Xây Dựng Mô Hình Phát Hiện Đèn Tín Hiệu Của Các Phương Tiện Giao Thông

Trần Đăng Khoa[1], Hoàng Đình Quang[1], Nguyễn Thế Mạnh[1]

Đại học Công nghệ Thông tin - Đại học Quốc gia Thành phố Hồ Chí Minh
{18520936, 18521294, 18521084}@gm.uit.edu.vn

Tóm tắt nội dung Phát hiện đèn tín hiệu của phương tiện phía trước là một nhiệm vụ cần thiết khi phát triển các hệ thống giao thông thông minh. Việc dự đoán chính xác hành động kế tiếp của các phương tiện phía trước thông qua đèn tín hiệu sẽ góp phần tăng tính liên kết giữa các phương tiện, từ đó nâng cao tính an toàn cho hệ thống giao thông. Trong những năm gần đây, khoa học và công nghệ phát triển bùng nổ, đặc biệt là trí tuệ nhân tạo. Các thuật toán phát hiện đối tượng với độ chính xác và tốc độ xử lí ấn tượng đã được đề xuất, điển hình là các thuật toán dựa trên phương pháp học sâu như YOLO4, YOLOv5, Mask-RCNN, CenterNet,... Mỗi thuật toán đều có các ưu và nhược điểm riêng, không có thuật toán nào thật sự là tốt nhất. Chúng tôi đã tiến hành thực nghiệm và đánh giá hiệu suất 3 thuật toán là YOLOv4, YOLOv5 và CenterNet trên bộ dữ liệu tự thu thập về đèn tín hiệu xe. Kết quả thực nghiệm được chúng tôi trình bày trong bài báo này.

Keywords: Đèn tín hiệu, deep learning.

1 Giới Thiệu

Trong thế giới tấp nập và hiện đại hoá ngày nay, số lượng phương tiện tham gia giao thông tăng một cách nhanh chóng theo từng năm, dẫn đến nguy cơ tai nạn giao thông ngày càng cao. Ở Việt Nam, trong năm 2020 đã có 14.510 vụ tai nạn giao thông, làm chết 6.700 người, bị thương 10.804 người. Đây là một vấn đề lớn của Việt Nam cả thế giới. Để đảm bảo an toàn giao thông, nếu chỉ đưa ra hàng loạt các luật, quy định, đặt thêm nhiều biển báo, đèn giao thông vẫn là chưa đủ. Hiện nay, với tốc độ phát triển vượt bậc của khoa học công nghệ, đặc biệt là lĩnh vực trí tuệ nhân tạo, một giải pháp mà chúng ta đang hướng đến là áp dụng công nghệ vào giao thông, tạo nên một hệ thống giao thông thông minh. Một hệ thống giao thông thông minh không chỉ mang đến sự an toàn mà còn có tác động rất lớn đến nền kinh tế và khuyến khích tăng trưởng kinh tế trong tương lai. Cụ thể, khi các phương tiện lưu thông trên đường được liên kết với nhau, có thể giao tiếp giữa chúng thì sẽ giảm tình trạng tắt nghẽn giao thông, giảm thời gian di chuyển và chi phí vận hành,chi phí về con người. Dựa trên thực tế đó, trong những thập kỉ gần đây, việc nghiên cứu và ứng dụng công nghệ thông tin vào các hệ thống giao thông đã được đẩy mạnh thực hiện

Trong lĩnh vực thị giác máy tính, các bài toán liên quan đến giao thông luôn là mối quan tâm hàng đầu hiện nay, đặc biệt là các bài toán về phát hiện đối tượng như phát hiện đèn giao thông, phát hiện các vật thể trên đường, phát hiện biển cáo giao thông,... Một trong những bài toán quan trọng phải kể đến là phát hiện đèn tín hiệu của phương tiện phía trước. Ứng dụng của bài toán này là vô cùng cần thiết, có thể cảnh báo người lái hoặc hệ thống tự lái tránh các nguy hiểm tiềm ẩn, xe thông minh có thể tự điều chỉnh tốc độ cho phù hợp như tốc độ bình thường, giảm tốc độ, phanh dựa theo đèn tín hiệu của các xe phía trước để đảm bảo an toàn. Tuy nhiện, với những ứng dũng quan trọng như vậy, bài toán này lại chưa được nghiên cứu nhiều. Vì vậy, trong bài báo này, mục tiêu của chúng tôi hướng tới xây đưng một mô hình phát hiện đèn tín hiệu của các phương tiện phía trước.

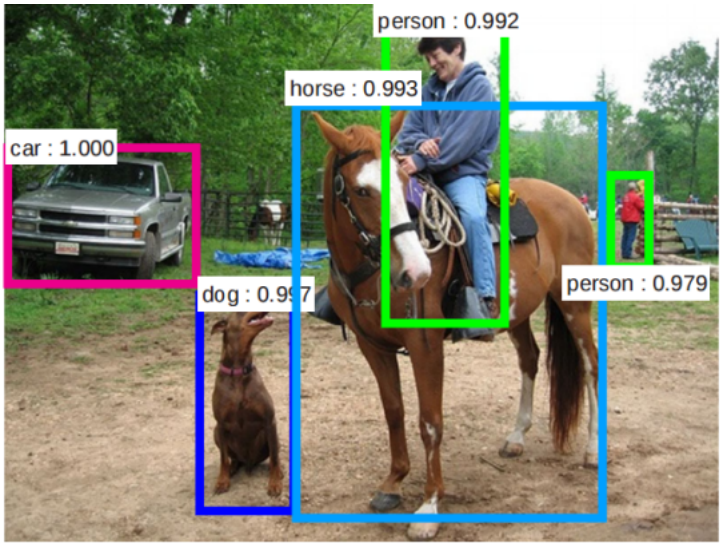
Với sự nổi lên của học sâu trong những năm gần đây, nhiều thuật toán phát hiện đối tượng được đề xuất. Điển hình như YOLO (You Only Look Once), R-CNN (Region-based Convolutional Network), Mask R-CNN (Mask Regionbased Convolutional Network), SSD (Single Shot Detection),... Không có mô hình nào được gọi là tốt nhất, tùy vào bài toán và mục đích mà chúng ta sẽ vận dụng cho phù hợp. Ở đây, chúng tôi sử dụng 3 mô hình YOLOv4, YOLOv5 và CenterNet để thực nghiệm cho bài toán của chúng tôi. Kết quả thực nghiệm được phân tích và đánh giá để chọn ra mô hình tốt nhất. Bộ dữ liệu cũng được chúng tôi tự thu thập để đáp ứng được yêu cầu của bài toán.

2 Cơ Sở Lí Thuyết

2.1 Bài Toán Phát Hiện Đối Tượng

Phát hiện đối tượng có lẽ là một trong những vấn đề được nghiên cứu nhiều nhất trong cả thị giác máy tính và trí tuệ nhân tạo. Điều này là do có rất nhiều ứng dụng phát hiện đối tượng, trải dài từ truyền thôn, tài chính, an ninh, giải trí, giáo dục, giao thông vận tải, sản xuất,... Một số ứng dụng phát hiện đối tượng thú vị bao gồm nhận diện khuôn mặt, phát hiện xe, phát hiện người đi bộ, đếm đối tượng, xe tự lái, phát hiện đối tượng bất thường. Là một trong những chủ đề quan trọng của thị giác máy tính và tí tuệ nhân tạo, phát hiện đối tượng có lịch sử lâu đời trong nghiên cứu học thuật. Phát hiện đối tượng có thể được phân tách thành hai vấn đề con, đó là định vị đối tượng và phân loại đối tượng. Định vị đối tượng là nhiệm vụ ước tính xem có đối tượng trong ảnh hay không và định vị vị trí của chúng. Phân loại đối tượng là nhiệm vụ xác định những đối tượng đó thuộc lớp nào. Kết hợp hai nhiệm vụ, phát hiện đối tượng có nghĩa là xác định vị trí ô tô, con người, động vật,.. như Hình 1.

Là một vấn đề lớn, làm cơ sở cho nhiều bài toán khác trong lĩnh vực thị giác máy tính nên có nhiều thuật toán đã được đề xuất cho bài toán phát hiện đối tượng. Đặc biệt, với sự bùng nổ của học sâu, nhiều thuật toán mới ra đời được gọi là các thuật toán state-of-the-art, với độ chính xác và tốc độ xử lí vượt trội hơn rất nhiều so với các thuật tóan truyền thống, thậm chí có thể phát hiện đối tượng trong thời gian thực (Real-Time Object Detection). Một số



Hình 1. Phát hiện đối tượng trong bức ảnh

kiến trúc mạng học sâu nổi bật hiện nay như YOLO (You Only Look Once), R-CNN (Region-based Convolutional Network), Mask R-CNN (Mask Regionbased Convolutional Network), SSD (Single Shot Detection),... Mỗi thuật toán đều có ưu nhược điểm riêng của chúng nên tùy vào mục đích sử dụng mà chúng ta áp dụng mô hình cho phù hợp để đạt kết quả tốt nhất.

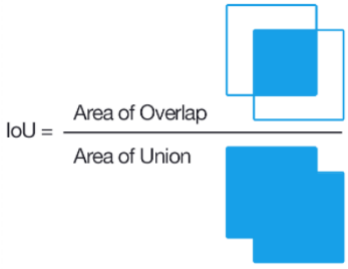
2.2 Các Độ Đo Đánh Giá

Mỗi mô hình được xây dựng đều có các độ đo đánh giá để đánh giá chất lượng cũng như hiệu suất mô hình đó. Đối với bài toán phát hiện đối tượng, chúng tôi sử dụng 2 độ đo chính để đánh giá là mAP (mean Average Precision) và IoU (Intersection over Union).

Intersection over Union (IoU) là chỉ số đánh giá được sử dụng để đo độ chính xác của phát hiện đối tượng trên tập dữ liệu cụ thể, là tỉ lệ giữa đo lường mức độ giao nhau giữa hai hộp giới hạn (bounding box) (thường là hộp giới hạn dự đoán và hộp giới hạn thực) để nhằm xác định hai khung hình có bị đè chồng lên nhau không. Tỷ lệ này được tính dựa trên phần diện tích giao nhau gữa 2 hộp giới hạn với phần tổng diện tích giao nhau và không giao nhau giữa chúng, như thể hiện ở Hình 2. Thông thường nếu IoU > 0.5 thì đối tượng được xem là nhận dạng đúng.

Mean Average Precision (mAP) hay còn được gọi là Area Under the Curve (AUC) là độ đo dùng để đánh giá mô hình dựa trên việc thay đổi một ngưỡng và quan sát giá trị Precision và Recall.

Giả sử có N ngưỡng để tính Precision và Recall, với mỗi ngưỡng cho một cặp giá trị Precision và Recall là P_n, R_n, n = 1, 2, .., N. Precision-Recall được vẽ



Hình 2. Cách xác định IoU

bằng cách vẽ từng điểm có tọa độ (P_n, R_n) trên trục tọa độ và nối chúng vơi nhau. Average Precision (AP) được tính bằng:

$$AP = \sum^{N}(R_n - R_{n-1})P_n \tag{1}$$

Trong đó, $(R_n - R_{n-1})P_n$ là diện tích hình chữ nhận có chiều rộng là $R_n - R_{n-1}$ và chiều cao là P_n.

mAP là trung bình AP cho tất cả các lớp.

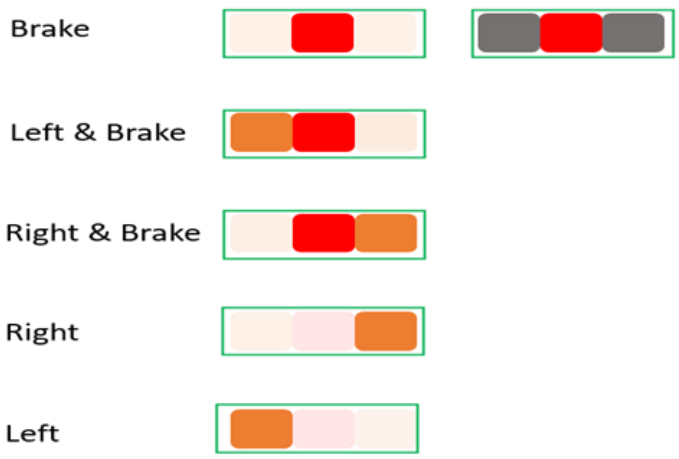
3 Hướng Tiếp Cận

3.1 Bộ Dữ Liệu

Để phục vụ cho bài toán đặt ra, chúng tôi đã xây dựng bộ dữ liệu về đèn tín hiệu xe. Bộ dữ liệu gồm 10 loại nhãn (lớp) được chú thích ở Hình 3, được chia theo tỉ lệ 8:2:1 với 4473 hình ảnh huấn luyện và 1280 ảnh kiểm định và 639 ảnh kiểm tra. Các hình ảnh được tách từ các video ngắn ở những địa điểm khác nhau ở Thành phố Hồ Chí Minh, Đồng Nai, Vũng Tàu, trung bình mỗi ảnh có từ 4-7 đối tượng. Cuối cùng, chúng tôi tiến hành gán nhãn cho các đèn tín hiệu xuất hiện trong ảnh theo qui tắc như ở Hình 4. Công cụ được sử dụng để gán nhã là LabelImg.

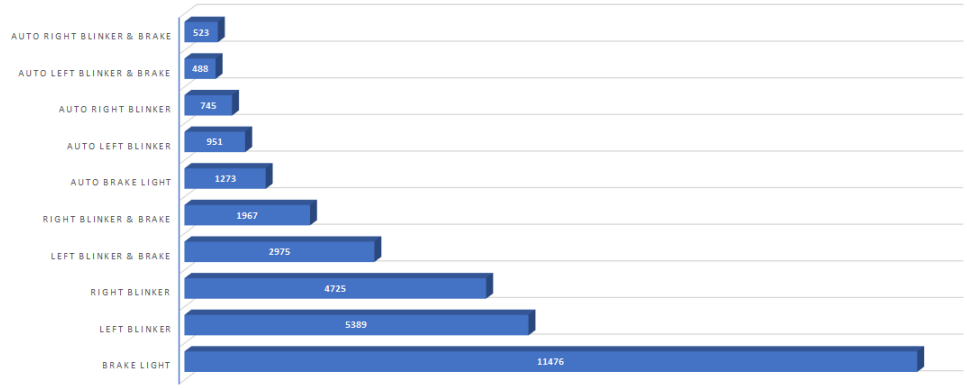
	Motobike	**Automobile**
Phanh	Brake Light	Auto Brake Light
Xi nhan trái	Left Blinker	Auto Left Blinker
Xi nhan phải	Right Blinker	Auto Right Blinker
Xi nhan trái và Phanh	Left Blinker & Brake	Auto Left Blinker & Brake
Xi nhan phải và Phanh	Right Blinker & Brake	Auto Right Blinker & Brake

Hình 3. Chú thích các lớp



Hình 4. Qui tắc gán nhãn

Các nhãn sau khi gán không đều nhau, rất ít nhãn về xe ô tô, cụ thể được chúng tôi thể hiện ở phân phối Hình 5.



Hình 5. Sự phân bố các nhãn trên tập dữ liệu

3.2 Thuật Toán

Để có được kết quả tốt nhất cho bài toán, chúng tôi sẽ sử dụng 3 mô hình học sâu được đánh giá là tốt nhất hiện nay là YOLOv4, YOLOv5 và Faster R-CNN.

Trong đó YOLOv4 đang là một thuật toán state-of-the-art ở hiện tại, YOLOv5 hiện vẫn chưa có một công trình nào nói về nó.

YOLOv4 (You Only Look Once version 4): YOLO là một mô hình mạng CNN cho việc phát hiện, nhận dạng, phân loại đối tượng. YOLO được tạo ra từ việc kết hợp giữa các convolutional layers và fully connected layers. Trong đó các convolutional layers sẽ trích xuất ra các feature của ảnh, còn fully connected layers sẽ dự đoán ra xác suất đó và tọa độ của đối tượng. Trải qua các phiên bản cải tiến, YOLOv4 hiện tại là một trong các thuật toán state-of-the-art với độ chính xác và tốc độ xử lí cực kì ấn tượng. Kiến trúc mô hinh YOLOv4 gòm có 3 phần chính:

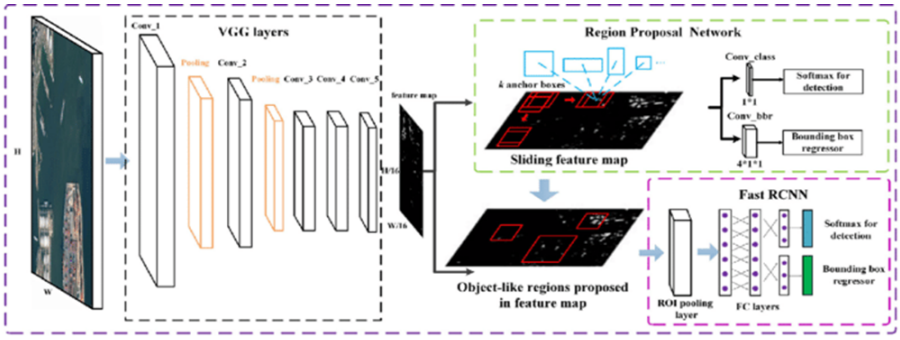
- Backbone: YOLO4 sử dụng CSPDarknet53 để làm backbone vì theo tác giả, CSPDarknet53 có độ chính xác trong task object detection cao hơn so với ResNet, và mặc dù ResNet có độ chính xác trong task classification cao hơn, hạn chế này có thể được cải thiện nhờ hàm activation Mish và một vài kỹ thuật sẽ được đề cập phía dưới.
- Neck: có nhiệm vụ trộn và kết hợp các features map đã học được thông qua quá trình trích xuất đặc trưng ở Backbone và Head. Tác giả của YOLOv4 đã cho phép tùy biến sử dụng các cấu trúc cho phần Neck như là FPN, PAN, NAS-FPN, BiFPN...
- Head: có nhiệm vụ phân loại các lớp và đưa ra dự đoán về vị trí các bounding box, có cấu trúc tương tự như Head của phiên bản YOLOv3.

Ngoài ra, YOLOv4 còn sử dụng những phương pháp trong Bag of Freebies (BoF) và Bag of Specials (BoS):

- Bag of Freebies: những phương pháp giúp cải thiện kết quả inference mà không làm ảnh hưởng tới tốc độ inference. Những phương pháp này thường là data augmentation, class imbalance, cost function, soft labeling...
- Bag of Specials: Những phương pháp hi sinh một chút tốc độ inference mà làm cải thiện độ chính xác của model đáng kể. Những phương pháp này bao gồm tăng receptive field, sử dụng attention, feature intergration (kết hợp thông tin của các feature maps với nhau) như skip-connection và FPN (Feature Paramyd Network), hậu xử lý như NMS (Non Maximum Suppression).

Faster R-CNN: Faster R-CNN là một phiên bản cải tiến, giải quyết một số hạn chế của 2 phiên bản trước đó là R-CNN và Fast R-CNN. Faster R-CNN có kiến trúc như Hình 6.

Faster R-CNN không dùng thuật toán selective search để lấy ra các region proposal, mà nó thêm một mạng CNN mới gọi là Region Proposal Network (RPN) để tìm các region proposal. Đầu tiên, cả bức ảnh được cho qua một mạng deep CNN để lấy feature map. Khác với Fast R-CNN, kiến trúc này không tạo RoI ngay trên feature map mà sử dụng feature map làm đầu vào để xác định các region proposal thông qua một RPN network. Đồng thời feature maps cũng là đầu vào cho classifier nhằm phân loại các vật thể của region proposal xác định được từ RPN network.



Hình 6. Kiến trúc mô hình Faster R-CNN

4 Kết Quã Thực Nghiệm Và Đánh giá

Sau khi áp dụng áp dụng 3 mô hình trên vào bộ dữ liệu đèn tín hiệu xe, chúng tôi thu được kết quả như Hình 7, 8 và 9 bên dưới:

Model	Average IOU
Faster R - CNN	57.23
YOLOv4	65.08
YOLOv5	62.20

Hình 7. Trung bình IoU

Ta nhận thấy, mô hình YOLOv4 cho kết quả IoU tốt hơn so với 2 mô hình còn lại, tuy nhiên, với giá trị IoU = 65.08% thì vẫn chưa thể áp dụng vào thực tế vì còn quá thấp, ít nhất phải trên 85%. Đối với mAP, với giá trị ngưỡng IoU

Model	mAP@0.5	mAP@0.75
Faster R-CNN	0.810	0.5677
YOLOv4	0.962	**0.717**
YOLOv5	**0.981**	0.609

Hình 8. mAP theo từng ngưỡng IoU

= 0.5 thì mô hình YOLOv5 cho kết quả cao nhất với mAP = 98.1% nhưng với ngưỡng IoU = 0.75 thì mô hình YOLOv4 lại cho kết quả cao nhất là 71.7%. Đi sâu vào từng lớp cụ thể, ta dễ dàng nhận thấy mô hình YOLOv4 cho kết quả

	Brake Light	Left Blinker	Right Blinker	Left Blinker & Brake	Right Blinker & Brake
Faster R-CNN	0.7317	0.6500	0.6140	0.5249	0.6710
YOLOv4	0.8268	0.8371	0.8005	0.8088	0.8234
YOLOv5	0.9730	0.9457	0.8714	0.8329	0.8544
	Auto Brake Light	**Auto Left Blinker**	**Auto Right Blinker**	**Auto Left Blinker & Brake**	**Auto Right Blinker & Brake**
Faster R-CNN	0.4483	0.4926	0.5270	0.4720	0.5459
YOLOv4	0.4886	0.5880	0.6872	0.6462	0.6633
YOLOv5	0.2370	0.3192	0.3478	0.3510	0.3640

Hình 9. mAP theo từng lớp

tốt và ổn định với tất cả các lớp. YOLOv5 thì chỉ cho kết quả tốt ở các lớp xe máy, đối với các lớp xe ô tô thì kết quả lại rất thấp. Từ tất cả những đánh giá trên, chúng tôi nhận định rằng mô hình YOLOv4 là mô hình tốt nhất cho bài toán của chúng tôi. Tuy nhiên kết quả đạt được vẫn chưa thật sự tốt, theo nhận định khách quan ban đầu, nguyên nhân có thể nằm ở vấn đề dữ liệu chưa đủ chất lượng, chưa cân bằng giữa các lớp, đặc biệt là các lớp xe ô tô còn rất ít.

5 Kết Luận Và Hướng Phát Triển

Trong bài báo này, chúng tôi đã trình bày chi tiết quá trình nghiên cứu, thực nghiệm và kết quả của bài toán phát hiện đèn tín hiệu xe. Với những quả đã được tương đối tốt ở mô hình được chọn là YOLOv4, ta có thể thấy được đây là một bài toán có tiềm năng phát triển và ứng dụng vào thực tế rất cao, đóng góp vào bài toán lớn về hệ thống giao thông thông minh. Trong tương lai, chúng tôi sẽ tiếp tục nghiên cứu, thực nghiện trên nhiều mô hình khác để cải thiện kết quả bài toán. Một trong những hướng phát triển chúng tôi sẽ nghiên cứu sắp tới là ứng dụng bộ lọc Kalman vào kết quả bài toán để làm mượt chuỗi kết quả các bouding box, gia tâng độ chính xác của mô hình.

Tài liệu

1. ntoniou, C., Ben-Akiva, M., and amp; N., H. (2010). Kalman Filter Applications for Traffic Management. Kalman Filter. https://doi.org/10.5772/9583.
2. u, X., Si, Y., and amp; Li, L. (2019). Pedestrian detection based on improved Faster RCNN algorithm. 2019 IEEE/CIC International Conference on Communications in China (ICCC). https://doi.org/10.1109/iccchina.2019.8855960 .
3. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
4. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014.

5. Redmon, J., Divvala, S., Girshick, R. Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
6. Z.-Y. Liu, Q. Ye, F. Li, M.-H. Zhao, J.-S. Nie, and X.-Q. Sun, "Taillight detection algorithm based on four thresholds of brightness and color," Computer Engineering, vol. 36, no. 21, pp. 202–206, 2010.
7. R. O'Malley, E. Jones, and M. Glavin, "Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions," IEEE Transactions on Intelligent Transportation Systems, vol. 11, no. 2, pp. 453–462, 2010.