Đại học Bách Khoa Hà Nội

**Trường Công nghệ thông tin và Truyền thông**

Trịnh Quyết Tiến

**Nghiên cứu đánh giá và tối ưu phương pháp nhận dạng biển số xe tự động trên hệ thống nhúng**

Chuyên ngành: Khoa học máy tính

Mã đề tài : 2021KHDL-E12

Luận văn thạc sĩ

Hà nội tháng 10 năm 2023

**LỜI CẢM ƠN**

**TÓM TẮT**

Phân biệt biển số xe tự động (ALPR - automatic license plate recognition) là một thách thức thu hút sự chú ý của nhiều nhà nghiên cứu do ứng dụng rộng rãi trong thực tế. Nhiều nghiên cứu đã thử nghiệm nhiều phương pháp khác nhau và đã đạt được kết quả ấn tượng, cho phép tích hợp nó vào các ứng dụng cụ thể. Tuy nhiên, tùy thuộc vào bối cảnh ứng dụng, xuất phát từ những yêu cầu cụ thể, đòi hỏi về độ chính xác và khả năng xử lý của hệ thống ALPR có thể biến đổi. Chẳng hạn, trong trường hợp nhận dạng biển số cho các phương tiện ra vào bãi đỗ xe có điều kiện đơn giản (xe dừng yên trong khung hình, thời gian xử lý không cần quá nhanh), hoặc nhận dạng biển số cho các phương tiện di chuyển trên cao tốc hoặc đường phố để sử dụng trong việc xử lý vi phạm và kiểm soát giao thông, điều này đòi hỏi ngữ cảnh phức tạp hơn (xe di chuyển nhanh, khoảng cách xa). Hơn nữa, sự gia tăng của tính năng xử lý trên thiết bị cuối (edge computing) cũng đã đánh dấu một xu hướng quan trọng trong triển khai hệ thống ALPR, giúp tối ưu hóa hiệu suất và hiệu quả của nó. Với sự phát triển này, việc lựa chọn phương pháp nhận dạng biển số xe đúng để phù hợp với các tình huống ứng dụng cụ thể đã trở nên càng quan trọng hơn. Nghiên cứu này tập trung vào việc đánh giá các phương pháp nhận dạng biển số xe và mục tiêu cuối cùng là tối ưu hóa chúng để triển khai trên hệ thống nhúng.

Các phương pháp nhận dạng biển số xe thực hiện tùy thuộc vào điều kiện ngữ cảnh, và các yếu tố như tốc độ cao của phương tiện, khoảng cách xa, hoặc hình ảnh không rõ gây khó khăn cho việc xác định và nhận dạng biển số. Điều này có thể dẫn đến sự giảm sút độ chính xác và yêu cầu việc liên tục đánh giá, cải tiến để phù hợp với các ứng dụng cụ thể. Triển khai xử lý nhận dạng biển số xe trên thiết bị nhúng có lợi thế là không cần truyền dữ liệu hình ảnh hoặc video lớn đến máy chủ, giảm bớt gánh nặng xử lý tại máy chủ. Tuy nhiên, thực hiện trên thiết bị nhúng đòi hỏi sự đánh giá kỹ lưỡng về hiệu năng và thời gian xử lý. Để giải quyết các thách thức nêu trên, cần thiết phải thực hiện sự đánh giá cặn kẽ của các phương pháp nhận dạng biển số xe và tối ưu hóa chúng để đáp ứng yêu cầu về hiệu năng và độ chính xác trên hệ thống nhúng. Sử dụng hệ thống edge computing có thể giải quyết nhiều vấn đề bởi vì nó cho phép xử lý dữ liệu tại nguồn mà không cần truyền dữ liệu đến một trung tâm xử lý tập trung ở xa. Kết quả là giảm thiểu độ trễ và tăng tốc độ xử lý, đồng thời giảm tải băng thông mạng và chi phí lưu trữ dữ liệu. Hơn nữa, hệ thống edge computing có khả năng hoạt động độc lập và đảm bảo tính sẵn sàng cao, ngay cả khi kết nối mạng bị gián đoạn. Do đó, việc áp dụng hệ thống edge computing vào bài toán nhận dạng biển số xe sẽ cải thiện hiệu suất và độ tin cậy của hệ thống.

**Lời cam đoan**

**MỤC LỤC**

**Danh mục hình vẽ**

**Danh mục bảng biểu**

**Danh mục từ viết tắt**

**CHƯƠNG 1:** **GIỚI THỆU**

* 1. **Giới thiệu đề tài**

Với sự phát triển nhanh chóng của kinh tế và xã hội, nước ta đang đối mặt với nhiều thách thức phức tạp, trong đó tình trạng giao thông quá tải và hỗn loạn đang là vấn đề nghiêm trọng tại các trung tâm kinh tế lớn của cả nước. Những tác động tiêu cực của vấn đề này đang gây ra những thiệt hại lớn cho nền kinh tế cũng như đời sống xã hội. Để giải quyết vấn đề này, cần có sự nâng cao ý thức và chấp hành giao thông của người dân, đồng thời cần thiết phải tập trung vào việc giám sát và quản lý giao thông.

Trong đó, bài toán nhận diện biển số xe là một trong những bài toán quan trọng để giúp quản lý phương tiện giao thông một cách hiệu quả. Tuy nhiên, để giải quyết bài toán này cần phải áp dụng các công nghệ mới như trí tuệ nhân tạo, học sâu, xử lý ảnh, và có những phương pháp đánh giá và xử lý dữ liệu để đạt được độ chính xác cao. Vì vậy, việc nghiên cứu và giải quyết bài toán này là cực kỳ cần thiết để đóng góp vào giải quyết vấn đề giao thông chung hiện nay.

Vấn đề phát hiện và nhận diện biển số xe là một trong những hướng nghiên cứu đã được quan tâm trong lĩnh vực thị giác máy tính từ lâu. Tuy nhiên, hiện nay với sự phát triển mạnh mẽ của phương pháp học sâu (deep learning), đã mở ra một hướng tiếp cận mới cho vấn đề này. Các mạng nơ-ron trong học sâu như Convolution Neural Networks (CNN), Recurrent Neural Networks (RNN) cùng với các kiến trúc mạng như LeNet, ImageNet, Fast R-CNN, ... đang được phổ biến và được sử dụng trong nhiều ứng dụng thực tế. Đặc biệt, các kết quả đạt được thông qua các cuộc thi và nghiên cứu trong giới học thuật đều rất ấn tượng. Những phương pháp này giúp cho việc phát hiện và nhận diện biển số xe trở nên chính xác hơn, nhanh hơn và hiệu quả hơn. Việc áp dụng phương pháp học sâu này không chỉ giải quyết được bài toán giám sát giao thông tự động mà còn có thể ứng dụng trong nhiều lĩnh vực khác.

Các phương pháp nhận dạng biển số xe có khả năng ứng dụng thực tiễn khác nhau phụ thuộc vào điều kiện ngữ cảnh đòi hỏi. Các yếu tố làm tăng độ phức tạp, như phương tiện di chuyển nhanh, khoảng cách xa, ảnh kém chất lượng, … gây ra khó khăn cho việc phát hiện và nhận dạng biển số, làm giảm độ chính xác, dẫn đến cần phân tích đánh giá, cải tiến để phù hợp với ứng dụng.

Việc đưa các xử lý của tác vụ nhận dạng biển số xe trên thiết bị nhúng giúp không phải truyền nhận giữ liệu hình ảnh, video lớn về server, giảm tải xử lý phía server. Tuy nhiên việc thực hiện trên hệ nhúng đòi hỏi những đánh giá, đáp ứng về hiệu năng, thời gian xử lý.

Với những vấn đề nêu trên, cần có sự đánh giá các phương pháp nhận dạng biển số xe, tối ưu cho phù hợp về hiệu năng, độ chính xác đáp ứng được trên hệ thống nhúng.

Sử dụng hệ thống edge computing có thể cải thiện các yếu tố nêu trên bởi vì nó cho phép xử lý dữ liệu cục bộ tại nguồn phát sinh mà không phải truyền dữ liệu tới một trung tâm xử lý tập trung (centralized processing center) ở xa. Kết quả là giảm thiểu độ trễ và tăng tốc độ xử lý, đồng thời giảm tải băng thông mạng và chi phí lưu trữ dữ liệu. Hơn nữa, hệ thống edge computing có thể hoạt động độc lập và đảm bảo tính sẵn sàng cao, ngay cả khi kết nối mạng bị gián đoạn. Từ đó, việc áp dụng hệ thống edge computing vào bài toán nhận dạng biển số xe sẽ giúp cải thiện hiệu suất và độ tin cậy của hệ thống.

* 1. **Mục tiêu đề tài**

Đề tài luận văn tập trung vào việc đánh giá và cải thiện phương pháp nhận dạng biển số xe tự động trên hệ thống nhúng. Đề tài có 3 mục tiêu chính:

* Nghiên cứu, tìm hiểu các phương pháp nhận dạng biển số xe hiện có.
* Thực nghiệm, đánh giá phương pháp nhận dạng biển số xe với một số tập dữ liệu biển số xe Việt Nam để đưa ra một mô hình nhận dạng biển số xe tối ưu (về hiệu năng, độ chính xác).
* Triển khai đánh giá mô hình nhận dạng trên một thiết bị nhúng (Raspberry Pi 4)
  1. **Ý nghĩa của đề tài**
     1. ***Ý nghĩa thực tiễn***

Phương pháp được đề xuất trong đề tài có thể áp dụng trong nhiều ứng dụng, như giám sát giao thông tự động, bãi giữ xe thông minh, trạm thu phí tự động và nhiều ứng dụng khác. Sử dụng phương pháp này có thể giúp giảm thời gian và công sức của con người, đồng thời giảm tình trạng kẹt xe và tăng sự tiện lợi cho người tham gia giao thông. Một trong những yếu tố quan trọng để đạt được hiệu quả cao của phương pháp là xây dựng được tập dữ liệu chính xác, đa dạng và phù hợp với các điều kiện thực tế tại Việt Nam. Các ứng dụng được phát triển từ phương pháp này sẽ mang lại lợi ích cho cả xã hội và góp phần nâng cao chất lượng cuộc sống của người dân.

* + 1. ***Ý nghĩa khoa học***

Đề tài đóng góp một phương pháp mới trong việc phát hiện và nhận diện biển số xe hiệu quả, với mục tiêu tăng độ chính xác và giải quyết các vấn đề còn tồn tại của các phương pháp đi trước. Nghiên cứu này có thể cung cấp cơ sở cho các nghiên cứu trong nước về sau, đặc biệt là trong lĩnh vực ứng dụng các hệ thống nhúng. Tập dữ liệu chính xác, đa dạng và sát với điều kiện thực tế của nước ta cũng được xem là một đóng góp đáng kể của nghiên cứu này.

* + 1. ***Phạm vi của đề tài***

Trong đề tài, tập dữ liệu được sử dụng để huấn luyện và kiểm tra mô hình chỉ bao gồm các hình ảnh biển số xe được chụp theo phương ngang, không bị che khuất, không bị hư hỏng, tróc sơn, rỉ sét và có độ mờ và độ biến dạng thấp. Vị trí của camera để thu thập hình ảnh là cố định hoặc di động với góc lệch không quá lớn. Hình ảnh có thể bị nhiễu từ thời tiết như mưa, sương mù, vv., nhưng vẫn có thể nhận dạng được đầy đủ các ký tự bằng mắt thường. Trong đề tài, chỉ sử dụng các biển số xe máy và ô tô phổ biến tại Việt Nam, tuân thủ theo Thông tư 24/2023/TT-BCA về kích thước, ký hiệu và bố trí của biển số xe. Các biển số xe của các nước khác sẽ không được xem xét trong đề tài.

* 1. **Phương pháp nghiên cứu**

Đề tài xoay quanh việc nghiên cứu, phát triển hệ thống nhận dạng biển số xe trên các thiết bị nhúng và dựa trên dàn ý sau:

* *Nghiên cứu khảo sát các phương pháp nhận dạng biển số xe*: Trước tiên, Đề tài sẽ tiến hành một khảo sát về các phương pháp đã được phát triển để nhận dạng biển số xe. Điều này bao gồm việc tìm hiểu về các phương pháp truyền thống cũng như các phương pháp hiện đại dựa trên học máy và trí tuệ nhân tạo. Đề tài sẽ xem xét các điểm mạnh và yếu của từng phương pháp, đánh giá khả năng áp dụng của chúng trong các tình huống ứng dụng cụ thể và nắm vững kiến thức về các tiến bộ trong lĩnh vực này.
* *Đề xuất 1 phương pháp tốt nhất để thực nghiệm, đánh giá*: Dựa trên kết quả của khảo sát, đề tài sẽ lựa chọn hoặc đề xuất một phương pháp nhận dạng biển số xe mà được xem xét là tốt nhất cho việc thực nghiệm và đánh giá. Lựa chọn này sẽ dựa trên nhiều yếu tố, bao gồm độ chính xác, hiệu năng xử lý, khả năng tùy chỉnh, và tính ứng dụng trong các tình huống khác nhau. Đề tài cũng sẽ trình bày lý do tại sao phương pháp được chọn hoặc đề xuất được coi là lựa chọn tốt nhất dựa trên mục tiêu của đề tài.
* *Triển khai trên hệ thống nhúng, tối ưu về hiệu năng:* Sau khi chọn được phương pháp, đề tài sẽ tiến hành triển khai nó trên hệ thống nhúng. Điều này bao gồm việc tối ưu hóa phương pháp để đảm bảo hiệu năng tốt nhất trên các thiết bị nhúng có tài nguyên hạn chế. Đề tài sẽ xem xét các yếu tố như tài nguyên xử lý, bộ nhớ, và khả năng tích hợp vào các ứng dụng thực tế. Đồng thời cũng sẽ đánh giá hiệu suất của hệ thống trong việc nhận dạng biển số xe trong các tình huống ngữ cảnh khác nhau, như đã thảo luận trong phần trước.

Tóm lại, đề tài sẽ bao gồm việc nghiên cứu, lựa chọn, và triển khai một phương pháp nhận dạng biển số xe trên hệ thống nhúng, tối ưu hóa nó về hiệu năng để đáp ứng được các yêu cầu của các ứng dụng thực tế.

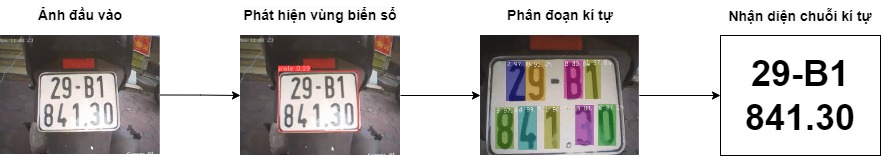
* 1. **Bố cục luận văn**

Báo cáo này gồm 4 chương, trong đó Chương 2 tập trung vào giới thiệu các nghiên cứu liên quan đến bài toán phát hiện và nhận dạng biển số xe, cách xây dựng bộ dữ liệu biển số xe, cùng với các phương pháp giải quyết được áp dụng trong lĩnh vực này, đồng thời cung cấp những kiến thức cơ bản để đặt nền móng cho các mô hình được đề xuất. Chương 3 bao gồm thông tin về các thí nghiệm thực hiện trên hệ thống nhúng, các chỉ tiêu đánh giá và kết quả đánh giá của các phương pháp, đồng thời cũng đưa ra một số nhận xét về hiệu năng của hệ thống. Chương 4 đưa ra kết luận chung của báo cáo và trình bày những hướng phát triển tiềm năng trong tương lai, kèm theo danh sách các tài liệu tham khảo được sử dụng trong đề tài.

**CHƯƠNG 2: NGHIÊN CỨU PHƯƠNG PHÁP NHẬN DẠNG BIỂN SỐ XE**

* 1. **Khảo sát các phương pháp nhận dạng biển số xe**
     1. ***Mô hình phương pháp nhận dạng biển số xe tổng quát***

Bài toán nhận dạng biển số xe là một trong những bài toán quan trọng trong lĩnh vực nhận diện văn bản ngoại cảnh. Vì vậy, các nghiên cứu liên quan đến hai bài toán này thường có sự tương quan mật thiết. Trong những năm gần đây, bài toán nhận diện biển số xe đã được quan tâm nhiều và có sự phát triển đáng kể. Thông thường, bài toán nhận dạng biển số xe bao gồm một số bước cơ bản như sau:



*Các bước cơ bản của một hệ thống nhận diện biển số xe*

Từ những bước trên, có ba bài toán chính cần được giải quyết trong việc nhận dạng biển số xe: phát hiện vị trí của biển số trên hình, phân đoạn các ký tự trong biển số và nhận diện chuỗi ký tự trên biển số. Dựa trên những nghiên cứu mới đây, đề tài nhận thấy rằng hướng tiếp cận nhận dạng chuỗi ký tự trên biển số mà không phải qua bước phân đoạn ký tự rất tiềm năng, vì có nhiều ưu điểm hơn. Do đó, đề tài tập trung vào khảo sát các công trình liên quan đến phát hiện vị trí của biển số và nhận diện ký tự trên biển số. Ngoài ra, trong số các nghiên cứu gần đây còn đề cập đến việc kết hợp xử lý cả hai quá trình phát hiện và nhận dạng một cách đồng thời, tạo ra một phương pháp tổng thể hiệu quả hơn để giải quyết bài toán này.

Ngoài ra độ chính xác của việc nhận dạng biển số xe còn phụ thuộc đặc điểm của dữ liệu đầu vào, các ảnh biển số xe thu thập từ nhiều tình huống và điều kiện khác nhau. Các đặc điểm quan trọng của dữ liệu bao gồm:

* Tình huống điều kiện qui định: Đây là ảnh chụp trong các điều kiện thông thường như ra vào bãi đỗ xe, ban ngày, ánh sáng tự nhiên tốt.
* Tình huống điều kiện phức tạp: Điều này bao gồm ảnh chụp từ xa, trong điều kiện thiếu sáng, hoặc trong các tình huống khó khăn khác nhau như mưa, tuyết, hay bị che khuất bởi các vật thể khác.

Dữ liệu đa dạng này giúp mô hình trở nên mạnh mẽ và có khả năng nhận dạng biển số xe trong nhiều tình huống thực tế khác nhau.

* + 1. ***Xác định vùng biển số xe trong ảnh***
       1. *Các phương pháp trích xuất đặc trưng thủ công bằng xử lý ảnh*

Trong quá khứ, các phương pháp trích đặc trưng thủ công được sử dụng phổ biến trong việc phân tích ảnh. Những đặc trưng thường được sử dụng bao gồm các đặc trưng cơ bản như góc, cạnh, vân ảnh, màu sắc và mức sáng. Các phương pháp phổ biến để trích xuất các đặc trưng này bao gồm việc nhị phân hóa ảnh, sử dụng thuật toán SIFT, HOG, phân tích thành phần liên thông và hình thái học. Sau đó, các đặc trưng này được đưa qua các bộ phân loại như AdaBoost, SVM để phân loại.

Các phương pháp trích xuất đặc trưng cơ bản thường đơn giản, dễ hiện thực và nhanh chóng, đặc biệt là khi kết hợp nhiều đặc trưng với nhau, kết quả có thể đạt được độ chính xác tương đối cao. Tuy nhiên, các phương pháp này có nhược điểm là khá nhạy cảm với sự thay đổi mức sáng, nhiễu và mờ. Hơn nữa, việc thiết kế bộ rút trích đặc trưng còn phụ thuộc khá nhiều vào tập dữ liệu, do đó cần phải đảm bảo tập dữ liệu đủ đa dạng để đảm bảo tính chính xác của phương pháp trích xuất.

Dưới đây là một số phương pháp và bài báo sử dụng phương pháp trích xuất đặc trưng trong lĩnh vực xử lý hình ảnh:

*A, Trích xuất biển số xe bằng cách sử dụng thông tin đường biên (License Plate Extraction using Boundary/Edge Information)*

Các phương pháp này bao gồm sử dụng bộ lọc Sobel để phát hiện các cạnh, kết hợp cạnh dọc để tạo ra các hình chữ nhật biển số, và sử dụng khối để xác định các vùng biển số ứng viên. Các phương pháp tương tự bao gồm sử dụng biến đổi Hough để phát hiện các đường thẳng và sử dụng biến đổi đối xứng tổng quát (GST) để trích xuất biển số xe. Các phương pháp trích xuất biển số xe bằng cách sử dụng thông tin đường biên là các phương pháp đơn giản và nhanh chóng. Tuy nhiên, chúng đòi hỏi tính liên tục của các cạnh. Khi kết hợp với các bước hình thái học để loại bỏ các cạnh không mong muốn, tỷ lệ trích xuất tương đối cao.

* ***Một số bài báo sử dụng phương pháp:***
* H. Bai and C. Liu, “A hybrid license plate extraction method based on edge statistics and morphology,” Int. Conf. Patt. Recog., vol. 2, pp. 831-834, 2004.
* D. Zheng et al, “An efficient method of license plate location,” Pattern recog. Letter, vol. 26, no. 15, pp. 2431-2438, 2005.
* S. Wang and H. Lee, “Detection and recognition of license plate characters with different appearances,” Int. Conf. Intell. Transp. Syst., vol. 2, pp. 979- 984, 2003.
* F. Faradji et al, “A morphological-based license plate location,” IEEE Int. Conf. Image Process., vol. 1, pp. 57-60, 2007.
* ***Ưu điểm:***
* Chính xác trong trích xuất biển số: Thông tin biên và đường biên có thể cung cấp các dấu hiệu mạnh mẽ về vị trí của biển số xe, giúp tăng độ chính xác trong việc trích xuất biển số.
* Kháng nhiễu tốt: Các thuật toán trích xuất biên thường kháng nhiễu tốt hơn so với các phương pháp dựa trên màu sắc, giúp loại bỏ các yếu tố gây nhiễu như ánh sáng yếu hoặc thay đổi màu sắc.
* Thích hợp cho các biển số có độ tương phản thấp: Khi biển số có độ tương phản thấp hoặc bị che khuất, thông tin biên có thể cung cấp dấu hiệu quan trọng để xác định vị trí của biển số.
* ***Nhược điểm:***
* Nhạy cảm với biến đổi ánh sáng: Phương pháp sử dụng biên thường nhạy cảm với biến đổi ánh sáng, đặc biệt là khi có sự thay đổi về cường độ ánh sáng hoặc điều kiện chiếu sáng không đồng đều.
* Khó xử lý trong các trường hợp biển số bị che khuất: Trong trường hợp biển số bị che khuất bởi các vật thể khác, việc trích xuất dựa trên biên có thể gặp khó khăn.
* Yêu cầu tính toán cao: Các thuật toán xử lý biên thường đòi hỏi tính toán phức tạp hơn so với một số phương pháp khác, đặc biệt là khi xử lý hình ảnh/video thời gian thực.
* Không thể xác định ký tự trực tiếp: Phương pháp này chỉ giúp trích xuất vị trí của biển số mà không thể xác định các ký tự trực tiếp. Sau khi trích xuất vị trí, cần một bước tiếp theo để nhận dạng ký tự.

Tổng quan, phương pháp trích xuất biển số xe sử dụng thông tin biên/đường biên có thể rất hữu ích trong nhiều tình huống, nhất là khi biển số xe có độ tương phản thấp hoặc trong điều kiện ánh sáng biến đổi. Tuy nhiên, nó cũng có nhược điểm và cần được kết hợp với các phương pháp khác để đảm bảo độ chính xác và hiệu quả trong ứng dụng thực tế.

*B, Trích xuất biển số xe bằng cách sử dụng Global Image Information*

Việc sử dụng phân tích thành phần kết nối (CCA) trong xử lý ảnh nhị phân để trích xuất biển số xe sẽ quét hình ảnh và gán nhãn cho các pixel dựa trên sự kết nối của chúng. Phương pháp này sử dụng các đặc điểm không gian như diện tích và tỷ lệ khía cạnh để trích xuất biển số xe. Ngoài ra, nó cũng có thể sử dụng đến các thuật toán phát hiện đường viền và tương quan 2D để tìm biển số xe. Tuy nhiên, một số phương pháp có thể gặp khó khăn trong trường hợp hình ảnh chất lượng kém hoặc đòi hỏi tính toán tốn thời gian.

* ***Một số bài báo sử dụng phương pháp:***
  + Z. Qin et al, “Method of license plate location based on corner feature,” in Proc. World Congr. Intell. Control Automat., vol. 2, pp. 8645-8649
  + J. Matas et al, “Unconstrained license plate and text loc recognition,” IEEE Int. Conf. intell. Transp. Syst., pp. 225-230
  + B.-F. Wu et al, “Extracting characters from real vehicle license doors,” IET Comput. Vis., vol. 1, no. 1, pp. 2-10, 2007.
  + M. M. I. Chacon and S. A. Zimmerman, “License plate locatio dynamic PCNN scheme,” Int. Joint Conf. Neur. Netw., vol. 2, p 2003.
* ***Ưu điểm***:
  + Khả năng tổng hợp thông tin global: Phương pháp này sử dụng thông tin global của hình ảnh, bao gồm cả cấu trúc và ngữ cảnh của bức tranh, để xác định vị trí của biển số xe. Điều này có thể giúp cải thiện độ chính xác của việc trích xuất.
  + Kháng nhiễu tốt: Bằng cách xem xét thông tin global, phương pháp này có thể giảm thiểu ảnh hưởng của nhiễu và sự biến đổi về ánh sáng lên việc trích xuất biển số xe.
  + Độ linh hoạt trong ứng dụng: Phương pháp này thường áp dụng một loạt các kỹ thuật xử lý hình ảnh và máy học sâu để phát hiện và trích xuất biển số xe, do đó nó có thể được điều chỉnh và tùy chỉnh cho các tình huống ứng dụng cụ thể.
* ***Nhược điểm***:
  + Yêu cầu tính toán cao: Xử lý thông tin toàn cầu của hình ảnh thường đòi hỏi nhiều tính toán, đặc biệt là khi áp dụng các thuật toán phức tạp như máy học sâu. Điều này có thể làm tăng thời gian xử lý.
  + Phụ thuộc vào chất lượng hình ảnh: Phương pháp này có thể không hiệu quả nếu hình ảnh có chất lượng kém hoặc nhiễu mạnh. Các biến đổi ánh sáng và nhiễu có thể làm giảm độ chính xác của việc trích xuất.
  + Khó khăn trong các tình huống phức tạp: Trong các tình huống phức tạp, chẳng hạn như khi có nhiều biển số xe trong cùng một khung hình hoặc các vật thể che khuất, phương pháp này có thể gặp khó khăn trong việc xác định đúng vị trí của biển số xe.

Tổng quan, "License Plate Extraction using Global Image Information" là một phương pháp tiềm năng để trích xuất biển số xe, nhưng nó cũng đối diện với các thách thức về tính toán và độ chính xác, đặc biệt trong điều kiện hình ảnh có chất lượng thấp.

*C, Trích xuất biển số xe bằng các đặc điểm về cấu trúc*

Các phương pháp trích xuất biển số xe bằng các đặc điểm về cấu trúc dựa vào sự thay đổi mức xám và cấu trúc của văn bản trên biển số xe. Dưới đây là tóm tắt các phương pháp nổi bật:

* Dò quét dòng (Scan-line techniques): Các phương pháp này dựa vào sự thay đổi mức xám để xác định số ký tự trên biển số xe. Khi sự thay đổi này xuất hiện trên dòng quét, số ký tự có thể được xác định.
  + Paper tham khảo: F. Yang and Z. Ma, “Vehicle license plate location based on h and mathematical morphology,” IEEE Workshop Automa. Ident Techn., pp. 89-94, 2005.
* Vector hóa (Vector Quantization - VQ): Phương pháp này sử dụng biểu đồ VQ để xác định vị trí văn bản trong hình ảnh. Kết quả thử nghiệm cho thấy tỷ lệ phát hiện là 98% và thời gian xử lý là 200ms.
  + Paper tham khảo: R. Zunino and S. Rovetta, “Vector quantization for license-plat image coding,” IEEE Trans. Ind. Electron., vol. 47, no. 1, pp 159-167, 2000.
* Cửa sổ trượt tập trung (Sliding Concentric Windows - SCW): Phương pháp này xem biển số xe như là sự không đồng đều trong cấu trúc hình ảnh và sử dụng sự thay đổi đột ngột trong các đặc điểm cục bộ để xác định vị trí tiềm năng của biển số xe.
  + Paper tham khảo: C.-N. E. Anagnostopoulos et al “A license plate-recognition intelligent transportation system applications,” IEEE Trans. Syst., vol. 7, no. 3, pp. 377-392, 2006.
* Bộ lọc Gabor và biến đổi Fourier (Gabor Filters và Discrete Fourier Transform - DFT): Các phương pháp này sử dụng bộ lọc Gabor và biến đổi DFT để xác định vị trí và cấu trúc của biển số xe. Tuy nhiên, phương pháp này tốn thời gian tính toán.
  + Paper tham khảo: K. Deb, H.-U. Chae, and K.-H. Jo, “Vehicle license plate dete based on sliding concentric windows and histogram,” J. of Co no. 8, pp. 771-777, 2009
* Biến đổi sóng (Wavelet Transform - WT): Phương pháp này sử dụng WT để trích xuất biển số xe dựa trên sự thay đổi trong cạnh dọc và cạnh ngang. Thời gian thực hiện nhanh và độ chính xác là 97.33%.
  + Paper tham khảo: C.-T. Hsieh et al, “Multiple license plate detection for complex background” Int. Conf. Adv. Inform. Netw. and Applicat., vol. 2, pp. 389-395, 2005.
* AdaBoost và đặc điểm Haar-like: Các phương pháp này kết hợp AdaBoost với các đặc điểm Haar-like để xây dựng bộ phân loại cho việc trích xuất biển số xe. Chúng không nhạy cảm với độ sáng, màu sắc, kích thước và vị trí của biển số xe.
  + Paper tham khảo: L. Dlagnekov, “License Plate Detection Using AdaBoost”, Com and Engineering Department, San Diego, 2004
* ***Ưu điểm***:
  + Khả năng xử lý trong các tình huống nhiễu: Texture features có thể giúp phát hiện biển số xe trong các hình ảnh có nhiễu hoặc điều kiện ánh sáng biến đổi, bởi vì chúng không phụ thuộc vào màu sắc hoặc cường độ pixel.
  + Phù hợp với các biển số xe có đặc điểm cấu trúc đặc biệt: Các biển số xe thường có các mẫu cấu trúc hoặc ký tự đặc biệt. Texture features có thể bắt được những đặc điểm này và giúp xác định biển số xe dựa trên chúng.
* Nhược điểm:
  + Yêu cầu tập dữ liệu lớn: Để xác định các texture features hiệu quả, phải có một tập dữ liệu lớn chứa các hình ảnh biển số xe với đa dạng về điều kiện ánh sáng, góc nhìn, và kích thước biển số.
  + Tính toán phức tạp: Phương pháp này thường đòi hỏi tính toán phức tạp hơn so với một số phương pháp khác, đặc biệt là khi sử dụng các thuật toán phân tích cấu trúc phức tạp.
  + Khó khăn trong việc xác định đặc điểm texture đặc trưng: Việc xác định các texture features phù hợp và hiệu quả có thể đòi hỏi sự nghiên cứu và tinh chỉnh đáng kể.

Phương pháp "License Plate Extraction using Texture Features" có tiềm năng trong việc trích xuất biển số xe, đặc biệt là trong các tình huống khó khăn về nhiễu và ánh sáng. Tuy nhiên, nó cũng có nhược điểm và đòi hỏi nhiều công sức trong việc xác định các đặc điểm texture thích hợp và xây dựng tập dữ liệu đủ lớn.

*D,* *Trích xuất biển số xe bằng tính năng màu sắc (License Plate Extraction using Color Features)*

Với ý tưởng cơ bản là màu sắc của biển số và ký tự trên biển số là duy nhất và thường chỉ xuất hiện trong khu vực của biển số xe. Các phương pháp dựa trên màu sắc bao gồm:

* Sử dụng mô hình màu HLS để phân loại từng điểm ảnh vào 13 loại màu sắc dựa trên định dạng cụ thể của biển số Trung Quốc.
  + Paper tham khảo: X. Shi et al, Automatic License Plate Recognition System Based on Color Image Processing, vol. 3483, O. Gervasi et al., Ed. New York: SpringerVerlag, pp. 1159-1168, 2005.
* Sử dụng mạng nơ-ron để phân loại màu sắc của từng điểm ảnh sau khi chuyển đổi hình ảnh RGB sang HLS. Màu sắc của biển số xe ở Hàn Quốc là xanh, đỏ và trắng. Sử dụng bộ cảm biến biên biến màu chỉ tập trung vào ba loại cạnh liên quan đến màu sắc biển số xe. Độ chính xác của việc xác định vị trí biển số xe đạt 97,9%.
  + Paper tham khảo: S.-L. Chang et al, “Automatic license plate recognition,” IEEE Trans. Intell. Transp. Syst., vol. 5, no. 1, pp. 42-53, 2004.
* Sử dụng giải thuật di truyền (GA) để xác định màu sắc của biển số xe. GA được sử dụng để xác định ngưỡng trên và ngưỡng dưới cho màu sắc của biển số xe dựa trên đào tạo với nhiều điều kiện ánh sáng khác nhau.
  + Paper tham khảo: S. Yohimori et al, “License plate detection system by using threshold function and improved template matching method,” IEEE Annu. Meet. of the Fuzzy Inform., vol. 1, pp. 357-362, 2004.
* Sử dụng Gaussian Weighted Histogram Intersection (GWHI) để phát hiện biển số xe bằng cách so khớp màu sắc. Các phương pháp này sửa đổi histogram thông thường bằng cách sử dụng hàm Gaussian để xử lý các tình huống ánh sáng khác nhau.
  + Paper tham khảo: W. Jia et al, “Gaussian weighted histogram intersection for license plate classification,” Int. Conf. Patt. Recog., vol. 3, pp. 574-577, 2006.
* Sử dụng sự kết hợp màu sắc giữa biển số xe và màu sắc của ký tự để tạo ra hình ảnh cạnh. Tất cả các cạnh trong hình ảnh mới được phân tích để tìm các vùng tiềm năng của biển số xe.
  + Paper tham khảo: X. Wan et al, “A vehicle license plate localization method using color barycenters hexagon model,” Proc. of SPIE, vol. 8009, pp. 80092O-1- 80092O-5, 2011.
* Trong trường hợp sử dụng mô hình màu HSI, các giá trị trung bình và độ lệch chuẩn của màu sắc được sử dụng để xác định các điểm ảnh của biển số màu xanh và vàng, cũng như biển số màu xanh, vàng và trắng từ hình ảnh của xe.
  + Paper tham khảo: K. Deb and K.-H. Jo, “A Vehicle license plate detection method for intelligent transportation system applications,” Cybern. and Syst.: An Int. J., vol. 40, pp. 689-705, 2009.
* ***Ưu điểm*:** 
  + Điểm mạnh trong việc nhận diện màu sắc đặc trưng: Phương pháp này dựa vào tính duy nhất của kết hợp màu sắc giữa biển số và ký tự, giúp xác định biển số xe một cách chính xác và hiệu quả.
  + Khả năng nhận diện biển số xe bất kể biển số bị nghiêng hoặc biến dạng: Phương pháp này có thể phát hiện biển số xe dù nó có hình dạng bị nghiêng hoặc biến dạng, giúp nâng cao độ linh hoạt trong ứng dụng thực tế.
  + Có thể áp dụng cho các quốc gia có quy định về màu sắc biển số: Trong trường hợp một số quốc gia có quy định cụ thể về màu sắc biển số xe, phương pháp này có thể dễ dàng thích nghi để trích xuất biển số.
* ***Nhược điểm*:** 
  + Nhạy cảm với điều kiện ánh sáng và môi trường: Phương pháp dựa vào màu sắc có thể gặp khó khăn khi đối mặt với biến đổi về ánh sáng, gây sai lầm trong việc nhận diện màu sắc.
  + Khó khăn trong định nghĩa màu sắc bằng giá trị RGB: Sử dụng mô hình màu RGB để xác định màu sắc của điểm ảnh có thể gặp khó khăn, đặc biệt trong các tình huống ánh sáng khác nhau.
  + Sai lầm trong việc phân loại màu sắc: Các phương pháp phân loại màu sắc có thể gặp sai lầm, đặc biệt khi có sự trùng lặp màu sắc giữa biển số xe và một phần của xe như thân xe, gây hiện tượng sai lầm.

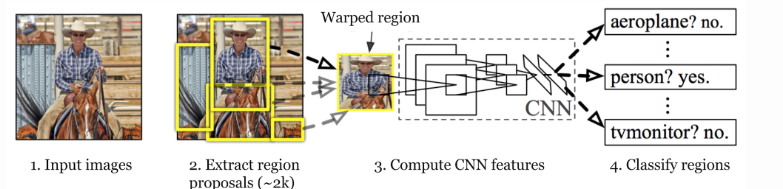
*E, Trích xuất biển số xe bằng cách sử dụng các đặc điểm ký tự (License Plate Extraction using Character Features)*

Các phương pháp trích xuất biển số xe dựa trên việc xác định các ký tự của biển số xem xét hình ảnh để tìm sự có mặt của các ký tự. Nếu các ký tự được tìm thấy, khu vực của chúng được trích xuất như vùng biển số xe. Một số phương pháp tiêu biểu:

* Thay vì sử dụng trực tiếp các thuộc tính của biển số xe, thuật toán cố gắng tìm tất cả các vùng giống ký tự trong hình ảnh bằng cách sử dụng phương pháp dựa trên vùng. Sau đó, các vùng này được phân loại bằng mạng nơ-ron, và nếu tìm thấy một sự kết hợp tuyến tính của các vùng giống ký tự, thì nó giả định có một biển số xe đầy đủ.
  + Paper tham khảo: J. Matas and K. Zimmermann, “Unconstrained license plate and text localization and recognition,” IEEE Conf. Intell. Transp. Syst., pp. 572-577, 2005.
* Phương pháp quét hình ảnh theo hướng ngang để tìm sự thay đổi đối phản lặp lại trên một tỷ lệ 15 điểm ảnh trở lên. Điều này giả định rằng độ tương phản giữa các ký tự và phông nền đủ tốt và có ít nhất ba đến bốn ký tự với kích thước dọc tối thiểu là 15 điểm ảnh. Đạt được độ chính xác 99% trong điều kiện ngoài trời.
  + Paper tham khảo: S. Draghici, “A neural network based artificial vision system for license plate recognition,” Int. J. on Neural Syst., vol. 8, no. 1, pp. 113-126, 1997
* Các đối tượng nhị phân có cùng tỷ lệ khung với ký tự và có hơn 30 điểm ảnh được gắn nhãn. Biến đổi Hough được áp dụng để tìm các đường thẳng, và khu vực giữa các đường thẳng song song với số lượng đối tượng giữa chúng tương tự với số lượng ký tự được coi là vùng biển số xe.
  + Paper tham khảo: F. Alegria and P. S. Girao, “Vehicle plate recognition for wireless traffic control and law enforcement system,” IEEE Int. Conf. Ind. Tech., pp. 1800- 1804, 2006.
* Các vùng cực độ ổn định tối đa (MSER) được sử dụng để có được một tập hợp các vùng ký tự. Các vùng không giống nhau được loại bỏ, và các vùng còn lại với đủ điểm chuyển đổi SIFT được giữ lại là vùng tiềm năng của biển số xe.
  + Paper tham khảo: H. W. Lim and Y. H. Tay, “Detection of license plate characters in natural scene with MSER and SIFT unigram classifier,” IEEE Conf. Sustainable Utilization and Development in Eng. and Tech., pp. 95-98, 2010.
* ***Ưu điểm*:** 
  + Chính xác với biển số xe hoàn chỉnh: Phương pháp này thường cho kết quả chính xác khi tìm kiếm và trích xuất biển số xe đầy đủ. Nếu tất cả các ký tự trên biển số xe có thể được tìm thấy, phương pháp này thường cho kết quả tốt.
  + Khả năng xử lý biển số xe nghiêng hoặc biến dạng: Phương pháp này có khả năng xử lý biển số xe nghiêng hoặc biến dạng mà không cần biết trước hình dạng cụ thể của biển số xe.
* ***Nhược điểm*:**
  + Tốn thời gian: Phương pháp này tốn nhiều thời gian khi xử lý hình ảnh vì phải quét toàn bộ hình ảnh để tìm các ký tự.
  + Khả năng nhận dạng sai khi có văn bản khác: Nếu trong hình ảnh có sự hiện diện của văn bản khác ngoài ký tự của biển số xe, phương pháp này có thể nhận dạng sai và trích xuất sai vùng của biển số xe.
  + Khả năng bị ảnh hưởng bởi độ tương phản hình ảnh: Phương pháp này yêu cầu độ tương phản tốt giữa ký tự và phông nền trên biển số xe, do đó nếu độ tương phản không đủ, phương pháp này có thể không hoạt động tốt.
  + Tốn nhiều bộ nhớ và tài nguyên tính toán: Phương pháp này cần nhiều bộ nhớ và tài nguyên tính toán để xử lý hình ảnh, điều này có thể đòi hỏi phần cứng mạnh và làm tăng chi phí xử lý hình ảnh.
    - 1. *Các phương pháp học sâu*

Hiện nay, phương pháp học sâu được sử dụng để phát hiện đối tượng rất đa dạng và hiệu quả. Có nhiều kiến trúc mạng được áp dụng, ví dụ như CNN, Fast R-CNN, Faster RCNN, SSD, YOLOv1-v8. Các kiến trúc này tỏ ra hiệu quả trong việc phát hiện nhiều đối tượng trong cùng một ảnh.

*A, Mạng CNN*



Kiến trúc tổng quan của mạng R-CNN

R-CNN (Region-based Convolutional Neural Network) là một phương pháp phát hiện đối tượng trong hình ảnh dựa trên các vùng quan tâm (RoIs) được đề xuất. Dưới đây là mô tả về kiến trúc của mạng R-CNN:

* Đề xuất RoIs: Bước đầu tiên của R-CNN là sử dụng một phương pháp để đề xuất các vùng quan tâm trong hình ảnh. Phương pháp được sử dụng thường là một thuật toán như Selective Search hoặc EdgeBoxes, và nó sinh ra một danh sách các RoIs, mỗi RoI là một vùng tiềm năng chứa đối tượng.
* Trích xuất đặc trưng: Mạng CNN được sử dụng để trích xuất đặc trưng từ mỗi RoI. Mạng CNN đã được huấn luyện trước (thường là các phiên bản của VGGNet hoặc ResNet) được sử dụng để trích xuất thông tin từ hình ảnh trong mỗi RoI. Hình ảnh trong mỗi RoI được điều chỉnh kích thước để phù hợp với mạng CNN, và sau đó được đưa vào mạng để trích xuất đặc trưng.
* Vector hóa đặc trưng: Đặc trưng từ mỗi RoI sau khi đi qua mạng CNN được vector hóa thành một biểu diễn số học. Điều này thường được thực hiện bằng cách sử dụng lớp Fully Connected (FC) hoặc một lớp Global Average Pooling (GAP).
* Phân loại đối tượng: Vector hóa đặc trưng được đưa vào một bộ phân loại, thường là một mạng neural network với một lớp Softmax ở cuối. Bộ phân loại này dự đoán xác suất rằng mỗi RoI chứa một đối tượng thuộc một trong các lớp cần phát hiện.
* Dự đoán hộp giới hạn: Ngoài việc phân loại đối tượng, mạng R-CNN cũng dự đoán hộp giới hạn cho mỗi RoI. Điều này thường được thực hiện bằng cách đưa vector hóa đặc trưng qua một lớp Fully Connected khác để dự đoán bốn giá trị: tọa độ của hộp giới hạn (x, y, width, height).
* Kết hợp các RoIs: Các RoIs sau đó được xử lý để loại bỏ các RoIs chồng lấn và lấy ra các RoIs có xác suất cao nhất. Các hộp giới hạn sau khi được điều chỉnh dựa trên dự đoán được kết hợp để tạo ra kết quả cuối cùng về các đối tượng được phát hiện trong hình ảnh.
* Paper tham khảo: X Xie, G Cheng, J Wang, X Yao, J Han, “Oriented R-CNN for object detection”, ICCV 2021. <https://doi.org/10.48550/arXiv.2108.05699>.

***Ưu điểm***:

* Độ chính xác cao: Mạng R-CNN có khả năng đạt được độ chính xác rất cao trong việc phát hiện đối tượng trong hình ảnh. Điều này là do nó sử dụng các đặc trưng từ mạng CNN đã được huấn luyện trước và có khả năng biểu diễn hình ảnh một cách rất hiệu quả.
* Tích hợp thông tin cục bộ: Mạng R-CNN xử lý mỗi RoI riêng lẻ, cho phép nó chú trọng vào các vùng cụ thể trong hình ảnh thay vì phải quan tâm đến toàn bộ hình ảnh. Điều này làm cho nó phù hợp với việc phát hiện các đối tượng nhỏ hoặc đối tượng nằm trong nhiễu.
* Điều chỉnh kích thước RoIs: Mạng R-CNN có thể điều chỉnh kích thước của mỗi RoI để phù hợp với mạng CNN đã được huấn luyện trước. Điều này giúp đảm bảo rằng bất kỳ hình ảnh nào cũng có thể được xử lý một cách hiệu quả.

***Nhược điểm***:

* Chậm: Mạng R-CNN tốn rất nhiều thời gian tính toán vì phải xử lý từng RoI một cách riêng lẻ. Điều này làm cho nó không phù hợp với các ứng dụng yêu cầu xử lý nhanh chóng và thời gian thực.
* Không hiệu quả trong việc đề xuất RoIs: Mạng R-CNN dựa vào các phương pháp đề xuất RoIs bên ngoài như Selective Search hoặc EdgeBoxes. Nhược điểm của điều này là các phương pháp đề xuất RoIs không được huấn luyện cùng mạng, làm giảm tính toàn vẹn của quá trình phát hiện.
* Khả năng tái sử dụng kém: Mạng R-CNN không thể tái sử dụng đặc trưng đã trích xuất cho việc phát hiện nhiều đối tượng. Điều này đồng nghĩa với việc trích xuất đặc trưng cho mỗi RoI gây lãng phí tài nguyên tính toán.
* Phụ thuộc vào đề xuất RoIs: Độ chính xác của mạng R-CNN phụ thuộc vào chất lượng của các RoIs được đề xuất bên ngoài. Nếu các RoIs không tốt, mạng R-CNN sẽ không thể thực hiện phát hiện đối tượng tốt.

*B, Mạng Fast RCNN*

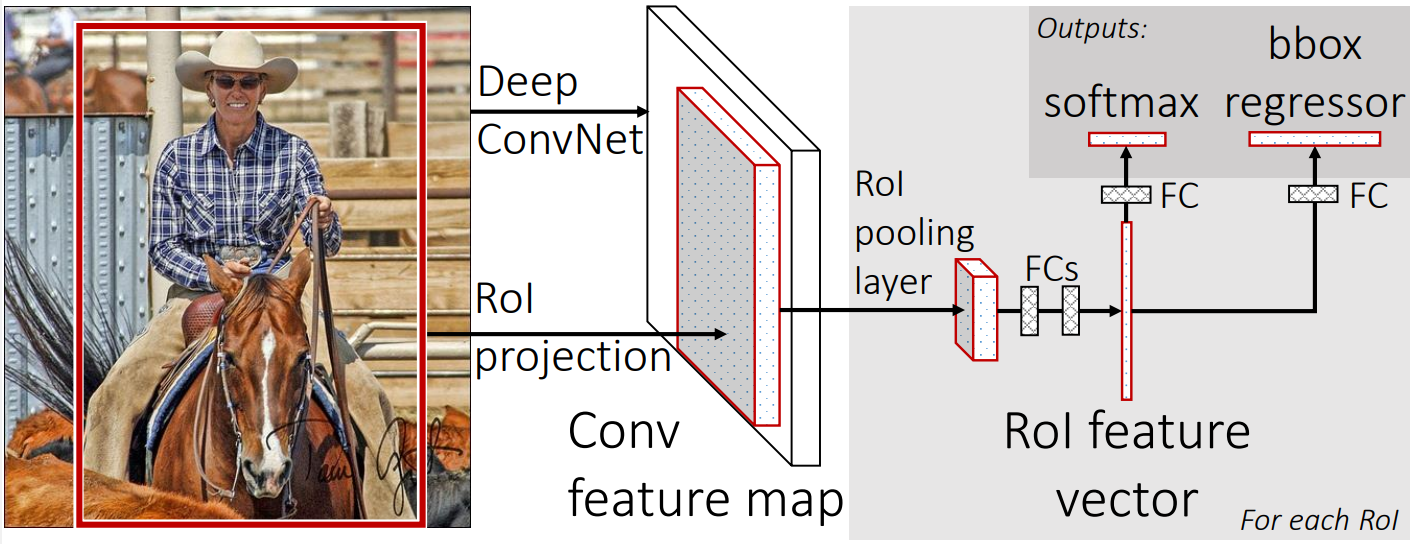
Fast R-CNN bắt đầu bằng việc đầu vào một hình ảnh (hoặc một batch hình ảnh) và truyền chúng qua một mạng ConvNet để trích xuất các đặc trưng hình ảnh. Đối với kiến trúc sử dụng mạng VGG-16, mạng ConvNet này bao gồm một loạt các lớp tích chập và lớp gộp. Mục tiêu của bước này là biến đổi hình ảnh thành một biểu diễn đặc trưng có chiều sâu cao và bao gồm thông tin đa dạng về các đặc điểm của hình ảnh.

Sau khi trích xuất đặc trưng, Fast R-CNN sử dụng một Region Proposal Network (RPN) để dự đoán vùng đề xuất (region proposals) trên ảnh. RPN này là một mạng neural convolutional dựa trên cùng một đặc trưng hình ảnh trích xuất từ bước trước. RPN dự đoán vùng đề xuất bằng cách tạo ra một bản đồ heatmap với các hộp giới hạn (bounding boxes) đề xuất cùng với điểm số đo chất lượng của mỗi hộp.

Khi đã có các vùng đề xuất, Fast R-CNN sử dụng một cơ chế "RoI pooling" để cắt và điều chỉnh các phần của đặc trưng ban đầu liên quan đến từng vùng đề xuất. Sau đó, các vùng đề xuất này được truyền qua một mạng neural nữa để thực hiện phân loại và dự đoán vị trí của đối tượng. Đầu ra của mạng Fast R-CNN chứa thông tin về các hộp giới hạn cũng như các xác suất liên quan đến mỗi lớp đối tượng.

Fast R-CNN có thể đạt được tốc độ xử lý nhanh với khả năng phát hiện đối tượng với độ chính xác cao. Phương pháp này đã đạt được nhiều kết quả xuất sắc trong các cuộc thi và được sử dụng rộng rãi trong ứng dụng thị giác máy tính.

Paper tham khảo: R. Girshick, “Fast R-CNN,” in IEEE International Conference on Computer Vision (ICCV), 2015.



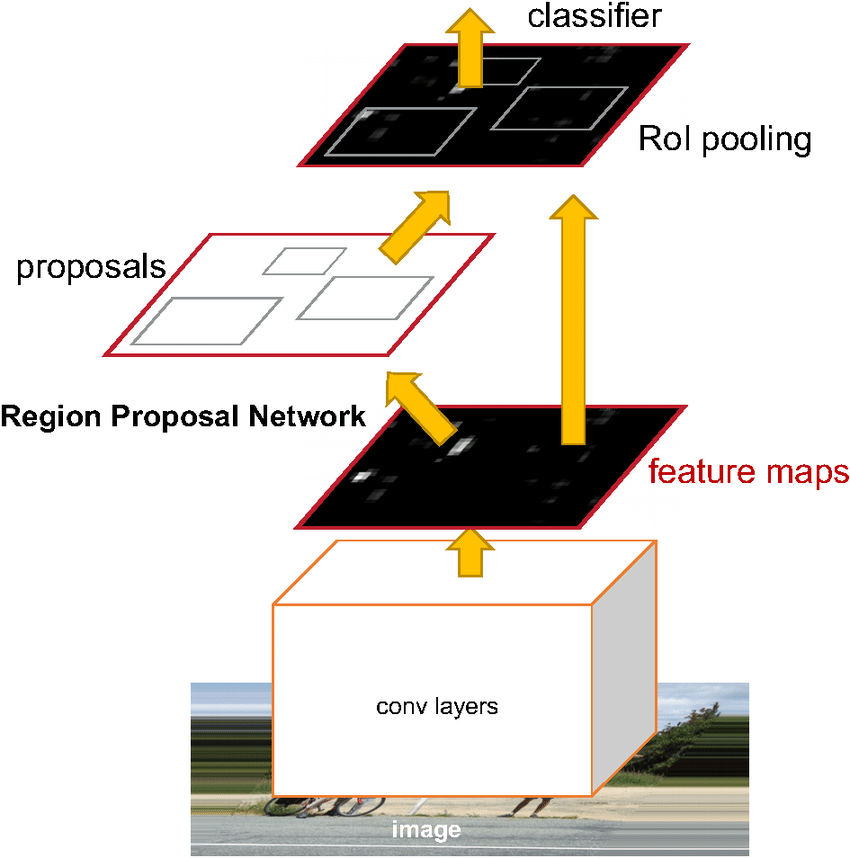
*Kiến trúc của mạng Fast R-CNN*

Fast R-CNN và R-CNN là hai phương pháp khác nhau cho việc phát hiện đối tượng trong hình ảnh. Sự khác biệt chính giữa chúng là trong cách chúng trích xuất và xử lý các khu vực quan tâm (RoIs - Regions of Interest).

|  |  |
| --- | --- |
| **R-CNN** | **Fast R-CNN** |
| - Sau khi sử dụng mạng CNN để trích xuất đặc trưng từ toàn bộ hình ảnh đầu vào, R-CNN sử dụng một thuật toán tiền đề (ví dụ: Selective Search) để tạo ra một danh sách các RoIs, các vùng hình ảnh có khả năng chứa đối tượng.  - Mỗi RoI được cắt ra từ hình ảnh gốc và đưa qua mạng CNN để trích xuất đặc trưng.  - Cuối cùng, các đặc trưng từ mỗi RoI được đưa qua một bộ phân loại (ví dụ: SVM) để xác định xem RoI đó chứa đối tượng gì và đặc trưng của RoI cũng được sử dụng để dự đoán hộp giới hạn cho đối tượng. | - Sau khi sử dụng mạng CNN để trích xuất đặc trưng từ toàn bộ hình ảnh đầu vào, thay vì cắt các RoIs từ hình ảnh gốc, Fast R-CNN sử dụng một lớp RoI Pooling để tạo ra các bản đồ đặc trưng có kích thước cố định từ toàn bộ bức ảnh.  - Các bản đồ đặc trưng này sau đó được đưa vào các lớp kết nối đầy đủ để phân loại và dự đoán hộp giới hạn cho đối tượng trong mỗi RoI.  - Fast R-CNN giúp giảm thời gian tính toán bằng cách chia sẻ đặc trưng giữa toàn bộ hình ảnh và các RoIs. |

C, *Mạng Faster RCNN*

Trong Faster R-CNN, kiến trúc tổng hợp cả một mạng neural phát hiện vùng đề xuất (Region Proposal Network - RPN) và mạng phát hiện đối tượng (object detection network). Mục tiêu chính của Faster R-CNN là đề xuất vùng dự kiến chứa đối tượng và sau đó phát hiện đối tượng trong các vùng này. Để làm được điều này, kiến trúc sử dụng mạng convolutional để trích xuất đặc trưng từ hình ảnh và sau đó sử dụng RPN để đề xuất các vùng đề xuất, được dự đoán dựa trên đặc trưng đã trích xuất.



Kiến trúc mạng Faster RCNN

Chi tiết kiến trúc:

* Trích xuất đặc trưng: Đầu tiên, Faster R-CNN sử dụng mạng convolutional (thường là mạng VGG hoặc ResNet) để trích xuất đặc trưng từ hình ảnh đầu vào. Đặc trưng này là một biểu diễn của hình ảnh có chiều sâu cao, giúp mã hóa thông tin về cạnh, góc, và các đặc điểm của hình ảnh.
* Region Proposal Network (RPN): RPN là một mạng convolutional đặc biệt để dự đoán các vùng đề xuất trên hình ảnh. RPN sử dụng các bộ lọc tích chập để quét toàn bộ hình ảnh và đưa ra các hộp giới hạn dự kiến chứa đối tượng. Mỗi hộp giới hạn được dự đoán cùng với một điểm số đo chất lượng, biểu thị mức độ tin cậy về việc vùng đó chứa đối tượng hay không. RPN sử dụng anchor boxes để dự đoán các vùng đề xuất và điểm số.
* RoI Pooling: Các vùng đề xuất từ RPN có thể có các kích thước khác nhau, do đó, Faster R-CNN sử dụng cơ chế RoI Pooling để chuyển đổi các vùng đề xuất thành các kích thước đồng nhất (thường là một kích thước cố định) để có thể đưa vào mạng phát hiện đối tượng.
* Mạng phát hiện đối tượng (Object Detection Network): Các vùng đề xuất đã được chuyển đổi thông qua RoI Pooling được đưa vào mạng phát hiện đối tượng để thực hiện phân loại đối tượng và dự đoán vị trí của đối tượng trong từng vùng. Mạng phát hiện đối tượng thường bao gồm một loạt các lớp tích chập và lớp kết nối đầy đủ để thực hiện nhiệm vụ này.

Faster R-CNN là một trong những kiến trúc đầu tiên đạt được độ chính xác cao trong việc phát hiện đối tượng trên ảnh, đồng thời tạo ra các vùng đề xuất một cách nhanh chóng. Phương pháp này đã tạo ra sự đột phá trong lĩnh vực thị giác máy tính và được sử dụng rộng rãi trong các ứng dụng như nhận dạng khuôn mặt, theo dõi đối tượng, và nhiều ứng dụng khác.

Paper tham khảo: S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Advances in neural information processing systems, pp. 91–99, 2015.

Faster R-CNN và Fast R-CNN đều là các phương pháp tiên tiến để phát hiện đối tượng trong hình ảnh, nhưng có một số sự khác biệt quan trọng:

|  |  |
| --- | --- |
| **Fast R-CNN** | **Faster R-CNN** |
| - Fast R-CNN sử dụng một phần mạng CNN để trích xuất đặc trưng từ toàn bộ hình ảnh và sau đó sử dụng lớp RoI Pooling để trích xuất đặc trưng từ các vùng quan tâm.  - Việc trích xuất RoIs và phân loại/dự đoán hộp giới hạn được thực hiện bởi hai phần riêng lẻ (mạng CNN và mạng phân loại).  - Fast R-CNN đã cải thiện tốc độ và hiệu suất so với R-CNN, nhưng việc tạo ra RoIs (vùng quan tâm) vẫn là một bước chậm trong quá trình. | - Faster R-CNN giới thiệu một mạng thứ hai, gọi là Region Proposal Network (RPN), để tự động tạo ra các RoIs. RPN sử dụng đặc trưng từ mạng CNN để đề xuất các vùng quan tâm trong hình ảnh.  - Cả quá trình trích xuất RoIs và phân loại/dự đoán hộp giới hạn được tích hợp vào một mạng duy nhất.  - Faster R-CNN điều chỉnh quá trình tạo ra RoIs bằng cách sử dụng RPN, giúp cải thiện tốc độ đáng kể và hiệu suất so với Fast R-CNN và R-CNN. |

- ***Ưu điểm***:

* Hiệu suất cao: Faster R-CNN đạt được hiệu suất phát hiện đối tượng tốt trên nhiều tập dữ liệu khác nhau, bao gồm PASCAL VOC và MS COCO. Điều này chứng minh rằng nó có khả năng phát hiện đối tượng đa dạng và có độ chính xác cao.
* Tích hợp: Mạng Faster R-CNN kết hợp việc trích xuất vùng đối tượng và phân loại trong một kiến trúc đơn, giúp tối ưu hóa việc huấn luyện và triển khai.
* Áp dụng được rộng rãi: Faster R-CNN là một mô hình tiêu chuẩn và được sử dụng rộng rãi trong nhiều ứng dụng thực tế, chẳng hạn như phát hiện đối tượng trong hình ảnh y tế, xe tự hành và nhiều lĩnh vực khác.
* Khả năng học đối tượng cơ động: Nó có khả năng phát hiện đối tượng cơ động và vùng đối tượng có thể thay đổi trong mỗi hình ảnh.

- ***Nhược điểm***:

* Cần tính toán phức tạp: Faster R-CNN có thể yêu cầu nhiều tài nguyên tính toán, đặc biệt là trong giai đoạn huấn luyện. Điều này làm cho nó không phù hợp cho các ứng dụng có tài nguyên tính toán hạn chế.
* Khó khăn trong việc triển khai trên thiết bị cấu hình thấp: Trong môi trường thời gian thực, đặc biệt là trên các thiết bị nhúng, Faster R-CNN có thể gặp khó khăn vì tính năng của nó.
* Phụ thuộc vào số lượng RoI: Hiệu suất của Faster R-CNN có thể phụ thuộc vào số lượng khu vực đề xuất (RoI) và việc thiết lập tham số liên quan đến việc chọn RoI. Việc thiết lập không tốt có thể ảnh hưởng đến hiệu suất phát hiện.
* Tốc độ chậm hơn so với mạng phát hiện đối tượng một giai đoạn (single-stage): So với mạng SSD hoặc YOLO, Faster R-CNN có thể chậm hơn về tốc độ phát hiện, đặc biệt là trong các tình huống đòi hỏi xử lý thời gian thực.

Mạng Faster R-CNN là một giải pháp mạnh mẽ cho việc phát hiện đối tượng trong hình ảnh với độ chính xác cao, nhưng cần xem xét kỹ lưỡng các ưu điểm và nhược điểm khi triển khai trong các tình huống cụ thể.

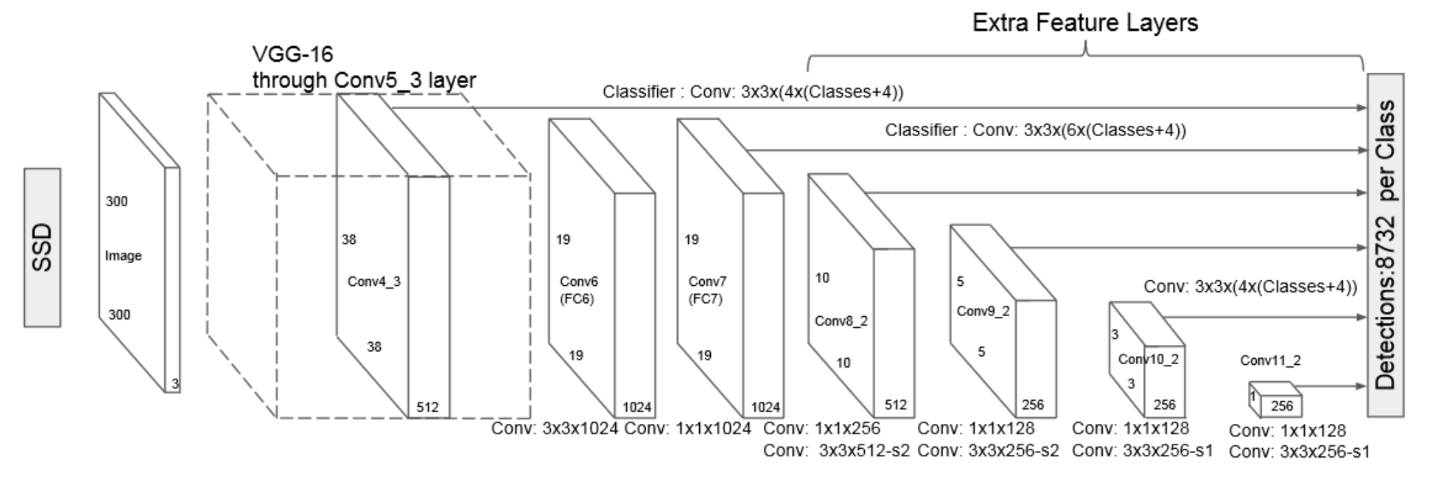
*D, Mạng Single Shot Multibox Detector (SSD)*

Mạng Single Shot Multibox Detector (SSD) là một mạng học sâu được thiết kế để phát hiện và phân loại đối tượng trong hình ảnh. Khác với phương pháp truyền thống sử dụng cửa sổ trượt với kích thước cố định, SSD tạo ra một tập hợp các ô chuẩn (default box) để xác định vị trí và lớp của các đối tượng trong quá trình huấn luyện. Việc sử dụng các ô chuẩn giúp cho mạng có khả năng dự đoán kích thước của các đường bao chữ nhật quanh vị trí của đối tượng một cách chính xác.

SSD áp dụng các ô chuẩn trên nhiều lớp bản đồ đặc trưng với các kích thước khác nhau để phát hiện được những đối tượng có kích thước lớn hoặc nhỏ khác nhau. Điều này giúp cho SSD có khả năng phát hiện đối tượng một cách hiệu quả và đồng thời giảm thiểu số lượng các phép tính cần thiết để xác định vị trí và lớp của đối tượng.

Mạng SSD được thiết kế với kiến trúc đầy đủ từ lớp mạng nền cho tới các lớp dự đoán vị trí và phân loại. Điều này giúp cho mạng có khả năng tích hợp việc phân loại và phát hiện vị trí trong cùng một mạng, tăng độ chính xác và giảm thời gian xử lý so với phương pháp truyền thống.

Paper tham khảo: W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in European conference on computer vision, pp. 21–37, Springer, 2016.



Kiến trúc mạng SSD

Kiến trúc của mạng Single Shot Multibox Detector (SSD) bao gồm một số thành phần chính:

* Backbone Network: Mạng nền (backbone network) trong SSD thường là một mạng nơ-ron tích chập sâu (deep convolutional neural network) như VGG16, ResNet hoặc MobileNet, được sử dụng để trích xuất đặc trưng từ hình ảnh đầu vào. Điều này giúp nắm bắt các đặc trưng thô từ hình ảnh.
* Multi-scale Feature Maps: SSD sử dụng các lớp tích chập sau cùng của mạng nền để tạo ra các bản đồ đặc trưng đa tỷ lệ (multi-scale feature maps). Các bản đồ này cung cấp thông tin về các đặc trưng ở nhiều tỷ lệ khác nhau trong hình ảnh, từ cận đến xa.
* Convolutional Layers for Object Detection: Trên mỗi bản đồ đặc trưng đa tỷ lệ, SSD áp dụng một số lớp tích chập để dự đoán vị trí và điểm số của các hộp giới hạn (bounding boxes) tương ứng với các đối tượng. Các thông số như kích thước của các hộp và số lượng lớp đối tượng được dự đoán có thể tùy chỉnh.
* Default Boxes (or Priors): Mạng SSD sử dụng các hộp mặc định (default boxes hoặc priors) tùy chỉnh cho mỗi lớp tích chập để xác định các hộp giới hạn ứng viên. Các hộp mặc định này có kích thước và tỷ lệ cố định.
* Non-maximum Suppression (NMS): Sau khi có dự đoán từ các lớp tích chập khác nhau và các hộp mặc định, SSD sử dụng thuật toán Non-maximum Suppression để loại bỏ các hộp trùng lặp và giữ lại các hộp có xác suất cao nhất.

- ***Ưu điểm***:

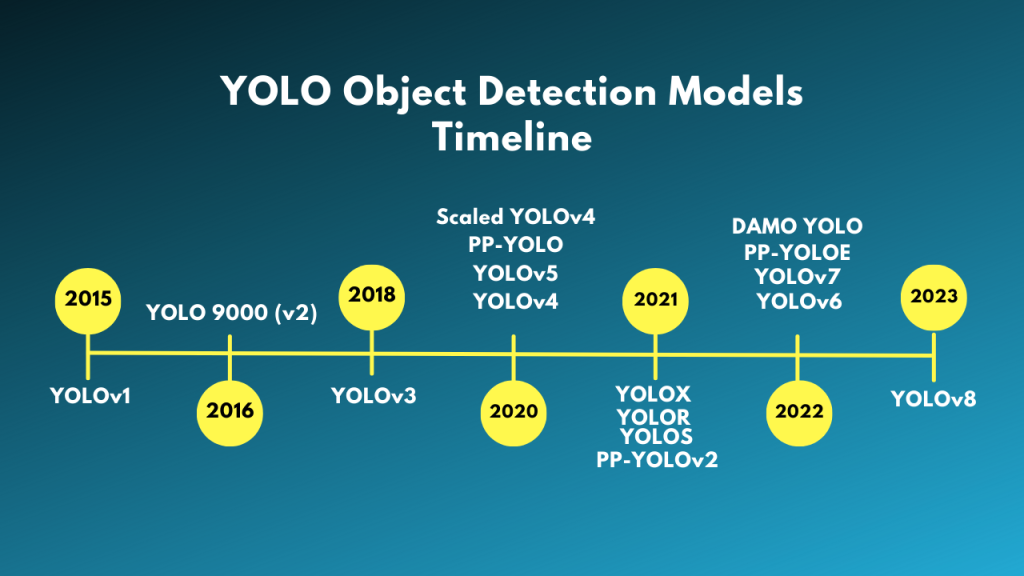
* Hiệu quả về mặt thời gian: SSD cho phép phát hiện đối tượng trên nhiều tỷ lệ khác nhau trong một lần chạy, giúp giảm thiểu thời gian tính toán. Nó nhanh hơn so với một số mô hình phát hiện đối tượng trước đây.
* Khả năng phát hiện đa tỷ lệ: SSD tạo ra các dự đoán đối tượng ứng với nhiều tỷ lệ khác nhau, từ cận đến xa, giúp phát hiện các đối tượng ở mọi kích thước và từ xa.
* Tích hợp đa tỷ lệ: SSD không yêu cầu phải thực hiện nhiều lần tính toán cho từng tỷ lệ, điều này giúp tiết kiệm tài nguyên tính toán và bộ nhớ.
* Hiệu suất tốt: SSD đạt được hiệu suất cao trên nhiều bộ dữ liệu thử nghiệm và trở thành một trong những mô hình phát hiện đối tượng phổ biến được sử dụng trong các ứng dụng thị giác máy tính.

- ***Nhược điểm***:

* Thiết kế phức tạp: SSD với nhiều lớp tích chập và hộp mặc định (default boxes) có thể trở nên phức tạp, đòi hỏi nhiều tài nguyên tính toán và bộ nhớ.
* Đòi hỏi nhiều dữ liệu đào tạo: Để đạt được hiệu suất tốt, SSD yêu cầu một lượng lớn dữ liệu đào tạo để đảm bảo rằng các lớp tích chập và hộp mặc định được học cách phát hiện đối tượng hiệu quả.
* Khó điều chỉnh: Đối với những người mới vào lĩnh vực thị giác máy tính, việc điều chỉnh và đào tạo mô hình SSD có thể khó khăn và đòi hỏi hiểu biết sâu về deep learning và thị giác máy tính.

*E, Mô hình YOLO (You Only Look Once)*

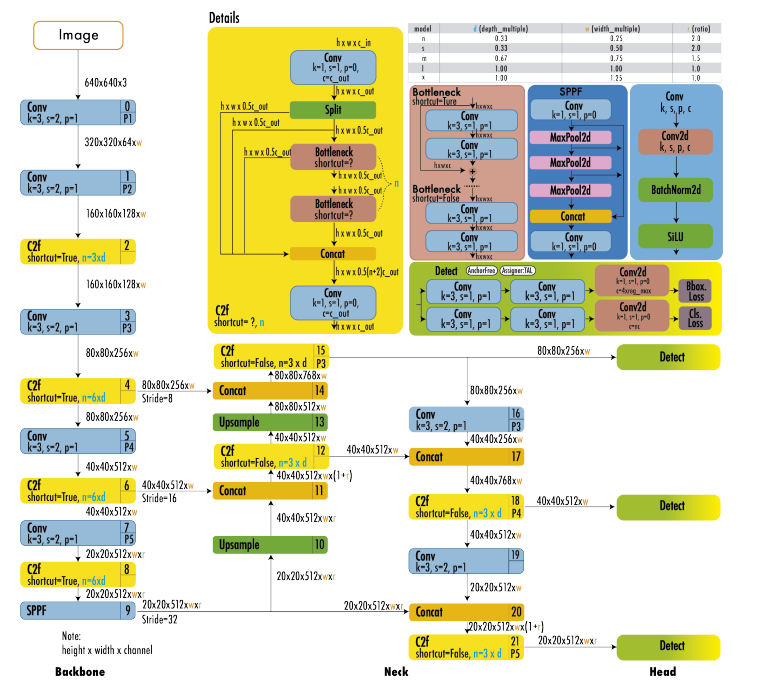
Mô hình YOLO (You Only Look Once) là một mô hình học sâu được sử dụng cho việc nhận diện đối tượng trong ảnh và video. Nó được thiết kế để nhận diện và phân loại các đối tượng đồng thời trong một khung hình. Mô hình YOLO sử dụng một mạng nơ-ron tích chập để xác định vị trí, kích thước và loại của các đối tượng trong ảnh và đưa ra các dự đoán trong thời gian thực. Với độ chính xác cao và khả năng xử lý nhanh, YOLO đã trở thành một trong những mô hình phổ biến nhất trong lĩnh vực nhận diện đối tượng. YOLO xuất hiện từ năm 2015, đã phát triển qua nhất nhiều phiên bản, chúng ta có thể nhìn thấy các phiên bản như trên hình.



Sự phát triển của YOLO

* YOLOv1: Phiên bản đầu tiên của mô hình YOLO, đó là YOLOv1 đã được xuất bản bởi Joseph Redmon et al. vào năm 2015. Đây là mô hình phát hiện đối tượng (SSD - Single Stage object Detection) single stage đầu tiên đã tạo ra các mô hình SSD và tất cả các mô hình YOLO tiếp theo.
* YOLOv2: còn được gọi là YOLO 9000 được xuất bản bởi tác giả gốc của YOLOv1, Joseph Redmon. Nó đã cải thiện YOLOv1 bằng cách giới thiệu khái niệm anchor boxes và backbone tốt hơn, đó là Darknet-19.
* YOLOv3: Năm 2018, Joseph Redmon và Ali Farhadi đã xuất bản YOLOv3. YOLOv3 sử dụng backbone Darknet-53, loại bỏ các kết nối dư thừa, pretrain tốt hơn và các kỹ thuật tăng cường hình ảnh để đem lại những cải tiến.
* Ultralytics YOLO Object Detection Models: Tất cả các mô hình phát hiện đối tượng YOLO cho đến YOLOv3 đều được viết bằng ngôn ngữ lập trình C và sử dụng framework Darknet. Những người mới học sẽ gặp khó khăn khi đọc code và tinh chỉnh các mô hình. Cùng khoảng thời gian với YOLOv3, Ultralytics đã phát hành mô hình YOLO (YOLOv3) đầu tiên được triển khai bằng framework PyTorch. Nó cũng dễ tiếp cận và dễ sử dụng hơn cho việc transfer learning. Ngay sau khi xuất bản YOLOv3, Joseph Redmon đã rời khỏi cộng đồng nghiên cứu Thị giác máy tính. YOLOv4 (của Alexey và cộng sự) là mô hình YOLO cuối cùng được viết trên Darknet. Sau đó đã có rất nhiều mô hình YOLO khác nhau. Trong đó YOLOv4, YOLOX, PP-YOLO, YOLOv6 và YOLOv7 là một số mô hình nổi bật. Sau YOLOv3, Ultralytics cũng phát hành YOLOv5 thậm chí còn tốt hơn, nhanh hơn và dễ sử dụng hơn tất cả các mô hình YOLO khác. Tính đến thời điểm hiện tại, Ultralytics đã xuất bản YOLOv8, đây có lẽ là mô hình YOLO tốt nhất cho đến nay.

**YOLOv8**



Kiến trúc Yolov8

YOLOv8 sử dụng một trọng điểm (backbone) tương tự như YOLOv5 với một số thay đổi trên CSPLayer, hiện được gọi là mô-đun C2f. Mô-đun C2f (mô-đun gắn kết gần giữa các giai đoạn với hai lớp tích chập) kết hợp các đặc trưng cấp cao với thông tin ngữ cảnh để cải thiện độ chính xác trong việc phát hiện.

YOLOv8 sử dụng một mô hình không sử dụng anchor (anchor-free) với một đầu độc lập để xử lý riêng lẻ các nhiệm vụ đối tượng, phân loại và hồi quy. Thiết kế này cho phép mỗi nhánh tập trung vào nhiệm vụ của nó và cải thiện độ chính xác tổng thể của mô hình. Ở lớp đầu ra của YOLOv8, họ sử dụng hàm sigmoid làm hàm kích hoạt cho điểm độ tin cậy đối tượng, đại diện cho xác suất rằng hộp giới hạn chứa một đối tượng. YOLOv8 sử dụng hàm softmax cho xác suất lớp, đại diện cho xác suất rằng đối tượng thuộc vào mỗi lớp có thể có.

YOLOv8 sử dụng các hàm mất mát CIoU [Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-iou loss: Faster and better learning for bounding box regression,” in Proceedings of the AAAI conference on artificial intelligence, vol. 34, pp. 12993–13000, 2020.] và DFL [X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, “Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection,” Advances in Neural Information Processing Systems, vol. 33, pp. 21002–21012, 2020.] cho mất mát hộp giới hạn và hàm mất mát binary cross-entropy cho mất mát phân loại. Các mất mát này đã cải thiện hiệu suất phát hiện đối tượng, đặc biệt là khi xử lý các đối tượng nhỏ.

YOLOv8 cung cấp một mô hình phân đoạn ngữ nghĩa gọi là YOLOv8-Seg. Trọng điểm (backbone) của nó là bộ trích xuất đặc trưng CSPDarknet53, theo sau là mô-đun C2f thay vì kiến trúc cổ điển YOLO neck. Mô-đun C2f được theo sau bởi hai đầu phân đoạn, học để dự đoán các mặt nạ phân đoạn ngữ nghĩa cho hình ảnh đầu vào. Mô hình có các đầu phát hiện tương tự với YOLOv8, bao gồm năm mô-đun phát hiện và một lớp dự đoán. Mô hình YOLOv8-Seg đã đạt được kết quả hàng đầu trên nhiều bài kiểm tra về phát hiện đối tượng và phân đoạn ngữ nghĩa trong khi duy trì tốc độ và hiệu suất cao.

Được đánh giá trên tập dữ liệu MS COCO test-dev 2017, YOLOv8x đã đạt được AP (Average Precision) là 53,9% với kích thước hình ảnh là 640 pixel (so với 50,7% của YOLOv5 với cùng kích thước đầu vào) với tốc độ là 280 hình trên giây trên GPU NVIDIA A100 và TensorRT.

- ***Ưu điểm***:

* Tốc độ cao: Mô hình YOLO nổi tiếng với tốc độ xử lý nhanh. Nó có khả năng xử lý ảnh thời gian thực, phù hợp cho các ứng dụng yêu cầu sự nhanh nhạy như xe tự lái, theo dõi video thời gian thực, và nhiều ứng dụng khác.
* Kiến trúc đơn giản: YOLO sử dụng một mạng nơ-ron duy nhất để dự đoán bounding box và lớp của các đối tượng. Điều này giúp giảm độ phức tạp của mô hình và dễ dàng để triển khai và sử dụng.
* Hiệu quả về tài nguyên: Mô hình YOLO sử dụng ít tài nguyên hơn so với nhiều mô hình phát hiện đối tượng khác, đặc biệt khi cần xử lý ảnh ở thời gian thực trên các thiết bị có tài nguyên hạn chế.
* Phát hiện đối tượng đa lớp: YOLO có khả năng phát hiện nhiều đối tượng thuộc nhiều lớp khác nhau trong một lần đánh giá, giúp cải thiện hiệu suất và độ chính xác.

- ***Nhược điểm***:

* Độ chính xác không cao bằng mô hình phức tạp hơn: So với một số mô hình phát hiện đối tượng phức tạp, YOLO có thể không cung cấp độ chính xác cao trong việc phát hiện đối tượng, đặc biệt là đối tượng nhỏ hoặc có hình dạng phức tạp.
* Khó khăn trong việc phát hiện đối tượng gần nhau: YOLO có thể gặp khó khăn khi phát hiện các đối tượng gần nhau hoặc chồng chéo lên nhau do kiến trúc lưới mạng và sự hạn chế về độ phân giải.
* Yêu cầu tập dữ liệu lớn: Để đảm bảo hiệu suất và độ chính xác, YOLO cần một tập dữ liệu huấn luyện lớn, đòi hỏi nhiều dữ liệu chú thích và tài nguyên tính toán để huấn luyện hiệu quả.
* Khó khảo sát đối tượng nhỏ: YOLO có thể gặp khó khăn trong việc phát hiện đối tượng nhỏ, nhưng các phiên bản sau của nó đã cố gắng cải thiện điều này.
  + 1. ***Phân đoạn ký tự trên vùng biển số xe***

*A, Tiền xử lý dữ liệu*

Biển số xe sau khi được trích xuất có thể gặp một số vấn đề như nghiêng và độ sáng không đồng đều. Vì vậy để tăng độ chính xác biển số xe cần được tiền xử lý trước khi đưa vào các thuật toán phân đoạn. Dưới đây trình bày một số các phương pháp tiền xử lý ảnh:

* Sử dụng phương pháp biến đổi tuyến tính để chuyển đổi biển số xe trích xuất bị nghiêng thành một hình chữ nhật thẳng.
  + Paper tham khảo: X. Xu et al, “A method of multi-view vehicle license plates location based on rectangle features,” Int. Conf. Signal Process., vol. 3, pp. 16-20, 2006.
* Áp dụng phương pháp bình phương tối thiểu để xử lý sự nghiêng theo chiều ngang và chiều dọc trong hình ảnh biển số xe.
  + Paper tham khảo: M.-S. Pan et al, “Vehicle license plate character segmentation,” Int. J. Automat. and Comput., vol. 5, no. 4, pp. 425-432, 2008
* Áp dụng biến đổi Karhunen-Loeve (K-L) để tổ chức tọa độ của các ký tự thành một ma trận hiệp phương đa chiều hai chiều. Vector riêng và góc quay α được tính toán tuần tự. Sau đó, thực hiện điều chỉnh nghiêng theo chiều ngang của hình ảnh. Đối với việc điều chỉnh nghiêng theo chiều dọc, đề xuất ba phương pháp: sử dụng biến đổi K-L, phương pháp phù hợp dựa trên gom cụm K-means và phù hợp dựa trên bình phương tối thiểu để tính toán góc nghiêng theo chiều dọc θ.
  + Paper tham khảo: M.-S. Pan et al, “A new method for correcting vehicle license plate tilt,” Int. J. of Automat. and Comput., vol. 6, no. 2, pp. 210-216, 2009.
* Sử dụng bình phương tối thiểu kèm theo khoảng cách vuông góc để điều chỉnh góc nghiêng của biển số xe theo chiều ngang. Đồng thời, điều chỉnh nghiêng theo chiều dọc bằng cách tối thiểu hóa sự biến động của tọa độ các điểm chiếu. Sau khi điều chỉnh nghiêng theo chiều ngang, phân đoạn các ký tự được thực hiện và tọa độ của các ký tự được chiếu theo chiều dọc sau khi áp dụng biến đổi cắt đối xứng.
  + Paper tham khảo: K. Deb et al, “Projection and least square fitting with perpendicular offsets based vehicle license plate tilt correction,” SICE Annu. Conf., pp. 3291-3298, 2010.

Việc lựa chọn một ngưỡng không phù hợp cho quá trình nhị phân hóa biển số xe trích xuất có thể dẫn đến các ký tự bị kết nối với nhau, làm cho quá trình phân đoạn trở nên khó khăn. Những ký tự này tạo ra khó khăn trong quá trình phân đoạn. Đối với biển số xe có khung xung quanh, sau quá trình nhị phân hóa, một số ký tự có thể bị kết nối với khung, tạo ra thách thức trong việc phân đoạn. Việc cải thiện chất lượng hình ảnh trước quá trình nhị phân hóa giúp lựa chọn ngưỡng thích hợp. Các phương pháp thường được sử dụng để cải thiện hình ảnh biển số xe bao gồm loại bỏ nhiễu, cân bằng biểu đồ xám và tăng cường độ tương phản. Dưới đây là một số nghiên cứu đề xuất tăng chất lượng ảnh:

* Thực hiện phân tích độ dốc trên toàn bức ảnh để phát hiện biển số xe, sau đó biển số xe đã được tối ưu hóa thông qua biến đổi mức xám.
  + Paper tham khảo: P. Comelli et al, “Optical recognition of motor vehicle license plates,” IEEE Trans. Veh. Tech., vol. 44, no. 4, pp. 790-799, 1995
* Kích thước của các ký tự chiếm khoảng 20% kích thước của biển số xe. Trước hết, mức xám được tỷ lệ từ 0-100, sau đó 20% pixel lớn nhất được nhân với 2,55. Chỉ có các ký tự được tối ưu hóa trong khi các pixel nhiễu được giảm bớt.
  + Paper tham khảo: Y. Zhang, and C. Zhang, “A new algorithm for character segmentation of license plate,” IEEE Intell. Veh. Symp., pp. 106-109, 2003.
* Phương pháp phân ngưỡng cục bộ được áp dụng cho từng pixel, trong đó ngưỡng được tính bằng cách trừ một hằng số c từ mức xám trung bình trong một cửa sổ m×n được tâm ở pixel đó.
  + Paper tham khảo: D. Llorens et al, Car License Plates Extraction and Recognition Based on Connected Components Analysis and HMM Decoding, J. S. Marques et al., Eds. New York: Springer-Verlag, vol. 3522, pp. 571-578, 2005
* Ngưỡng được tính dựa trên công thức nhị phân hóa Niblack để biến đổi ngưỡng dựa trên giá trị trung bình cục bộ và độ lệch chuẩn.
  + Paper tham khảo: C. Coetzee et al, “PC based number plate recognition system,” IEEE Int. Symp. Ind. Electron., pp. 605-610, 1998.

*B, Phân đoạn biển số xe bằng kết nối Pixel (License Plate Segmentation using Pixel Connectivity)*

Phân đoạn được thực hiện bằng cách đánh dấu các pixel kết nối trong hình ảnh biển số xe nhị phân. Các pixel được đánh dấu được phân tích và những pixel có cùng kích thước và tỷ lệ khía cạnh giống với các ký tự biển số được xem xét như các ký tự biển số. Tuy nhiên, phương pháp này thất bại trong việc trích xuất tất cả các ký tự khi có ký tự bị nối hoặc bị rời rạc.

- Paper tham khảo:

* K. Kanayama et al, “Development of vehicle-license number recognition system using real-time image processing and its application to travel-time measurement,” IEEE Veh. Tech. Conf., pp. 798-804, 1991.
* S.-L. Chang et al, “Automatic license plate recognition,” IEEE Trans. Intell. Transp. Syst., vol. 5, no. 1, pp. 42-53, 2004.
* T. Nukano, M. Fukumi, and M. Khalid, “Vehicle license plate character recognition by neural networks,” Proc. of Int. Symp. on Intell. Signal Process. and Commun. Syst., pp. 771-775, 2004.
* V. Shapiro and G. Gluhchev, “Multinational license plate recognition system: segmentation and classification,” Proc. of Int. Conf. on Patt. Recog., vol. 4, pp. 352-355, 2004.
* B.-F. Wu et al, “Extracting characters from real vehicle license plates out-ofdoors,” IET Comput. Vision, vol. 1, no. 1, pp. 2-10, 2007.

- ***Ưu điểm***:

* Đơn giản và hiệu quả: Phương pháp này dựa vào việc kết nối các pixel để phân đoạn biển số xe. Điều này tạo ra một giải pháp đơn giản và hiệu quả, đặc biệt là đối với các biển số xe có ký tự rõ ràng và không bị nhiễu nhiều.
* Tích hợp dữ liệu hình ảnh: Bằng cách sử dụng thông tin về kết nối giữa các pixel, phương pháp này có thể tích hợp dữ liệu hình ảnh để xác định vị trí và biên giới của biển số xe.
* Hiệu quả trong nhiều trường hợp: Phương pháp này hoạt động tốt trong nhiều trường hợp khi biển số xe không có ký tự bị nối hoặc bị rời rạc, và hình ảnh biển số xe không bị nhiễu quá nhiều.

- ***Nhược điểm***:

* Khả năng phân đoạn hạn chế: Phương pháp này gặp khó khăn trong việc phân đoạn các biển số xe có ký tự bị nối hoặc bị rời rạc. Trong trường hợp này, phương pháp có thể bỏ lỡ hoặc phân đoạn sai các ký tự.
* Nhạy cảm với nhiễu hình ảnh: Nếu hình ảnh biển số xe bị nhiễu quá nhiều hoặc có điểm sáng và tối không đồng đều, phương pháp này có thể không hoạt động hiệu quả.
* Yêu cầu tiền xử lý: Đôi khi, cần thực hiện các bước tiền xử lý để làm sạch hình ảnh trước khi áp dụng phương pháp này. Điều này có thể tạo thêm công việc và độ phức tạp cho quá trình nhận diện biển số xe.

*C, Phân đoạn biển số xe bằng cách sử dụng Projection Profiles (License Plate Segmentation using Projection Profiles)*

Vì ký tự và nền biển số xe có màu sắc khác nhau, chúng có giá trị nhị phân trái ngược trong hình ảnh nhị phân. Do đó, một số phương pháp được đề xuất sử dụng việc chiếu thẳng đứng hình ảnh nhị phân của biển số xe để xác định các vị trí bắt đầu và kết thúc của các ký tự, sau đó chiếu ngang các ký tự đã trích xuất để tách từng ký tự riêng lẻ.

Trong nghiên cứu [S. Zhang et al, “Car plate character extraction under complicated environment,” IEEE Int. Conf. Syst. Man Cybern., vol. 5, pp. 4722-4726, 2004], phương pháp chiếu thẳng đứng được sử dụng để trích xuất các ký tự, kết hợp với việc loại bỏ nhiễu và phân tích trình tự ký tự. Thông qua việc kiểm tra trên hơn 30.000 hình ảnh, phương pháp này đã đạt được độ chính xác 99,2% với tốc độ xử lý từ 10-20ms.

Trong [E. R. Lee et al, “Automatic recognition of a car license plate using color image processing,” IEEE Int. Conf. Image Process., vol. 2, pp. 301-305, 1994.] và [C. A. Rahman et al, “A real time vehicle’s license plate recognition system,” IEEE Conf. Adv. Video and Signal Based Surveillance, pp. 163-166, 2003.], thông tin về màu sắc của ký tự được sử dụng trong quá trình chiếu thay vì sử dụng hình ảnh nhị phân của biển số xe. Tổng cộng, dựa trên tài liệu nghiên cứu, phương pháp sử dụng chiếu thẳng đứng và chiếu ngang của các pixel là phương pháp phổ biến và đơn giản nhất.

- ***Ưu điểm***:

* Tính độc lập vị trí: Phương pháp chiếu cho phép trích xuất các ký tự trên biển số xe mà không cần biết trước vị trí cụ thể của chúng. Điều này giúp phương pháp này đảm bảo tính linh hoạt và có khả năng xử lý biển số xe có thể bị xoay hoặc dịch chuyển trong hình ảnh.
* Độ chính xác cao: Nếu được cài đặt và điều chỉnh đúng cách, phương pháp chiếu có thể đạt được độ chính xác cao trong việc trích xuất ký tự từ biển số xe. Các nghiên cứu đã thể hiện rằng nó có thể đạt được tỷ lệ chính xác lên đến 99,2% với tốc độ xử lý nhanh chóng.
* Độc lập với màu sắc: Phương pháp này không phụ thuộc vào màu sắc của ký tự hay biển số xe, điều này có thể giúp nó hoạt động tốt trong nhiều tình huống ánh sáng và màu sắc khác nhau.

***- Nhược điểm***:

* Chất lượng hình ảnh ảnh hưởng đến độ chính xác: Phương pháp chiếu dựa vào giá trị pixel trong hình ảnh, do đó nó nhạy cảm với nhiễu và chất lượng hình ảnh kém. Nếu hình ảnh biển số xe bị nhiễu hoặc mờ, độ chính xác có thể bị giảm.
* Yêu cầu thông tin trước về số lượng ký tự: Phương pháp này đòi hỏi biết trước số lượng ký tự trên biển số xe để có thể tạo ra các phần chiếu thích hợp. Nếu số lượng ký tự thay đổi (ví dụ, biển số có thể có 7 hoặc 8 ký tự), phương pháp này có thể gặp khó khăn.
* Tác động của nhiễu: Bất kỳ nhiễu nào trong hình ảnh cũng sẽ ảnh hưởng đến giá trị chiếu, và có thể dẫn đến việc trích xuất ký tự không chính xác.

*D, Phân đoạn biển số xe sử dụng Prior Knowledge of Characters (License Plate Segmentation using Prior Knowledge of Characters)*

Trong nghiên cứu [C. Busch et al, “Feature based recognition of traffic video streams for online route tracing,” IEEE Veh. Tech. Conf., vol. 3, pp. 1790-1794, 1998], hình ảnh nhị phân được quét bởi một đường ngang để tìm vị trí bắt đầu và kết thúc của các ký tự. Khi tỷ lệ giữa số pixel của các ký tự so với nền trên đường ngang này vượt qua một ngưỡng nhất định, điều này được xem xét là vị trí bắt đầu của các ký tự. Còn tỷ lệ ngược lại được thực hiện để tìm vị trí kết thúc của các ký tự.

Trong nghiên cứu [I. Paliy, et al, “Approach to recognition of license plate numbers using neural networks,” IEEE Int. Joint Conf. Neur. Netw., vol. 4, pp. 2965-2970, 2004], biển số xe trích xuất được điều chỉnh kích thước thành kích thước mẫu đã biết trước. Trong kích thước mẫu này, tất cả vị trí của các ký tự đã được biết trước. Sau khi điều chỉnh kích thước, các vị trí giống nhau được trích xuất để tạo thành các ký tự. Phương pháp này có lợi thế về tính đơn giản. Tuy nhiên, trong trường hợp có sự thay đổi trong biển số xe trích xuất, việc trích xuất kết quả sẽ là nền chứ không phải là các ký tự.  
Trong nghiên cứu [Q. Gao et al, “License plate recognition based on prior knowledge,” IEEE Int. Conf. Automat. and Logistics, pp. 2964-2968, 2007.], phương pháp đề xuất giải pháp cho biển số xe của các phương tiện bị hỏng nặng. Sự phối hợp màu sắc được sử dụng để xác định vị trí của biển số xe trên hình ảnh. Kích thước của mỗi ký tự được sử dụng để phân đoạn các ký tự. Cấu trúc của biển số xe Trung Quốc được sử dụng để xây dựng một bộ phân loại cho quá trình nhận dạng.

Biển số xe tại Đài Loan đều có cùng một phân phối màu sắc là ký tự màu đen trên nền trắng. Bài báo [J.-M. Guo and Y.-F. Liu, “License plate localization and character segmentation with feedback self-learning and hybrid binarization techniques,” IEEE Trans. Veh. Tech., vol. 57, no. 3, pp. 1417-1424, 2008] đề xuất nếu biển số xe được quét bằng một đường ngang, số lượng sự chuyển đổi từ màu đen sang màu trắng (hoặc từ màu trắng sang màu đen) ít nhất là 6 và nhiều nhất là 14. Biến đổi Hough được sử dụng để sửa chế nghiêng; kỹ thuật nhị phân hóa kết hợp được sử dụng để phân đoạn các ký tự trong biển số xe bẩn; và quy trình học tự động phản hồi được sử dụng để điều chỉnh các thông số. Trong thí nghiệm sử dụng tổng cộng 332 hình ảnh khác nhau được chụp dưới các điều kiện chiếu sáng khác nhau và ở các khoảng cách khác nhau. Tỉ lệ tổng thể về vị trí và phân đoạn là 97,1% và 96,4%.

- ***Ưu điểm***:

* Độ chính xác cao: Phương pháp này có khả năng đạt được độ chính xác cao trong việc phân đoạn các ký tự của biển số xe, đặc biệt khi có sẵn thông tin về vị trí bắt đầu và kết thúc của các ký tự. Điều này đặc biệt hữu ích khi ứng dụng trong hệ thống nhận dạng biển số xe tự động.
* Độ tin cậy cao với màu sắc: Phương pháp này sử dụng thông tin về màu sắc của các ký tự để xác định vị trí của chúng trong biển số xe. Điều này giúp tăng cường độ tin cậy của phân đoạn, đặc biệt trong trường hợp biển số xe có ký tự màu đen trên nền trắng, như trong trường hợp của biển số tại Đài Loan.
* Tính đơn giản: Phương pháp này thường có tính đơn giản trong việc triển khai và thực hiện. Việc sử dụng kiến thức trước về các ký tự giúp giảm độ phức tạp trong quá trình phân đoạn và nhận dạng.

- ***Nhược điểm***:

* Phụ thuộc vào kiến thức trước: Phương pháp này yêu cầu có kiến thức trước về vị trí và tính chất của các ký tự trong biển số xe. Điều này đồng nghĩa rằng nếu có sự biến đổi hoặc địa chỉ thay đổi trong biển số xe, phân đoạn có thể trở nên không hiệu quả và dẫn đến việc phân đoạn các phần nền thay vì ký tự.
* Khả năng xử lý nhiễu: Mọi nhiễu hoặc sự biến đổi trong hình ảnh có thể ảnh hưởng đến độ tin cậy của phương pháp này. Điều này đặc biệt đúng khi sử dụng phương pháp dựa trên màu sắc.
* Giới hạn trong việc áp dụng rộng rãi: Phương pháp này thường dựa vào cụ thể về biển số xe hoặc màu sắc, vì vậy nó có thể không phù hợp cho các trường hợp biển số có thiết kế hoặc màu sắc khác biệt.