

Xây Dựng Bộ Dữ Liệu Đèn Sau Xe Bốn Bánh Phục Vụ Quản Lí Giao Thông Ban Đêm Qua Camera Giám Sát

Đỗ Trọng Hợp*, Trần Đăng Khoa[†], Hoàng Đình Quang[†], Nguyễn Thế Mạnh[†]

Email: *{hopdt}@uit.edu.vn

[†]{18520936,18521294, 18521084}@gm.uit.edu.vn

Tóm tắt nội dung—Thế giới ngày nay đang trong thời kì công nghệ 4.0, dữ liệu là một trong những yếu tố quan trọng góp phần vào sự phát triển của công nghệ thông tin. Tuy nhiên, số lượng bộ dữ liệu có chất lượng và được công bố rộng rãi phục vụ cho nghiên cứu lại rất ít. Vì vậy, trong bài báo này, chúng tôi cung cấp một bộ dữ liệu phục vụ cho quản lí giao thông qua camera giám sát - bộ dữ liệu đèn sau xe bốn bánh (UIT-TLD). Chúng tôi thu thập, xử lí, tạo bộ dữ liệu sau đó tiến hành thực nghiệm và đánh giá bằng mô hình YOLOv4 - một mô hình state-of-the-art hiện nay. Kết quả thu được khá khả quan với IoU 78.76% và mAP 98.78%. Từ đó chúng tôi chỉ ra ưu nhược điểm của bộ dữ liệu và đưa ra hướng phát triển trong tương lai.

Từ khóa—Xây dựng bộ dữ liệu, đèn sau xe, UIT-TLD.

I. GIỚI THIỆU

Trong những năm gần đây, với sự tiến bộ vượt bậc của khoa học và công nghệ, đặc biệt là trong lĩnh vực trí tuệ nhân tạo (AI), nhiều mô hình AI được tạo ra hỗ trợ con người trong hầu hết các lĩnh vực như công nghiệp, nông nghiệp, y tế, giáo dục,... đem lại hiệu suất, độ chính xác cao hơn con người rất nhiều lần, thậm chí có thể làm được những công việc mà con người không thể làm. Một trong những yếu tố quan trọng góp phần tạo nên sự thành công của mô hình AI chính là dữ liệu. Mô hình sẽ hoạt động kém hiệu quả nếu được xây dựng trên một bộ dữ liệu không phù hợp. Do đó, tạo bộ dữ liệu phù hợp là một trong những bước chiếm nhiều thời gian và công sức nhất trong chu trình AI.

Như chúng tôi đã nói ở trên, các ứng dụng của AI đang lan rộng ra hầu hết các lĩnh vực, và giao thông là một trong những lĩnh vực đang được quan tâm hàng đầu. Hiện nay, quản lí giao thông luôn là một vấn đề quan trọng ở bất kì quốc gia nào trên thế giới. Số lượng các phương tiện giao thông gia tăng một cách nhanh chóng, hoạt động liên tục cả ngày lẫn đêm khiến cho việc quản lí giao thông trở nên ngày càng khó khăn. Sự ra đời của các hệ thống quản lí giao thông thông minh để hỗ trợ con người là một điều tất yếu. Tuy nhiên, dữ liệu có sẵn để phục vụ cho các bài toán giao thông hiện nay không nhiều. Ngoài ra, với mỗi một bài toán cụ thể, chúng ta có rất nhiều hướng tiếp cận, mỗi hướng lại đòi hỏi một bộ dữ liệu riêng để đem lại hiệu quả tối ưu nhất. Vì vậy dữ liệu càng trở nên khan hiếm. Để góp phần cung cấp thêm dữ liệu về giao thông, chúng tôi đã xây dựng một bộ dữ liệu về đèn sau xe bốn bánh (UIT-TLD) cho bài toán phát hiện phương tiện giao thông ban đêm thông qua đèn sau xe, phục vụ cho quản lí giao thông ban đêm thông qua camera giám sát. Một số hình ảnh mẫu về bộ dữ liệu của chúng tôi được thể hiện ở Hình 1.

- Bộ dữ liệu UIT-TLD gồm 6018 hình ảnh về đèn sau xe bốn bánh được chúng tôi thu thập ở các con đường ở thành phố Hồ Chí Minh (Xa lộ Hà Nội, đường Mai Chí Thọ, đường Điện Biên Phủ, đường Võ Văn Kiệt) vào ban đêm từ 18 giờ - 23 giờ. Bộ dữ liệu sẽ được công bố công khai phục vụ cho việc nghiên cứu.



Hình 1: Một số hình ảnh mẫu trong bộ dữ liệu UIT-TLD

- Để có những cái nhìn bao quát, chúng tôi tiến hành phân tích bộ dữ liệu ở các khía cạnh như kích thước ảnh, số lượng đối tượng trong ảnh, định dạng nhãn.
- Bộ dữ liệu sẽ được tiến hành thực nghiệm trên mô hình YOLOv4 (You Only Look Once version 4), một mô hình state-of-the-art ở thời điểm hiện tại.

Các phần còn lại trong bài báo sẽ được cấu trúc như sau: Phần II chúng tôi sẽ giới thiệu các công trình liên quan, cụ thể là các bộ dữ liệu hiện có về giao thông, sau đó là quá trình tạo nên bộ dữ liệu đèn sau xe bốn bánh (UIT-TLD) cũng như một số phân tích cơ bản về bộ dữ liệu ở Phần III. Phần IV chúng tôi sẽ giới thiệu khái quát về phương pháp tiếp cận, tiến hành thực nghiệm trên bộ dữ liệu UIT-TLD và đánh giá ở Phần V và đưa ra kết luận cũng như những công việc trong tương lai ở Phần VI.

II. CÁC CÔNG TRÌNH LIÊN QUAN

Trong quá trình xây dựng bộ dữ liệu đèn sau xe bốn bánh (UIT-TLD), chúng tôi đã tìm hiểu và tham khảo một số bộ dữ liệu về giao thông hiện nay như BIT - Vehicle Dataset, KITTI Object Detection with Bounding Boxes, LISA Traffic Light Dataset, Nepalese Vehicles, Rain and Snow Traffic Surveillance, Stanford Cars Dataset, Semantic Segmentation for Self Driving Cars, TME Motorway Dataset. Các bộ dữ liệu trên sẽ được chúng tôi mô tả bên dưới:

- BIT - Vehicle Dataset: Từ Phòng thí nghiệm Công nghệ Thông tin thông minh Bắc Kinh, bộ dữ liệu này bao gồm 9,850 hình ảnh phương tiện. Hình ảnh được chia thành 6 loại theo loại xe: xe buýt, xe buýt nhỏ, xe tải nhỏ, xe sedan, SUV và xe tải.
- KITTI Object Detection with Bounding Boxes: được trích từ bộ dữ liệu tiêu chuẩn của Viện Công nghệ Karlsruhe, phần phát

hiện đối tượng. Bộ dữ liệu gồm 14,999 hình ảnh các phương tiện chia thành 7,481 ảnh huấn luyện và 7,518 ảnh thử nghiệm với các tệp chứa nhãn riêng biệt.

- **LISA Traffic Light Dataset:** là bộ dữ liệu về đèn giao thông, gồm 430,007 hình ảnh tách từ các đoạn video ban đêm, trong đó có 113,888 hình ảnh các đèn giao thông.
- **Nepalese Vehicles:** Bao gồm 30 video giao thông được quay trên đường phố Kathmandu, tập dữ liệu này bao gồm hình ảnh của 4,800 phương tiện được cắt từ các video đó. Trong số 4,800 hình ảnh, 1,811 hình ảnh là xe hai bánh và 2,989 hình ảnh là xe bốn bánh.
- **Rain and Snow Traffic Surveillance:** Tập dữ liệu này bao gồm 22 video về giao thông trong mưa và tuyết, mỗi video khoảng 5 phút. Các video được quay bằng cả camera màu RGB và camera nhiệt hồng ngoại. Do đó, dữ liệu bao gồm hơn 130,000 cặp hình ảnh.
- **Stanford Cars Dataset:** từ Phòng thí nghiệm AI của Stanford, bộ dữ liệu này bao gồm 16,185 hình ảnh với 196 loại ô tô khác nhau.
- **TME Motorway Dataset:** Bao gồm 28 video có thời lượng 27 phút, bộ dữ liệu này bao gồm hơn 30,000 khung hình xe mô tô.

Cho tới hiện tại, bài toán phát hiện đèn xe ban đêm đã được nghiên cứu và công bố như Study on Leading Vehicle Detection at Night Based on Multisensor and Image Enhancement Method. Mathematical Problems in Engineering (Chen và các cộng sự, 2016), Study on Leading Vehicle Detection at Night Based on Multisensor and Image Enhancement Method. Mathematical Problems in Engineering (Z.-Y. Liu và các cộng sự, 2010), Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions (R. O'Malley và các cộng sự, 2010),... Tuy nhiên, vẫn chưa có một bộ dữ liệu về đèn sau xe được công bố rộng rãi phục vụ cho nghiên cứu. Đó là động lực để chúng tôi tiến hành xây dựng và phát triển bộ dữ liệu đèn sau xe bốn bánh (UIT-TLD).

III. BỘ DỮ LIỆU

Trong phần này, chúng tôi trình bày quy trình thu thập, các thông tin cơ bản về bộ dữ liệu và các thách thức mà chúng tôi phải đối mặt khi xây dựng bộ dữ liệu UIT-TLD. Quy trình xây dựng bộ dữ liệu được thể hiện ở Hình 2.



Hình 2: Quy trình tạo bộ dữ liệu UIT-TLD

A. Thu thập dữ liệu

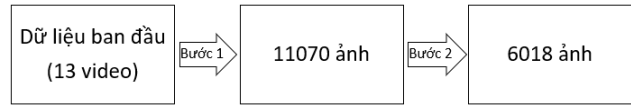
- **Địa điểm:** các con đường chỉ có xe bốn bánh ở thành phố Hồ Chí Minh (đường Mai Chí Thọ, xa lộ Hà Nội, đường Điện Biên Phủ, đường Võ Văn Kiệt)
- **Thời gian:** toàn bộ quá trình thu thập dữ liệu đều được thực hiện vào ban đêm trong khung giờ 18 giờ - 23 giờ.
- **Công cụ:** chúng tôi sử dụng camera điện thoại Apple iPhone X làm công cụ thu thập dữ liệu.
- **Cách thức thu thập:** dùng camera để quay các video hình ảnh các phương tiện bốn bánh vào thời gian và địa điểm đã được xác định. Góc quay phía sau từ trên xuống ở giữa hoặc bên trái hoặc bên phải làn đường.
- **Kết quả:** sau bước thu thập dữ liệu, chúng tôi thu được 13 video kích thước 608x1080, tốc độ khung hình là 30fps và tổng độ dài các video là 369 giây.

B. Tiền xử lý dữ liệu

Sau khi có được dữ liệu, chúng tôi tiến hành tiền xử lý dữ liệu với 2 bước như qui trình ở Hình 3:

- **Bước 1:** Tách các video thành các ảnh riêng biệt.
- **Bước 2:** Loại bỏ các ảnh không chứa đối tượng.

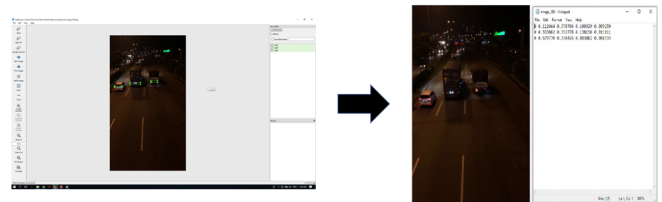
Kết quả sau khi tiền xử lý chúng tôi thu được 6018 ảnh.



Hình 3: Quy trình tiền xử lý dữ liệu

C. Gán nhãn dữ liệu

Cuối cùng, với 6018 ảnh thu được sau khi tiền xử lý chúng tôi tiến hành dùng công cụ LabelImg để gán nhãn cho chúng. Nhãn được gán theo định dạng đầu vào của thuật toán YOLO gồm 5 thông số lần lượt là a, x, y, w, h. Trong đó a là nhãn của đối tượng (trong bộ dữ liệu này chỉ có một nhãn là Light), x, y là tọa độ tâm của hộp giới hạn, w, h là tỉ lệ chiều rộng và chiều cao của hộp giới hạn so với toàn bộ hình ảnh. Hình 4 thể hiện bước gán nhãn cho dữ liệu dùng công cụ LabelImg.



Hình 4: Gán nhãn dữ liệu

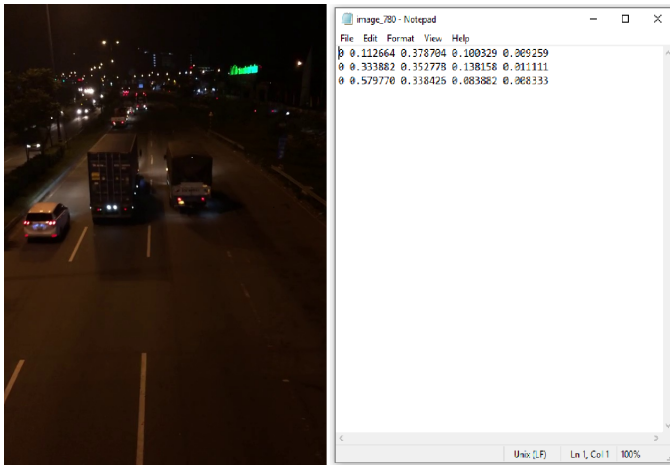
D. Thông tin về bộ dữ liệu

Bộ dữ liệu hoàn chỉnh gồm 6,018 ảnh kích thước 608x1080, 6,180 tệp chứa thông tin nhãn và vị trí hộp giới hạn của các đối tượng trong ảnh tương ứng và 1 tệp chứa thông tin các nhãn của bộ dữ liệu (ở đây chỉ có 1 nhãn là Light). Tên tệp ảnh và tệp nhãn giống nhau hoàn toàn. Số đối tượng trong 1 ảnh trong khoảng từ 1-7, tập trung chủ yếu khoảng 3-4 đối tượng trong 1 ảnh. Chúng tôi tiến hành chia bộ dữ liệu thành 2 tập huấn luyện và kiểm tra với 5,000 ảnh huấn luyện và 1,018 ảnh kiểm tra. Một ảnh mẫu cùng với tệp nhãn tương ứng được chúng tôi thể hiện ở Hình 5.

E. Thách thức của bộ dữ liệu

Trong quá trình xây dựng bộ dữ liệu UIT-TLD, chúng tôi đã gặp phải một số thách thức:

- **Thiếu ngữ cảnh:** là một bài toán về giao thông, bộ dữ liệu cần phải bao quát được tất cả các tình huống có thể xảy ra và ở mọi góc độ. Tuy nhiên, bộ dữ liệu UIT-TLD chỉ được thu thập ở 4 con đường chính là đường Mai Chí Thọ, đường Võ Văn Kiệt, đường Điện Biên Phủ, xa lộ Hà Nội cũng như chưa sự đa dạng về tình huống và góc quay nên đã bộc lộ một số hạn chế ảnh hưởng đến chất lượng bộ dữ liệu.
- **Chất lượng hình ảnh thấp:** do điều kiện cơ sở vật chất còn hạn chế, sử dụng camera trên iPhone X để thu thập nên chất lượng hình ảnh chưa cao, có thể ảnh hưởng xấu đến quá trình huấn luyện mô hình.
- **Dữ liệu nhiễu:** do đối tượng của bộ dữ liệu là đèn sau xe mà hiện nay, kiểu dáng và vị trí của đèn xe rất đa dạng nên dễ gây nhầm lẫn cho việc phát hiện của mô hình. Ngoài ra, khi ở góc



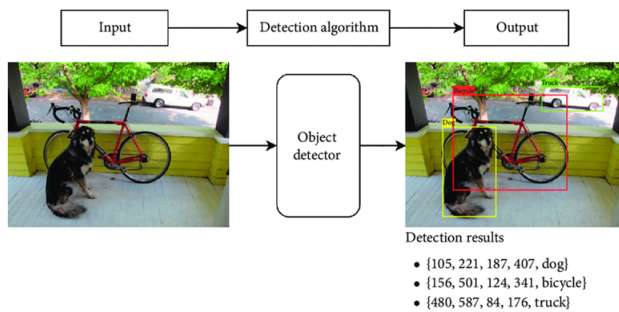
Hình 5: Ảnh dữ liệu mẫu và tệp nhãn tương ứng

quay từ 2 bên, mô hình thường nhầm lẫn đèn bên hông xe là đèn sau, gây ảnh hưởng đến kết quả phát hiện của mô hình.

IV. PHƯƠNG PHÁP TIẾP CẬN

A. Bài toán phát hiện vật thể (Object Detection)

Object Detection là một bài toán nằm trong lĩnh vực thị giác máy tính, nhằm xác định vị trí và gán nhãn tất cả các đối tượng được quan tâm có trong bức hình. Theo đó, đầu vào của bài toán này là ảnh và danh sách các đối tượng quan tâm. Đầu ra của bài toán là các vị trí của các đối tượng kèm theo nhãn của đối tượng đó, như Hình 6. Nói cách khác, Object Detection là sự kết hợp của Image Classification (phân loại hình ảnh) và Object Localization (định vị vật thể). Object Detection đã có rất nhiều ứng dụng khác nhau: Image Annotation (Chú thích hình ảnh), Face Detection (Phát hiện khuôn mặt), License Plate Identification (Nhận diện biển số xe), People Counting (Đếm số lượng người),...



Hình 6: Bài toán Object Detection

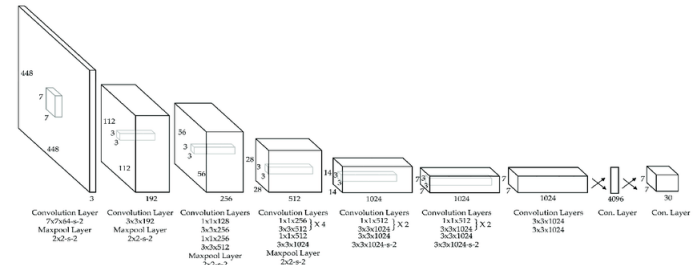
B. Các thuật toán phát hiện đối tượng

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 Object Detection. Đặc biệt, với sự bùng nổ của Deep Learning, nhiều thuật toán mới ra đời được gọi là các thuật toán state-of-the-art, với độ chính xác và tốc độ xử lý vượt trội hơn rất nhiều so với các thuật toán truyền thống, thậm chí có thể phát hiện đối tượng trong thời gian thực (Real-Time Object Detection). Một số kiến trúc mạng học sâu nổi bật hiện nay như YOLO (You Only Look Once), R-CNN (Region-based Convolutional Neural Network), Mask R-CNN

(Mask Region-based Convolutional Neural 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.

C. Thuật toán đề xuất

Trong bài báo này, thuật toán được chúng tôi đề xuất để tiến hành thực nghiệm là YOLOv4 (You Only Look Once version 4) vì trong một bài toán về giao thông, tốc độ xử lý yêu cầu càng nhanh càng tốt để có thể giải quyết kịp thời các vấn đề xảy ra. Tốc độ xử lý của các mô hình YOLO4 nhanh hơn rất nhiều so với các thuật toán hiện nay trong khi độ chính xác bằng hoặc chỉ thấp hơn một chút.



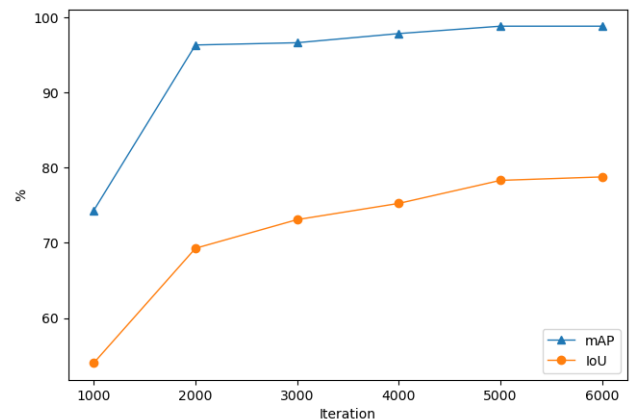
Hình 7: Kiến trúc thuật toán YOLO

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. Kiến trúc cơ bản của YOLO được thể hiện ở Hình 7. 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 lý cực kì ấn tượng. Ngoài ra, YOLOv4 còn có các phương pháp giúp thuật toán tăng độ chính xác mà không làm giảm tốc độ chạy (Bag of Freebies), các phương pháp hoặc các mô-đun ảnh hưởng nhẹ đến tốc độ chạy, nhưng đem cải thiện độ chính xác cho model một cách đáng kể (Bag of Special).

V. KẾT QUẢ THỰC NGHIỆM VÀ ĐÁNH GIÁ

A. Kết quả thực nghiệm

Sau khi áp dụng mô hình YOLOv4 cho bộ dữ liệu UIT-TLD, kết quả được chúng tôi thu thập và thể hiện ở Hình 8. Toàn bộ kết quả này đều được đánh giá trên dữ liệu kiểm tra.



Hình 8: Kết quả IoU và mAP thực nghiệm

Với mAP (mean Average Precision) 98.78% thì mô hình này được chúng tôi đánh giá khá tốt, tuy nhiên IoU (Intersection over Union)

= 78.76% cho thấy khả năng xác định vị trí của đối tượng còn hạn chế. Hình ảnh thử nghiệm trên hình ảnh kiểm tra cũng được thể hiện ở Hình 9.



Hình 9: Kết quả trên ảnh kiểm tra

B. Phân tích đánh giá

Với kết quả thực nghiệm khá tốt khi sử dụng mô hình YOLOv4 ta có thể thấy bộ dữ liệu tương đối tốt nhưng vẫn còn một vài hạn chế, chưa có nhiều thách thức để có thể nghiên cứu thêm trong tương lai. Chúng tôi có một số nhận định về nguyên nhân hạn chế của bộ dữ liệu như sau:

- Chỉ có một nhãn là "Light": đối với các mô hình học sâu, việc xác định sai nhãn cho đối tượng là một trong những nguyên nhân chủ yếu làm cho kết quả thấp. Vì vậy, trong bộ dữ liệu UIT-TLD, mô hình YOLOv4 cho kết quả cao cũng có thể phần nào giải thích được nguyên nhân là do chỉ có duy nhất 1 nhãn, mô hình chỉ cần phát hiện đúng vật thể là xem như thành công.
- Toàn bộ dữ liệu đều tương đối giống nhau ở cả 2 tập huấn luyện và kiểm tra làm cho kết quả cho ra khá tốt.
- Bộ dữ liệu còn đơn giản, chưa có nhiều tình huống gây khó khăn, thách thức cho mô hình như kẹt xe, xe trùng nhau,...

VI. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong bài báo này, chúng tôi đã trình bày quá trình xây dựng bộ dữ liệu đèn sau xe bốn bánh (UIT-TLD). Sau đó tiến hành thực nghiệm sử dụng thuật toán YOLOv4, kết quả thu được tương đối tốt với IoU và mAP lần lượt là 78.76% và 98.78%. Tuy còn nhiều hạn chế nhưng những hạn chế đó hoàn toàn có thể khắc phục được trong tương lai nếu cung cấp thêm nhiều bối cảnh thách thức cho dữ liệu.

Trong tương lai, chúng tôi sẽ tiến hành thu thập thêm nhiều dữ liệu trong các tình huống phức tạp, cũng như có thể phát triển bộ dữ liệu thêm đèn trước hoặc áp dụng lên với các loại phương tiện giao thông khác để tăng tính thách thức cho bộ dữ liệu, phục vụ cho việc nghiên cứu và có thể ứng dụng vào thực tế để phục vụ công đồng.

TÀI LIỆU

- [1] 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.
- [2] 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.
- [3] Wang, Y., Wang, L., Jiang, Y., and Li, T. (2020). Detection of Self-Build Data Set Based on YOLOv4 Network. 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE).

- [4] Zheng, Zhaohui and Wang, Ping and Liu, Wei and Li, Jinze and Ye, Rongguang and Ren, Dongwei. (2020). Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. Proceedings of the AAAI Conference on Artificial Intelligence.
- [5] Bochkovskiy, Alexey and Wang, Chien-Yao and Liao, Hong-yuan. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection.
- [6] 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
- [7] Chen, M., Jin, L., Jiang, Y., Gao, L., Wang, F., and Xie, X. (2016). Study on Leading Vehicle Detection at Night Based on Multisensor and Image Enhancement Method. Mathematical Problems in Engineering, 2016, 1–13.
- [8] 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.
- [9] 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.
- [10] H. Wu, H. Huo, T. Fang et al., "Nighttime video vehicle detection in complex environment," Application Research of Computers, vol. 24, no. 12, pp. 386–389, 2007.
- [11] Kiet Van Nguyen, Duc-Vu Nguyen, Anh Gia-Tuan Nguyen, Ngan Luu-Thuy Nguyen. A Vietnamese Dataset for Evaluating Machine Reading Comprehension, 2020.