

NGHIÊN CỨU ỨNG DỤNG MẠNG HỌC SÂU TRONG BÀI TOÁN XÁC ĐỊNH
LÀN ĐƯỜNG TRONG CÁC HỆ THỐNG XE TỰ LÁIRESEARCH ON APPLYING DEEP LEARNING NETWORKS IN LANE
DETECTION FOR SELF-DRIVING VEHICLE SYSTEMS

LÊ QUYẾT TIẾN*, TRẦN THỊ HƯƠNG

Khoa Công nghệ Thông tin, Trường Đại học Hàng hải Việt Nam

*Email liên hệ: tienlqcnt@vimaru.edu.vn

Tóm tắt

Công nghệ phát hiện làn đường tự động đóng vai trò quan trọng trong việc giúp xe tự lái định vị chính xác trong môi trường giao thông đô thị có nhiều làn đường. Tuy nhiên, việc nhận diện vạch kẻ đường dưới các điều kiện thời tiết khác nhau vẫn là một thách thức lớn đối với các phương pháp xử lý hình ảnh truyền thống và thị giác máy tính. Nghiên cứu này đóng góp đáng kể bằng cách áp dụng mạng học sâu để giải quyết bài toán xác định làn đường trong hệ thống xe tự lái. Cụ thể, nhóm tác giả đã sử dụng mô hình Encoder-Decoder với mạng tích chập hoàn toàn nhằm phát hiện làn đường một cách hiệu quả. Quá trình nghiên cứu và thực nghiệm được tiến hành kỹ lưỡng, mang lại kết quả ấn tượng, với độ chính xác đạt tới 97,82%.

Từ khóa: Phát hiện làn đường, học sâu.**Abstract**

Automatic lane detection technology plays a crucial role in enabling self-driving cars to accurately position themselves in multi-lane urban traffic environments. However, recognizing lane markings under different weather conditions remains a significant challenge for traditional image processing methods and computer vision techniques. This study makes a substantial contribution by applying deep learning networks to address the problem of lane detection in autonomous vehicle systems. Specifically, the authors employed an Encoder-Decoder model with a fully convolutional network to effectively detect lanes. The research and experimental processes were conducted rigorously, yielding impressive results with an accuracy of up to 97.82%.

Keywords: Lane detection, deep learning.**1. Giới thiệu**

Hiện nay, sự gia tăng nhanh chóng của số lượng ô tô đang đặt ra những thách thức lớn đối với an toàn giao thông đường bộ. Cơ sở hạ tầng giao thông không phát triển đồng bộ với sự gia tăng của phương tiện, dẫn đến tình trạng ùn tắc và tai nạn ngày càng gia tăng. Do đó, việc kiểm soát và hỗ trợ phương tiện tham gia giao thông trở thành một vấn đề quan trọng, thu hút sự quan tâm lớn từ cộng đồng nghiên cứu. Các giải pháp thông minh ứng dụng trí tuệ nhân tạo đang dần chứng tỏ ưu thế trong việc giải quyết những thách thức này.

Trong đó, phát hiện làn đường đóng vai trò then chốt trong việc phát triển các hệ thống xe tự hành cũng như nâng cao an toàn giao thông. Đây là một bài toán quan trọng, thu hút sự nghiên cứu của giới khoa học trong suốt thời gian qua. Các phương pháp học máy truyền thống trước đây đạt hiệu suất cao trong điều kiện lý tưởng nhưng lại gặp nhiều hạn chế khi áp dụng vào thực tế, tiềm ẩn nguy cơ mất an toàn cho người điều khiển phương tiện. Trong khi đó, với sự phát triển của học sâu, bài toán phát hiện làn đường đã được cải thiện đáng kể, giúp nâng cao độ chính xác và hiệu quả một cách rõ rệt.

2. Bối cảnh nghiên cứu

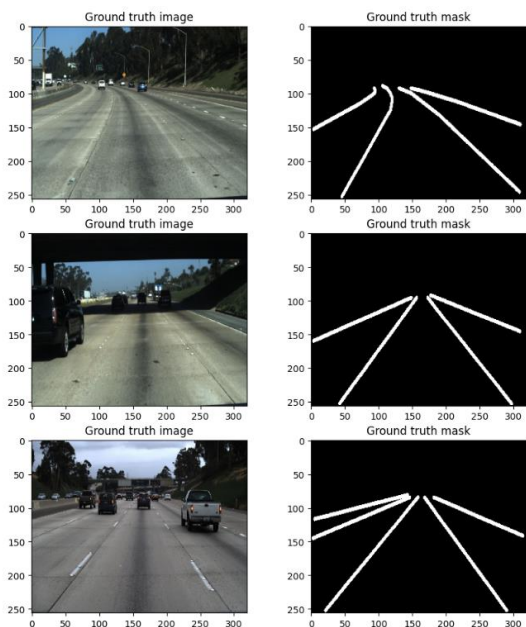
Trong những năm gần đây, nhiều kỹ thuật tiên tiến đã được phát triển để cải thiện khả năng nhận diện làn đường. Qin Z [1] đã nghiên cứu cách tận dụng chuỗi khung hình liên tiếp từ cảnh lái xe để xác định làn đường, đồng thời đề xuất một mô hình học sâu kết hợp giữa Mạng nơ-ron tích chập (CNN) và Mạng nơ-ron hồi quy (RNN). Han Ma [2] đưa ra một phương pháp phát hiện làn đường dựa trên việc tối ưu hóa bản đồ chênh lệch dày đặc. Tuy nhiên, những phương pháp này gặp khó khăn khi làm việc trong điều kiện phức tạp như ánh sáng yếu hoặc khi có sự can thiệp mạnh như giao cắt và rẽ.

Davy Neven [3] tiếp cận bài toán phát hiện làn đường theo hướng phân đoạn thể hiện, trong đó mỗi làn đường được xử lý như một thực thể độc lập và có thể huấn luyện từ đầu đến cuối. Để biểu diễn các làn

đường sau khi phân đoạn, ông đã đề xuất một phương pháp biến đổi phối cảnh có điều kiện trên hình ảnh, thay vì sử dụng góc nhìn cố định từ trên xuống. Ze Wang [4] giới thiệu LaneNet, một mô hình học sâu phân tách quá trình nhận diện làn đường thành hai bước: xác định lề làn đường và định vị chính xác làn đường. Tuy nhiên, phương pháp này chỉ tập trung vào nhận diện làn đường cùng với các ký hiệu đặc biệt trên mặt đường như mũi tên hoặc chữ cái.

Một hướng tiếp cận khác là sử dụng Mạng tích chập toàn cầu (GCN) [5], trong đó phân đoạn dựa trên màu sắc được áp dụng để thực hiện cả nhiệm vụ phân loại và xác định vị trí làn đường. Giải pháp này cung cấp khả năng xử lý video theo thời gian thực nhưng đòi hỏi tài nguyên tính toán rất lớn. Đối với các hệ thống hỗ trợ lái xe tiên tiến (ADAS), P.R. Chen [6] đã phát triển Máy dò vạch kẻ làn đường (LMD), một mô hình CNN sâu được thiết kế để trích xuất đặc trưng làn đường bằng cách sử dụng cấu trúc nhẹ hơn nhằm giảm chi phí tính toán.

Mặc dù những phương pháp tiên tiến này đã đạt được kết quả đáng khích lệ trong một số tình huống thực tế, chúng vẫn bộc lộ hạn chế khi đối mặt với những điều kiện khắc nghiệt như bóng tối dày đặc, vạch kẻ đường bị mờ nghiêm trọng hoặc tình trạng giao thông đông đúc [7, 9, 10].



Hình 1. Hình ảnh và nhãn làn đường

3. Nghiên cứu bài toán phát hiện làn đường

3.1. Giới thiệu tập dữ liệu

Trong thí nghiệm này, nhóm tác giả sử dụng bộ dữ

liệu TuSimple đã qua tiền xử lý được tải từ Kaggle [8]. Đây là một nguồn tài nguyên quan trọng dành cho các nhà nghiên cứu và nhà phát triển trong lĩnh vực xe tự hành. Bộ dữ liệu TuSimple gốc bao gồm 3626 đoạn video ngắn, trong đó khung hình cuối cùng của mỗi đoạn được gắn nhãn với thông tin về làn đường (xem Hình 1). Các hình ảnh trong tập dữ liệu được ghi lại trong các điều kiện thời tiết ánh sáng khác nhau đa dạng (xem Hình 2). Được xây dựng chuyên biệt nhằm hỗ trợ quá trình đào tạo và đánh giá các thuật toán nhận diện làn đường, bộ dữ liệu này đóng vai trò thiết yếu trong việc nâng cao khả năng điều hướng của hệ thống lái xe tự động. Các hình ảnh dùng để huấn luyện và đánh giá được chuẩn hóa về kích thước 1280 x 720.



Hình 2. Hình ảnh trong tập dữ liệu với những điều kiện ánh sáng khác nhau

3.2. Mô hình Encoder - Decoder

Mô hình được sử dụng trong nghiên cứu này có kiến trúc Encoder - Decoder, xem Hình 3 và Hình 4. Phần encoder được thử nghiệm sử dụng mạng Resnet50 [11] và Xception [12] với tham số đã được huấn luyện trước (sử dụng bộ tham số imagenet) nhằm đánh giá kết quả trên các mô hình khác nhau. Mục đích của mã hóa là giảm kích thước của hình ảnh trong khi thu thập các đặc điểm chính của hình ảnh vào nhiều kênh hơn. Ngoài ra các tham số của các lớp đầu tiên của bộ encoder sẽ được bảo toàn trong quá trình huấn luyện. Mặt khác, thành phần Decoder của mô

hình bao gồm các lớp mạng tích chập đơn giản. Mô hình lấy mẫu các vector đặc điểm lên mặt nạ hình ảnh, có cùng chiều cao và chiều rộng với hình ảnh gốc, ở định dạng thang độ xám. Ý tưởng chính đằng sau việc lựa chọn kiến trúc mạng này là để tăng tốc quá trình đào tạo khi xử lý hình ảnh có độ phân giải cao hơn và phù hợp với số lượng dữ liệu và tài nguyên hiện có.

Layer (type)	Output Shape	Param #
input_layer_1 (InputLayer)	(None, 256, 320, 3)	0
functional_1 (Functional)	(None, 32, 40, 512)	1,460,096
conv2d_1 (Conv2D)	(None, 32, 40, 256)	1,179,904
conv2d_transpose_1 (Conv2DTranspose)	(None, 64, 80, 256)	262,400
conv2d_2 (Conv2D)	(None, 64, 80, 128)	295,040
conv2d_transpose_2 (Conv2DTranspose)	(None, 128, 160, 128)	65,664
conv2d_3 (Conv2D)	(None, 128, 160, 64)	73,792
conv2d_transpose_3 (Conv2DTranspose)	(None, 256, 320, 64)	16,448
conv2d_4 (Conv2D)	(None, 256, 320, 1)	65

Total params: 3,353,409 (12.79 MB)
Trainable params: 2,735,809 (10.44 MB)
Non-trainable params: 617,600 (2.36 MB)

Hình 3. Kiến trúc mô hình 1 sử dụng encoder ResNet50

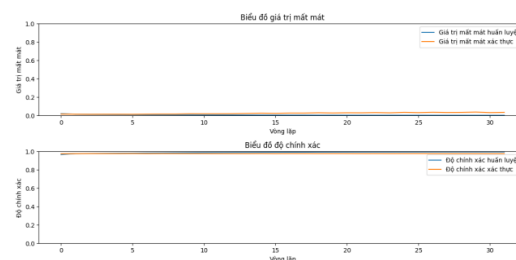
Layer (type)	Output Shape	Param #
input_layer_11 (InputLayer)	(None, 256, 320, 3)	0
functional_10 (Functional)	(None, 8, 10, 2048)	20,861,480
conv2d_46 (Conv2D)	(None, 8, 10, 256)	4,718,848
conv2d_transpose_21 (Conv2DTranspose)	(None, 16, 20, 256)	262,400
conv2d_47 (Conv2D)	(None, 16, 20, 128)	295,040
conv2d_transpose_22 (Conv2DTranspose)	(None, 32, 40, 128)	65,664
conv2d_48 (Conv2D)	(None, 32, 40, 64)	73,792
conv2d_transpose_23 (Conv2DTranspose)	(None, 64, 80, 64)	16,448
conv2d_49 (Conv2D)	(None, 64, 80, 32)	18,464
conv2d_transpose_24 (Conv2DTranspose)	(None, 128, 160, 32)	4,128
conv2d_50 (Conv2D)	(None, 128, 160, 16)	4,624
conv2d_transpose_25 (Conv2DTranspose)	(None, 256, 320, 16)	1,040
conv2d_51 (Conv2D)	(None, 256, 320, 1)	17

Total params: 26,321,945 (100.41 MB)
Trainable params: 5,460,465 (20.83 MB)
Non-trainable params: 20,861,480 (79.58 MB)

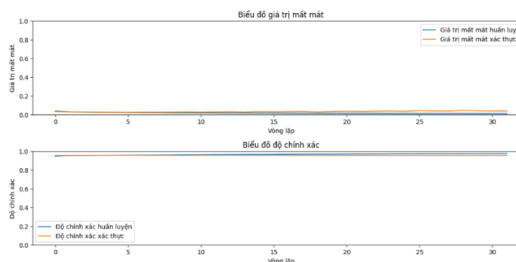
Hình 4. Kiến trúc mô hình 2 sử dụng encoder Xception

Thí nghiệm được thực hiện trên môi trường Kaggle Notebook, sử dụng RAM 30GB, CPU Intel Xeon 4 CPU 2.2 GHz và NVIDIA T4 x 2 GPU. Do số lượng dữ liệu không phải quá nhiều nên mô hình được huấn luyện theo hướng học chuyển tiếp khi sử dụng bộ Resnet50 được huấn luyện trước làm bộ Encoder. Hướng tiếp cận học chuyển tiếp sẽ giúp tinh chỉnh lại các thông số của mô hình để phù hợp với bài toán. Trong quá trình huấn luyện kích thước batch được sử dụng là 8. Bộ bốn chỉ số cơ bản trong bài toán được sử dụng đánh giá mô hình là độ chính xác (accuracy), giá trị mất mát (loss), F1-Score, số khung hình xử lý trên giây (FPS). Ngoài ra các giá trị trong ma trận sai

số (Confusion Matrix) cũng được đưa ra để xem xét.



Hình 5. Biểu đồ giá trị mất mát và độ chính xác trong quá trình huấn luyện mô hình 1



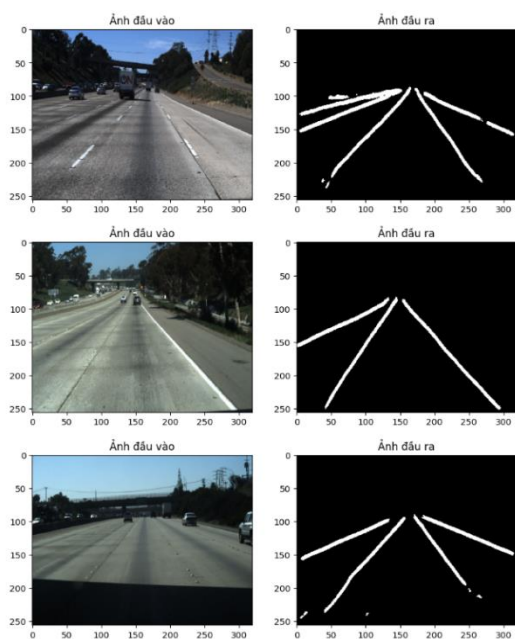
Hình 6. Biểu đồ giá trị mất mát và độ chính xác trong quá trình huấn luyện mô hình 2

Bảng 1. Kết quả thực nghiệm của hai mô hình

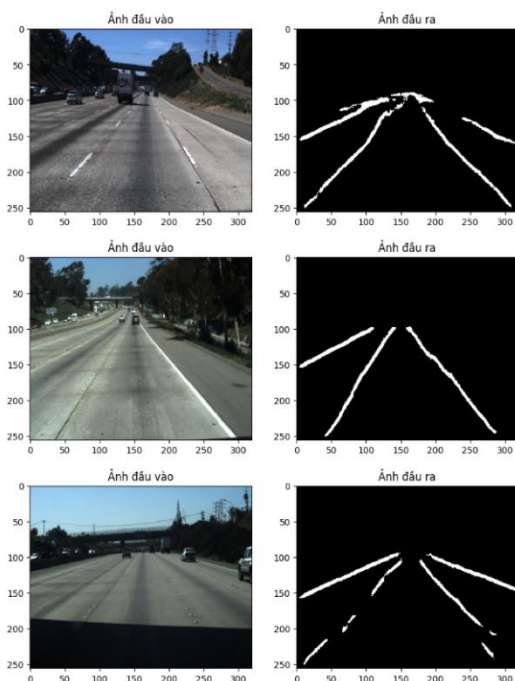
Mô hình sử dụng Encoder ResNet	
TP (True Positive)	511,865
TN (True Negative)	15,514,963
FP (False Positive)	214,303
FN (False Negative)	142,869
Độ chính xác (Accuracy)	0.9782
Độ mất mát (Loss)	0.034
F1-Score	0.741
Số khung hình trên giây (FPS)	122.9
Mô hình sử dụng Encoder Xception	
TP (True Positive)	495,738
TN (True Negative)	15,214,879
FP (False Positive)	404,029
FN (False Negative)	269,354
Độ chính xác (Accuracy)	0.9589
Độ mất mát (Loss)	0.035
F1-Score	0.596
Số khung hình trên giây (FPS)	106.7

Thí nghiệm tiến hành huấn luyện mô hình và sau 32 vòng (epochs) huấn luyện, nhận thấy không có nhiều sự thay đổi trong hiệu suất nên việc huấn luyện đã dừng lại (xem Hình 5 và Hình 6). Mô hình 1 với bộ encoder ResNet50 đã đạt được các kết quả khá ấn tượng khi tiến hành phát hiện các đối tượng trên tập

đánh giá. Chi tiết kết quả thể hiện trong Bảng 1. Một cách tổng quan, mô hình hoạt động tốt khi các chỉ số độ chính xác, giá trị mất mát, F1-score và chỉ số FPS (Frame Per Second - số khung hình xử lý trong 1 giây) khá tốt lần lượt là 0,9782, 0,034, 0,741 và 122,9. Điều này cho thấy hướng tiếp cận của bài toán đã đi đúng hướng. Một cách trực quan có thể thấy được các kết quả đạt được là khá tốt trong Hình 7.



Hình 7. Một số kết quả thử nghiệm với mô hình 1



Hình 8. Một số kết quả thử nghiệm với mô hình 2

Mô hình 2 sử dụng bộ encoder Xception cho kết quả tương đối tốt với mức chính xác ở mức 0.9589, độ mất mát 0.035, F1-score 0.596 và FPS 106.7. Các kết quả của mô hình số 2 kém hơn 1 chút so với mô hình số 1 và trong Hình 7 và Hình 8 cũng có thể thấy được phần nào điều đó. Điều này thể hiện kiến trúc bộ encoder ResNet50 phù hợp hơn trong bài toán này.

4. Kết luận

Nhằm phát triển một hệ thống phát hiện làn đường từ tín hiệu hình ảnh phục vụ xe tự lái, nghiên cứu này áp dụng mô hình Encoder-Decoder để giải quyết bài toán. Tập dữ liệu sử dụng trong nghiên cứu được lấy từ bộ dữ liệu TuSimple và đã trải qua quá trình tiền xử lý. Do kích thước bộ dữ liệu ban đầu còn hạn chế, kỹ thuật học chuyển tiếp được áp dụng để cải thiện quá trình huấn luyện mô hình. Kết quả thử nghiệm trên tập đánh giá sử dụng bộ encoder ResNet cho thấy độ chính xác đạt 0,9782, trong khi giá trị mất mát đạt 0,034 tốt hơn 0.9589 và 0,035 của mô hình sử dụng bộ encoder Xception. Trong các nghiên cứu tiếp theo, nhóm tác giả dự kiến sẽ mở rộng tập dữ liệu CULane1, BDD100K và nâng cấp tài nguyên phần cứng để đào tạo mô hình trên quy mô lớn hơn, nhằm nâng cao độ tin cậy của hệ thống.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Hàng hải Việt Nam trong đề tài mã số: DT24-25.73.

TÀI LIỆU THAM KHẢO

- [1] Qin Zou, Hanwen Jiang, Qiyu Dai, Yuanhao Yue, Long Chen, Qian Wang (2019). *Robust Lane Detection from Continuous Driving Scenes Using Deep Neural Networks*. arXiv:1903.02193.
- [2] Han Ma, Yixin Ma, Jianhao Jiao, M Usman Maqbool Bhutta (2018). *Multiple Lane Detection Algorithm Based on Optimised Dense Disparity Map Estimation*. arXiv:1808.09128v1 [cs.CV].
- [3] D. Neven, B. D. Brabandere, S. Georgoulis, M. Proesmans, and L. V.Gool (2018), *Towards end-to-end lane detection: an instance segmentation approach*, CoRR, vol. abs/1802.05591.
- [4] Ze Wang, Weiqiang Ren, Qiang Qiu (2018). *LaneNet: Real-Time Lane Detection Networks for Autonomous Driving*. arXiv:1807.01726v1 [cs.CV] 4 Jul 2018.
- [5] Mahale Chaoran Chen and Wenhui Zhang (2018). *End to End Video Segmentation for Driving : Lane Detection For Autonomous Car*. arXiv:1812.05914v1 cs.CV] 13 Dec 2018.

- [6] Ping-Rong Chen, Shao-Yuan Lo, Hsueh-Ming Hang National Chiao. *Efficient Road Lane Marking Detection with Deep Learning*. arXiv:1809.03994v1.
- [7] Mohsen Ghafoorian, Cedric Nugteren, Nora Baka, Olaf Booij, Michael Hofmann (2018). *EL-GAN: Embedding Loss Driven Generative Adversarial Networks for Lane Detection*. arXiv:1806.05525v2 Jul 2018.
- [8] Shirke, Suvarna, and R. Udayakumar (2019). *Lane datasets for lane detection*. 2019 International Conference on Communication and Signal Processing (ICCSP). IEEE.
- [9] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla (2016). *SegNet: A Deep Convolutional EncoderDecoder Architecture for Image Segmentation*. IEEE Transactions on Pattern Analysis and Machine, 2016. DOI:10.1109/TPAMI.2016.2644615.
- [10] Shriyash Chougule, Nora Koznek, Asad Ismail, Ganesh Adam1, Vikram Narayan and Matthias Schulze (2018). *Reliable multilane detection and classification by utilizing CNN as a regression network*. ECCV 2018. <https://link.springer.com/conference/eccv>
- [11] He, Kaiming, et al (2016). *Deep residual learning for image recognition*. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [12] Chollet, François (2017). *Xception: Deep learning with depthwise separable convolutions*. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

Ngày nhận bài:	25/02/2025
Ngày nhận bản sửa:	25/03/2025
Ngày duyệt đăng:	26/03/2025