



COMBINING YOLOV4-TINY AND LSTM MODELS TO ENHANCE ACCURACY OF THE WARNING SYSTEM OF SMOKING BEHAVIOR IN PUBLIC PLACE

Vo Thien Linh

Division of electrical and electronics engineering, Campus in Ho Chi Minh City, University of Transport and Communications, No. 450 – 451 Le Van Viet Street, Tang Nhon Phu A Ward, Thu Duc City, Ho Chi Minh City, Vietnam

ARTICLE INFO

TYPE: Research Article

Received: 24/08/2022

Revised: 27/09/2022

Accepted: 14/10/2022

Published online: 15/10/2022

<https://doi.org/10.47869/tcsj.73.8.4>

* *Corresponding author*

Email: linhvt_ph@utc.edu.vn; Tel: +84 907001184

Abstract. Tobacco smoke contains many chemicals, which cause typical dangerous diseases such as lung cancer, cardiovascular disease, infertility and many other incurable diseases. Monitoring and warning smokers in public places are one of the important jobs in propaganda and prevention of the harmful effects of tobacco. A smoking monitoring system and smart warnings have been researched and implemented by several groups so far. The objective of this study is to improve the previous system using Yolov4 and LSTM combination to give highly accurate prediction results. Instead of just using Yolov4 as a method of detecting people holding a cigarette, this proposed method uses Yolov4 to extract features from the frames of the videos. This sequence of consecutive feature frames is fed into the LSTM network for prediction. To evaluate the method's performance, the presented study performed on the group's own collected data set: 5000 photos and 120 videos containing smoking behaviour. The results show that the proposed approach is successful in detecting human smoking actions on this dataset with higher accuracy in comparison with that in conventional method.

Keywords: Smoking behavior; convolutional neural network; Deep learning; YOLOv4-tiny; Long Short Term Memory.



KẾT HỢP MÔ HÌNH YOLOV4-TINY VÀ LSTM ĐỂ NÂNG CAO ĐỘ CHÍNH XÁC CỦA HỆ THỐNG CẢNH BÁO NGƯỜI CÓ HÀNH VI HÚT THUỐC LÁ NƠI CÔNG CỘNG

Võ Thiện Linh

Bộ môn Điện-Điện tử, Phân Hiệu tại Thành phố Hồ Chí Minh, Trường Đại học Giao thông vận tải, 450 - 451 Lê Văn Việt, Tăng Nhơn Phú A, TP. Thủ Đức, TP. Hồ Chí Minh, Việt Nam

THÔNG TIN BÀI BÁO

CHUYÊN MỤC: Công trình khoa học

Ngày nhận bài: 24/08/2022

Ngày nhận bài sửa: 27/09/2022

Ngày chấp nhận đăng: 14/10/2022

Ngày xuất bản Online: 15/10/2022

<https://doi.org/10.47869/tcsj.73.8.4>

* Tác giả liên hệ

Email: linhvt_ph@utc.edu.vn; Tel: +84 907001184

Tóm tắt. Khói thuốc lá chứa nhiều chất hóa học, trong đó có nhiều chất gây ra những căn bệnh nguy hiểm điển hình như: ung thư phổi, bệnh tim mạch, vô sinh và nhiều bệnh nan y khác. Nhắc nhở và cảnh báo người hút thuốc lá nơi công cộng là một trong những việc làm quan trọng trong việc tuyên truyền và phòng chống tác hại của thuốc lá. Các hệ thống giám sát người hút thuốc lá và cảnh báo thông minh đã được quan tâm nghiên cứu bởi cộng đồng khoa học trong thời gian qua. Mục đích của nghiên cứu này là cải thiện hệ thống trước đây bằng cách sử dụng Yolov4 kết hợp LSTM để cho ra kết quả dự đoán chính xác cao. Thay vì chỉ sử dụng Yolov4 như là phương pháp phát hiện người cầm điếu thuốc lá, phương pháp đề xuất này sử dụng Yolov4 để trích các đặc trưng từ các khung hình của video, chuỗi các khung hình đặc trưng liên tiếp này được đưa vào mạng LSTM để dự đoán. Để đánh giá hiệu suất của phương pháp, nghiên cứu đã thực hiện trên bộ dữ liệu được thu thập riêng của nhóm: 5000 ảnh và 120 video chứa hành vi hút thuốc. Kết quả cho thấy rằng cách tiếp cận đề xuất đã thành công khi nhận ra các hành động hút thuốc lá của con người trên bộ dữ liệu này với độ chính xác cao hơn phương pháp truyền thống.

Từ khóa: Hành vi hút thuốc; mạng nơ-ron tích chập; Deep learning; YOLOv4-tiny; Long Short Term Memory.

1. ĐẶT VẤN ĐỀ

Khói thuốc lá rất độc hại, là tác nhân gây ra biết bao nhiêu bệnh tật. Hiện nay, công tác về phòng chống tác hại thuốc lá còn không ít những khó khăn, thách thức và rất lâu dài. Việt Nam vẫn là nước có tỷ lệ người hút thuốc lá cao nhất thế giới. Tỷ lệ hút thuốc lá ở nam giới có xu hướng gia tăng; tỷ lệ trẻ em tiếp xúc thụ động với khói thuốc lá còn rất cao; hành vi hút thuốc lá nơi công cộng vẫn phổ biến. Bên cạnh đó, tình trạng quảng cáo, khuyến mại, tiếp thị và tài trợ của một số công ty thuốc lá vẫn được thực hiện dưới nhiều hình thức. Lời cảnh báo sức khỏe trên vỏ bao thuốc lá không gây ấn tượng mạnh với người tiêu dùng, vì vậy việc cảnh báo không hiệu quả đối với người nghiện thuốc lá và ngăn ngừa người bắt đầu hút thuốc lá, nhất là trẻ em. Không ít những trường hợp còn hút thuốc lá nơi đông người, nơi cấm hút một cách tự nhiên mà không bị một phản ứng hay xử lý nào...

Hút thuốc lá là một hoạt động phổ biến của con người, vì vậy để phát hiện hành vi hút thuốc lá có thể dựa trên việc phát hiện khói thuốc hoặc dựa trên các chuyển động của cơ thể thông qua các cảm biến đeo lên người [1-4], các nhóm tác trình bày một hệ thống dựa trên mạng không dây giữa các thiết bị mang trên người, bao gồm hai thiết bị có sẵn, một điện thoại thông minh và một đồng hồ thông minh, để phát hiện các trường hợp hút thuốc bằng cách khai thác dữ liệu cảm biến quán tính từ cả hai thiết bị. Hệ thống sử dụng tập dữ liệu cảm biến trên để xác định hành động hút thuốc hay không hút thuốc trên cơ sở bộ phân loại rừng ngẫu nhiên để phát hiện sự kiện hút thuốc. Các phương pháp này có thể áp dụng cho việc phát hiện và giám sát cho từng cá nhân riêng lẻ, không khả thi khi áp dụng cho một đám đông. Đồng thời khi triển khai các phương pháp này khá bất tiện cho con người, độ chính xác còn phụ thuộc vào các cảm biến.

Các phương pháp hiện đại hơn cũng được triển khai để phát hiện hành vi hút thuốc lá như là phương pháp xử lý ảnh trên cơ sở học máy và học sâu như minh họa tại Hình 1. Bài báo [5] sử dụng dữ liệu cảm biến thô làm đầu vào cho mô hình huấn luyện mạng để nhận dạng các hoạt động của con người. Từ đó có thể sử dụng trong nhiều ứng dụng khác nhau như theo dõi hoạt động thể chất đến các ứng dụng giám sát an toàn. Các mô hình này có thể dễ dàng mở rộng cho việc được huấn luyện với các nguồn dữ liệu khác nhau để tăng độ chính xác hoặc mở rộng phân loại cho các lớp dự đoán khác nhau. Ngoài ra, phương pháp kết hợp nhận dạng quỹ đạo bàn tay và khẩu hình của miệng được đề xuất để phát hiện về hành động hút thuốc. Thêm vào đó, bộ phát hiện khói được sử dụng để cho ra kết luận có hút thuốc hay không [6]. Trong một nghiên cứu khác, hai tác giả Tse-Lun Bien và Chang Hong Lin dựa trên tư thế con người để phát hiện các sự kiện hút thuốc lá [7]. Trong tài liệu [8], nhóm tác giả trình bày một phương pháp phát hiện các sự kiện hút thuốc lá thông qua việc theo dõi nhiều người, nhận dạng hành động và phát hiện khói. Trong bài báo này, nhóm tác giả sử dụng phương pháp HOG để phát hiện đối tượng con người trước tiên. Sau đó, sử dụng phương pháp bám điểm tâm để bám các đối tượng là con người. Tiếp theo, việc phát hiện sự kiện hút thuốc lá hội đủ hai yếu tố. Đầu tiên là đối tượng con người được phát hiện có đang hiển thị các hành vi hút thuốc lá bằng CNN từ thư viện Keras và yếu tố thứ hai là sử dụng SVM để xác định có khói trong khung hình chứa đối tượng đầu tiên không. Bài báo [7] và [8] kết hợp nhiều giải thuật phức tạp do đó tốc độ xử lý rất chậm. Hơn nữa, việc phát hiện khói thuốc là không khả thi khi ánh sáng môi trường thay đổi hay tầm nhìn camera cách xa.

Phương pháp Yolov4-tiny cũng được triển khai nghiên cứu với độ chính xác cao [9], phương pháp này phát hiện đối tượng nhanh và đưa ra kết quả thời gian thực. Để phát hiện được hành vi hút thuốc, trong phương pháp này, tác giả đề xuất 2 đối tượng cần được phát hiện, một là tay cầm điều thuốc, hai là tay đặt lên miệng. Kết hợp kết quả dự đoán 2 trường hợp này thì mới cho ra kết luận về hành vi hút thuốc. Tuy nhiên, có một vấn đề cần giải quyết ở đây, trong một khung hình nếu có một người tay cầm điều thuốc và một người khác đưa tay lên miệng thì đôi khi hệ thống dự đoán sai. Trong bài báo [10], nhóm tác giả sử dụng mô hình Yolov5 để phát hiện hành vi hút thuốc lá, mô hình này chỉ thực hiện phát hiện một nhãn là điều thuốc từ đó đưa ra kết quả dự đoán về hành vi hút thuốc lá nơi công cộng. Hệ thống [9, 10] chỉ xác định hành vi trong một khung hình, chưa xét đến chuỗi các khung liên tiếp.



Hình 1. Minh họa phát hiện hành vi hút thuốc bằng xử lý ảnh. (Hình của nhóm sinh viên tình nguyện Phân Hiệu tại Thành phố Hồ Chí Minh, Trường Đại học Giao thông vận tải).

Trong nghiên cứu này, tôi muốn cải thiện hệ thống phát hiện người hút thuốc lá nơi công cộng bằng phương pháp kết hợp giữa mạng Yolov4-tiny và mạng LSTM (Long Short-Term Memory network) [11]. Nội dung Nghiên cứu bao gồm: tổng quan về mạng Yolov4-tiny và mô hình dự báo dữ liệu chuỗi thời gian LSTM. Tiếp theo, xây dựng kiến trúc mạng nơ-ron tích chập và mô hình dự báo hành vi phù hợp với bài toán đặt ra. Công việc tiến hành song song trong quá trình này là thu thập tập mẫu cho việc huấn luyện. Sau khi xây dựng được kiến trúc mạng hoàn chỉnh, huấn luyện mạng và triển khai phần mềm. Cuối cùng là thử nghiệm, hiệu chỉnh, hoàn thiện thiết bị và đánh giá hệ thống cảnh báo hút thuốc lá.

Bố cục bài báo được trình bày như sau: phần 1 là phần giới thiệu về tổng quan tình hình nghiên cứu; trình bày các phương pháp nhận biết hành vi hút thuốc lá. Phần 2 trình bày phương pháp nghiên cứu đề xuất của đề tài: mô tả cấu trúc mạng kết hợp YOLOv4-tiny + LSTM; mô tả sơ đồ tổng quan hệ thống cảnh báo hành vi hút thuốc; mô hình hệ thống và thuật toán. Phần 3: trình bày kết quả thử nghiệm và đánh giá kết quả. Phần 4 trình bày kết luận.

2. MÔ HÌNH HỆ THỐNG

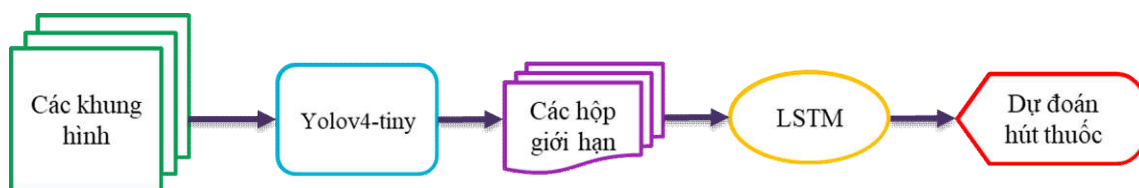
Hành vi hút thuốc lá của con người không chỉ chứa thông tin không gian mà còn chứa thông tin thời gian. Phương pháp Yolov4-tiny [9] không thể sử dụng đầy đủ thông tin thời gian của các video. Nếu thông tin thời gian giữa các khung hình video không được tính đến, việc dự đoán sai dễ xảy ra.

Trong tài liệu [9], việc phát hiện các sự kiện hút thuốc bằng cách nhận biết hành vi bằng cách phân tích các hành động của con người từ tư thế của họ. Phương pháp ước lượng tư thế con người Bài báo đề xuất phương pháp kết hợp giữa mạng Yolov4-tiny và mạng LSTM cho việc dự đoán hành vi hút thuốc lá. Phương pháp này sử dụng Yolov4-tiny để trích xuất các đặc trưng từ các khung hình riêng lẻ của một video, chuỗi các đặc trưng của các khung hình được đưa vào mạng LSTM để dự đoán.

2.1. Ý tưởng của phương pháp

Để thực hiện ghi nhận hành động hút thuốc lá của con người từ các video đã cho, các khung hình làm dữ liệu đầu vào được cắt bớt từ đoạn video. Phương pháp Yolov4-tiny được áp dụng để tạo bản đồ đặc trưng nhằm ghi nhận các nhãn hành động của con người. Trong suốt quá trình huấn luyện mạng, các hành động của con người được nhận ra và xuất các lớp đối tượng, như được hiển thị trong Hình 2. Tổng thể của hệ thống được thể hiện như sau:

- Bước đầu tiên là nhập video vào, tiền xử lý dữ liệu video đầu vào và phân tích khung hình video: quá trình trích xuất các khung hình và thay đổi kích thước ảnh;
- Bước thứ hai là phân tích các đặc điểm của từng khung hình và trích xuất các lớp đối tượng;
- Bước thứ ba là tạo bản đồ đặc trưng của hành động từ các đặc trưng được trích xuất;
- Bước thứ tư là sử dụng bản đồ đặc trưng theo thời gian để xác định các hành động;
- Bước cuối cùng là xuất ra kết quả.



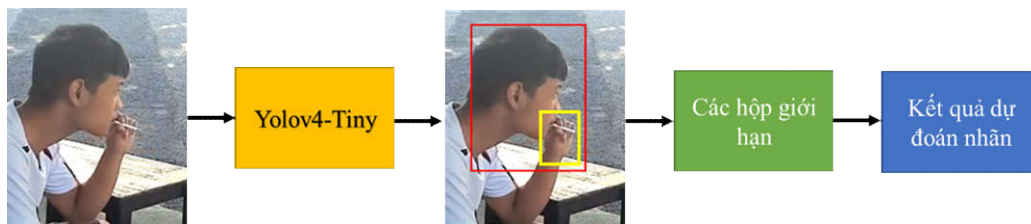
Hình 2. Các bước thực hiện hệ thống.

Yolov4-tiny

Trong bài toán phát hiện hành vi hút thuốc, yêu cầu cần phát hiện ra hai lớp đối tượng: một là bàn tay cầm điều thuốc và hai là bàn tay đặt lên miệng. Do đó, tập mẫu huấn luyện khi đánh nhãn cần đầy đủ hai lớp này. YOLOv4-tiny là một mô hình mạng nơ-ron tích chập có 38 lớp trong đó có 29 lớp tích chập được thiết kế để phát hiện các đối tượng mà có ưu điểm nổi trội là nhanh hơn nhiều so với những mô hình cũ, phù hợp cho hệ thống real-time và có thể chạy tốt trên những máy tính nhúng như Jeston nano hay Raspberry Pi 4.

Trong bài báo này, ảnh đầu vào của mô hình là 416x416, được chia thành lưới có kích thước 13x13 ô, số lớp phân loại là 2, đầu ra mô hình là một ma trận 3 chiều có kích thước

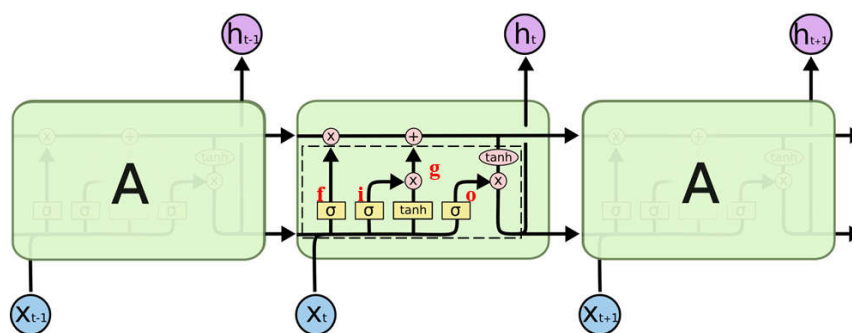
$13 \times 13 \times (5 \times N + M)$ với số lượng tham số mỗi ô là $(5 \times N + M)$ với N và M lần lượt là số lượng hộp giới hạn và lớp mà mỗi ô cần dự đoán. Dự đoán mỗi hộp giới hạn gồm 5 thành phần: $(x, y, w, h, \text{prediction})$ với (x, y) là tọa độ tâm của hộp giới hạn, (w, h) lần lượt là chiều rộng và chiều cao của hộp giới hạn. Trong đó h là chiều cao của hộp giới hạn và w là chiều rộng.



Hình 3. Minh họa Yolov4-tiny trong bài toán phát hiện người hút thuốc lá.

LSTM

Mạng bộ nhớ dài-ngắn (Long Short Term Memory networks – LSTM) là một dạng đặc biệt của mạng nơ-ron hồi quy. LSTM được thiết kế để có khả năng ghi nhớ chuỗi dữ liệu quá khứ trong thời gian dài. LSTM có kiến trúc dạng chuỗi, các mô-đun trong nó có cấu trúc 4 lớp tương tác với nhau một cách rất đặc biệt.



Hình 4. Cấu trúc của một mạng LSTM.

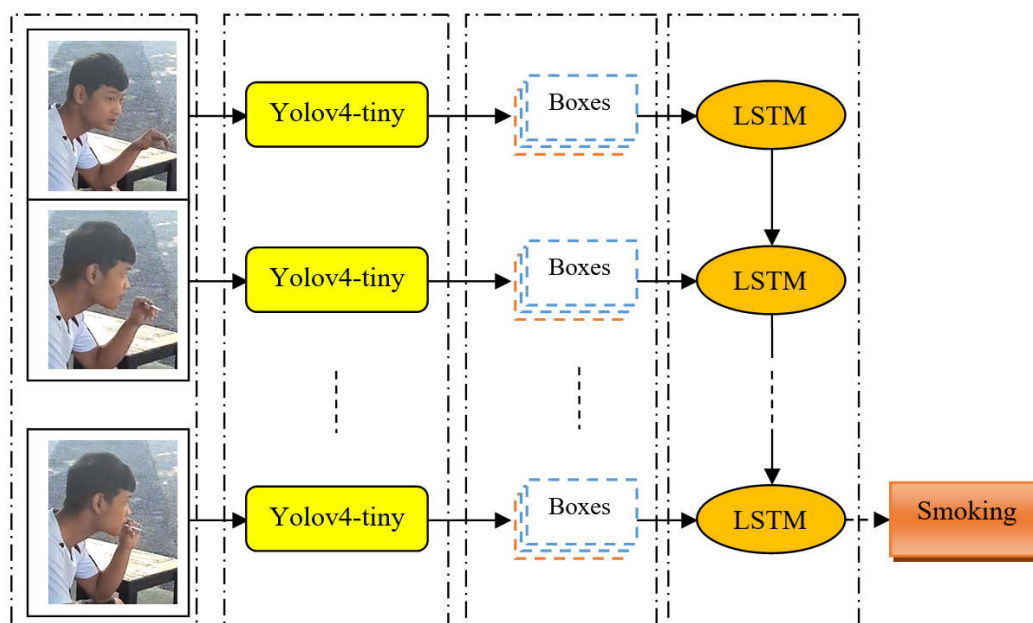
Mạng LSTM có thể bao gồm nhiều tế bào LSTM liên kết với nhau và kiến trúc cụ thể của mỗi tế bào được biểu diễn như trong Hình 4. Ý tưởng của LSTM là bổ sung thêm trạng thái bên trong tế bào và ba cổng sàng lọc các thông tin đầu vào và đầu ra cho tế bào bao gồm f (forget gate), i (input gate) và o (output gate). Tại mỗi bước thời gian t , các cổng đều lần lượt nhận giá trị đầu vào x_t (đại diện cho một phần tử trong chuỗi đầu vào) và giá trị h_t có được từ đầu ra của tế bào từ bước thời gian trước đó $t-1$. Các cổng đều đóng vai trò có nhiệm vụ sàng lọc thông tin với mỗi mục đích khác nhau, cụ thể:

- Forget gate: Có nhiệm vụ loại bỏ những thông tin không cần thiết nhận được khỏi trạng thái ẩn.

- Input gate: Có nhiệm vụ chọn lọc những thông tin cần thiết nào được thêm vào trạng thái ẩn.
- Output gate: Có nhiệm vụ xác định những thông tin nào từ trạng thái ẩn được sử dụng như đầu ra.

Kết hợp Yolov4-tiny và LSTM

Với ngõ vào là ảnh được trích xuất đặc trưng qua mô hình Yolov4-tiny đã huấn luyện trước với tập dữ liệu tự thu thập, quá trình được gọi là nhúng ảnh và ngõ ra là một vector. Ý tưởng sẽ là dùng các đặc trưng nhúng của ảnh và dùng các đối tượng nhân liên tiếp để dự đoán hành vi hút thuốc.



Hình 5. Cấu trúc của mô hình Yolov4-tiny + LSTM.

Trước hết, video thử nghiệm được chia thành các khung để tạo thành tập dữ liệu hình ảnh. Tập dữ liệu này được sử dụng làm đầu vào của Yolov4-tiny để dự đoán các Hộp giới hạn. Kết quả bước này được lưu trữ trong tập dữ liệu đặc trưng và bản đồ đặc trưng. Sau đó, tập dữ liệu đặc trưng được đưa vào mạng LSTM làm dữ liệu đầu vào. Mạng LSTM được sử dụng để huấn luyện dữ liệu chuỗi. Sau khi huấn luyện mô hình, các tham số mạng cố định của mạng được sử dụng để nhận dạng hành động của con người.

2.2. Huấn luyện

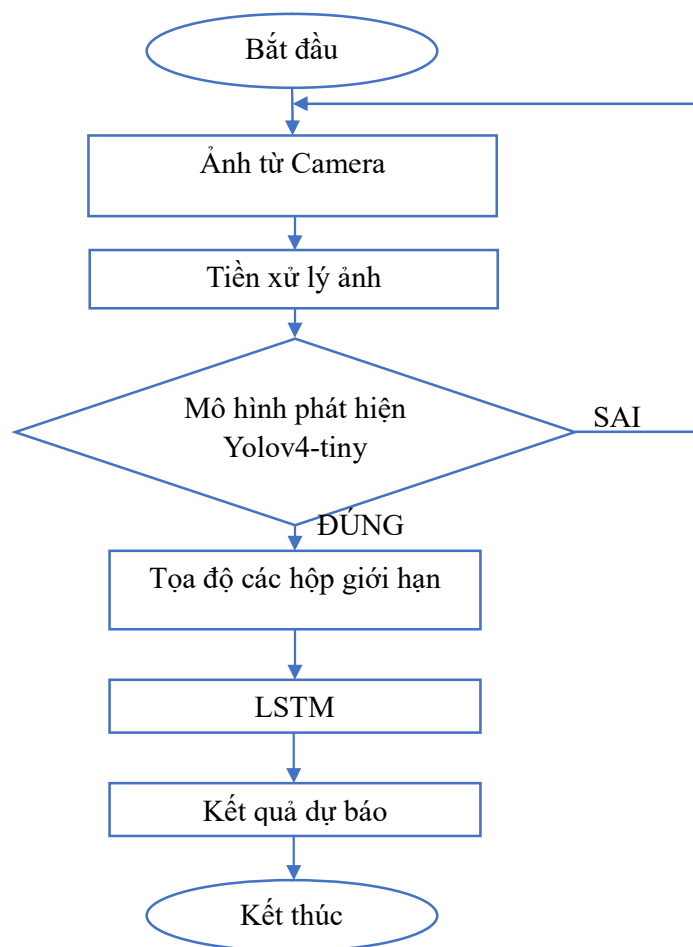
Để nhận biết cho hành vi hút thuốc, với mỗi chuỗi video, số khung hình video được trích xuất tại một khoảng thời gian làm đầu vào. Các đặc trưng không gian được trích xuất từ các khung video đã cho. Đặc trưng không gian ở đây là tọa độ các hộp giới hạn cuối cùng có xác suất cao được đưa vào LSTM để liên kết mối quan hệ thời gian giữa chuỗi các khung hình video. Tôi nhận được các tham số mạng cố định sau khi đào tạo mô hình. Cụ thể các bước thực hiện như sau:

Bước 1: Sử dụng Yolov4-tiny đã được huấn luyện trước đó để phát hiện các lớp đối tượng và nhãn trong các khung hình của một video. Các lớp ở đây là “Thuốc” và hành động đang “Hút”. Kết quả là ta có được tọa độ các Hộp giới hạn của các đối tượng.

Bước 2: Phân tích chuyển động của các lớp đối tượng theo thời gian và dự đoán bằng mô hình LSTM. Mô hình LSTM sẽ dự đoán được các hộp giới hạn của khung hình tiếp theo. Đầu vào của mô hình LSTM là tọa độ các hộp giới hạn trên một khung hình và nhãn của nó. Lấy các khung hình nối tiếp nhau theo thứ tự hành vi hút thuốc.

Trong quá trình nghiên cứu, bài báo sử dụng tập dữ liệu tự thu thập gồm 5000 mẫu có chứa hành vi hút thuốc tương đối đa dạng trong các môi trường trong nhà và ngoài trời khác nhau và 120 video ngắn có chứa hành vi hút thuốc lá. Trong huấn luyện, tập dữ liệu được chia thành 80% và 20% tương ứng để huấn luyện và thử nghiệm. Mô hình mạng được triển khai sử dụng Python và các thư viện darknet trên máy tính DELL Intel(R) Core(TM) i7-4800MQ CPU @ 2.70GHz(8core), 8GB RAM, NVIDIA Quadro K2100M.

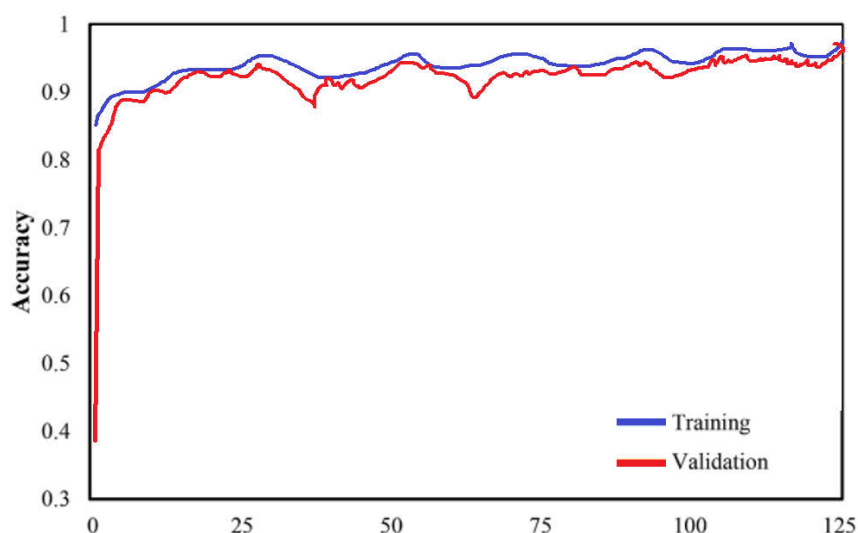
2.3. Lưu đồ giải thuật của hệ thống



Hình 6. Lưu đồ hệ thống.

Ở giai đoạn thử nghiệm, các khung hình video được trích xuất với khoảng cách tương đương được lấy làm dữ liệu đầu vào của mô hình LSTM. Sau khi trích xuất đặc điểm không gian và phân loại mối quan hệ thời gian, giá trị đầu ra dự đoán của LSTM được lấy làm kết quả phân loại cuối cùng.

Hình 7 biểu thị đánh giá hiệu suất của bộ phân loại LSTM với độ chính xác trong quá trình huấn luyện và xác nhận. Độ chính xác của quá trình huấn luyện là 95,8% ở 125 lần học.



Hình 7. Hai hình ảnh kết quả huấn luyện LSTM.

3. THỬ NGHIỆM

Sự kết hợp giữa YOLOv4-tiny và LSTM là một lựa chọn tốt hơn so với phương pháp chỉ sử dụng Yolov4-tiny để phát hiện đối tượng có hành vi hút thuốc lá nơi công cộng.

Thử nghiệm với các trường hợp với tổng số 10 video ngắn chứa 10 tình huống có hút thuốc và 10 tình huống khác:

Với mô hình kết hợp này, hệ thống phát hiện có độ chính xác cao hơn so với Yolov4-tiny. Mô hình ít bị nhiễu và quan sát trên máy tính hình ảnh ít chậm chạp hơn. Thông qua việc lấy các tọa độ hộp giới hạn từ mô hình Yolov4-tiny và dự đoán theo chuỗi thời gian tuần tự các khung hình liên tiếp nhau. Kết quả phát hiện của mô hình Yolov4-tiny chỉ cho ra được các hộp giới hạn và các lớp riêng lẻ do đó việc đưa ra kết luận cho hành vi hút thuốc lá còn nhiều hạn chế và không chính xác. Kết quả của mô hình Yolov4-tiny như hình 7. Hệ thống được đề xuất có thể theo dõi thành công các vị trí thời gian để bù lại các hộp giới hạn bị bỏ sót cho các đối tượng không thể phát hiện và biến mất hay che khuất.

Trong bài báo này, tôi đánh giá độ chính xác của mô hình thông qua việc xây dựng một tập hợp các thông tin phân loại được dự đoán và thực tế. Trong khi thực hiện phân tích dự đoán, một ma trận vuông nhầm lẫn được tạo ra có chứa tỷ lệ dương và âm (cả đúng và sai), còn được gọi là ma trận lỗi [12-13]. Tỷ lệ tương ứng cho tất cả các trường hợp đó được tính toán riêng. Ngoài ra, độ chính xác, độ nhạy đánh giá từ tỷ lệ dương tính và âm tính.

Quá trình thử nghiệm với 20 video có chứa hành vi hút thuốc lá và 20 video có chứa hành vi gần giống hút thuốc lá (ăn kẹo mút, ngậm viết, ...) được tổng hợp như Bảng 1.



Hình 8. Các video được thử nghiệm theo nguồn từ youtube.com và tự quay thực tế nhóm sinh viên tình nguyện của Phân Hiệu tại Thành phố Hồ Chí Minh, Trường Đại học Giao thông vận tải.



Hình 9. Hình ảnh thử nghiệm với mô hình kết hợp Yolov4-tiny + LSTM (hình ảnh nhóm sinh viên tình nguyện của Phân Hiệu tại Thành phố Hồ Chí Minh, Trường Đại học Giao thông vận tải).



Hình 10. Hai hình ảnh minh họa được thử nghiệm với mô hình Yolov4-tiny ((hình ảnh nhóm sinh viên tình nguyện của Phân Hiệu tại Thành phố Hồ Chí Minh, Trường Đại học Giao thông vận tải).

Kết quả thử nghiệm với mỗi phương pháp cho ma trận đầu ra có bốn ô: dương tính thật (TP), âm tính thật (TN), dương tính giả (FP) và âm tính giả (FN). TP có nghĩa là giá trị thực tế và giá trị dự đoán đều dương, FN nghĩa là giá trị thực tế là dương nhưng giá trị dự đoán của mô hình là âm, FP có nghĩa là giá trị thực là âm nhưng giá trị dự đoán của mô hình là dương và cuối cùng, TN có nghĩa là cả hai các giá trị thực tế và dự đoán là âm.

Với phương pháp kết hợp Yolov4-tiny + LSTM:

Bảng 1. Thống kê kết quả thử nghiệm của 2 phương pháp: Yolov4-tiny+LSTM và Yolov4-tiny.

Yolov4-tiny + LSTM		Thực tế		Yolov4-tiny		Thực tế	
		P	N			P	N
Dự đoán	P	TP=18	FP=3	Dự đoán	P	TP=16	FP=5
	N	FN=2	TN=17		N	FN=4	TN=15

$$\text{Độ chính xác} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{18 + 17}{40} = 87,5\% \quad (1)$$

$$\text{Độ nhạy} = \frac{TP}{TP + FN} = \frac{18}{18 + 2} = 0.9 \quad (2)$$

Với phương pháp Yolov4-tiny:

$$\text{Độ chính xác} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{16 + 15}{40} = 77,5\% \quad (3)$$

$$\text{Độ nhạy} = \frac{TP}{TP + FN} = \frac{16}{16 + 4} = 0.8 \quad (4)$$

Phương pháp kết hợp Yolov4-tiny+LSTM cho kết quả chính xác 87,5% và phương pháp Yolov4-tiny cho kết quả chính xác 77,5% với tập các video thử nghiệm có chứa hành vi hút thuốc và không có hành vi hút thuốc. Quá trình thử nghiệm cho thấy, với các trường hợp nhiều, các hành vi gần giống hành vi hút thuốc lá, phương pháp đề xuất cho kết quả dự đoán chính xác hơn so với phương pháp Yolov4-tiny. Điều này có thể được giải thích: phương pháp Yolov4-tiny có tỷ lệ dự đoán các hộp giới hạn tại các thời điểm tức thời và cho ra kết luận, nếu kết quả dự đoán các hộp giới hạn sai dẫn đến kết luận sai; ngược lại, với phương pháp đề xuất nếu kết quả dự đoán các hộp giới hạn sai nhưng sâu chuỗi các khung hình liên tiếp sẽ đưa ra kết luận chính xác hơn. Cả hai mô hình kết hợp YOLOV4-tiny + LSTM phát hiện rất tốt trong môi trường ban ngày, hoạt động kém trong môi trường thiếu ánh sáng. Tác giả đã thử nghiệm mô hình trong các điều kiện khác nhau về môi trường (thời gian, địa điểm, ánh sáng), khác nhau về đối tượng (có cầm thuốc lá, không có cầm thuốc lá). Theo kết quả thử nghiệm, kết quả của hệ thống đề xuất trong nghiên cứu này có độ chính xác và ổn định hơn phương pháp tổng hợp kết quả dự đoán đồng thời hai nhãn của mô hình Yolov4-tiny trong công bố trước đây [9]. Thông qua các trường hợp thử nghiệm, tác giả đã chọn ra ngưỡng phù hợp để xác định các hộp giới hạn cho các nhãn để tăng độ ổn định cho hệ thống.

Bằng cách phân tích kết quả, bài báo này cho thấy rằng sự kết hợp giữa Yolov4-tiny và LSTM có cải thiện và ảnh hưởng đáng kể đến việc phát hiện người hút thuốc lá dựa trên việc trích xuất tự động và sâu chuỗi các đặc trưng từ hình ảnh video.

4. KẾT LUẬN

Kết quả thử nghiệm cho thấy, với mô hình kết hợp YOLOv4-tiny + LSTM, hệ thống phát hiện rất tốt trong môi trường ban ngày, hệ thống hoạt động kém trong môi trường thiếu ánh sáng. Nghiên cứu đã thử nghiệm mô hình trong các điều kiện khác nhau về môi trường (thời gian, địa điểm, ánh sáng), khác nhau về đối tượng (có cầm thuốc lá, không có cầm thuốc lá). Theo kết quả thống kê từ thử nghiệm, kết quả của hệ thống đề xuất cải thiện hơn so với hệ thống áp dụng mô hình Yolov4-tiny.

Kết quả của nghiên cứu này sẽ là tiền đề để ứng dụng trong thực tế nơi công cộng với yêu cầu hạ tầng đơn giản gồm một máy tính có kết nối với camera. Trong tương lai, nghiên cứu sẽ được mở rộng theo hướng tối ưu hóa kiến trúc mạng và triển khai hệ thống trên máy tính nhúng Jetson nano hay Raspberry cũng như bổ sung thiết bị cảnh báo một cách phù hợp với môi trường và địa điểm sử dụng.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Trường Đại học Giao thông vận tải trong đề tài mã số T2022-PHII_DDT-002.

Tác giả xin chân thành cảm ơn Bộ môn Kỹ thuật điện tử, Trường Đại học Giao thông vận tải đã đóng góp ý chuyên môn để hoàn thiện nghiên cứu này; cảm ơn nhóm sinh viên chuyên ngành Kỹ thuật điện tử và tin học công nghiệp, Phân Hiệu tại Thành phố Hồ Chí Minh, Trường Đại học Giao thông vận tải đã hỗ trợ trong quá trình thu thập dữ liệu mẫu.

TÀI LIỆU THAM KHẢO

- [1]. C. Fan, F. Gao, A New Approach for Smoking Event Detection Using a Variational Autoencoder and Neural Decision Forest, IEEE Access, 8 (2020) 120835-120849. <https://doi.org/10.1109/ACCESS.2020.3006163>
- [2]. A. L. Skinner, C. J. Stone, H. Doughty, M. R. Munafo, StopWatch: The Preliminary Evaluation of a Smartwatch-Based System for Passive Detection of Cigarette Smoking, Nicotine & Tobacco Research, 21 (2019) 257–261. <https://doi.org/10.1093/ntr/nty008>
- [3]. Abhinav Parate, Meng-Chieh Chiu, Chaniel Chadowitz, Deepak Ganesan, Evangelos Kalogerakis, RisQ: recognizing smoking gestures with inertial sensors on a wristband. In Proceedings of the 12th annual international conference on Mobile systems, applications, and services, Association for Computing Machinery, New York, USA, (2014) 149–161. <https://doi.org/10.1145/2594368.2594379>
- [4]. Chrisogonas O Odhiambo, Casey A Cole, Alaleh Torkjazi, Homayoun Valafar, State Transition Modeling of the Smoking Behavior Using LSTM Recurrent Neural Networks, International Conference on Computational Science and Computational Intelligence (CSCI), 2019, pp. 898-904. <https://doi.org/10.1109/CSCI49370.2019.00171>
- [5]. S. W. Pienaar and R. Malekian, Human Activity Recognition using LSTM-RNN Deep Neural Network Architecture, IEEE 2nd Wireless Africa Conference (WAC), 2019, pp. 1-5. <https://doi.org/10.1109/AFRICA.2019.8843403>
- [6]. Zhenkai Lin; Changfeng Lv; Yimin Dou; Jinping Li, Smoking Behavior Detection Based on Hand Trajectory Tracking and Mouth Saturation Changes, 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2018, pp. 326-331. <https://doi.org/10.1109/SPAC46244.2018.8965455>
- [7]. Tse-Lun Bien and Chang Hong Lin, Detection and recognition of indoor smoking events, Proc. SPIE 8784, Fifth International Conference on Machine Vision: Algorithms, Pattern Recognition, and Basic Technologies, 2013, 878424. <https://doi.org/10.1117/12.2020967>

- [8]. Waynebert Jan D. Cabanto, Aira Danielle B. Jocson, Renzel Laurence T. Lateo, and Joel C. De Goma, Real-Time Multi-Person Smoking Event Detection, In Proceedings of the 2nd International Conference on Computing and Big Data (ICCBD 2019), Association for Computing Machinery, New York, NY, USA, 2019, pp.126–130. <https://doi.org/10.1145/3366650.3366678>
- [9]. Võ Thiện Lĩnh, Triển khai hệ thống phát hiện và cảnh báo hút thuốc lá nơi công cộng dựa trên cấu trúc mạng YOLOv4-tiny, Hội nghị quốc gia lần thứ XXIV về điện tử, truyền thông và công nghệ thông tin REV – ECIT, 2021, pp. 182-186.
- [10]. Xiangkui Jiang, Haochang Hu, Xun Liu, Rui Ding, Yuanbo Xu, Jianxu Shi, Yaoyao Du and Chunlin Da, A smoking behavior detection method based on the YOLOv5 network, Journal of Physics: Conference Series, 2021 International Conference on Control Theory and Applications, 2232 (2021) 012001. <https://doi.org/10.1088/1742-6596/2232/1/012001>
- [11]. Almeida, Aitor, Gorka Azkune, Predicting Human Behaviour with Recurrent Neural Networks, Applied Sciences, 8 (2018) 305. <https://doi.org/10.3390/app8020305>
- [12]. Joy Bhattacharjee, Soumen Santra, Arpan Deyasi, Chapter 10 - Novel detection of cancerous cells through an image segmentation approach using principal component analysis, Editor(s): Siddhartha Bhattacharyya, Paramartha Dutta, Debabrata Samanta, Anirban Mukherjee, Indrajit Pan, Recent Trends in Computational Intelligence Enabled Research, Academic Press, 2021, pp. 171-195. <https://doi.org/10.1016/B978-0-12-822844-9.00035-9>
- [13]. Chandramouli Das, Abhaya Kumar Sahoo, Chittaranjan Pradhan, Chapter 12 - Multicriteria recommender system using different approaches, Editor(s): Sushruta Mishra, Hrudaya Kumar Tripathy, Pradeep Kumar Mallick, Arun Kumar Sangaiah, Gyoo-Soo Chae, In Cognitive Data Science in Sustainable Computing, Cognitive Big Data Intelligence with a Metaheuristic Approach, Academic Press, 2022, pp. 259-277. <https://doi.org/10.1016/B978-0-323-85117-6.00011-X>