

DỰ ĐOÁN GÓC QUAY VÔ LĂNG CỦA XE TỰ LÁI SỬ DỤNG MẠNG NƠ-RON TÍCH CHẬP

Bùi Văn Hiệu

Trường Đại học Thủy lợi, email: hieubv@tlu.edu.vn

1. GIỚI THIỆU

Các xe vận chuyển có vai trò quan trọng trong nhiều lĩnh vực như vận chuyển người, vận chuyển nguyên vật liệu trong xây dựng và công nghiệp. Nghiên cứu về xe tự lái đã phát triển rất mạnh ở trên thế giới với nhiều thành tựu đột phá [1]. Tuy nhiên, xây dựng các mô hình xe tự lái để vận chuyển người, hàng hóa phù hợp với các điều kiện đặc trưng của thực tế như trong các nhà máy công nghiệp, công trường, khuôn viên trường học còn chưa được phát triển nhiều.

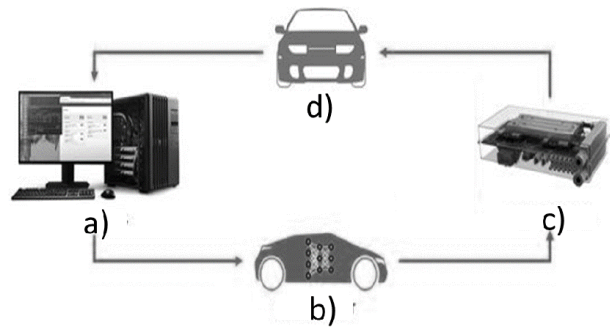
Phát triển xe tự lái có rất nhiều cách tiếp cận khác nhau, một trong các cách tiếp cận phổ biến nhất đó là dùng các mạng nơ-ron nhân tạo để xử lý hình ảnh thu được từ camera [1]. Để có thể xử lý các thông tin hình ảnh và đưa ra các đặc trưng của hình ảnh cho quá trình dự báo cần có các mạng nơ-ron phù hợp. Quá trình xây dựng cấu trúc mạng nơ-ron phù hợp với sự đa dạng và tính chất dữ liệu đóng vai trò quan trọng trong phát triển mô hình xe tự lái.

Bài báo này thể hiện nỗ lực của tác giả trong việc phát triển cấu trúc mạng nơ-ron tích chập (CNNs) để có thể dự báo trực tiếp góc quay vô lăng của xe tự lái từ hình ảnh. Trong đó, các lớp tích chập dùng để trích xuất các đặc trưng của dữ liệu ảnh thu thập từ camera. Các đặc trưng này được phẳng hóa và đưa đến các lớp kết nối đầy đủ dùng để dự báo góc quay vô lăng. So sánh kết quả dự báo của mạng nơ-ron tích chập và kết quả thực tế ở bộ dữ liệu cho thấy mô hình đã dự báo thành công góc quay vô lăng từ dữ liệu hình ảnh.

2. PHƯƠNG PHÁP

2.1. Mô hình xe tự lái

Một nền tảng xe tự lái cơ bản được cấu thành từ các thành phần: kết cấu xe, máy tính trên xe, mạng nơ-ron nhân tạo, hệ thống máy tính cấu hình cao (Hình 1).



Hình 1. Nền tảng xe tự lái phát triển bởi NVIDIA

Phần đưa ra tín hiệu điều khiển xe tự lái là một máy tính được tích hợp các mô hình học sâu là các mạng nơ-ron nhiều lớp. Các tín hiệu đầu vào của hệ thống được thu thập từ các camera, các cảm biến khoảng cách bằng các tia hồng ngoại (LiDAR), các cảm biến sóng âm. Các tín hiệu này sẽ được chuyển đến các bộ xử lý tín hiệu là các bộ xử lý đồ họa được thiết kế để đảm nhiệm vai trò xử lý các khối dữ liệu hình ảnh, đồ họa và video cùng với các thông tin đa luồng khác ở tốc độ cao.

Các mạng nơ-ron nhân tạo trên xe sẽ nhận các trọng số ban đầu đã được huấn luyện từ các bộ dữ liệu lớn bằng hệ thống siêu máy tính. Sự kết hợp của các trọng số đã được huấn luyện và các tín hiệu thu được theo thời gian thực cho phép các mạng nơ-ron nhanh

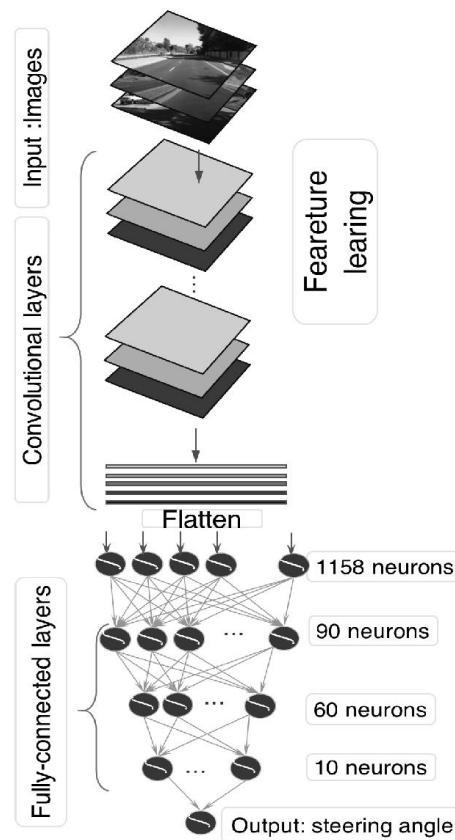
chóng tìm hiểu cách giải quyết những thách thức trong việc lái xe hàng ngày như các mảnh vỡ bất ngờ trên đường, các chướng ngại vật ở khu vực xây dựng. Mạng nơ-ron nhân tạo cũng giải quyết nhiều vấn đề mà các kỹ thuật thị giác máy tính truyền thống không đủ, chẳng hạn như điều kiện thời tiết xấu như mưa, tuyết và sương mù cũng như các điều kiện ánh sáng khó khăn như bình minh, hoàng hôn và bóng tối. Tín hiệu đầu ra của hệ thống máy tính trên xe là các tín hiệu điện để điều khiển vô lăng, phanh, động cơ và các hệ thống khác trên xe.

2.1. Mạng nơ-ron tích chập

Mạng nơ-ron trong nghiên cứu này xây dựng trên nền tảng là thư viện mã nguồn mở Tensorflow được phát triển bởi các nhà nghiên cứu từ Google. Tensorflow tạo ra các luồng dữ liệu dưới dạng biểu đồ với các nút là các toán tử các cạnh là các mảng dữ liệu đa chiều 'tensor' và các phép toán được khái quát hóa. Tensorflow giúp tăng tốc và đơn giản quá trình tính toán các mảng dữ liệu với số chiều và kích thước khác nhau. Xây dựng cấu trúc CNNs là quá trình xác định số lượng các lớp và lựa chọn ma trận lọc phù hợp cũng như chọn các phần đệm. Quá trình lựa chọn này phụ thuộc nhiều vào đặc trưng của bộ dữ liệu. Phương pháp thử dần sẽ được áp dụng để tìm ra cấu trúc mạng phù hợp [1,2]. Qua đó, một cấu trúc của mạng đã được phát triển gồm 4 lớp nơ-ron tích chập, 5 nơ-ron kết nối đầy đủ lớp với số lượng nơ-ron các lớp lần lượt là 1158, 90, 60, 10 nơ-ron (Hình 2).

Huấn luyện CNNs là quá trình cập nhật và tối ưu các trọng số khi dữ liệu đi qua các lớp mạng [2]. Một ảnh thông thường (là các không gian véc tơ 3 chiều $W \times H \times 3$ với chiều rộng (W), chiều cao (H), chiều sâu (3- ứng với 3 màu cơ bản) ban đầu được chuẩn hóa để đưa về giá trị từ (0:1) trước khi vào các lớp tích chập. Tiếp đó, ma trận lọc là các nhân với kích thước (3×3 , 5×5) sẽ quét trên toàn bộ bức ảnh để trích xuất các đặc trưng thông qua phép tích chập của vùng ảnh bị quét và ma trận lọc. Giá trị đạt được của phép tích chập này sẽ được đưa qua hàm kích hoạt

để chuẩn hóa tín hiệu chuẩn bị cho lớp tích chập tiếp. Khi các ảnh đi qua các lớp tích chập, đặc trưng thu được vẫn là không gian véc tơ 3 chiều nhưng có sự giảm về chiều cao chiều rộng và tăng chiều sâu so với ảnh đầu vào. Các tín hiệu đặc trưng sẽ được phẳng hóa thành ma trận một cột để có thể kết nối với mạng kết nối đầy đủ (FCN). Các đặc trưng này sẽ được đi qua mạng kết nối đầy đủ 3 lớp nơ-ron ẩn và đưa ra kết quả dự đoán góc quay vô lăng. Sai số dự đoán có được từ sự so sánh kết quả dự đoán và dữ liệu chuẩn sẽ được truyền lại các lớp trước và cập nhật lại trọng số của các lớp. Quá trình huấn luyện lặp lại với các tập dữ liệu khác nhau và trọng số của mô hình được cập nhật đến khi kết quả dự đoán và dữ liệu có sự sai khác nhỏ ở mức chấp nhận được. Trọng số của mô hình đạt được bởi quá trình huấn luyện thể hiện mối quan hệ giữa các đặc trưng của ảnh và giá trị góc quay vô lăng, trọng số này sẽ được sử dụng để dự đoán cho các hình ảnh thu thập được từ camera trên đường (Hình 2).



Hình 2. Cấu trúc của mạng CNNs gồm 2 lớp chính lớp tích chập và lớp kết nối đầy đủ.

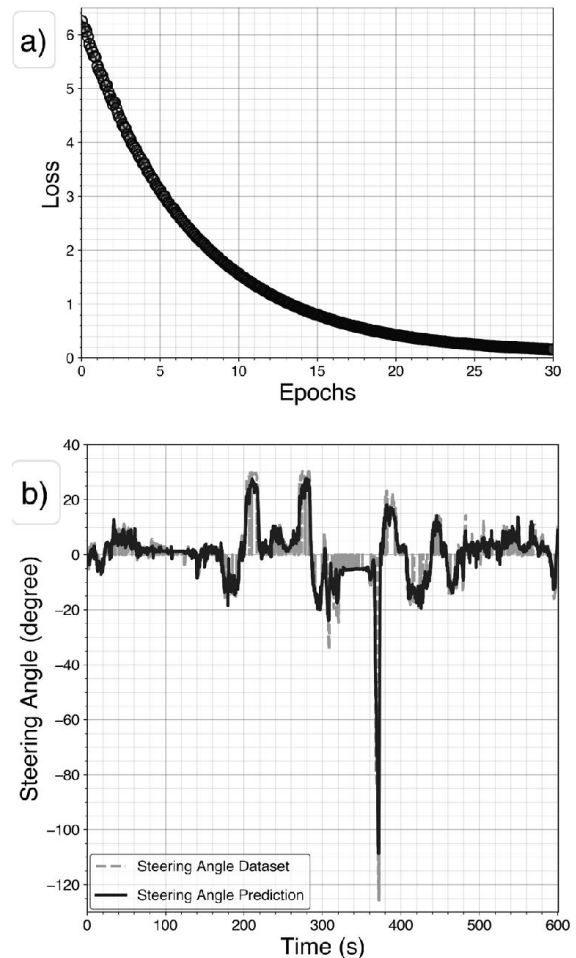
3. KẾT QUẢ

Hình 3a thể hiện kết quả huấn luyện của mô hình với hàm mất mát (loss) trả về thể hiện sự chênh lệch giữa góc quay vô lăng dự đoán và góc quay vô lăng của tập dữ liệu. Biểu đồ cho thấy ở giai đoạn đầu sự chênh lệch này giảm rất nhanh, ở giai đoạn sau sự chênh lệch giảm chậm dần và tiệm cận về 0. Giá trị nhỏ nhất của hàm mất mát đạt được là 0.164 qua 30 lần duyệt dữ liệu. Những kết quả này thể hiện sự phù hợp của cấu trúc mạng và tập dữ liệu huấn luyện.

Hình 3b thể hiện kết quả dự báo góc quay vô lăng trong 600 giây với 10000 ảnh được cắt ra từ camera với bước thời gian 0.06 giây giữa 2 ảnh. Sự so sánh giá trị góc quay vô lăng từ mô hình dự báo (đường xanh thẫm) và góc quay thực tế từ tập dữ liệu (đường màu cam) được thể hiện trên hình 3b. Từ hình ảnh ta thấy mô hình CNNs đã cho kết quả dự báo tốt kể cả với những điểm có sự thay đổi đột ngột như tại vị trí -120 độ. Tuy nhiên có một số điểm mô hình chưa dự báo tốt đó là những điểm góc quay trả về 0 đột ngột. Các chỉ số để đánh giá mô hình dự báo cho kết quả tốt (hệ tương quan và các sai số có giá trị lần lượt là: $R = 0.97$, $RMSE = 3.63$, $MAE = 2.28$). Xét một cách tổng quát, đường dự báo góc quay vô lăng của mô hình có sự phù hợp tốt với đường kết quả từ tập dữ liệu.

4. KẾT LUẬN

Trong nghiên cứu này, tôi đã xây dựng cấu trúc mạng CNNs phù hợp để dự báo góc quay vô lăng từ hình ảnh. Kết quả dự báo của mô hình với độ tương quan cao, sai số nhỏ so với góc quay vô lăng thật của bộ dữ liệu. Sử dụng mạng CNNs để dự báo kết quả góc quay vô lăng từ dữ liệu của camera là phương pháp ít tốn kém. Phương pháp này có thể chuyển các xe có sẵn thành xe tự lái ở cấp thấp giúp giảm chi phí, sức người và giảm tiếp từ đó xúc nâng cao hiệu quả vận chuyển. Trong bước tiếp theo, kết quả của nghiên cứu sẽ được so sánh với các mô hình khác và áp dụng để làm các xe tự lái trong khuôn viên trường học, nhà máy, công trường.



Hình 3. (a) Hàm mất mát. (b) Kết quả dự báo góc quay vô lăng của CNNs

5. TÀI LIỆU THAM KHẢO:

- [1] Mariusz Bojarski and Davide Del Testa and Daniel Dworakowski and Bernhard Firner and Beat Flepp and Praseon Goyal and Lawrence D. Jackel and Mathew Monfort and Urs Muller and Jiakai Zhang and Xin Zhang and Jake Zhao and Karol Zieba. End to End Learning for Self-Driving Cars. arXiv, 2016, 1604.07316.
- [2] Bùi Văn Hiệu; Bùi Minh Đức; Peter Rutschmann. 2020. The Prediction of Fine Sediment Distribution in Gravel-Bed Rivers Using a Combination of DEM and FNN. Water , 12(6), 10.3390/w12061515.