vì nó dựa trên khoảng cách Euclid giữa tâm các đối tượng được phát hiện (tức là giữa vùng mới được phát hiện và các vùng đã được phát hiện trước đó), và thực hiện loại bỏ lưu trữ của đối tượng sau một thời gian không xuất hiện trong khung hình. Những đối tượng chưa nhận dạng thành công sẽ tiếp tục được đưa qua bước phát hiện biển số, số còn lại sẽ bị bỏ qua. Những đối tượng chưa nhận dạng thành công gồm có: đối tượng mới xuất hiện; đối tượng bất biến số chưa thành công; kết quả nhận dạng OCR sai so với quy định của biển số Vietnam mà chính phủ ban hành, cái được cụ thể hóa ở thuật toán 2.

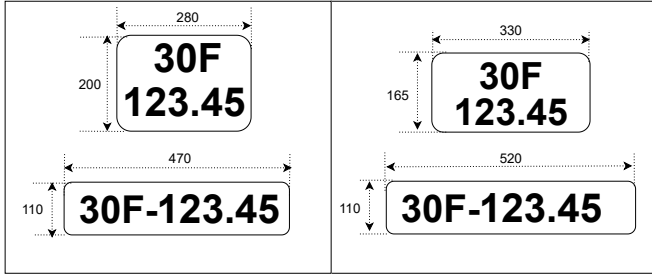
Thuật toán tracking chúng tôi đề xuất có ba ý nghĩa. Thứ nhất, các đối tượng ô tô chạy khi thu từ camera giao thông ở Vietnam chỉ di chuyển một chiều. Vì thế, chúng tôi xây dựng thuật toán tracking đơn giản nhưng đủ để giải quyết vấn đề, đảm bảo tài nguyên sử dụng. Thứ hai, mỗi đối tượng thực hiện nhận dạng thành công đúng một lần, làm tăng hiệu năng chung. Thứ ba, các đối tượng chưa nhận dạng thành công sẽ có cơ hội nhận dạng lại vào các khung hình sau đó. Điều này sẽ làm tăng độ chính xác và sự mạnh mẽ của hệ thống.

B. License plate detection

Trong bài báo này, chúng tôi tập trung vào phần license plate OCR III-C, và các cải tiến về mặt hệ thống với thuật toán tracking và thuật toán ocr-check 2. Vì vậy, đối với phần phát hiện biển số, chúng tôi tận dụng kết quả của nghiên cứu [34] với mô hình phát hiện biển số WPOD-NET. Lý do chúng tôi sử dụng mô hình này là nó có thể phát hiện và căn chỉnh được các biển số khi chúng bị nghiêng ở các góc khác nhau.

C. License plate OCR

1) **Đặc điểm biển số xe Vietnam:** Biển số xe ô tô Việt Nam được thiết kế theo quy định của chính phủ. Về kích thước kiểu cũ và kiểu mới (xem hình 2). Về màu sắc, biển số Việt Nam gồm biển nền xanh hoặc đỏ với chữ trắng, biển nền vàng hoặc trắng với chữ đen. Về quy định các ký tự trong biển số, các ký



Hình 2: Quy định kích thước biển số cũ (left) và mới (right).

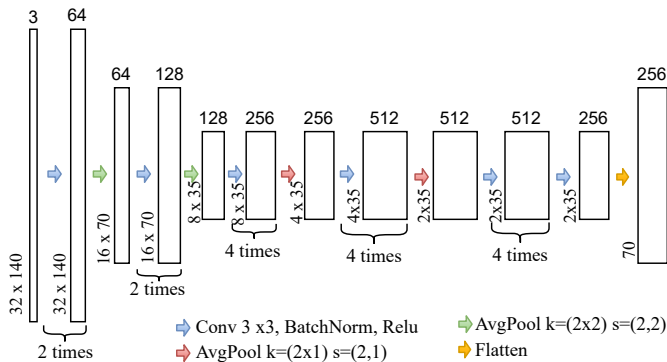
tự biển số xe Việt Nam đều nằm trong tập hợp các ký tự gồm chữ số từ 0 đến 9 và chữ cái latin từ A đến Z cùng với dấu gạch ngang và dấu chấm. Các ký tự này không phải được sắp xếp ngẫu nhiên mà đều thuộc một trong năm dạng template (xem hình 3), đối với biển số được lưu dưới dạng một dòng thì được chèn thêm một dấu gạch ngang giữa dòng trên và dòng dưới như đối với biển hai dòng.



Hình 3: Các định dạng văn bản của biển số.

2) **Tiền xử lý:** Đối với các biển số hai dòng, chúng tôi xây dựng thuật toán 1 dựa trên phép chiếu để có thể tìm ra vị trí chính xác của vị trí phân tách.

3) **Vocr network:** Đầu vào của Vocr là ảnh biển số có 1 dòng có kích thước 32x140 (trường hợp biển số hai dòng đã được xử lý bởi 1). Phần thứ nhất của Vocr network (gọi là Vocr-cnn) trích xuất đặc trưng local của ảnh đầu vào dựa trên VGG16 [11] (xem hình 4). Do biển số thường có kích thước



Hình 4: Kiến trúc của Vocr-cnn

khá nhỏ. Vì thế, để tránh mất thông tin, chúng tôi lựa chọn

Algorithm 1 Two-row license splitting algorithm

Input: LPR image

Output: position to split license plate horizontally

$img \leftarrow \text{convert } img \text{ to gray}$

$temp \leftarrow 255 - img$

for $i = 1$ to 2 **do**

$img \leftarrow \text{remove border}$

$img \leftarrow \text{equalize}$

$thresh \leftarrow \text{turn } img \text{ into binary image}$

$indexes \leftarrow \text{horizontally distributed count of thresh}$

for $i = 0$ to $len(indexes)$ **do**

if $indexes[i] = 0$ **then**

$indexes \leftarrow \text{remove indexes}[i]$

else

break

end if

end for

for $i = len(indexes)$ downto 0 **do**

if $indexes[i] = 0$ **then**

$indexes \leftarrow \text{remove indexes}[i]$

else

break

end if

end for

for $i = 0$ to $len(indexes)$ **do**

if $indexes[i] = 0$ **then**

return i

end if

end for

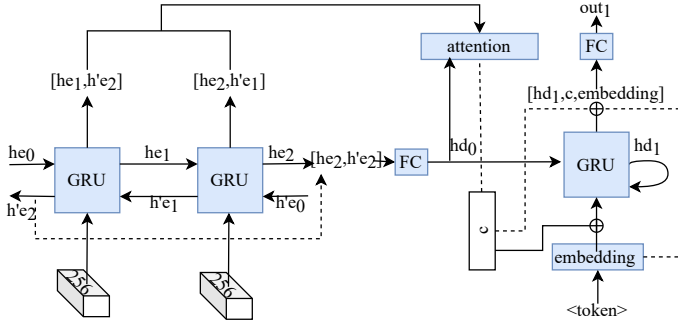
$img \leftarrow temp$

end for

return none

Average pooling thay cho max pooling như trong kiến trúc gốc. Đồng thời, chiều dài thường lớn hơn nhiều so với chiều rộng, vì thế, 2 lớp Average pooling cuối chúng tôi thay đổi kích thước kernel $k=(2,1)$, stride $s=(2,1)$. Kỹ thuật BatchNorm [35] được sử dụng nhằm tăng tốc độ huấn luyện cho mô hình. Đầu ra của Vocr-net có 256 feature maps, mỗi feature map có kích thước 1×70 . Chúng tôi xem xét số chiều của mỗi feature map chính là số lượng time steps, như vậy time steps $T=70$. Tại mỗi time step t , vector đặc trưng có kích thước 1×256 chứa đặc trưng của tất cả các feature maps tại time step đó. Phần thứ hai của Vocr network (gọi là Vocr-rnn) có cấu trúc encoder-decoder sử dụng GRU kết hợp cơ chế attention (xem hình 5). Trong đó, $he, h'e, hd$ là các hidden output feature tại mỗi time step của encoder và decoder, có số chiều là 256.

Phần encoder, chúng tôi lựa chọn GRU-bidirectional nhằm trích xuất được quan các quan hệ phụ thuộc lẫn nhau trong chuỗi đầu vào. Đầu ra của encoder tại mỗi time step là việc concatenate $[he, h'e]$ có số chiều 512. Đầu ra của encoder thông qua lớp Fully connected nhằm giảm số chiều bằng với số chiều của hd . Cơ chế attention được lấy ý tưởng từ [1] xem mỗi hd tại mỗi time step là query và tất cả $[he, h'e]$ đóng



Hình 5: Kiến trúc của Voc-rnn

vai trò là keys và values. Vì số chiều của keys và query là không giống nhau, do đó chúng tôi sử dụng attention scoring function là Additive attention (xem công thức 1) [36].

$$a(Q, K) = w_v^T \tanh(W_q Q + W_k K) \in R^{1 \times T} \quad (1)$$

Trong đó, $Q \in R^{q \times T}$, $W_q \in R^{h \times q}$, $K \in R^{k \times T}$, $W_k \in R^{h \times k}$, $w_v \in R^h$. Trong bài toán này, $Q = hd$, $K = [he, h'e]$, $h = 256$, $q = 256$, $k = 512$, $T = 70$. Khi đó context vector c tổng hợp như sau (xem công thức 4).

$$w_i = \frac{e^{a_i}}{\sum_{i=1}^T e^{a_i}} \in R \quad (2)$$

$$w = [w_1, w_2, \dots, w_T] \in R^{1 \times T} \quad (3)$$

$$C = V \times w \in R^{v \times 1} \quad (4)$$

Trong đó, w là attention weights của các time steps, $V = K$, $v = k$. Context vector c thu được chứa thông tin đặc trưng quan trọng từ chuỗi đầu vào mà bước decode hiện tại cần quan tâm. Điều này rất có ý nghĩa, chẳng hạn, dự đoán nhãn là chữ cái đầu tiên của biển số, attention weights w sẽ có trọng số cao hơn đối với những time steps chứa thông tin của chữ cái đó.

Chúng tôi huấn luyện decoder sử dụng chiến lược huấn luyện teacher forcing [37]. Decoder network nhận đầu vào là hd ở time step trước và concatenate [context vector c , embedding] của nhãn ở time step hiện tại. Đầu ra của decoder là concatenate của $[hd, c, embedding]$ đi qua lớp FC có số chiều đầu ra $N = 40$ (tương ứng với số class của kí tự xuất hiện trong biển số Vietnam cộng thêm 3 kí tự $< sos >$, $< eos >$, $< pad >$).

Hàm mất mát chúng tôi sử dụng là Cross-entropy loss kết hợp với kỹ thuật Label smoothing [38] nhằm tăng tính tổng quát cho mô hình.

$$L = - \sum_{i=1}^N (l_i \times y_i \times (\log(\hat{y}_i)))$$

Trong đó, y_i là groundtruth, \hat{y}_i là dự đoán của mô hình, $l_i = |y_i - \delta|$ là trọng số của label smoothing tại lớp thứ i . Chúng tôi lựa chọn $\delta = 0.1$ dựa trên kết quả thực nghiệm ở các ngưỡng δ khác nhau đã được thực hiện trong [38].

Để tối ưu hàm loss L , chúng tôi sử dụng phương pháp cập nhật tham số dựa trên gradient descent với thuật toán Adam

[39] tự động điều chỉnh learning rate và tránh rơi vào local minimum.

4) *Kiểm tra kết quả Ocr*: Trong kiến trúc hệ thống 1, kết quả của Voc-rnn sau khi dự đoán được kiểm tra lại thông qua thuật toán Ocr-check dưới đây. Thuật toán Ocr-check 2 dựa trên những quy định về biển số Vietnam mà chính phủ ban hành (xem mục III-C1).

Algorithm 2 Ocr-check

Input: Text of LP

Output: legal or illegal

$line1 \leftarrow$ regular expression of LPR in first line

$line2 \leftarrow$ regular expression of LPR in second line

one line exists

$l1 \leftarrow$ text of first line

$l2 \leftarrow$ text of second line

for $i = 0$ to 5 **do**

if $line1[i].match(l1)$ AND $line2[i].match(l2)$ **then**
 return true

end if

end for

return false

two lines exists

$l1 \leftarrow$ text of two line

for $i = 0$ to 5 **do**

if $(line1[i] + " - " + line2[i]).match(l1)$ **then**
 return true

end if

end for

return false

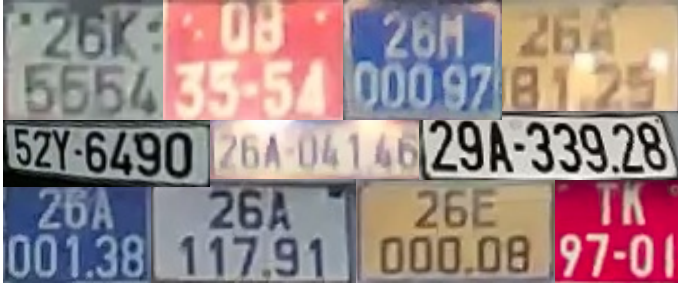
IV. THỰC NGHIỆM

A. Tập dữ liệu

Tại Vietnam các nghiên cứu trước chỉ tự xây dựng bộ dữ liệu cá nhân nhỏ tính đến thời điểm hiện tại. Vinh và cộng sự [40] chỉ kiểm tra trên 700 ảnh, Duan và cộng sự [41] chỉ kiểm tra trên 805 ảnh. Xác định vấn đề này, chúng tôi xây dựng một tập dữ liệu lớn có tên là VLP dataset. Chúng tôi lắp đặt 2 camera giao thông ở hai cung đường khác nhau. Chúng tôi cài đặt hệ thống 1, ngoại trừ bước license plate OCR, lên một máy chủ. Quá trình thu thập diễn ra trong vòng 3 tuần, từ 6h sáng tới 6h chiều hằng ngày, trong các điều kiện thời tiết, ánh sáng và phương tiện di chuyển bất kỳ dẫn tới loại biển cũ và mới, kích thước, kiểu chữ, độ mờ, nhiễu rất đa dạng (xem hình 6). Bộ dataset gồm 11305 ảnh cars cropped bởi Yolov2 và 11198 ảnh plate thu được bởi Wpod-net. Sau khi kiểm tra lại, chúng tôi có 11073 ảnh biển số. Phân bố của các loại biển theo màu sắc của chúng được tổng hợp ở bảng I.

B. Kết quả thực nghiệm

Trong phần kết quả thực nghiệm, chúng tôi tập trung đánh giá kết quả phần licenseplate OCR sử dụng kiến trúc Voc-rnn.



Hình 6: Một số ảnh biển số trong tập dữ liệu VLP.

Bảng I: Số lượng biển theo màu sắc

Số lượng biển	Trắng	Vàng	Xanh	Đỏ	Tổng
Biển 1 dòng	3323	0	13	7	3343
Biển 2 dòng	6620	952	142	16	7730
Tổng	9943	952	155	23	11073
Tổng Splitted	18803				

1) *Cài đặt thực nghiệm*: Với dữ liệu như trên, chúng tôi chia tập 18803 biển số phân bố đều ở 3 tập: train, validate, test. Với tập train là 14103 ảnh chiếm tỉ lệ 75%, tập validate là 1880 tương đương 10% và tập test là 2828 với 15%. Để so sánh, chúng tôi thử nghiệm Vocr, LPRnet [42] is also a segmentation-free approach, Yolov2 [34] is a segmentation-based object detection approach.

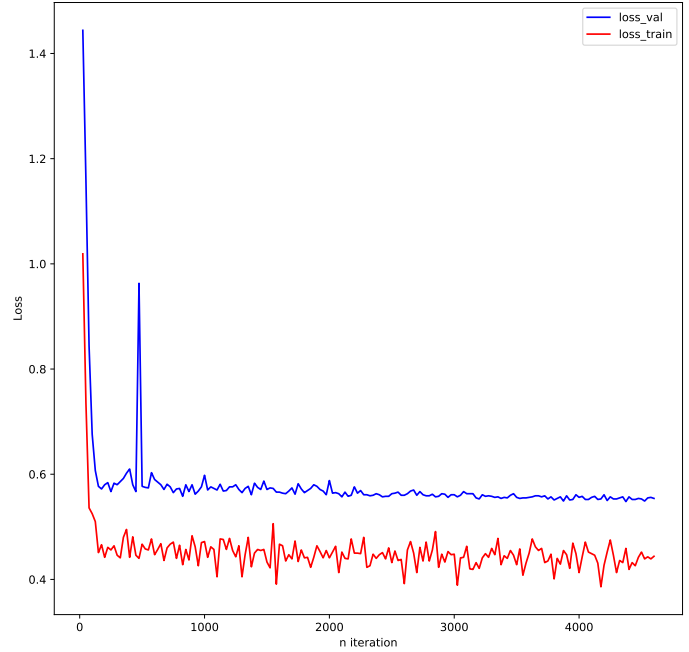
2) *Phương pháp đánh giá*: Chúng tôi đánh giá kết quả nhận dạng theo tỷ lệ phần trăm LP được nhận dạng chính xác (Acc_{seq}), trong đó LP được coi là đúng nếu tất cả các ký tự được nhận dạng chính xác và không phát hiện thấy ký tự bổ sung nào [34]. Tuy nhiên, có thể xảy ra trường hợp nếu kết quả dự đoán chỉ nhầm một ký tự và số còn lại vẫn đúng thì cách đánh giá trên chưa thể hiện được. Do đó, chúng tôi đưa thêm đánh giá độ chính xác trung bình của các ký tự (Acc_{char}). Một ký tự được cho là nhận dạng đúng, nếu vị trí của nó trong nhân và trong kết quả dự đoán là như nhau.

3) *Training*: Nhằm tăng tính tổng quát cho mô hình, chúng tôi thực hiện ngẫu nhiên các phép agumentation trong quá trình huấn luyện, gồm có:

- Làm mờ ảnh: Gaussian Blur ($\sigma \in [0, 1.0]$), Motion Blur ($kernel - size = [3, 3]$).
- Tăng hoặc giảm giá trị kênh Hue và Saturation theo các giá trị ngẫu nhiên.
- Điều chỉnh độ tương phản của hình ảnh.
- Đảo ngược các giá trị của hình ảnh.
- Áp dụng phép biến đổi như (xoay, crop, translation) cho ảnh.
- Giảm chất lượng hình ảnh.
- Lọc bỏ bớt thông tin hình ảnh bằng cách cho các giá trị tại đó bằng 0.
- Nhân tất cả các pixel trong một hình ảnh với giá trị ngẫu nhiên.
- Tăng, giảm độ sáng của hình ảnh.

Quá trình huấn luyện chúng tôi sử dụng batch size 32, max iterations 4600, learning rate 0.001. Việc cập nhật tham số được thực hiện bởi thuật toán Adam [39].

Dựa vào giá trị hàm loss trong quá trình huấn luyện (xem hình 7), chúng ta thấy mô hình học khá ổn định. Tại $Iteration = 4600$, $L_{train} = 0.444$, $L_{valid} = 0.554$. Do chúng tôi sử dụng nhiều phương pháp agumentation khác nhau trong quá trình training nên giá trị hàm loss trên tập train biến động nhiều hơn so với giá trị hàm loss trên tập validate.



Hình 7: Giá trị hàm loss trên tập train và valid của VLP dataset.

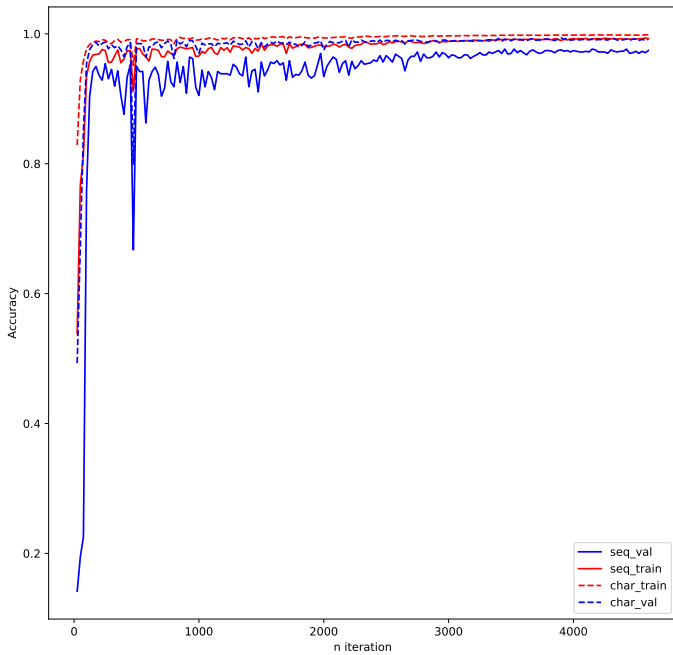
Trong quá trình huấn luyện, giá trị accuracy Full sequence của tập train và accuracy Mean char trên tập train và valid (xem hình 8) cho thấy độ chính xác tốt của mô hình. Tại $Iteration = 4600$, độ chính xác trên tập train là $Acc_{seq-train} = 0.9926$, $Acc_{char-train} = 0.9984$, trên tập valid là $Acc_{seq-valid} = 0.9743$, $Acc_{char-valid} = 0.9913$.

Bảng II: Kết quả kiểm tra trên VLP dataset

Method	Acc_{seq}	Acc_{char}
Vocr	98,36%	99,65%
Yolov2 [34]	83,12%	96,01%
LPRnet [42]	88,61%	97,11%

4) *Testing results and comparable*: Chúng tôi lựa chọn hai kiến trúc đại diện cho hai cách tiếp cận khác nhau để kiểm thử và so sánh. Đối với segmentation free method, chúng tôi lựa chọn LPRnet [42] với độ chính xác $Acc_{seq} = 95\%$ trên tập dữ liệu hơn 1000 ảnh biển số Trung Quốc. Đối với segmentation method, chúng tôi sử dụng Yolov2 [34] với độ chính xác $Acc_{seq} = 89.33\%$ trên tập dữ liệu 1729 ảnh biển số từ các tập dataset khác nhau. Để so sánh công bằng, chúng tôi huấn luyện lại các kiến trúc này trên tập dữ liệu của chúng tôi theo các tham số như trong bài báo gốc đã nêu.

Bảng II cho thấy kết quả của Vocr $Acc_{seq} = 98.36\%$, $Acc_{char} = 99.65\%$ cạnh tranh và có tiềm năng trong ứng



Hình 8: Giá trị accuracy trên tập train và valid của VLP dataset.

dụng thực tế. Phương pháp Yolov2 [34] cho kết quả không tốt trên tập dữ liệu của chúng tôi vì chất lượng biển số trong tập VLP khá đa dạng gồm có cả các chữ nghiêng, khoảng cách các chữ khá gần nhau và còn có thêm các ký tự ".", "-". Do đó, dẫn tới các bounding box phát hiện và dự đoán sai. LPRnet [42] cho kết quả kém trên tập VLP dataset của chúng tôi, đặc biệt là tỉ lệ chính xác Full sequence do một số nhược điểm của hàm CTC loss [9] và cách thiết kế chỉ dựa trên CNN của LPRnet khiến nó khó tổng quát hóa trên VLP dataset. Vocx xử lý tốt hầu hết các trường hợp, trong trường hợp ảnh có độ nhiễu cao, mô hình gặp khó khăn trong việc nhận dạng (xem hình 9).

V. KẾT LUẬN

Trong công trình này, bộ dữ liệu VLP lớn với hơn 18803 ảnh biển số xe Vietnam trong các điều kiện và chủng loại khác nhau được xây dựng. Chúng tôi đề xuất kiến trúc Vocx đánh giá trên tập dữ liệu VLP với hơn với tỉ lệ nhận diện trên tập test 2828 ảnh với độ chính xác Full sequence là 98.36%, Mean char là 99.65% cho thấy tính tổng quát và tiềm năng của cách tiếp cận này. Ngoài ra, một số cải tiến về mặt hệ thống bằng cách tích hợp thêm thuật toán tracking và ocr-check phù hợp với hệ thống sử dụng camera giao thông cũng được đề xuất. Công việc tương lai, chúng tôi sẽ xây dựng bộ dữ liệu và đánh giá bước phát hiện ô tô, phát hiện biển số cũng như hiệu năng chung của hệ thống. Đồng thời, chúng tôi sẽ mở rộng bài toán đối với biển số xe máy Vietnam.

Samples	Vocr	Yolov2 [34]	LPRnet [42]
	KB 54-30	B 54-30	K8 574-530
	26A 029.87	26 47	26A 020.87
	26A 067.64	2 067.64	26A 067.64
	26F 000.30	2 100.30	26F 000.30
	26A-003.94	26A-003.94	26A-003.94
	30A-203.88	30A-203.88	30A-203.88
	30F-215.10	30F-215.10	30F-215.19
	VT-25-47	VI-25-47	T-25-47
	20A-036.78	20-360.78	6A-360.78

Hình 9: Một số kết quả dự đoán của các thử nghiệm.

REFERENCES

- [1] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. "Effective Approaches to Attention-based Neural Machine Translation". In: *CoRR* (2015).
- [2] R. Laroca et al. "A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector". In: (2018).
- [3] A. M. Al-Ghaili et al. "A new vertical edge detection algorithm and its application". In: (2008).
- [4] M. Sarfraz, M. J. Ahmed, and S. A. Ghazi. "Saudi Arabian license plate recognition system". In: (2003).
- [5] Y. Wen et al. "An Algorithm for License Plate Recognition Applied to Intelligent Transportation System". In: *IEEE Transactions on Intelligent Transportation Systems* 3 (2011).
- [6] A. Rio-Alvarez et al. "Effects of Challenging Weather and Illumination on Learning-Based License Plate Detection in Noncontrolled Environments". In: *Scientific Programming* (June 2019).
- [7] Christos-Nikolaos Anagnostopoulos et al. "License Plate Recognition From Still Images and Video Sequences: A Survey". In: *Intelligent Transportation Systems, IEEE Transactions on* (Oct. 2008).
- [8] S. Du et al. "Automatic License Plate Recognition (ALPR): A State-of-the-Art Review". In: *IEEE Transactions on Circuits and Systems for Video Technology* 2 (2013).
- [9] Alex Graves et al. "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks". In: Jan. 2006.
- [10] Junyoung Chung et al. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". In: *CoRR* (2014).

- [11] Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *arXiv 1409.1556* (Sept. 2014).
- [12] Danian Zheng, Yan Zhao, and Jiaxin Wang. "An efficient method of license plate location". In: *Pattern Recognition Letters* (Nov. 2005).
- [13] S. Yohimori et al. "License plate detection system by using threshold function and improved template matching method". In: (July 2004).
- [14] Dong-Su Kim and Sung-Il Chien. "Automatic car license plate extraction using modified generalized symmetry transform and image warping". In: (Feb. 2001).
- [15] Hong-ke Xu et al. "A New Approach of the Vehicle License Plate Location". In: *Parallel and Distributed Computing, Applications and Technologies, PDCAT Proceedings* (Jan. 2006).
- [16] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: (June 2016).
- [17] Joseph Redmon and Ali Farhadi. "YOLO9000: Better, Faster, Stronger". In: (July 2017).
- [18] Gee-Sern Hsu et al. "Robust license plate detection in the wild". In: (Aug. 2017).
- [19] Lele Xie et al. "A New CNN-Based Method for Multi-Directional Car License Plate Detection". In: *IEEE Transactions on Intelligent Transportation Systems* (Jan. 2018).
- [20] Max Jaderberg et al. "Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition". In: (June 2014).
- [21] Andrea Vedaldi and Andrew Zisserman. "Synthetic Data for Text Localisation in Natural Images". In: June 2016.
- [22] Fangfang Wang et al. "Geometry-Aware Scene Text Detection with Instance Transformation Network". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018).
- [23] Anuja Nagare and Shalini Bhatia. "License Plate Character Recognition System using Neural Network". In: *International Journal of Computer Applications* (July 2011).
- [24] Ihor Paliy et al. "Approach to recognition of license plate numbers using neural networks". In: (Aug. 2004).
- [25] C.A. Rahman, Wael Badawy, and Ahmad Radmanesh. "A real time vehicle's license plate recognition system". In: (Aug. 2003).
- [26] Hui Li et al. "Reading Car License Plates Using Deep Neural Networks". In: *Image and Vision Computing* (Mar. 2018).
- [27] Zied Selmi, Mohamed Ben Halima, and Adel Alimi. "Deep Learning System for Automatic License Plate Detection and Recognition". In: (Nov. 2017).
- [28] Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (June 2015).
- [29] Teik Cheang, Yong Shean Chong, and Yong Haur Tay. "Segmentation-free Vehicle License Plate Recognition using ConvNet-RNN". In: (Jan. 2017).
- [30] Hui Li, Peng Wang, and Chunhua Shen. "Toward End-to-End Car License Plate Detection and Recognition With Deep Neural Networks". In: *IEEE Transactions on Intelligent Transportation Systems* (Sept. 2017).
- [31] Mark Everingham et al. "The Pascal Visual Object Classes (VOC) Challenge". In: *International Journal of Computer Vision* 2 (June 2010). ISSN: 1573-1405.
- [32] Olga Russakovsky et al. "ImageNet Large Scale Visual Recognition Challenge". In: *International Journal of Computer Vision* (Sept. 2014).
- [33] Tsung-Yi Lin et al. *Microsoft COCO: Common Objects in Context*. 2015.
- [34] Sérgio Montazzolli and Claudio Jung. "License Plate Detection and Recognition in Unconstrained Scenarios". In: (Sept. 2018).
- [35] Sergey Ioffe and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *CoRR* (2015).
- [36] Dzmitry Bahdanau, Kyunghyun Cho, and Y. Bengio. "Neural Machine Translation by Jointly Learning to Align and Translate". In: *ArXiv* (Sept. 2014).
- [37] Alex Lamb et al. *Professor Forcing: A New Algorithm for Training Recurrent Networks*. 2016.
- [38] Rafael Müller, Simon Kornblith, and Geoffrey E. Hinton. "When Does Label Smoothing Help?" In: *CoRR* (2019).
- [39] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017.
- [40] Vinh Mai, Duoqian Miao, and Ruizhi Wang. "Building a license plate recognition system for Vietnam tollbooth". In: *ACM International Conference Proceeding Series* (Aug. 2012).
- [41] Tran Duan et al. "Building an Automatic Vehicle License-Plate Recognition System". In: *Proc. Int. Conf. Comput. Sci. RIVF* (Feb. 2005).
- [42] Sergey Zherzdev and Alexey Gruzdev. "LPRNet: License Plate Recognition via Deep Neural Networks". In: (June 2018).