


# THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút):  
<https://youtu.be/IoTLNu0f9K8>
- Link slides (dạng .pdf đặt trên Github):  
<https://github.com/ESTA-2509/CS2205.DeCuong.FinalReport/blob/main/H%C6%B0ng%20Ng%C3%B4%20Th%C3%A1i%20-%20xCS2205.DeCuong.FinalReport.Template.Slide.pdf>
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*

<ul style="list-style-type: none"><li>● Họ và Tên: Ngô Thái Hưng</li><li>● MSSV: 230202006</li></ul> 	<ul style="list-style-type: none"><li>● Lớp: CS2205.MAR2024</li><li>● Tự đánh giá (điểm tổng kết môn): 7.5/10</li><li>● Số buổi vắng: 2</li><li>● Số câu hỏi QT cá nhân: 0</li><li>● Link Github: <a href="https://github.com/mynameuit/CS2205.APR2023/">https://github.com/mynameuit/CS2205.APR2023/</a></li></ul>
--	---

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

PHÁT HIỆN GIẢ MẠO TÍN HIỆU GPS TRONG PHƯƠNG TIỆN TỰ HÀNH SỬ DỤNG MÔ HÌNH HỌC MÁY BÁN GIÁM SÁT CNN

## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

DETECTING GPS SPOOFING SIGNALS IN AUTONOMOUS VEHICLES USING SEMI-SUPERVISED CNN MODEL

## TÓM TẮT (Tối đa 400 từ)

Xe ô tô tự hành (Autonomous Vehicles - AVs) đang là một lĩnh vực nghiên cứu và phát triển sôi động với tiềm năng to lớn trong việc thay đổi ngành giao thông vận tải. Tuy nhiên, an ninh mạng của AVs, đặc biệt là vấn đề giả mạo tín hiệu GPS, vẫn còn là một thách thức lớn. Kẻ tấn công có thể dễ dàng xâm nhập và thao túng dữ liệu GPS, dẫn đến các hậu quả nghiêm trọng như tai nạn giao thông hoặc mất kiểm soát phương tiện. Nghiên cứu này trình bày mô hình học máy bán giám sát dựa trên mạng nơ-ron tích chập (CNN) để phát hiện giả mạo tín hiệu GPS trong xe ô tô tự hành (Autonomous Vehicles - AVs). Mô hình được huấn luyện bằng bộ dữ liệu tín hiệu GPS thực và giả mạo từ A DATASET for GPS Spoofing Detection on Autonomous Vehicles [1]. Đồng thời, mô hình cũng được huấn luyện bằng bộ dữ liệu không được gắn nhãn, được trích xuất từ ứng dụng CARLA Simulation. CARLA là một nền tảng mô phỏng mạnh mẽ cho nghiên cứu và phát triển trong lĩnh vực tự động hóa lái xe, cho phép tạo ra các tình huống giao thông phức tạp để kiểm tra mô hình của bạn.

## GIỚI THIỆU (Tối đa 1 trang A4)

Trong kỷ nguyên công nghệ 4.0, xe ô tô tự hành (Autonomous Vehicles - AVs) đã trở thành một lĩnh vực nghiên cứu và phát triển sôi động với tiềm năng to lớn [2], hứa hẹn mang lại những lợi ích to lớn về an toàn giao thông, tiết kiệm năng lượng và hiệu quả vận hành. Tuy nhiên, sự phụ thuộc lớn vào Hệ thống Định vị Toàn cầu (GPS) để

điều hướng và kiểm soát khiến các CAVs dễ bị tổn thương trước các cuộc tấn công giả mạo GPS (GPS spoofing). Các cuộc tấn công này có thể khiến phương tiện tự hành xác định sai vị trí hiện tại, dẫn đến những hậu quả nghiêm trọng như mất kiểm soát, gây tai nạn hoặc bị chiếm quyền điều khiển [3]. Việc phát hiện và ngăn chặn các cuộc tấn công giả mạo GPS do đó trở thành một vấn đề cấp bách và cần thiết.



Hình 1. Cuộc tấn công giả mạo GPS nhằm vào xe tự hành [4]

Nghiên cứu này đề xuất sử dụng mô hình học máy bán giám sát dựa trên mạng nơ-ron tích chập (CNN) để phát hiện giả mạo tín hiệu GPS trong phương tiện tự hành (AVs) [5]. Mô hình được huấn luyện bán giám sát bằng bộ dữ liệu tín hiệu GPS thực và giả mạo có gắn nhãn, kết hợp với dữ liệu mô phỏng không gắn nhãn để nâng cao khả năng tổng quát.

Mô hình CNN sẽ trích xuất đặc trưng từ dữ liệu tín hiệu GPS. Sau đó, thuật toán học máy bán giám sát được áp dụng để tận dụng tối đa thông tin từ cả dữ liệu có nhãn và

không nhận, giúp mô hình học hỏi hiệu quả hơn và tăng khả năng phát hiện chính xác tín hiệu GPS giả mạo.

Dataset bao gồm:

- Dữ liệu có nhãn: Gồm các mẫu tín hiệu GPS thực và giả mạo từ A DATASET for GPS Spoofing Detection on Autonomous Vehicles [1]. Bộ dữ liệu kết quả chứa tổng cộng 158,170 mẫu, bao gồm 55% các trường hợp hợp pháp và 45% mẫu tương ứng với ba loại tấn công giả mạo tín hiệu GPS được mô phỏng, tất cả trong phân phối cân bằng. Dữ liệu có sẵn trong các tệp Excel ở hai định dạng: định dạng 8 kênh và định dạng dễ đọc (đơn giản/hiệu quả). Định dạng 8 kênh bao gồm 158.170 mẫu dữ liệu với 13 đặc trưng từ tám kênh được biểu diễn dưới dạng mở rộng 3 chiều của các kênh. Định dạng Excel dễ đọc là một chuyển đổi của dữ liệu đầu ra theo chuỗi thời gian từ tám kênh của bộ thu thành một bản phân phối bản đồ đặc trưng hai chiều (2D).
- Dữ liệu không nhãn: Bao gồm các mẫu tín hiệu GPS mô phỏng từ công cụ mô phỏng CARLA Simulation, với các cột dữ liệu tương ứng với bộ dữ liệu "A DATASET for GPS Spoofing Detection on Autonomous Vehicles".

Input:

- Dữ liệu tín hiệu GPS thu thập từ các cảm biến GPS trên AVs, bao gồm vị trí, tốc độ, thời gian, độ cao, v.v.

Output:

- Phân loại tín hiệu GPS thành "thực" hoặc "giả mạo". Báo động khi phát hiện tín hiệu GPS giả mạo.

## MỤC TIÊU

*(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)*

- Phát hiện và xác định giả mạo tín hiệu GPS: Phát triển và đánh giá một hệ thống có khả năng phát hiện và xác định các tín hiệu GPS bị giả mạo sử dụng

bộ dataset có gán nhãn, phân phối cân bằng để huấn luyện.

- Nâng cao khả năng tổng quát của mô hình thông qua việc sử dụng dữ liệu mô phỏng không nhãn, giúp mô hình học hỏi tốt hơn và tăng khả năng thích ứng với các tình huống mới.
- Kiểm thử và áp dụng hệ thống trong môi trường thực tế: Thực hiện các thử nghiệm và đánh giá hiệu suất của hệ thống trong các tình huống và môi trường thực tế khác nhau, đảm bảo hệ thống có thể hoạt động ổn định và hiệu quả trên các phương tiện tự hành.

## **NỘI DUNG VÀ PHƯƠNG PHÁP**

*(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)*

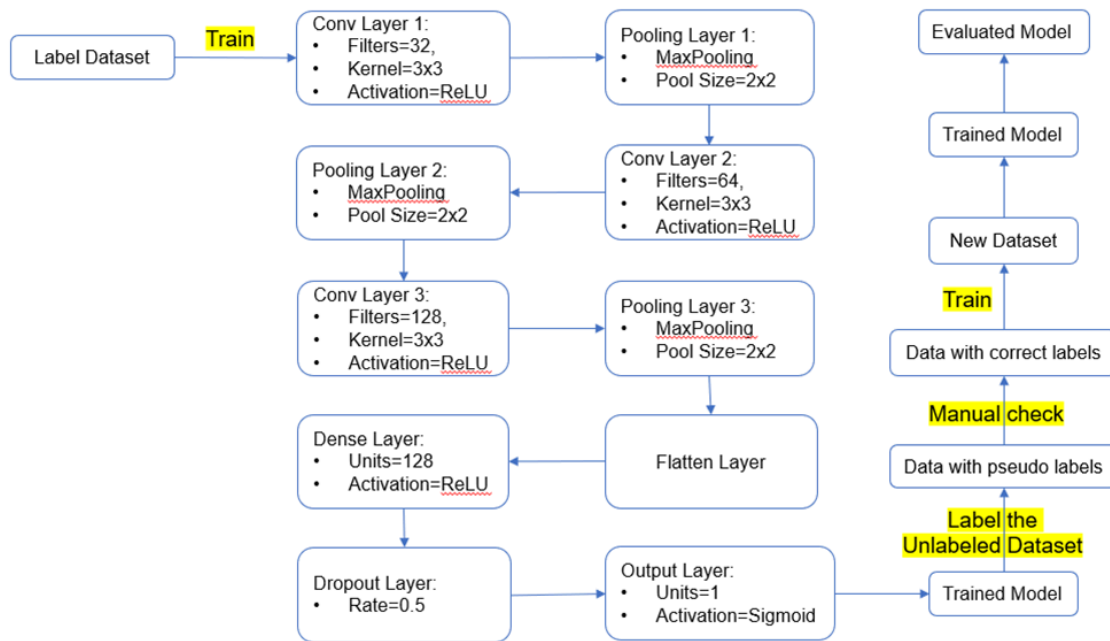
Mục tiêu 1: Phát triển mô hình học máy học CNN hiệu quả cho việc phát hiện giả mạo tín hiệu GPS trong phương tiện tự hành.

- Sử dụng dữ liệu tín hiệu GPS có nhãn từ A DATASET for GPS Spoofing Detection on Autonomous Vehicles [1]. Sau đó, Chia dữ liệu thành tập huấn luyện, tập xác nhận và tập đánh giá.
- Sử dụng thuật toán học máy học CNN để tận dụng tối đa thông tin từ cả dữ liệu có nhãn. Theo dõi và điều chỉnh các tham số huấn luyện để tối ưu hóa hiệu suất mô hình.
- Đánh giá hiệu suất mô hình trên tập xác nhận và tập đánh giá bằng các chỉ số như độ chính xác, độ nhạy, độ đặc hiệu.
- Phân tích kết quả đánh giá để xác định điểm mạnh và điểm yếu của mô hình.

Mục tiêu 2: Nâng cao khả năng tổng quát của mô hình thông qua việc sử dụng dữ liệu mô phỏng không nhãn.

- Tạo dữ liệu mô phỏng sử dụng các công cụ mô phỏng như CARLA để tạo ra dữ liệu tín hiệu GPS mô phỏng. Thiết kế các kịch bản mô phỏng đa dạng, bao gồm các trường hợp bình thường, bất thường và giả mạo tín hiệu GPS. Thêm nhiễu và biến đổi vào dữ liệu mô phỏng để tăng độ thực tế.

- Kết hợp dữ liệu mô phỏng không nhãn với dữ liệu thực có nhãn để huấn luyện mô hình. Điều chỉnh trọng số giữa dữ liệu thực và dữ liệu mô phỏng để đảm bảo hiệu quả huấn luyện.
- Điều chỉnh phương pháp tạo dữ liệu mô phỏng và huấn luyện mô hình để tối ưu hóa hiệu quả.



Hình 2. Mô hình học máy bán giám sát CNN đề xuất

Mục tiêu 3: Kiểm chứng và áp dụng hệ thống trong môi trường thực tế:

- Thiết kế kịch bản thử nghiệm: Lựa chọn các môi trường thử nghiệm khác nhau bao gồm khu vực đô thị, ngoại ô, và khu vực nông thôn để đảm bảo tính đa dạng của dữ liệu thử nghiệm.
- Thu thập dữ liệu trong môi trường thực tế: Thực hiện các chuyến đi thử nghiệm trong các môi trường đã chọn và ghi nhận các tín hiệu GPS thực tế.
- Áp dụng mô hình đã huấn luyện: Sử dụng mô hình học máy bán giám sát CNN đã được huấn luyện để phát hiện giả mạo tín hiệu GPS trên dữ liệu thu thập được.

- Đánh giá hiệu suất của hệ thống: Sử dụng các chỉ số đánh giá như độ chính xác (accuracy), độ nhạy (recall), độ đặc hiệu (specificity), và F1-score để đánh giá hiệu suất của hệ thống.
- Tối ưu hóa mô hình và Huấn luyện lại mô hình nếu cần thiết: Sử dụng các phương pháp tối ưu hóa như điều chỉnh siêu tham số, tăng cường dữ liệu, và cải thiện thuật toán để nâng cao hiệu suất của mô hình.

## KẾT QUẢ MONG ĐỢI

*(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)*

- Mô hình học máy bán giám sát CNN có khả năng phân biệt chính xác giữa tín hiệu GPS thực và giả mạo với độ chính xác cao, độ nhạy cao và độ đặc hiệu cao.
- Hệ thống có khả năng thu thập, xử lý và phân loại tín hiệu GPS theo thời gian thực và có thể được tích hợp dễ dàng vào các hệ thống lái tự động hiện có.
- Phát triển các thuật toán học máy bán giám sát mới hiệu quả hơn cho các bài toán phân loại khác.

## TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

- [1] S. B. H. E. A. a. N. K. Ghilas Aissou, "A DATASET for GPS Spoofing Detection on Autonomous Vehicles," 19 11 2022. [Online]. Available: <http://ieee-dataport.org/documents/dataset-gps-spoofing-detection-autonomous-vehicles>.
- [2] X. Sun, F. R. Yu, P. Zhang, "A Survey on Cyber-Security of Connected and

Autonomous Vehicles (CAVs)," IEEE Transactions on Intelligent Transportation Systems, vol. 23, 2022.

[3] S. V. T. S. P. Vipin Kumar Kukkala, Roadmap for Cybersecurity in Autonomous Vehicles, 2022.

[4] H. W. C. X. Peng Jiang, "DeepPOSE: Detecting GPS spoofing attack via deep recurrent neural network," Digital Communications and Networks, 2021.

[5] P. S. Pranav Singh Chib, "Recent Advancements in End-to-End Autonomous Driving using Deep Learning: A Survey," IEEE Transactions on Intelligent Vehicles, vol. 9, no. 1, pp. 103 - 118, January 2024.

[6] C. Guo, H. Yang, D. Huang, J. Zhang, N. Dong, J. Xu and J. Zhu, "Review Sharing via Deep Semi-Supervised Code Clone Detection," IEEE Access, 2020.