

Mục lục

1	Mở đầu	5
1.1	Lý do chọn đề tài	5
1.2	Mục đích đề tài	6
1.3	Đối tượng nghiên cứu	6
1.4	Tóm tắt kết quả	6
2	Tổng quan về bài toán nhận diện biển số xe	8
2.1	Multi-stage license plate recognition systems	8
2.1.1	License plate detection	8
2.1.2	License plate recognition	8
2.2	Single-stage license plate recognition systems	9
3	Tổng quan về hệ thống giám sát và truy vết ô tô trên diện rộng	10
4	Phương pháp giải quyết bài toán nhận diện biển số xe	11
4.1	Car detection	12
4.2	License plate detection	13
4.3	License plate OCR	15
4.3.1	Đặc điểm biển số xe Vietnam	15
4.3.2	Tiền xử lý	16
4.3.3	Vocr network	16
4.3.4	Kiểm tra kết quả Ocr	19
5	Xây dựng hệ thống	20
5.1	Các thành phần và chức năng chính	20
5.1.1	Phần mềm phía client sử dụng để quản lý hệ thống, giám sát hành trình xe trên bản đồ số	20
5.1.2	Server quản lý và điều khiển luồng dữ liệu từ các camera	21
5.1.3	Module nhận diện biển số xe và trả kết quả về server . .	21
5.2	Mô hình hệ thống	21
5.3	Module nhận diện biển số xe	22
5.4	Server chính quản lý và xử lý dữ liệu	22
5.5	Phần mềm phía client	22
6	Thực nghiệm	22
6.1	Tập dữ liệu	22
6.2	Đánh giá hiệu năng của hệ thống	23
6.2.1	Phát hiện biển số	23
6.2.2	Nhận diện biển số	25
6.3	Hiệu năng chung của hệ thống	28
6.3.1	Phần cứng triển khai	28
6.3.2	Thời gian phản hồi	28

6.3.3	Khả năng chịu tải	28
7	Kết luận và kiến nghị	28
8	Phụ lục	33

Danh sách hình vẽ

1	Sự minh họa mô hình nhận diện biển số xe.	12
2	Cách tiếp cận top-down HPE dựa trên heatmap regression điển hình.	14
3	Kiến trúc mạng semantic segmentation DDRNet [8].	15
4	Quy định kích thước biển số cũ (left) và mới (right).	15
5	Các định dạng văn bản của biển số.	16
6	Kiến trúc của Vocr-cnn	16
7	Kiến trúc của Vocr-rnn	18
8	Mô hình hệ thống.	21
9	Một số ảnh ô tô thu được trong tập dữ liệu VLP.	22
10	Một số ảnh biển số trong tập dữ liệu VLP.	23
11	Thiết kế các lớp keypoints của biển số (màu sắc tương ứng với lớp).	24
12	Kết quả của hệ thống sử dụng DDRNet23sh và Vocr.	33
13	Kết quả của hệ thống sử dụng DDRNet23sh và Vocr.	34

Danh sách bảng

1	Số lượng biến theo màu sắc	23
2	LP detection results ở các cách xem xét lớp của các điểm keypoints	24
3	LP detection results ở các kích thước khác nhau	25
4	LP detection results ở các kiến trúc khác nhau	25
5	Kết quả kiểm thử trên VLP dataset	27
6	Thực nghiệm sự ảnh hưởng màu sắc biến số tới Vocr	27

1 Mở đầu

1.1 Lý do chọn đề tài

Ngày nay, trước sự gia tăng nhanh chóng của các phương tiện giao thông thì công tác quản lý, kiểm soát các phương tiện giao thông một cách thông minh ngày càng thu hút được sự quan tâm. Hệ thống theo dõi và giám sát xe tự động đã được cài đặt ở nhiều quốc gia để phục vụ cho các nhiệm vụ giám sát giao thông.

Hệ thống theo dõi và giám sát xe tự động 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 và tiến hành lưu trữ, truy vết lộ trình của phương tiện đó trên bản đồ tại các điểm có camera giao thông. Các kỹ thuật để giải quyết bài toán nhận dạng biển số được đưa ra như object detection [1], image processing [2], pattern recognition [3]. Hệ thống theo dõi và giám sát xe tự động được thiết kế để triển khai ngoài trời do đó 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. Ngoài ra, 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 [4], [5].

Theo khảo sát [6], [7] 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 [6]. 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. Xuất phát từ các lý do trên, tôi chọn đề tài: *Xây dựng hệ thống giám sát và truy vết ô tô trên diện rộng sử dụng mô hình học sâu.*

1.2 Mục đích đề tài

Tìm hiểu về các công nghệ nhận dạng biển số xe, áp dụng vào xây dựng hệ thống giám sát và truy vết ô tô trên diện rộng.

1.3 Đối tượng nghiên cứu

- Các công nghệ nhận dạng biển số xe.
- Các loại camera hiện có.

1.4 Tóm tắt kết quả

Bài toán nhận diện biển số xe ô tô tự động (Automatic license plate recognition - ALPR) là một trong những bài toán có nhiều ứng dụng thực tế và đã thu hút được nhiều nghiên cứu trong những năm gần đây. Hầu hết các nghiên cứu trước đây đối với biển số xe Việt Nam tập trung vào nhận diện biển số xe khi xe ra vào bãi đỗ hoặc trạm thu phí. Khi đó góc nhu nhận hình ảnh từ camera tới vùng biển số, và ánh sáng tại khu vực xe ra vào có thể sắp đặt để có được môi trường lý tưởng cho việc nhận dạng (constrained environment). Các nghiên cứu gần đây được thực hiện trên tập dữ liệu nhỏ, chưa đa dạng, chưa bao phủ được hết các trường hợp biển số xe ở Việt Nam và chưa giải quyết được các thách thức đối với việc nhận dạng biển số trong unconstrained environment. Ngoài ra, vấn đề non-rectangular plate detection chưa được chú trọng dẫn tới độ chính xác của license plate detection chưa cao. Từ thực tế đó, trong báo cáo này chúng tôi đề xuất một mô hình cải tiến, trong đó tập trung vào nâng cao độ chính xác trong giai đoạn phát hiện biển số và nhận diện biển số xe trong các ngữ cảnh không ràng buộc. Chúng tôi cũng xây dựng bộ dữ liệu biển số xe ô tô Việt Nam đa dạng về chủng loại, màu sắc, thời gian thu thập, điều kiện thời tiết và góc bố trí camera để phục vụ việc huấn luyện và kiểm thử. Để 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 mean IOU $mIOU = 95\%$ và precision $P_{75} = 99,2\%$ dựa trên thay đổi từ kiến trúc segmantic segmentation DDRNet [8]. Phần nhận diện biển số sử dụng cách tiếp cận segmentation free dựa trên kiến trúc mạng encoder decoder. Ưu điểm chính của phương pháp 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 mức sequence là $Acc_{seq} = 99.4\%$, mức kí tự là $Acc_{char} = 99.8\%$ dựa trên kiến trúc [9] đã được áp dụng thành công trong bài toán dịch máy.

Hệ thống giám sát và truy vết ô tô trên diện rộng có chức năng theo dõi và giám sát xe dựa trên hệ thống camera giao thông ở các cung đường khác

nhau, trong các điều kiện khác nhau. Hệ thống được triển khai với kiến trúc client-server. Hiệu năng ...

Đóng góp của chúng tôi gồm có :

- Chúng tôi đề xuất mô hình hệ thống giám sát và truy vết ô tô trên diện rộng hoàn chỉnh cho triển khai thực tế. Hệ thống được tích hợp thêm công nghệ bản đồ số, thuật toán motion detection, object tracking, kiểm tra kết quả nhận dạng 2 có đúng theo quy định của biển số Việt Nam mà chính phủ ban hành.
- Giải quyết vấn đề non-rectangular object detection, chúng tôi đề xuất hướng tiếp dựa trên keypoints detection. Chúng tôi xem mỗi góc của biển số là một keypoint cần phát hiện. Kết quả tốt nhất thu được từ việc sửa đổi kiến trúc segmantic segmentation DDRNet [8] cho phù hợp với bài toán.
- 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 [10]. Tuy nhiên, cách tiếp cận này sẽ gặp phải một số nhược điểm nhất định (xem mục 2.1.2). Khắc phục những hạn chế đó, chúng tôi sử dụng kiến trúc encoder-decoder sử dụng GRU [11], tích hợp thêm cơ chế attention lấy ý tưởng từ [9]. Phần trích xuất đặc trưng được thiết kế dựa trên VGG-19 [12] với một số tinh chỉnh.
- Chúng tôi đề xuất bộ dữ liệu VLP gồm có 15571 ảnh ô tô chứa biển số và 16012 ảnh biển số xe ô tô Việt Nam ở các điều kiện khác nhau và các kiểu biển số khác nhau. Kết quả phát hiện biển số đạt $mIOU = 95\%$ và $Precision_{75} = 99.2\%$ trên 1558 ảnh kiểm thử. Độ chính xác nhận diện biển số Full sequence là $Acc_{seq} = 99.4\%$, Mean char là $Acc_{char} = 99.8\%$ trên 2472 ảnh kiểm thử.

Phần còn lại của báo cáo này được tổ chức như sau. Chương 2 sẽ trình bày tổng quan về bài toán nhận diện biển số xe. Chương 3 sẽ trình bày tổng quan về hệ thống theo dõi và giám sát xe trên diện rộng. Chương 4 mô tả chi tiết phương pháp giải quyết bài toán nhận diện biển số xe ô tô mà chúng tôi đề xuất. Bước thiết kế và xây dựng hệ thống được trình bày trong chương 5. Phần đánh giá hiệu năng của hệ thống được thể hiện trong chương 6. Kết luận của báo cáo với sự tóm lược và các gợi mở cho các nghiên cứu tiếp theo được thể hiện trong chương 7.

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

2 Tổng quan về bài toán nhận diện biển số xe

Hệ thống ALPR (Automatic license plate recognition) thường có 2 cách tiếp cận chính: multi-stage và single stage license plate recognition

2.1 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 [13], màu sắc [14], đối xứng [15], kết cấu [16]. 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 [17], [13].

2.1.1 License plate detection

Sự thành công của YOLO đã truyền cảm hứng cho nhiều nghiên cứu gần đây [18], các tác giả đã mở rộng mức độ chi tiết của output để tăng số lượng detection của mạng và đặt 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 [19], 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 [20], 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 [21],[22]. Gần đây hơn nghiên cứu [23] 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.

2.1.2 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 kích bản nhất định và phụ thuộc rất lớn vào đặc điểm đầu vào, ngưỡng được áp dụng [24], [25], [26]. Ngày nay các tiếp cận sử dụng CNN để phân đoạn ký tự trở nên phổ biến [27], 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 [1], character recognition using feature extractors sử dụng kỹ thuật xử lý ảnh và machine learning để trích xuất đặc trưng trước khi phân loại [3] hay sử dụng kỹ thuật học sâu để phân loại [24].

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 [28], [21] dựa trên variable length sequence decoding được điều khiển bằng connectionist temporal classification (CTC) loss [10]. Mô hình [29] tuân theo cách tiếp cận được mô tả trong [30] 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 [11] 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 [9] 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.

2.2 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 [29], [31]. Nghiên cứu [31] 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) [29]. 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 [29] 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

3 Tổng quan về hệ thống giám sát và truy vết ô tô trên diện rộng

Tại Hà Nội, riêng Phòng Cảnh sát giao thông Hà Nội đang quản lý, sử dụng gần 600 camera các loại, trong đó có hơn 300 camera để theo dõi lưu lượng phương tiện, hơn 100 camera phục vụ xử phạt vi phạm giao thông và gần 100 thiết bị giám sát giao thông. Ngoài ra, tại nhiều nút giao thông còn có camera của công an các quận; camera của VOV giao thông; camera theo dõi điểm ngập úng của ngành thoát nước... Nhờ hệ thống này, tình trạng người tham gia giao thông vi phạm các lỗi như lấn làn, vượt đèn đỏ, dừng đỗ không đúng vị trí... đã giảm đáng kể. Hà Nội đang khẩn trương nghiên cứu tích hợp không chỉ camera giám sát của các cơ quan nhà nước, mà còn của người dân, doanh nghiệp (sân bay, nhà ga, bến xe, các cơ quan hành chính, ngõ phố...) để hỗ trợ công tác điều hành giao thông chung..

Tại TP Hồ Chí Minh, trung tâm giám sát và điều khiển giao thông giai đoạn 1 (thuộc dự án Nâng cấp hệ thống điều khiển giao thông hiện hữu) đã hoàn thành, tổng mức đầu tư khoảng 250 tỷ đồng, Trung tâm có nhiệm vụ kiểm soát giao thông trên các tuyến đường chính, nút giao thông khu vực trung tâm thành phố, với 4 chức năng: điều khiển đèn tín hiệu; giám sát và cung cấp thông tin giao thông; hỗ trợ xử lý vi phạm. Trung tâm giám sát và điều khiển giao thông vận hành 24/7 trong cả 365 ngày có thể được xem như "mắt thần". Thông tin sẽ lập tức được chia sẻ trên cổng thông tin giao thông cho người dân cập nhật, tránh các điểm kẹt xe, chống ùn tắc, xử lý các sự cố hạ tầng hữu hiệu. Ở giai đoạn 2, sau năm 2020 sẽ hoàn thành Trung tâm điều hành giao thông thông minh có quy mô toàn thành phố, tổng vốn đầu tư 250 triệu USD. Trung tâm điều hành giao thông thông minh sẽ quản lý trực tiếp toàn bộ hệ thống điều khiển giao thông của thành phố cũng như cung cấp thông tin dữ

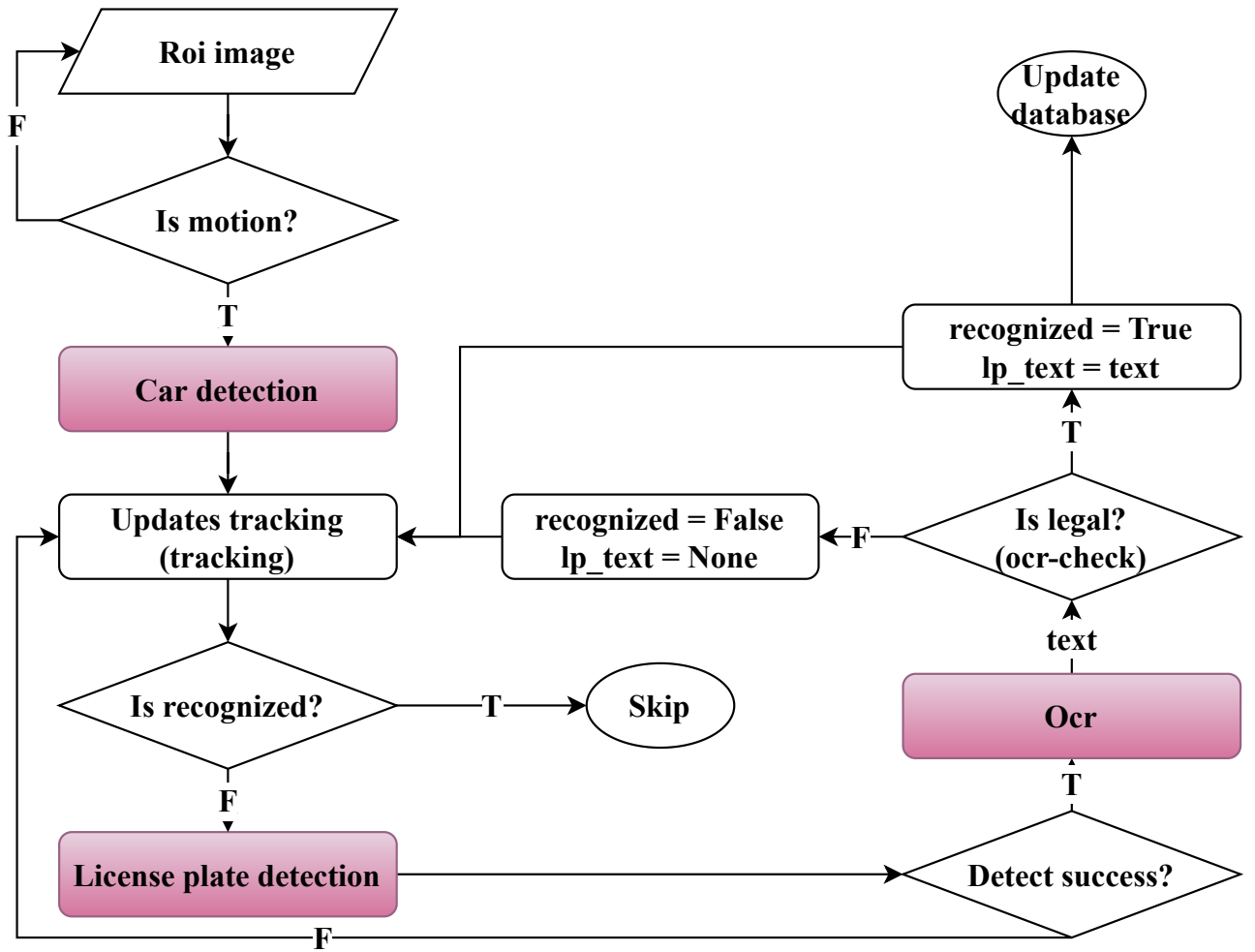
liệu, phục vụ cho việc xử lý các vi phạm; giám sát, đảm bảo an ninh trật tự đô thị; xử lý các sự cố khẩn cấp; phòng chống cháy nổ, ngập lụt; quản lý bến bãi.

Huế sẽ xử phạt vi phạm đồng bộ trên phạm vi toàn tỉnh kể từ ngày 01/11/2019. Các hành vi vi phạm về trật tự an toàn giao thông như: xe dừng, đỗ, đón, trả khách không đúng quy định; xe chạy vào đường cấm, vào thời gian cấm; xe không được phép, không có tuyến vào thành phố; xe chạy sai luồng, sai tuyến, xe vượt đèn đỏ, xe quá khổ, xe quá tải; lấn chiếm lòng, lề đường trái phép, thi công công trình trên đường bộ không đảm bảo an toàn giao thông... Các hành vi này được ghi nhận trực tiếp từ hệ thống ghi nhận hình ảnh do Trung tâm Giám sát, điều hành đô thị thông minh quản lý (bao gồm cả các thiết bị ghi nhận hình ảnh của các tổ chức, cá nhân, đơn vị khác đầu tư nhưng đã được phép đấu nối vào hệ thống giám sát điều hành đô thị thông minh và hình ảnh được các cá nhân, tổ chức phản ánh qua kênh tiếp nhận của Trung tâm Giám sát, điều hành đô thị thông minh đã được Trung tâm xác minh thông tin) theo quy định sẽ bị xử lý và xử phạt hành chính theo quy định của pháp luật.

Qua khảo sát có thể thấy rằng, tình trạng chung tại các thành phố lớn đều đã và đang được trang bị các camera giám sát. Tuy nhiên, chưa có các phần mềm xử lý ảnh chuyên dụng để xử lý dữ liệu từ các camera này hỗ trợ cho các lực lượng chức năng quản lý các vi phạm và tra cứu khi cần. Đặc biệt, trong các trường hợp khẩn cấp, với các camera được lắp đặt hiện nay chưa thể truy vết các xe vi phạm khi lưu hành qua các trạm quan sát của camera. Do đó, rất cần thiết xây dựng hệ thống nhận diện biển số xe để giám sát được các xe lưu thông trên địa bàn một cách trực quan và chính xác, hỗ trợ các cơ quan chức năng truy vết được các xe khi cần, nâng cao hiệu quả quản lý.

4 Phương pháp giải quyết bài toán nhận diện biển số xe

Mô hình nhận diện biển số xe 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 8. Đầu vào là một khung hình thu được từ camera giám sát trên đường quốc lộ có thể xác định trước vùng quan tâm (ROI). Thuật toán phát hiện chuyển động dựa trên frame difference. Nếu có chuyển động, module car detection phát hiện khung bao đóng của các ô tô. Mỗi khung bao đóng 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 tiến hành phát hiện biển số. Module ocr 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.



Hình 1: Sự minh họa mô hình nhận diện biển số xe.

4.1 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 [32], ImageNet [33] và COCO [34], 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 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 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 bắt biển 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 di chuyển 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 chỉ 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.

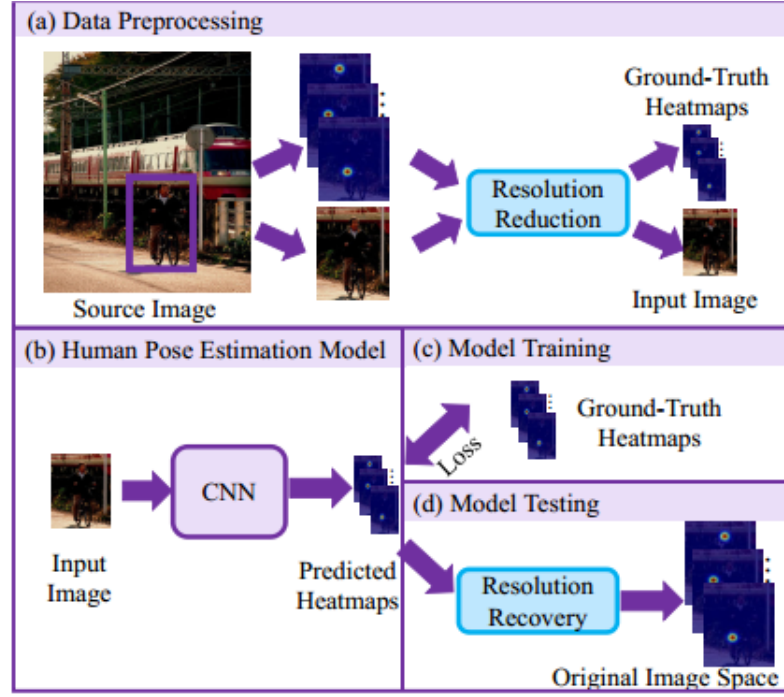
4.2 License plate detection

Đối với phần phát hiện biển số, chúng tôi tiếp cận bài toán theo hướng keypoints detection, mỗi góc của biển số được xem là một keypoint. Dựa vào các điểm keypoint, chúng tôi có thể căn chỉnh được các biển số khi chúng bị biến dạng, hoặc nghiêng ở các góc khác nhau, điều mà các phương pháp sử dụng các khung bao đóng dạng hình chữ nhật thuần túy để phát hiện đối tượng không giải quyết được.

Vấn đề phát hiện các điểm keypoints của biển số xe rất gần với human pose estimation problem. Có hai cách truyền thống để giải quyết bài toán này là coordinate regression và heatmap regression [35]. Cách tiếp cận heatmap regression đều đạt state-of-the-art trên các benchmark dataset, cho thấy việc huấn luyện mô hình hồi quy trực tiếp tọa độ của các điểm keypoints khó khăn hơn. Vì thế chúng tôi tiếp cận với cách thứ hai heatmap regression.

Tương tự như các bước trong bài toán human pose estimation (xem hình 2). Quá trình huấn luyện và inference được thực hiện như [36]. Tuy nhiên, chúng tôi nhận thấy ngoài việc phát hiện các điểm keypoint của đối tượng, Zhou [36] phát hiện thêm tâm của đối tượng và hồi quy các offset tới các keypoint của đối tượng như là một cách để kết nối các điểm keypoint khác của đối tượng là không cần thiết trong bài toán của chúng tôi. Vì mỗi ô tô chúng tôi phát hiện được chỉ có duy nhất một biển số và có thể dựa vào hình dạng của biển số để kiểm soát được các điểm keypoints nếu có nhiều hơn 4 điểm keypoints phát hiện được. Do đó, chúng tôi chỉ phát hiện 4 điểm keypoints đủ để giải quyết vấn đề và hạn chế tỉ lệ lỗi khi mô hình phải học nhiều thông tin hơn. Đồng thời, các nghiên cứu trước xem xét các điểm keypoints như là các lớp khác nhau, tuy nhiên, quá trình thực nghiệm cho thấy các keypoints của biển số có thể xem là một lớp duy nhất.

Các nghiên cứu trước huấn luyện heatmap regression có hai kiểu kiến trúc chính. Dạng kiến trúc duy trì high-resolution và các kĩ thuật feature fusion khác nhau như HRNet [37], DLA network [38]. Dạng kiến trúc có dạng high-to-low-

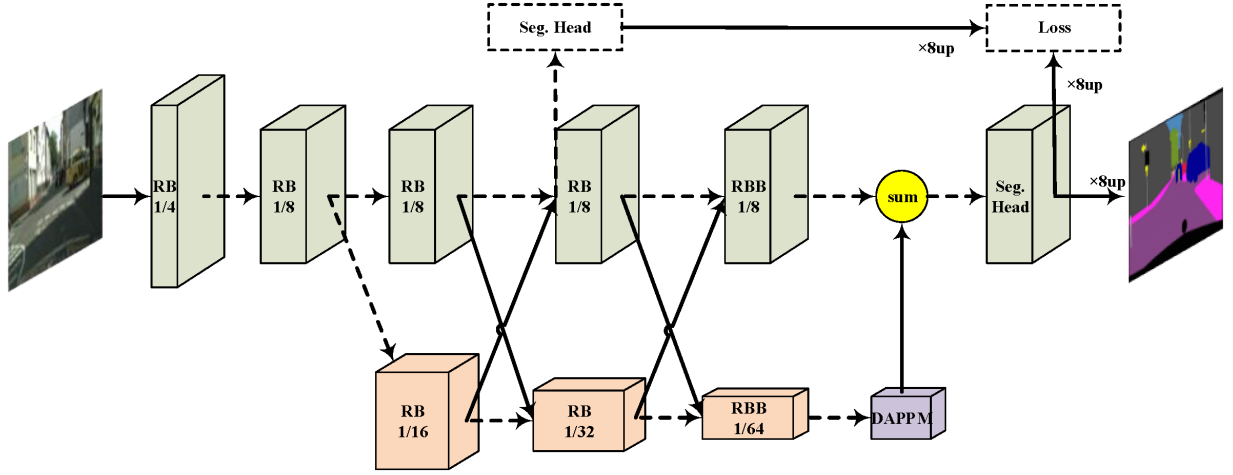


Hình 2: Cách tiếp cận top-down HPE dựa trên heatmap regression điển hình.

to-high resolution bằng việc down convolution và up convolution đối xứng kết hợp với các lateral connection như kiến trúc dựa trên Restnet [39], Hourglass Network [40]. Dạng kiến trúc thứ nhất cho độ chính xác tốt trong tác vụ này, [37] đang đứng đầu trên tập dữ liệu COCO test-dev human pose estimation với $AP = 74.4$. Chúng tôi nhận thấy các kiến trúc trên có mối liên hệ rất cao với các kiến trúc mạng segmentation, do đó chúng tôi thay đổi các kiến trúc mạng segmentation cho phù hợp với bài toán.

Hệ thống nhận diện biển số xe tự động từ các camera giao thông yêu cầu thời gian xử lý càng nhanh càng tốt, do các phương tiện không ngừng chuyển động với tốc độ lớn. Do đó, chúng tôi ưu tiên thử nghiệm các kiến trúc lightweight gồm có DLA-34 [38], HRNet-18 [37], Restnet18 [39]. Đối với các kiến trúc segmentation, chúng tôi lựa chọn DDRNet [8] và các biến thể của nó (xem hình 3). Do đây là kiến trúc segmantic segmentation đạt được state-of-the-art trade-off giữa độ chính xác và tốc độ khi kiểm thử trên tập cityscapes and camvid, mà không sử dụng tensorRT và thêm dữ liệu khác để huấn luyện.

Kích thước heatmap đầu ra mà chúng tôi lựa chọn có tỉ lệ so với đầu vào là $R = \frac{1}{4}$. Tuy nhiên DDRNet lại duy trì high-resolution feature map có kích thước là $R = \frac{1}{8}$ so với ảnh đầu vào. Cách thứ nhất, chúng tôi thay đổi mức độ down convolution ngay từ block convolution đầu tiên chỉ còn $R = \frac{1}{2}$. Cách thứ hai, chúng tôi thêm 1 block sử dụng Transpose convolution ở lớp cuối cùng. Để có thêm thông tin về lower feature maps, chúng tôi sử dụng residual connection ở block convolution thứ 4 và block convolution cuối cùng trong kiến trúc gốc



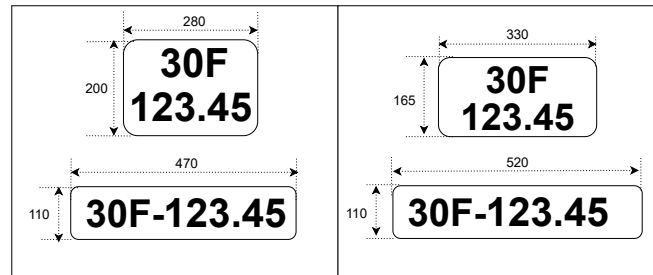
Hình 3: Kiến trúc mạng semantic segmentation DDRNet [8].

để thay thế cho việc huấn luyện và tổng hợp thông tin từ hai block này riêng biệt.

4.3 License plate OCR

4.3.1 Đặc điểm biển số xe Vietnam

Biển số xe ô tô Vietnam được thiết kế theo quy định của chính phủ. Về kích thước kiểu cũ và kiểu mới (xem hình 4). Về màu sắc, biển số Vietnam gồm biển nền xanh hoặc đỏ với chữ trắng, biển nền vàng hoặc trắng với chữ đen. Về quy



Hình 4: Quy định kích thước biển số cũ (left) và mới (right).

định các ký tự trong biển số, các ký 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 5), 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.



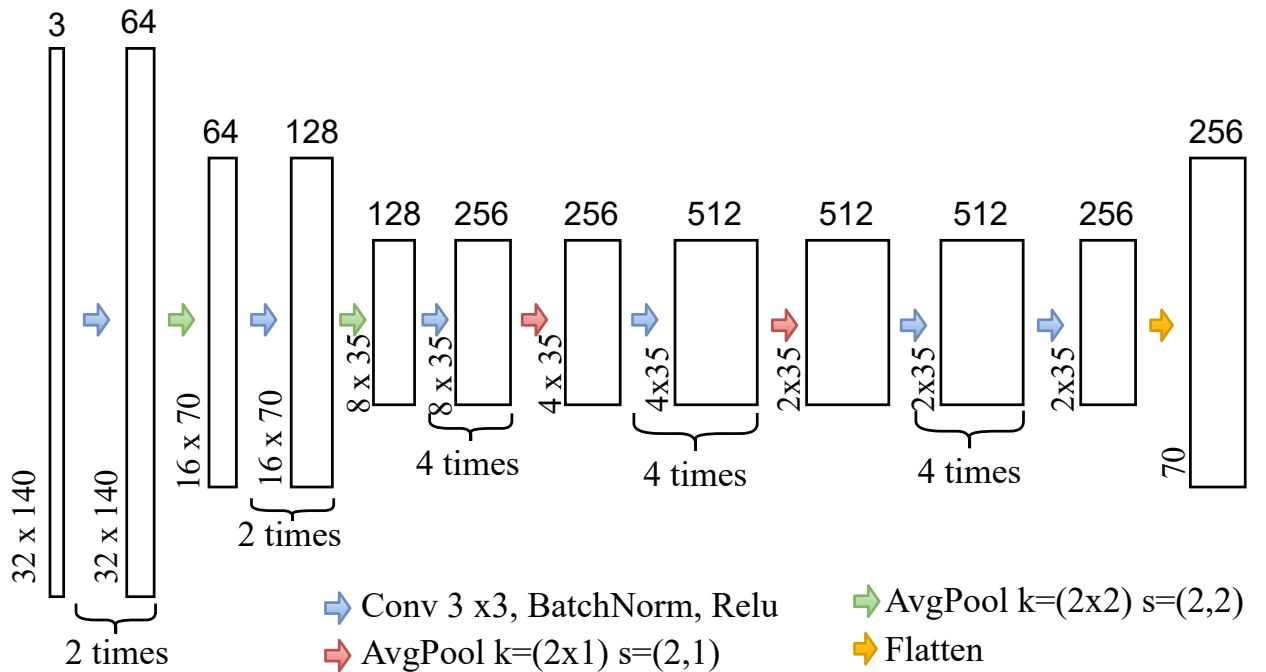
Hình 5: Các định dạng văn bản của biển số.

4.3.2 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.

4.3.3 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 [12] (xem hình 6). Do biển số thường có kích thước khá nhỏ. Vì thế, để tránh



Hình 6: Kiến trúc của Vocr-cnn

mất thông tin, chúng tôi lựa chọn 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 [41] 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 1x70. 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

Algorithm 1 Two-row license splitting algorithm

Input: LPR image

Output: position to split license plate horizontally

$img \leftarrow$ convert img to gray

$temp \leftarrow 255 - img$

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

$img \leftarrow$ remove border

$img \leftarrow$ equalize

$thresh \leftarrow$ turn img into binary image

$indexs \leftarrow$ horizontally distributed count of thresh

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

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

$indexs \leftarrow$ remove $indexs[i]$

else

 break

end if

end for

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

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

$indexs \leftarrow$ remove $indexs[i]$

else

 break

end if

end for

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

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

 return i

end if

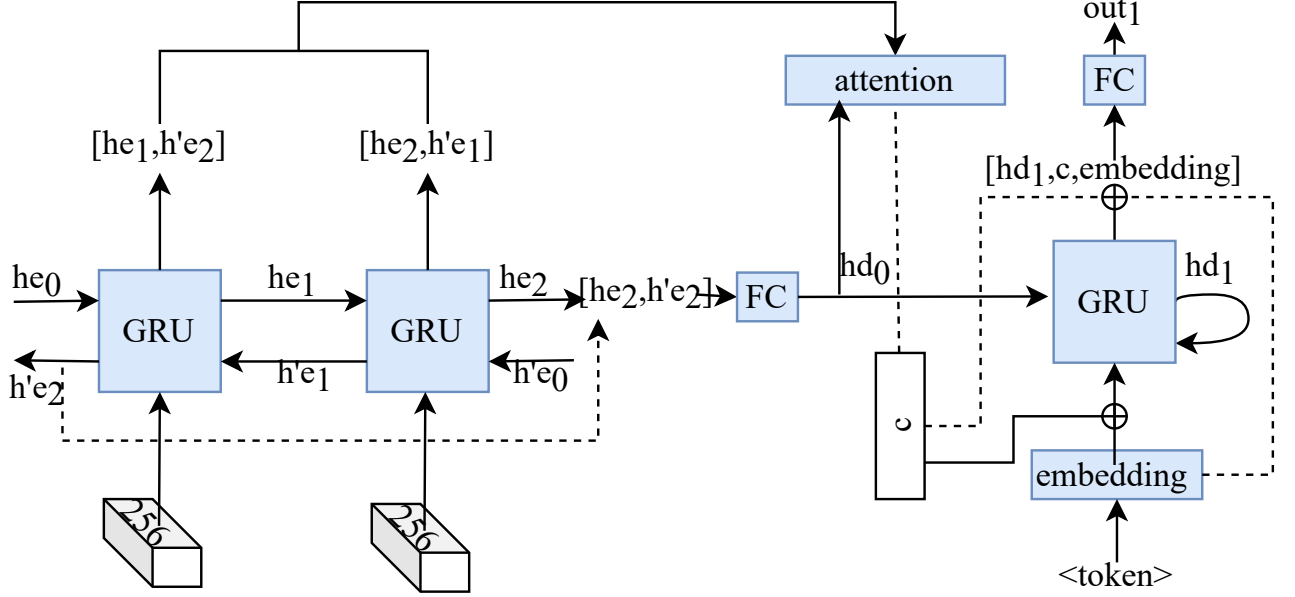
end for

$img \leftarrow temp$

end for

return none

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 7). Trong đó, he , $h'e$, hd là các hidden output feature tại mỗi time step của encoder



Hình 7: Kiến trúc của Vocr-rnn

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 [9] xem mỗi hd tại mỗi time step là query và tất cả $[he, h'e]$ đóng 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) [42].

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

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 [43]. 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 [44] 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 [44].

Để tối ưu hàm loss L , chúng tôi sử dụng phương pháp cập nhật tham số dựa trên gradient descent với thuật toán Adam [45] tự động điều chỉnh learning rate và tránh rơi vào local minimum.

4.3.4 Kiểm tra kết quả Ocr

Trong kiến trúc hệ thống 8, kết quả của Vocr sau khi dự đoán được kiểm tra lại thông qua thuật toán Ocr-check dưới đây. Thuật toán Ocr-check dựa trên những quy định về biển số Vietnam mà chính phủ ban hành (xem mục 4.3.1).

Algorithm 2 Ocr-check

Input:Text of LP

Output:legal or illegal

line1 \leftarrow regular expression of LPR in first line

line1 \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

5 Xây dựng hệ thống

5.1 Các thành phần và chức năng chính

Hệ thống được xây dựng gồm 3 thành phần bao gồm:

- Phần mềm phía client sử dụng để quản lý hệ thống, giám sát hành trình xe trên bản đồ số.
- Server quản lý và điều khiển luồng dữ liệu từ các camera.
- Module nhận diện biển số xe và trả kết quả về server.

5.1.1 Phần mềm phía client sử dụng để quản lý hệ thống, giám sát hành trình xe trên bản đồ số

Quản lý thiết bị camera (Địa chỉ IP, Port, chuỗi kết nối, Vị trí camera , loại camera,...).

Quản lý dữ liệu biển số xe.

Quản lý dữ liệu biển số xe theo camera.
Tra cứu lộ trình của xe trên bản đồ số.
Xem hình ảnh trực tiếp camera dưới dạng lưới bình thường hoặc ở chế độ nhận diện biển số.

5.1.2 Server quản lý và điều khiển luồng dữ liệu từ các camera

Thu nhận dữ liệu từ các server chứa các module nhận dạng biển số xe và lưu trữ vào cơ sở dữ liệu.

Trả về kết quả truy vấn cơ sở dữ liệu và hình ảnh cho client.

5.1.3 Module nhận diện biển số xe và trả kết quả về server

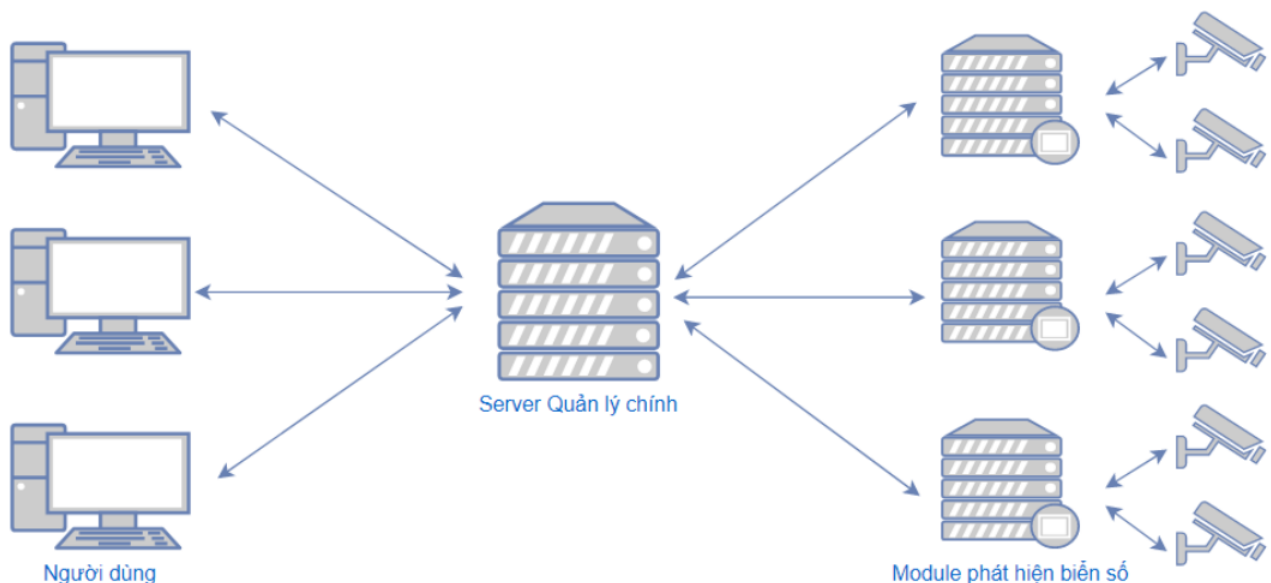
Mỗi module đảm nhận thu thập dữ liệu từ các camera.

Thu nhận hình ảnh từ các camera do mình quản lý sau đó xử lý và trả về server kết quả là hình ảnh và chuỗi từ các biển số xe.

5.2 Mô hình hệ thống

Hệ thống được chia thành 3 thành phần chính:

- Module nhận diện biển số xe.
- Server chính quản lý và xử lý dữ liệu.
- Phần mềm phía client.



Hình 8: Mô hình hệ thống.

5.3 Module nhận diện biển số xe

Sơ đồ luồng dữ liệu

5.4 Server chính quản lý và xử lý dữ liệu

Sơ đồ luồng dữ liệu

5.5 Phần mềm phía client

Quy trình thực hiện các chức năng và giao diện chức năng tương ứng

6 Thực nghiệm

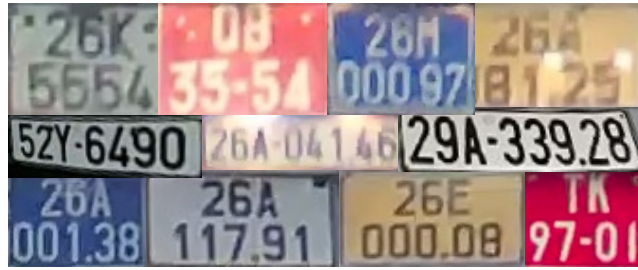
6.1 Tập dữ liệu

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 [46] chỉ kiểm tra trên 700 ảnh, Duan và cộng sự [47] 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 8, 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 khác nhau với các phương tiện bất kỳ di chuyển cùng với các loại biển cũ và mới, kích thước, kiểu chữ, độ mờ, nhiễu rất đa dạng (xem hình 9,10).



Hình 9: Một số ảnh ô tô thu được trong tập dữ liệu VLP.

Chúng tôi thu thập thêm bộ dataset hơn 3000 ảnh ô tô việt nam tại trạm thu phí của hầm giữ xe và trong các điều kiện ngoài trời được publish bởi [48], nhằm tăng tính đa dạng về phân bố của tập dữ liệu VLP. Kết hợp với bộ dữ liệu chúng tôi tự thu thập, kết quả tập dữ liệu xe ô tô là 15571 ảnh và 10773



Hình 10: Một số ảnh biển số trong tập dữ liệu VLP.

ảnh biển số. Số lượng biển số ít hơn số lượng xe ô tô, do chúng tôi bỏ bớt các biển số thu thập ở cùng khoảng thời gian tránh dữ liệu bias vào một phân bố cụ thể. Các loại biển theo màu sắc được tổng hợp ở bảng 1. Chúng tôi sử dụng thuật toán tách biển 1 để tách 2 dòng thành 1 dòng, tổng số biển 1 dòng thu được là 18809. Chúng tôi nhận thấy số lượng biển hai dòng tách ra, dòng thứ nhất chứa 3297 biển "26A", do ở cùng khu vực. Chúng tôi bỏ ngẫu nhiên 2797 ảnh. Kết quả thu được bộ dữ liệu biển số VLP gồm 16012 biển số 1 dòng.

Bảng 1: Số lượng biển theo màu sắc

Số lượng biển	Trắng	Vàng	Xanh	Đỏ	Tổng
Biển 1 dòng	2717	0	13	7	2737
Biển 2 dòng	6933	950	143	10	8036
Tổng	9650	950	156	17	10773
Tổng Splitted	18809				
Bỏ bớt 26A	2797				
VLP	16012				

6.2 Đánh giá hiệu năng của hệ thống

6.2.1 Phát hiện biển số

Cài đặt thực nghiệm: Với dữ liệu như trên, chúng tôi chia ngẫu nhiên tập 15571 ảnh ô tô chứa biển số thành 3 tập: train, validate, test với tỉ lệ 0.8:0.1:0.1.

Phương pháp đánh giá: Chúng tôi đánh giá dựa trên 2 thông số mean IOU ($mIOU$) và độ chính xác Precision với ngưỡng $mIOU \geq 0.75$ như [49]. Chúng tôi lựa chọn ngưỡng $mIOU \geq 0.75$ thay vì $mIOU \geq 0.7$ khẳng định tính mạnh mẽ của mô hình.

Training: Quá trình training, tập dữ liệu được agumentation online. Chúng tôi sử dụng thuật toán tối ưu Adam [45] với learning rate khởi tạo là 0.001, batch size là 8 và training tối đa 140 epochs với early stopping có patience là 10.

Testing results and comparable: Trước hết, chúng tôi xem xét việc nên xem các điểm keypoints của biển số là các class riêng biệt hay chỉ là một class.

Chúng tôi thực nghiệm trên kiến trúc Resnet18 [39] với các cách tiếp cận khác nhau: Resnet18_1c chỉ xem xét các điểm keypoints là 1 lớp, Resnet18_2c xem xét 2 điểm keypoints ở trên là 1 lớp và 2 điểm keypoints còn lại là 1 lớp do sự tương đồng của chúng, Resnet18_4c xem xét các điểm keypoints là 4 lớp riêng biệt, Resnet18_1c3 xem 2 điểm keypoints dưới và tâm của biển số thuộc cùng 1 lớp, bỏ qua hai điểm keypoints phía trên do background của nó thay đổi nhiều ở các loại phương tiện (xem hình 11).



Hình 11: Thiết kế các lớp keypoints của biển số (màu sắc tương ứng với lớp).

Bảng 2: LP detection results ở các cách xem xét lớp của các điểm keypoints

Phương pháp	Training size	Testing size	$mIOU$
Resnet18_1c	384	384	93.8
Resnet18_2c	384	384	93.797
Resnet18_4c	384	384	91.78
Resnet18_1c3	384	384	90.5

Xem bảng 2 ta thấy, Rest18_1c cho kết quả $mIOU = 93.8$ cao nhất giải thích cho việc các điểm keypoints của biển số đều có thể xem là 1 đối tượng. Ngoài ra Resnet18_2c có kết quả tương đương do độ tương đồng của các cặp keypoints phía trên cũng như phía dưới là tương tự nhau, đồng thời cặp keypoints phía trên có những đặc trưng khác hơn so với cặp keypoints phía dưới do đó có thể xem xét chúng như 2 lớp riêng biệt. Việc xem xét keypoints thành các class riêng giống như bài toán human pose estimation cho kết quả không tốt do các điểm keypoints có những đặc trưng chung khiến mô hình bị confuse trong quá trình huấn luyện. Việc xem xét 3 điểm keypoints cũng không cho kết quả khả quan do tâm của biển số nhập nhằng hơn trong heatmap regression. Các kết quả thực nghiệm dưới đây chúng tôi đều xem xét 4 điểm keypoints của biển số đều thuộc 1 lớp.

Chúng tôi phát hiện các keypoints của biển số từ ảnh ô tô thu được từ mô hình phát hiện ô tô, do đó kích thước ảnh gốc không quá lớn. Do đó, chúng tôi xem xét độ chính xác của mô hình ở hai kích thước ảnh là 384×384 và 512×512 (xem bảng 4) với mô hình DDRNet23 slim [8] với chỉnh sửa upconv mà chúng tôi đề xuất. Việc kiểm thử ở kích thước 384×384 thấp hơn lúc huấn luyện là 512×512 khiến $Precision_{75}$ giảm 1.3%. Training và testing ở cùng kích thước 512×512 có $Precision_{75} = 99.4$ cao nhất, tuy nhiên không cao hơn nhiều so với kích thước 384×384 , trong khi $mIOU$ lại thấp hơn 0.2% - dễ hiểu

do kích thước lớn hơn dẫn tới sai số cao hơn khi xét trên thông số $mIOU$. Do

Bảng 3: LP detection results ở các kích thước khác nhau

Phương pháp	Training size	Testing size	$mIOU$	$Precision_{75}$
DDRNet23sup	512	512	94.3	99.4
DDRNet23sup	512	384	93.1	98.1
DDRNet23sup	384	512	94.4	99.3
DDRNet23sup	384	384	94.5	99.2

đó, khi triển khai hệ thống tùy theo ưu tiên về mặt thời gian hay độ chính xác chúng ta có thể trade-off giữa kích thước đầu vào với sự thay đổi về độ chính xác.

Với kích thước training và testing là 384×384 , chúng tôi tiến hành huấn luyện và đánh giá hiệu năng của các kiến trúc khác nhau gồm có: DDRNet23sup, DDRNet23sh là phiên bản sửa đổi của chúng tôi từ kiến trúc DDRNet23-slim [8] bằng cách sử dụng transpose convolution và giảm mức độ down convolution tương ứng; DDRNet23up là phiên bản sửa đổi từ DDRNet23 [8] bằng transpose convolution; HRNet18_4s và HRNet18_3s tương ứng là phiên bản 4 stages và phiên bản chỉnh sửa 3 stages bỏ đi stage cuối cùng của HRNet18 [37]; Resnet18 [39], DLA34 [38] được giữ nguyên.

Bảng 4: LP detection results ở các kiến trúc khác nhau

Phương pháp	Training size	Testing size	$mIOU$	$Precision_{75}$
DDRNet23sup	384	384	94.5	99.2
DDRNet23sh	384	384	95.01	99.5
DDRNet23up	384	384	94.4	99.0
HRNet18_4s	384	384	94.2	98.58
HRNet18_3s	384	384	94.1	98.9
Resnet18	384	384	93.8	99.1
DLA34	384	384	90.5	93.6
DLA34	512	512	93.0	97.4

DDRNet23sh cho độ chính xác tốt nhất $mIOU = 95.01$, $Precision_{75} = 99.5$ mặc dù xuất phát điểm là kiến trúc dành cho bài toán segmentation. Mặc dù DDRNet23sh có hơn 6M tham số ít hơn nhiều so với DLA34 20M tham số, HRNet18_4s gần 10M tham số. Tuy nhiên cách thiết kế kiến trúc mạng tỉ mỉ vừa duy trì được high-resolution vừa có kiến trúc đa nhánh và có thêm module DAPPM làm giàu thông tin trích xuất và tăng độ phức tạp cho mô hình.

6.2.2 Nhận diện biển số

Cài đặt thực nghiệm: Với dữ liệu như trên, chúng tôi chia ngẫu nhiên 16012 biển số thành 3 tập train, validate, test theo tỉ lệ 0.75:0.1:0.15. Để chứng

minh sự mạnh mẽ của kiến trúc Vocr, chúng tôi so sánh với các cách tiếp cận điển hình cho bài toán nhận diện ký tự biến số. Với cách tiếp cận segmentation-free, chúng tôi thử nghiệm với LPRnet [50] là kiến trúc lightweight chỉ sử dụng CNN và CTC-loss [10]; kiến trúc CRNN [51] là sự kết hợp của CNN và RNN tối ưu bằng CTC-loss [10].

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 [52]. Tuy nhiên, có thể xảy ra trường hợp nếu kết quả dự đoán chỉ nhằm 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 nhãn và trong kết quả dự đoán là như nhau.

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

- Làm mờ ảnh: Gauussian 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 có sử dụng Early stopping. Việc cập nhật tham số được thực hiện bởi thuật toán Adam [45].

Kết quả kiểm thử và so sánh: Kiến trúc CRNN (VGG-19) tương tự như CRNN [51], tuy nhiên chúng tôi thay phần backbone CNN là VGG-19 của Vocr nhằm so sánh hiệu năng của cách tiếp cận CRNN kết hợp CTC-loss công bằng hơn với Vocr. Tuy nhiên, Vocr cho kết quả vượt trội trong tác vụ này $Acc_{seq} = 99.28\%$, $Acc_{char} = 99.7\%$ trên tập kiểm thử, do việc trích chọn đặc trưng từ CNN feature map có sử dụng cơ chế attention khiến mô hình học kết hợp được các đặc trưng cần thiết để decode ra ký tự (xem bảng 5). Mặc dù

Bảng 5: Kết quả kiểm thử trên VLP dataset

Phương pháp	Agu	Training set		Validation set		Testing set	
		Acc_{seq}	Acc_{char}	Acc_{seq}	Acc_{char}	Acc_{seq}	Acc_{char}
Vocr	✓	99.34%	99.8%	99.25%	99.9%	99.28%	99.7%
LPRnet-STN [50]	✓	92.3%	96.9%	94.1%	96.4%	94.8%	96.8%
CRNN [51]	✓	92.9%	96.9%	93.6%	97.2%	93.4%	97.1%
CRNN (VGG-19)	✓	92.3%	95.5%	92.5%	95.4%	93.0%	96.2%

có kiến trúc lightweight nhưng LPRnet-STN [50] cho kết quả tương đối cao do được thiết kế riêng cho bài toán nhận diện biển số với kích thước đầu vào cố định 24×94 đi cùng với các kích thước kernel, stride được thiết kế với kích thước cụ thể nhằm trích xuất được thông tin các kí tự của biển số. Tuy nhiên, việc cố định kích thước cũng chính là một hạn chế của kiến trúc mạng này bởi nó sẽ gặp phải vấn đề scale variant rất lớn khi resize các ảnh có kích thước khác nhau về cùng một kích thước cố định ở cả chiều rộng và chiều cao của ảnh.

Khoảng 90% tập dữ liệu VLP là biển trắng, số còn lại là biển xanh, vàng, đỏ do biển trắng là biển số xe dành cho cá nhân nên phổ biến hơn các biển dành cho các tổ chức, đơn vị kinh doanh. Một câu hỏi đặt ra là liệu mô hình có đủ tổng quát và hiệu năng có bị ảnh hưởng nhiều khi kiểm thử trên tập dữ liệu có nhiều biển màu xanh, vàng, đỏ hơn? Xác định vấn đề trên, chúng tôi sử dụng thuật toán image negative và histogram matching để tạo ra bộ dữ liệu các biển màu xanh, đỏ, vàng từ biển trắng trên cả 3 tập training, testing và validation. Chúng tôi không huấn luyện lại mà sử dụng mô hình Vocr đã huấn luyện ở tập dữ liệu Vocr gốc, để kiểm thử trên tập dữ liệu vừa sinh. Kết quả cho thấy, độ chính xác chỉ giảm chưa tới 1% ở tất cả các tập. Chúng tôi nhận thấy, hầu hết các ảnh có màu sắc xanh đỏ vàng sai là những ảnh xanh đỏ vàng đã sai trong tập gốc, hoặc là những ảnh do quá trình sinh dữ liệu khiến chữ không rõ ràng. Do đó, có thể khẳng định sự thay đổi màu sắc của biển số có ảnh hưởng rất

Bảng 6: Thực nghiệm sự ảnh hưởng màu sắc biển số tới Vocr

Tập dữ liệu	Color plate		Original	
	Acc_{seq}	Acc_{char}	Acc_{seq}	Acc_{char}
Training set	98.8%	99.6%	99.34%	99.8%
Validation set	98.7%	99.7%	99.25%	99.9%
Testing set	98.5%	99.5%	99.28%	99.7%

nhỏ tới kết quả của nhận dạng biển số.

6.3 Hiệu năng chung của hệ thống

6.3.1 Phần cứng triển khai

6.3.2 Thời gian phản hồi

6.3.3 Khả năng chịu tải

7 Kết luận và kiến nghị

Trong công trình này, chúng tôi đã xây dựng bộ dữ liệu VLP nhằm đánh giá hiệu năng của phương pháp đề xuất trong phát hiện biển số và nhận diện biển số xe ô tô Việt Nam. Giải quyết vấn đề non-rectangular object detection, chúng tôi đề xuất hướng tiếp dựa trên keypoints detection lấy ý tưởng từ bài toán human pose estimation. Chỉnh sửa kiến trúc xuất phát từ bài toán segmentation, DDRNet23sh cho kết quả phát hiện biển số tốt nhất với $mIOU = 95\%$, $Precision_{75} = 99.2\%$. Chúng tôi đề xuất nhận diện biển số segmentation free sử dụng kiến trúc CNN kết hợp với encoder-decoder RNN và attention mechanism. Kết quả nhận diện biển số sử dụng Vocr đạt $Acc_{seq} = 98.36\%$, $Acc_{char} = 99.65\%$. Các kết quả trên cho thấy tiềm năng của cách tiếp cận của chúng tôi. Ngoài ra, chúng tôi xây dựng *hệ thống giám sát và truy vết ô tô trên diện rộng* hoàn chỉnh tích hợp thêm công nghệ bản đồ số, thuật toán motion detection, tracking và ocr-check nâng cao hiệu năng của hệ thống trong triển khai thực tế. Trong công trình này, chúng tôi chỉ mới tập trung vào điều kiện ban ngày. Công việc tương lai, chúng tôi sẽ giải quyết bài toán nhận diện biển số xe vào cả ban đêm và mở rộng bài toán đối với giám sát và truy vết xe máy Vietnam.

References

- [1] R. Laroca et al. “A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector”. In: (2018).
- [2] A. M. Al-Ghaili et al. “A new vertical edge detection algorithm and its application”. In: (2008).
- [3] M. Sarfraz, M. J. Ahmed, and S. A. Ghazi. “Saudi Arabian license plate recognition system”. In: (2003).
- [4] Y. Wen et al. “An Algorithm for License Plate Recognition Applied to Intelligent Transportation System”. In: *IEEE Transactions on Intelligent Transportation Systems* 3 (2011).
- [5] 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).

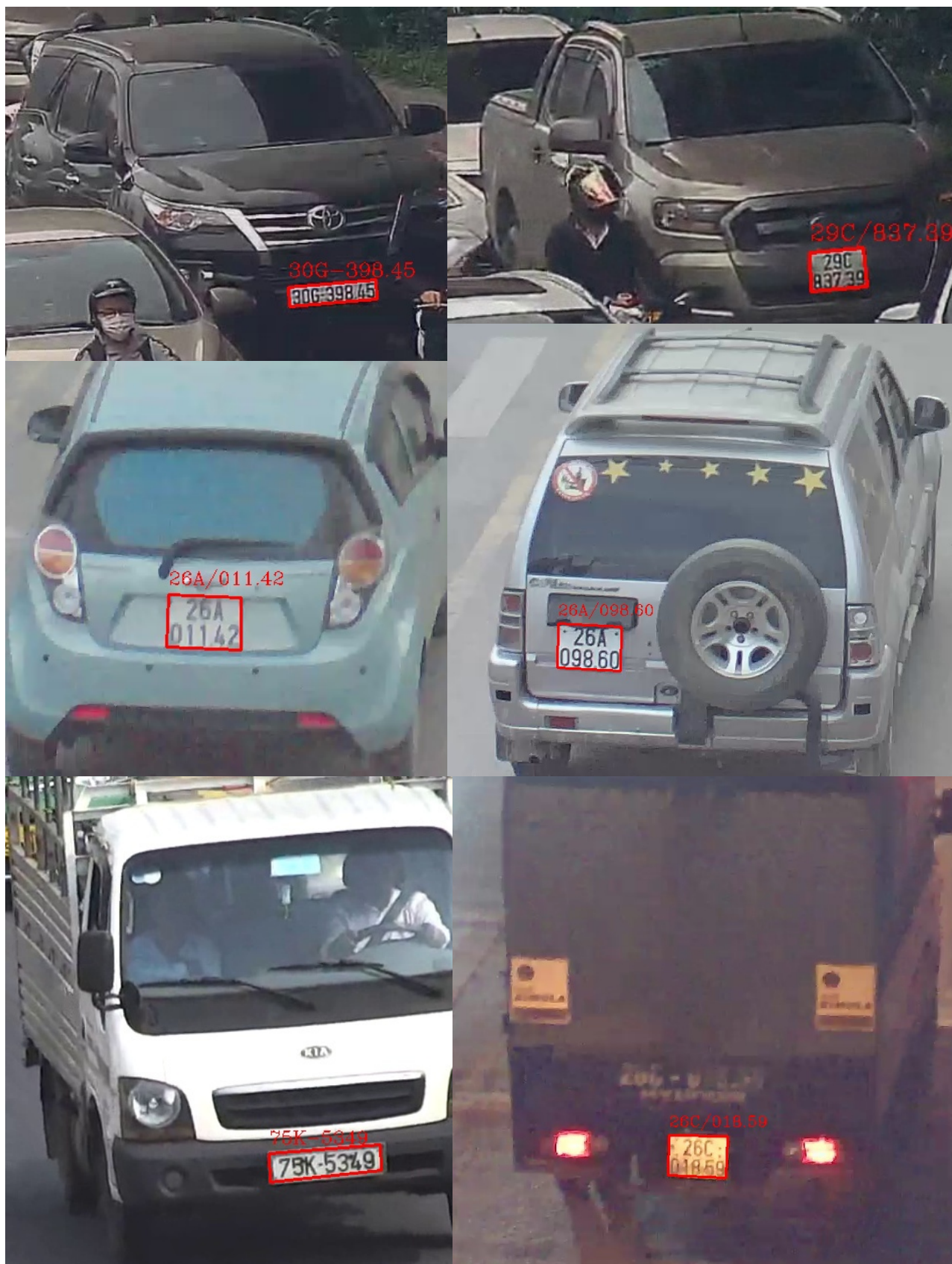
- [6] Christos-Nikolaos Anagnostopoulos et al. “License Plate Recognition From Still Images and Video Sequences: A Survey”. In: *Intelligent Transportation Systems, IEEE Transactions on* (Oct. 2008).
- [7] S. Du et al. “Automatic License Plate Recognition (ALPR): A State-of-the-Art Review”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 2 (2013).
- [8] Yuanduo Hong et al. “Deep Dual-resolution Networks for Real-time and Accurate Semantic Segmentation of Road Scenes”. In: *arXiv preprint arXiv:2101.06085* (2021).
- [9] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. “Effective Approaches to Attention-based Neural Machine Translation”. In: *CoRR* (2015).
- [10] Alex Graves et al. “Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks”. In: Jan. 2006.
- [11] Junyoung Chung et al. “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling”. In: *CoRR* (2014).
- [12] Karen Simonyan and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *arXiv 1409.1556* (Sept. 2014).
- [13] Danian Zheng, Yan Zhao, and Jiaxin Wang. “An efficient method of license plate location”. In: *Pattern Recognition Letters* (Nov. 2005).
- [14] S. Yohimori et al. “License plate detection system by using threshold function and improved template matching method”. In: (July 2004).
- [15] Dong-Su Kim and Sung-Il Chien. “Automatic car license plate extraction using modified generalized symmetry transform and image warping”. In: (Feb. 2001).
- [16] 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).
- [17] Joseph Redmon et al. “You Only Look Once: Unified, Real-Time Object Detection”. In: (June 2016).
- [18] Joseph Redmon and Ali Farhadi. “YOLO9000: Better, Faster, Stronger”. In: (July 2017).
- [19] Gee-Sern Hsu et al. “Robust license plate detection in the wild”. In: (Aug. 2017).
- [20] 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).

- [21] Max Jaderberg et al. “Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition”. In: (June 2014).
- [22] Andrea Vedaldi and Andrew Zisserman. “Synthetic Data for Text Localisation in Natural Images”. In: June 2016.
- [23] 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).
- [24] Anuja Nagare and Shalini Bhatia. “License Plate Character Recognition System using Neural Network”. In: *International Journal of Computer Applications* (July 2011).
- [25] Ihor Paliy et al. “Approach to recognition of license plate numbers using neural networks”. In: (Aug. 2004).
- [26] C.A. Rahman, Wael Badawy, and Ahmad Radmanesh. “A real time vehicle’s license plate recognition system”. In: (Aug. 2003).
- [27] Hui Li et al. “Reading Car License Plates Using Deep Neural Networks”. In: *Image and Vision Computing* (Mar. 2018).
- [28] Zied Selmi, Mohamed Ben Halima, and Adel Alimi. “Deep Learning System for Automatic License Plate Detection and Recognition”. In: (Nov. 2017).
- [29] 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).
- [30] Teik Cheang, Yong Shean Chong, and Yong Haur Tay. “Segmentation-free Vehicle License Plate Recognition using ConvNet-RNN”. In: (Jan. 2017).
- [31] 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).
- [32] Mark Everingham et al. “The Pascal Visual Object Classes (VOC) Challenge”. In: *International Journal of Computer Vision* 2 (June 2010). ISSN: 1573-1405.
- [33] Olga Russakovsky et al. “ImageNet Large Scale Visual Recognition Challenge”. In: *International Journal of Computer Vision* (Sept. 2014).
- [34] Tsung-Yi Lin et al. *Microsoft COCO: Common Objects in Context*. 2015.
- [35] Feng Zhang et al. “Distribution-Aware Coordinate Representation for Human Pose Estimation”. In: *CoRR* abs/1910.06278 (2019). arXiv: 1910.06278. URL: <http://arxiv.org/abs/1910.06278>.

- [36] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. “Objects as Points”. In: *CoRR* abs/1904.07850 (2019). arXiv: 1904.07850. URL: <http://arxiv.org/abs/1904.07850>.
- [37] Feng Zhang et al. *Distribution-Aware Coordinate Representation for Human Pose Estimation*. 2019. arXiv: 1910.06278 [cs.CV].
- [38] Fisher Yu et al. *Deep Layer Aggregation*. 2019. arXiv: 1707.06484 [cs.CV].
- [39] Bin Xiao, Haiping Wu, and Yichen Wei. *Simple Baselines for Human Pose Estimation and Tracking*. 2018. arXiv: 1804.06208 [cs.CV].
- [40] Alejandro Newell, Kaiyu Yang, and Jia Deng. *Stacked Hourglass Networks for Human Pose Estimation*. 2016. arXiv: 1603.06937 [cs.CV].
- [41] Sergey Ioffe and Christian Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: *CoRR* (2015).
- [42] Dzmitry Bahdanau, Kyunghyun Cho, and Y. Bengio. “Neural Machine Translation by Jointly Learning to Align and Translate”. In: *ArXiv* (Sept. 2014).
- [43] Alex Lamb et al. *Professor Forcing: A New Algorithm for Training Recurrent Networks*. 2016.
- [44] Rafael Müller, Simon Kornblith, and Geoffrey E. Hinton. “When Does Label Smoothing Help?” In: *CoRR* (2019).
- [45] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017.
- [46] Vinh Mai, Duoqian Miao, and Ruizhi Wang. “Building a license plate recognition system for Vietnam tollbooth”. In: *ACM International Conference Proceeding Series* (Aug. 2012).
- [47] Tran Duan et al. “Building an Automatic Vehicle License-Plate Recognition System”. In: *Proc. Int. Conf. Comput. Sci. RIVF* (Feb. 2005).
- [48] <https://thigiactmaytinh.com/tai-nguyen-xu-ly-anh/tong-hop-data-xu-ly-anh/>. 2014.
- [49] Jing Han et al. “Multi-Oriented and Scale-Invariant License Plate Detection Based on Convolutional Neural Networks”. In: *Sensors* 19 (Mar. 2019), p. 1175. DOI: 10.3390/s19051175.
- [50] Sergey Zherzdev and Alexey Gruzdev. “LPRNet: License Plate Recognition via Deep Neural Networks”. In: (June 2018).
- [51] Baoguang Shi, Xiang Bai, and Cong Yao. *An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition*. 2015. arXiv: 1507.05717 [cs.CV].

- [52] Sérgio Montazzoli and Claudio Jung. “License Plate Detection and Recognition in Unconstrained Scenarios”. In: (Sept. 2018).

8 Phụ lục



Hình 12: Kết quả của hệ thống sử dụng DDRNet23sh và Vocr.



Hình 13: Kết quả của hệ thống sử dụng DDRNet23sh và Vocr.