Journal of educational equipment: Applied research, Volume 1, Issue 324 (November 2024)
ISSN 1859 - 0810

Đề xuất mô hình mạng Nơ-ron nhân tạo và thư viện Python trong nhận dạng phương tiện giao thông qua Video phục vụ giảng dạy môn Trí tuệ nhân tạo bậc đại học

Nguyễn Văn Hách*, Lê Phú Hưng*

**Khoa Công nghệ thông tin, Trường Đại học Tài nguyên và Môi trường Hà Nội*
Received: 9/10/2024; Accepted: 21/10/2024; Published:29/10/2024

Abstract: This article focuses on the application of artificial neural networks to build applications to identify people and vehicles through video devices, to support traffic monitoring and management systems. smart in urban areas. With the rapid increase in traffic and the need for security control, smart identification systems not only help reduce traffic congestion and minimize accidents but also play an important role in surveillance activities and public safety supervision. The research team's content also uses the Python programming language and specialized libraries such as TensorFlow and OpenCV, combining modern methods in computer vision to build image and video recognition applications that serve for smart traffic.

Keywords: *Artificial intelligence, ANN, deep learning, Video recognition systems, Computer vision*

1. Giới thiệu

Trong bối cảnh Cách mạng Công nghiệp 4.0, trí tuệ nhân tạo (AI) và học máy (Machine Learning) đã đạt được những tiến bộ vượt bậc, đặc biệt là các ứng dụng học sâu (Deep Learning) trong phân tích dữ liệu phức tạp như hình ảnh và video. Mô hình mạng nơ-ron nhân tạo (ANN) đã thể hiện tiềm năng mạnh mẽ trong việc nhận dạng và xử lý dữ liệu, trở thành công cụ quan trọng cho các hệ thống an ninh, giảm tắc nghẽn giao thông và quản lý đô thị thông minh. Nhóm tác giả tập trung vào nghiên cứu về ANN ứng dụng vào nhận dạng người và phương tiện không chỉ nâng cao an toàn công cộng mà còn góp phần vào xây dựng thành phố thông minh. Ngoài ra, nhóm nghiên cứu đã ứng dụng các kỹ thuật hiện đại như mạng nơ-ron tích chập (CNN), sử dụng Python và các thư viện học máy tiên tiến để phát triển hệ thống giám sát giao thông thông minh, đáp ứng nhu cầu an ninh và giám sát giao thông trong các đô thị lớn, góp phần giải quyết những thách thức của thành phố thông minh trong tương lai.

2. Nội dung nghiên cứu

2.1. Hướng tiếp cận xây dựng nhận diện hình ảnh qua video

Trước khi nói về phân loại video, hãy cùng tìm hiểu nhận dạng hoạt động của con người là gì. Nói một cách đơn giản, nhiệm vụ phân loại hoặc dự đoán hoạt động/hành động đang được ai đó thực hiện được gọi là nhận dạng hoạt động [1][3]. Vấn đề là mô hình không phải lúc nào cũng hoàn toàn tự tin về khả năng dự đoán của từng khung hình video, do đó các dự đoán sẽ thay đổi và dao động nhanh chóng. Điều này là do mô hình không xem xét toàn bộ chuỗi video mà chỉ phân loại từng khung hình một cách độc lập.

Một giải pháp dễ dàng cho vấn đề này là thay vì phân loại và hiển thị kết quả cho một khung hình duy nhất, tại sao không tính trung bình kết quả trên 5, 10 hoặc n khung hình. Điều này sẽ loại bỏ hiệu quả hiện tượng nhấp nháy đó. Khi đã quyết định được giá trị n, chúng ta có thể sử dụng một kỹ thuật đơn giản như kỹ thuật trung bình động/trung bình lăn để đạt được điều này. Vậy hãy giả sử:

n = Số khung hình để tính trung bình

P_f = Xác suất dự đoán cuối cùng

P = Xác suất dự đoán của khung hiện tại

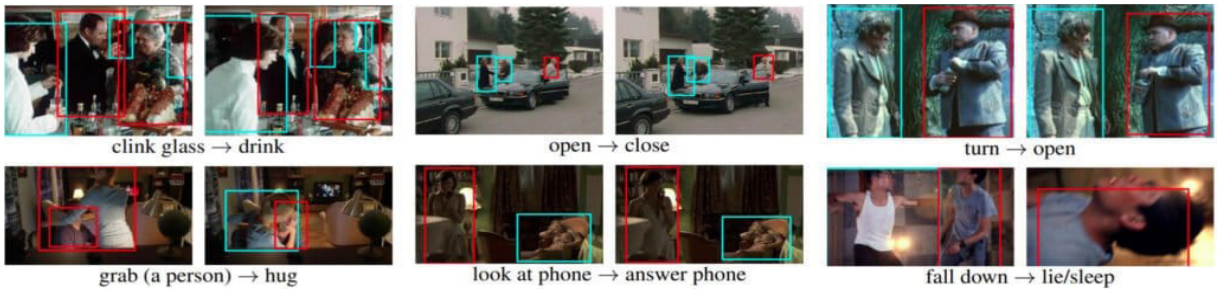
P_{-1} = Xác suất dự đoán của khung cuối cùng

P_{-2} = xác suất dự đoán của khung cuối cùng thứ 2

…

P_{-n+1} = (n-1) xác suất dự đoán của khung cuối cùng

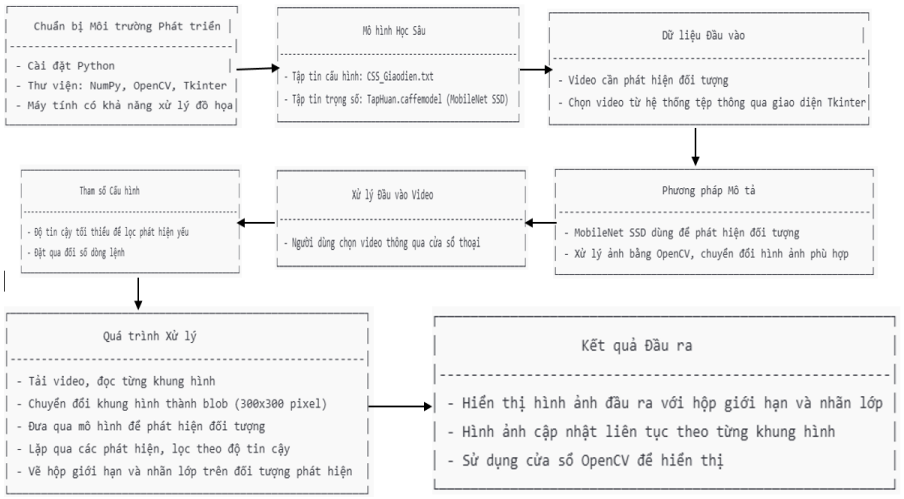
Ý tưởng trong phương pháp này là sử dụng mạng tích chập để trích xuất các đặc điểm cục bộ của từng khung. Đầu ra của các mạng tích chập độc lập này được đưa vào mạng LSTM (Long Short-Term Memory) [5] là một loại mạng nơ-ron hồi quy (Recurrent Neural Network - RNN) nhiều lớp nhiều-một để hợp nhất thông tin được trích xuất này tạm thời.



Ảnh 2.1: Các hành động thay đổi theo thời gian [5]

2.2. Phương pháp thực nghiệm

2.2.1. Đề xuất mô hình thực nghiệm. Mô hình đề xuất học máy mạng nơ ron xử lý sau:



Hình 2.1. Mô hình đề xuất học máy nhận diện hình ảnh qua video

a. Chuẩn bị: Môi trường phát triển như cài đặt Python và các thư viện cần thiết như NumPy, OpenCV, và Tkinter. Cần có một máy tính với khả năng xử lý đồ họa tốt để xử lý video.

b. Mô hình học sâu: Hai tập tin cần thiết là CSS_Giaodien.txt (tập tin cấu hình mô hình) và TapHuan.caffemodel (tập tin trọng số của mô hình MobileNet SSD).

Dữ liệu đầu vào: Video cần phát hiện đối tượng, có thể chọn từ hệ thống tệp bằng giao diện người dùng Tkinter.

c. Phương pháp mô tả: Sử dụng mô hình MobileNet SSD, một mô hình mạng nơ-ron đã được huấn luyện sẵn cho nhiệm vụ phát hiện đối tượng. Thực hiện các thao tác xử lý ảnh bằng OpenCV để chuyển đổi hình ảnh thành định dạng phù hợp cho mô hình.

d. Xử lý đầu vào video: Người dùng chọn video từ máy tính thông qua một cửa sổ thoại.

e. Tham số cấu hình: Độ tin cậy tối thiểu để lọc các phát hiện yếu, được đặt thông qua đối số dòng lệnh.

g. Quá trình xử lý: Tải video và đọc từng khung hình (frame). Sau đó chuyển đổi mỗi khung hình thành blob (đối tượng ảnh) với kích thước 300x300 pixel và chuẩn hóa giá trị pixel. Đưa qua mô hình để nhận diện đối tượng và nhận các phát hiện. Lặp qua các phát hiện, lọc theo độ tin cậy, và vẽ hộp giới hạn quanh các đối tượng phát hiện được cùng với nhãn lớp.

h. Kết quả đầu ra: Hình ảnh đầu ra sẽ hiển thị các đối tượng phát hiện với hộp giới hạn và nhãn lớp tương ứng. Hình ảnh sẽ được cập nhật liên tục theo từng khung hình của video và được hiển thị trong một cửa sổ OpenCV.

i. Đánh giá kết quả phản ánh trong học sâu: Chương trình này minh họa khả năng của học sâu trong việc phát hiện đối tượng trong ảnh/video với độ chính xác cao. Nó cho thấy sức mạnh của các mô hình học sâu được huấn luyện sẵn (pre-trained models) trong các ứng dụng thực tế. Kết quả đầu ra phản ánh hiệu quả của mô hình trong việc nhận diện và phân loại các đối tượng, góp phần vào các ứng dụng như giám sát video, xe tự lái, và tương tác người-máy.

2.2.2. Thực nghiệm

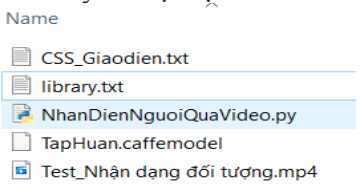
Để thực nghiệm được chương trình chạy, chúng ta cần có 04 file sau:

a. File dữ liệu xử lý giao diện: File CSS_Giaodien.txt trong đoạn mã Python mà cung cấp là một tập tin cấu hình được sử dụng để mô tả cấu trúc của mô hình học sâu đang sử dụng với OpenCV. Với mục đích File này chứa thông tin về cấu trúc của mô hình, bao gồm các lớp (layers) và cách mà các lớp này liên kết

với nhau. Nó cho phép OpenCV biết cách xử lý dữ liệu đầu vào và cách thức để tạo ra các đầu ra dự đoán.

b. File học máy chạy chương trình: Mục đích của file này là lưu trữ trạng thái của mô hình sau khi được huấn luyện trên một tập dữ liệu nhất định. Điều này cho phép người dùng sử dụng mô hình này để dự đoán trên dữ liệu mới mà không cần phải tái huấn luyện. Với tập tin có đuôi TapHuan.caffemodel là định dạng tập tin được sử dụng bởi Caffe, một framework học sâu (deep learning) mã nguồn mở được phát triển bởi Berkeley Vision and Learning Center (BVLC). Tập tin này chứa các trọng số (weights) của mô hình đã được huấn luyện và có thể được sử dụng để thực hiện dự đoán (inference) mà không cần huấn luyện lại mô hình.

c. File code python mô tả chương trình chính: Ngoài thực hiện kê khai một số thư viện để viết code không bị lỗi, chương trình được mô tả thực hiện gọi các tập tin và xử lý các sự kiện

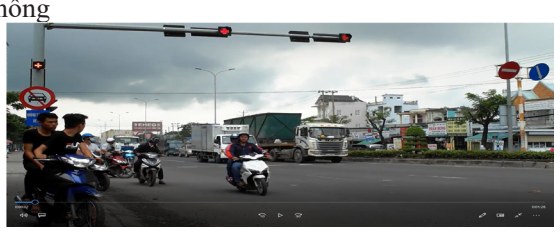


Hình 2.2. Các file mô phỏng của chương trình

d. File Test_Nhận dạng đối tượng.mp4 là video đầu vào có dữ liệu người và các phương tiện tham gia giao thông để thực hiện nhận dạng

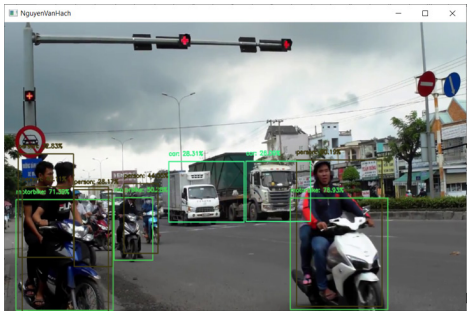
2.3. Kết quả chương trình chạy

Đầu vào là một file video các hình ảnh chuyển động trong một giao thông hỗn độn giữa người đi đường, xe máy, ô tô… các phương tiện tham gia giao thông



Hình 2.3. Hình ảnh giao thông chưa qua nhận diện

Kết quả hình trả về nhận dạng các đối tượng theo tỷ lệ phân tích học máy



Hình 2.4. Hình ảnh sau khi được phát hiện

Như vậy sau khi chương trình chạy và nhận biết được hình ảnh phát hiện ra người, xe máy, ô tô,… các phương tiện tham gia giao thông được nhận diện rõ theo tỷ lệ ngưỡng đến 100% đạt nhận biết rõ nhất

3. Kết luận

Trong bài báo này, nhóm tác giả nghiên cứu trí tuệ nhân tạo và mô hình học máy để xây dựng ứng dụng nhận diện hình ảnh người và các phương tiện tham gia giao thông từ dữ liệu thu được video bằng Python và một số thư viện học máy của Python. Kết quả đã xây dựng được phần mềm dựa vào mô hình đề xuất và kỹ thuật học máy để nhận dạng được người và phương tiện tham gia giao thông qua video đầu vào, không chỉ tách được từng đối tượng mà còn bám sát theo đối tượng với phát hiện tỷ lệ cao nhất của mạng nơ ron đạt đến 100% trong quá trình xử lý.

Tài liệu tham khảo

[1].Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013). "A Public Domain Dataset for Human Activity Recognition Using Smartphones." 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013.

[2]. Halachev, P. (2012). "Prediction of e-Learning Efficiency by Neural Networks." Cybernetics And Information Technologies, Vol. 12, Bulgarian Academy Of Sciences.

[3]. Zhu, X., Pang, J., & Ouyang, W. (2017). "Deep Learning in Object Detection." Proceedings of the IEEE International Conference on Computer Vision.

[4]. Simonyan, K., & Zisserman, A. (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv preprint arXiv:1409.1556.

[5]. Hochreiter, S., & Schmidhuber, J. (1997). "Long Short-Term Memory." Neural Computation, 9(8), 1735-1780.

[6].Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). "SSD: Single Shot MultiBox Detector." European Conference on Computer Vision (ECCV).