



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0036608

(51)^{2020.01} G06K 9/00

(13) B

(21) 1-2021-01281

(22) 10/03/2021

(45) 25/08/2023 425

(43) 25/05/2021 398

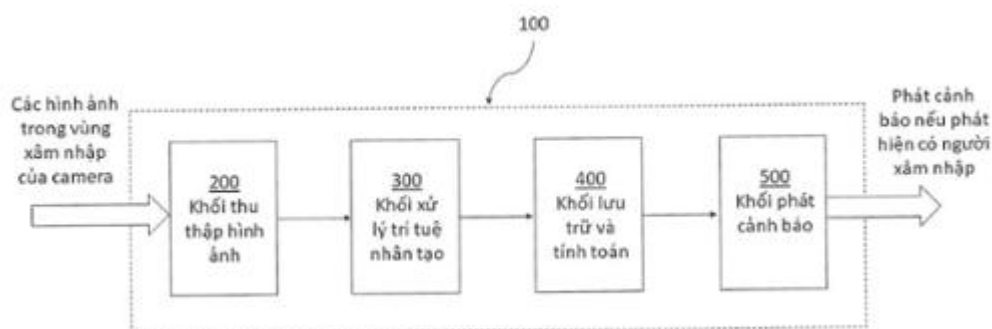
(73) Công ty cổ phần LUMI Việt Nam (VN)

Số 38, đường Đỗ Đức Dục, Phường Mễ Trì, Quận Nam Từ Liêm, Thành phố Hà Nội

(72) Nguyễn Tuấn Anh (VN); Nguyễn Đức Tài (VN); Bùi Mạnh Linh (VN); Lại Thế Hoàng (VN); Nguyễn Hải Nam (VN).

(54) PHƯƠNG PHÁP VÀ HỆ THỐNG PHÁT HIỆN ĐỐI TƯỢNG XÂM NHẬP BẰNG HÌNH ẢNH VÀ VẬT GHI ĐỂ LƯU TRỮ CHƯƠNG TRÌNH MÁY TÍNH

(57) Sáng chế đề cập đến phương pháp và hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh, sử dụng camera và trí tuệ nhân tạo (AI) để phát hiện người xâm nhập và cảnh báo khi có xâm nhập, bao gồm các đặc trưng: thu nhận ảnh hoặc chuỗi ảnh từ camera hoặc đầu thu là dữ liệu đầu vào, để nhận diện bằng AI, cụ thể là mô hình YOLO v5 để phát hiện sự có mặt của người, đầu người, và mặt người một cách đồng thời trong ảnh hoặc chuỗi được nhận diện, và trả về kết quả là khung hình kết quả có chứa người, đầu người, mặt người; kết hợp phát hiện chuyển động để giảm thiểu các cảnh báo nhầm; trong đó việc phát hiện chuyển động là nhận diện chuyển động của người thông qua việc thu nhận và so sánh chuỗi các khung hình kết quả để nâng cao độ chính xác phát hiện, giảm thiểu cảnh báo nhầm.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến lĩnh vực kỹ thuật phát hiện đối tượng xâm nhập bằng cách phân tích hình ảnh thu được từ camera, sử dụng trí tuệ nhân tạo để phát hiện có hay không có người xâm nhập. Và cụ thể hơn, sáng chế đề cập đến phương pháp và hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh, và vật ghi để lưu trữ chương trình máy tính.

Tình trạng kỹ thuật của sáng chế

Với sự phát triển của công nghệ máy tính và trí tuệ nhân tạo, công nghệ nhận dạng hình ảnh được ứng dụng ngày càng rộng rãi, trong đó công nghệ phát hiện sự có mặt của người được sử dụng phổ biến trong đời sống, đặc biệt là trong lĩnh vực an ninh, chống trộm và theo dõi mục tiêu.

Mục đích của việc phát hiện người là để xác định sự có mặt của người lạ xâm nhập vào trong một phạm vi hoặc khu vực nhất định, để đưa ra các cảnh báo cho người dùng (chủ nhà chẳng hạn) biết được có người đang ở trong khu vực cần phát hiện xâm nhập, chẳng hạn như để đảm bảo an ninh, cụ thể hơn là đưa ra các quyết định an ninh cần thiết, ví dụ phát cảnh báo dưới dạng còi, đèn, hoặc đưa ra các cảnh báo trên hệ thống an ninh chung, v.v..

Hiện tại, các giải pháp xác định sự có mặt của người trong ảnh nói chung là như sau: đối với hình ảnh thu được, thực hiện trích xuất hình ảnh, thực hiện đối sánh tương đồng về đặc điểm của các ảnh thu được với ảnh mẫu, và xác định có người trong ảnh. Tuy nhiên, trong các hệ thống hiện nay đều chỉ tập trung vào nhận diện người riêng biệt ở trạng thái tĩnh, chỉ dựa vào kết quả đơn lẻ nên khả năng phát hiện nhằm gây báo động giả.

Do vậy, có nhu cầu về giải pháp phát hiện đối tượng xâm nhập một cách nhanh chóng và có độ chính xác được nâng cao.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất phương pháp phát hiện đối tượng xâm nhập một cách nhanh chóng bằng cách xác định có người trong ảnh thu được từ camera, và giảm thiểu sự phát hiện nhầm đối tượng, tránh việc đưa ra các cảnh báo sai của hệ thống.

Để đạt được ít nhất mục đích trên, sáng chế đề xuất phương pháp phát hiện đối tượng xâm nhập sử dụng mô hình trí tuệ nhân tạo YOLO v5.

Theo một khía cạnh, sáng chế đề xuất phương pháp phát hiện đối tượng xâm nhập bằng hình ảnh bao gồm:

thu thập dữ liệu hình ảnh, bởi khối thu thập hình ảnh, để thu thập các hình ảnh trong vùng xâm nhập của camera;

xử lý, bởi khối xử lý trí tuệ nhân tạo, sử dụng dữ liệu hình ảnh nêu trên làm dữ liệu đầu vào, để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra liên quan đến đối tượng được nhận diện gồm có dữ liệu nhận diện người, dữ liệu nhận diện đầu người, và dữ liệu nhận diện mặt người, nhờ đó xác định có hay không sự có mặt của đối tượng xâm nhập, trong đó:

khối xử lý trí tuệ nhân tạo sử dụng mô hình YOLO v5 để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra nêu trên, mô hình này được huấn luyện để đưa ra các tham số mô hình dựa trên tập dữ liệu huấn luyện đầu vào gồm có các dữ liệu có người, đầu người, và mặt người được dán nhãn, được trộn cùng với các dữ liệu không có người, đầu người, và mặt người,

việc nhận diện người, đầu người, mặt người nêu trên được thực hiện một cách đồng thời, và

dữ liệu đầu ra có dạng khung hình bao quanh đối tượng được nhận diện và có chứa ít nhất là thông tin về tọa độ vị trí của khung hình này;

lưu trữ các dữ liệu đầu ra liên quan đến ít nhất là một đối tượng được nhận diện là các khung hình có thông tin về tọa độ vị trí theo mảng dữ liệu, và

tính toán để xác định có hay không có sự di chuyển của đối tượng được nhận diện, dựa trên mức độ sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện ở hai hoặc nhiều hơn hai thời điểm liên tiếp nhau;

phát cảnh báo, bởi khối phát cảnh báo, để phát cảnh báo có đối tượng xâm nhập trong trường hợp phát hiện sự có mặt của đối tượng xâm nhập và phát hiện có sự di chuyển của đối tượng được nhận diện tương ứng,

trong đó, khối phát cảnh báo tiếp tục phát cảnh báo trong trường hợp tỉ lệ số lần phát hiện sự có mặt của đối tượng xâm nhập đối với số các lần phát hiện trong một khoảng thời gian nhất định lớn hơn hoặc bằng tỉ lệ duy trì cảnh báo được xác định trước, và khối phát cảnh báo không phát cảnh báo hoặc không duy trì phát cảnh báo trong các trường hợp khác với các trường hợp được nêu trên.

Theo một phương án, tập dữ liệu huấn luyện đầu vào bao gồm các ảnh để bị nhận diện nhằm để nâng cao độ chính xác, với số lượng ảnh có thể lên tới 50% so với các ảnh có người.

Theo một phương án, tập dữ liệu huấn luyện đầu vào có thể được làm phong phú bằng cách thay đổi kích thước ảnh đầu vào, trộn lẫn các ảnh với nhau hoặc tương tự, và tốc độ học có thể nhỏ hơn hoặc bằng 0,0001 để nâng cao độ chính xác của mô hình sau khi được huấn luyện.

Tốt hơn là, tỉ lệ duy trì cảnh báo được xác định trước nêu trên là 60%.

Tốt hơn là, hai thời điểm liên tiếp nhau nêu trên, mà dùng để tính toán sự sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện, cách nhau khoảng 1 giây.

Theo một phương án, một hoặc nhiều camera để thu thập các hình ảnh có thể cùng được kết nối với một khối xử lý trí tuệ nhân tạo chung để giảm yêu cầu tài nguyên tính toán, và mỗi camera có thể được thiết lập hai vùng xâm nhập để nâng cao khả năng phát hiện.

Theo một phương án, việc phát cảnh báo có thể bao gồm bật đèn, bật còi hú, thông báo cho người dùng thông qua điện thoại, tin nhắn, thư điện tử, hoặc tương tự khi có người di chuyển tới trong vùng xâm nhập.

Tốt hơn là, phương pháp nêu trên còn bao gồm bước tự động tắt đèn khi không phát hiện thấy có người trong vùng xâm nhập.

Theo một khía cạnh khác, sáng chế đề xuất hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh bao gồm:

khối thu thập hình ảnh để thu thập dữ liệu hình ảnh là các hình ảnh trong vùng xâm nhập của camera;

khối xử lý trí tuệ nhân tạo, sử dụng dữ liệu hình ảnh nêu trên làm dữ liệu đầu vào, để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra liên quan đến đối tượng được nhận diện gồm có dữ liệu nhận diện người, dữ liệu nhận diện đầu người, và dữ liệu nhận diện mặt người, nhờ đó xác định có hay không sự có mặt của đối tượng xâm nhập, trong đó:

khối xử lý trí tuệ nhân tạo sử dụng mô hình YOLO v5 để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra nêu trên, mô hình này được huấn luyện để đưa ra các tham số mô hình dựa trên tập dữ liệu huấn luyện đầu vào gồm có các dữ liệu có người, đầu người, và mặt người được dán nhãn, được trộn cùng với các dữ liệu không có người, đầu người, và mặt người,

việc nhận diện người, đầu người, mặt người nêu trên được thực hiện một cách đồng thời, và

dữ liệu đầu ra có dạng khung hình bao quanh đối tượng được nhận diện và có chứa ít nhất là thông tin về tọa độ vị trí của khung hình này;

khối lưu trữ và tính toán để lưu trữ các dữ liệu đầu ra liên quan đến ít nhất là một đối tượng được nhận diện là các khung hình có thông tin về tọa độ vị trí theo mảng dữ liệu, và tính toán để xác định có hay không có sự di chuyển của đối tượng được nhận diện, dựa trên mức độ sự sai khác tọa độ vị trí giữa các

khung hình của cùng một đối tượng được nhận diện ở hai hoặc nhiều hơn hai thời điểm liên tiếp nhau;

khởi phát cảnh báo để phát cảnh báo có đối tượng xâm nhập trong trường hợp phát hiện sự có mặt của đối tượng xâm nhập và phát hiện có sự di chuyển của đối tượng được nhận diện tương ứng,

trong đó, khởi phát cảnh báo tiếp tục phát cảnh báo trong trường hợp tỉ lệ số lần phát hiện sự có mặt của đối tượng xâm nhập đối với số các lần phát hiện trong một khoảng thời gian nhất định lớn hơn hoặc bằng tỉ lệ duy trì cảnh báo được xác định trước, và khởi phát cảnh báo không phát cảnh báo hoặc không duy trì phát cảnh báo trong các trường hợp khác với các trường hợp được nêu trên.

Theo một phương án, tập dữ liệu huấn luyện đầu vào bao gồm các ảnh để bị nhận diện nhằm để nâng cao độ chính xác, với số lượng ảnh có thể lên tới 50% so với các ảnh có người.

Theo một phương án, tập dữ liệu huấn luyện đầu vào có thể được làm phong phú bằng cách thay đổi kích thước ảnh đầu vào, trộn lẫn các ảnh với nhau hoặc tương tự, và tốc độ học có thể nhỏ hơn hoặc bằng 0,0001 để nâng cao độ chính xác của mô hình sau khi được huấn luyện.

Tốt hơn là, tỉ lệ duy trì cảnh báo được xác định trước nêu trên là 60%.

Tốt hơn là, hai thời điểm liên tiếp nhau nêu trên, mà dùng để tính toán sự sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện, cách nhau khoảng 1 giây.

Theo một phương án, một hoặc nhiều camera để thu thập các hình ảnh có thể cùng được kết nối với một khối xử lý trí tuệ nhân tạo chung để giảm yêu cầu tài nguyên tính toán, và mỗi camera có thể được thiết lập hai vùng xâm nhập để nâng cao khả năng phát hiện.

Theo một phương án, việc phát cảnh báo có thể bao gồm bật đèn, bật còi hú, thông báo cho người dùng thông qua điện thoại, tin nhắn, thư điện tử, hoặc tương tự khi có người di chuyển tới trong vùng xâm nhập.

Tốt hơn là, phương pháp nêu trên còn bao gồm bước tự động tắt đèn khi không phát hiện thấy có người trong vùng xâm nhập.

Theo một khía cạnh khác, sáng chế đề xuất vật ghi bất khả biến có chứa các lệnh máy tính mà khi được thực thi sẽ làm cho máy tính thực hiện phương pháp phát hiện đối tượng xâm nhập bằng hình ảnh nêu trên.

Mô tả vắn tắt các hình vẽ

Hình 1 là sơ đồ khối thể hiện các thành phần chính của hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh theo một phương án ưu tiên của sáng chế;

Hình 2 là lưu đồ thể hiện phương pháp phát hiện đối tượng xâm nhập bằng hình ảnh theo một phương án ưu tiên của sáng chế;

Hình 3 là lưu đồ mô tả giải thuật theo một ví dụ cụ thể của sáng chế.

Mô tả chi tiết sáng chế

Dưới đây, các ưu điểm, hiệu quả và bản chất của sáng chế có thể được hiểu rõ hơn thông qua việc mô tả chi tiết các phương án ưu tiên thực hiện có dựa vào các hình vẽ kèm theo. Tuy nhiên, cần hiểu rằng các phương án này chỉ được mô tả với mục đích làm ví dụ giúp cho việc hiểu rõ hơn về bản chất và các ưu điểm của sáng chế, mà không giới hạn phạm vi của sáng chế theo các phương án được mô tả này.

Trước tiên, các thành phần chính của hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh sẽ được mô tả dựa trên Hình 1.

Như được thể hiện trên hình vẽ, hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh 100 bao gồm: khối thu thập hình ảnh 200, khối xử lý trí tuệ nhân tạo 300, khối lưu trữ và tính toán 400, và khối phát cảnh báo 500.

Khối thu thập hình ảnh 200 để thu thập dữ liệu hình ảnh là các hình ảnh trong vùng xâm nhập của camera.

Khối xử lý trí tuệ nhân tạo 300, sử dụng dữ liệu hình ảnh nêu trên làm dữ liệu đầu vào, để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra liên quan đến đối tượng được nhận diện gồm có dữ liệu nhận diện người, dữ liệu nhận diện đầu người, và dữ liệu nhận diện mặt người, nhờ đó xác định có hay không sự có mặt của đối tượng xâm nhập. Khối xử lý trí tuệ nhân tạo 300 sử dụng mô hình YOLO v5 để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra nêu trên, mô hình này được huấn luyện để đưa ra các tham số mô hình dựa trên tập dữ liệu huấn luyện đầu vào gồm có các dữ liệu có người, đầu người, và mặt người được dán nhãn, được trộn cùng với các dữ liệu không có người, đầu người, và mặt người. Việc nhận diện người, đầu người, mặt người nêu trên được thực hiện một cách đồng thời, và dữ liệu đầu ra có dạng khung hình bao quanh đối tượng được nhận diện và có chứa ít nhất là thông tin về tọa độ vị trí của khung hình này.

Khối lưu trữ và tính toán 400 để lưu trữ các dữ liệu đầu ra liên quan đến ít nhất là một đối tượng được nhận diện là các khung hình có thông tin về tọa độ vị trí theo mảng dữ liệu, và tính toán để xác định có hay không có sự di chuyển của đối tượng được nhận diện, dựa trên mức độ sự sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện ở hai hoặc nhiều hơn hai thời điểm liên tiếp nhau.

Khối phát cảnh báo 500 để phát cảnh báo có đối tượng xâm nhập trong trường hợp phát hiện sự có mặt của đối tượng xâm nhập và phát hiện có sự di chuyển của đối tượng được nhận diện tương ứng. Khối phát cảnh báo 500 này sẽ tiếp tục phát cảnh báo trong trường hợp tỉ lệ số lần phát hiện sự có mặt của đối tượng xâm nhập đối với số các lần phát hiện trong một khoảng thời gian nhất định lớn hơn hoặc bằng tỉ lệ duy trì cảnh báo được xác định trước, và khối phát cảnh báo không phát cảnh báo hoặc không duy trì phát cảnh báo trong các trường hợp khác với các trường hợp được nêu trên.

Mô hình YOLO v5 (YOLO - You Only Look Once phiên bản thứ năm) là đã được biết đến trong lĩnh vực kỹ thuật tương ứng, việc vẽ hình và mô tả chi tiết về mô hình YOLO này được dự định lược bỏ và chỉ nêu ra một số đặc trưng chính để tránh các thông tin dư thừa.

Nói chung, mô hình YOLO là một mô hình trí tuệ nhân tạo ứng dụng mạng nơ ron tích chập (CNN - Convolutional Neural Networks) trong việc phát hiện, nhận dạng, phân loại đối tượng, mô hình này được tạo ra nhờ kết hợp giữa các lớp tích chập và các lớp được liên kết. Trong đó các lớp tích chập sẽ trích xuất ra các đặc trưng của ảnh thông qua các phép tích chập, và các lớp được liên kết sẽ dự đoán ra xác suất sự có mặt của đối tượng và tọa độ của đối tượng đó.

Theo một ví dụ cụ thể, mô hình YOLO có kiến trúc gồm ba phần chính: phần xương sống (Backbone), phần đầu (Head), và phần kết nối (Neck) kết nối phần xương sống và phần đầu.

Phần xương sống đóng vai trò tổng hợp và trích xuất đặc trưng của ảnh, sử dụng kiến trúc CSPDenseNet (Cross Stage Partial DenseNet) cho phần xương sống này và kết hợp với các lớp tích chập. Kiến trúc này giúp cho việc đạo hàm ngược không bị triệt tiêu và giúp mô hình YOLO nhanh hội tụ hơn.

Phần kết nối sử dụng mạng có đặc trưng hình chóp (Feature Pyramid Network) có mục tiêu làm đa dạng hóa dữ liệu trước khi đưa vào nhận diện. Mạng này kết hợp các đặc trưng cấp thấp có độ phân giải cao với những đặc trưng cấp cao mang thông tin rõ ràng về đối tượng. Từ đó có thể giúp mô hình có thể nhận diện được đối tượng trên một phạm vi rộng và sâu đồng thời tăng tốc độ học của mô hình.

Phần đầu là bộ phận nhận đầu vào từ phần kết nối thực hiện phân loại và xác định vị trí của các đối tượng cần nhận diện. Quá trình dự đoán, từ các tham số được thiết lập trong quá trình học (huấn luyện), mỗi bức ảnh làm dữ liệu đầu vào sẽ đi qua mạng YOLO và được trích xuất các đặc trưng rồi được tổng hợp và quyết định xem trong khung hình có phải là người không qua mô hình xác

xuất. Đầu ra sẽ bao gồm các thông số gồm có: xác xuất dự báo vật thể có trong khung hình p_0 , tọa độ của khung hình (tx , ty , tw , th), trong đó (tx , ty) là tọa độ tâm của khung hình; (tw , th) là kích thước dài và rộng của khung hình và p_0 là xác xuất dự đoán trong khung hình đó là nhân nào (người, đầu người, hay là mặt người). Với xác xuất dự đoán p_0 lớn hơn một giá trị nhất định, mô hình sẽ xác định là trong khung hình có xuất hiện đối tượng cần nhận diện, cụ thể là có xuất hiện người, đầu người, hoặc mặt người.

Đối với mô hình YOLO nói chung, cơ chế học dữ liệu của mô hình này dựa trên tập dữ liệu đầu vào được dán nhãn, cụ thể hơn là bao gồm các nhãn xác định người, đầu người, mặt người, hoặc tương tự để tạo ra một bộ dữ liệu hoàn chỉnh bao gồm các ảnh và các nhãn tương ứng. Bộ dữ liệu này được đưa vào mô hình YOLO để huấn luyện mạng này học được/nhận diện được các khái niệm/các đối tượng gồm có người, đầu người, và mặt người thông qua việc điều chỉnh tham số các tham số của mô hình và các lớp mạng với mục tiêu là có thể trích xuất ra đặc trưng của gồm có người, đầu người, và mặt người và so sánh nó với các nhãn đã được đánh dấu sao cho có độ khớp lớn nhất, quá trình học từ tập dữ liệu đầu vào này sẽ giúp tối ưu hoá bộ tham số của mô hình YOLO. Mô hình YOLO với bộ tham số được tối ưu hoá này sẽ được sử dụng trong việc nhận diện các hình ảnh trong vùng xâm nhập của camera.

Tiếp theo, một phương án ưu tiên của phương pháp phát hiện đối tượng xâm nhập bằng hình ảnh được thực hiện bởi hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh 100 nêu trên sẽ được mô tả dựa trên Hình 2.

Theo phương án ưu tiên này, phương pháp phát hiện đối tượng xâm nhập bằng hình ảnh bao gồm các bước:

Bước 1: thu thập dữ liệu hình ảnh, bởi khối thu thập hình ảnh 200, để thu thập các hình ảnh trong vùng xâm nhập của camera;

Bước 2: xử lý, bởi khối xử lý trí tuệ nhân tạo 300, sử dụng dữ liệu hình ảnh nêu trên làm dữ liệu đầu vào, để xử lý dữ liệu đầu vào và đưa ra kết quả là

các dữ liệu đầu ra liên quan đến đối tượng được nhận diện gồm có dữ liệu nhận diện người, dữ liệu nhận diện đầu người, và dữ liệu nhận diện mặt người, nhờ đó xác định có hay không sự có mặt của đối tượng xâm nhập, trong đó:

khởi xử lý trí tuệ nhân tạo 300 sử dụng mô hình YOLO v5 để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra nêu trên, mô hình này được huấn luyện để đưa ra các tham số mô hình dựa trên tập dữ liệu huấn luyện đầu vào gồm có các dữ liệu có người, đầu người, và mặt người được dán nhãn, được trộn cùng với các dữ liệu không có người, đầu người, và mặt người,

việc nhận diện người, đầu người, mặt người nêu trên được thực hiện một cách đồng thời, và

dữ liệu đầu ra có dạng khung hình bao quanh đối tượng được nhận diện và có chứa ít nhất là thông tin về tọa độ vị trí của khung hình này;

Bước 3: lưu trữ, bởi khối lưu trữ và tính toán 400, các dữ liệu đầu ra liên quan đến ít nhất là một đối tượng được nhận diện là các khung hình có thông tin về tọa độ vị trí theo mảng dữ liệu, và tính toán để xác định có hay không có sự di chuyển của đối tượng được nhận diện, dựa trên mức độ sự sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện ở hai hoặc nhiều hơn hai thời điểm liên tiếp nhau;

Bước 4: phát cảnh báo, bởi khối phát cảnh báo 500, để phát cảnh báo có đối tượng xâm nhập trong trường hợp phát hiện sự có mặt của đối tượng xâm nhập và phát hiện có sự di chuyển của đối tượng được nhận diện tương ứng,

trong đó, khối phát cảnh báo 500 tiếp tục phát cảnh báo trong trường hợp tỉ lệ số lần phát hiện sự có mặt của đối tượng xâm nhập đối với số các lần phát hiện trong một khoảng thời gian nhất định lớn hơn hoặc bằng tỉ lệ duy trì cảnh báo được xác định trước, và khối phát cảnh báo không phát cảnh báo hoặc không duy trì phát cảnh báo trong các trường hợp khác với các trường hợp được nêu trên.

Theo một phương án ưu tiên, tập dữ liệu huấn luyện đầu vào bao gồm các ảnh dễ bị nhận diện nhầm để nâng cao độ chính xác, với số lượng ảnh có thể lên tới 50% so với các ảnh có người.

Các ảnh dễ bị nhận diện nhầm này có thể được xác định là một số đồ vật trong nhà có cấu tạo giống hình dáng của người, chẳng hạn như có đầu tròn và thân như là quạt, ghế, tay vịn cầu thang, v.v.. Với dạng này, tăng cường số lượng ảnh chỉ chứa các đồ nội thất trong nhà (không có người) với tỉ lệ số lượng ảnh dễ bị nhận diện nhầm so với số lượng ảnh có người là 2:1. Các ảnh dễ bị nhận diện nhầm này có thể được xác định là những con vật trong nhà như chó, mèo, v.v.. Với dạng này, mô hình YOLO có thể được bổ sung thêm một lớp phân loại với các loại động vật trong nhà như chó, mèo, trong đó các ảnh chứa chó, mèo được thêm vào với tỉ lệ so với số ảnh có người là 1:2. Các ảnh chó, mèo được thêm vào cần mô phỏng đúng hình ảnh chó, mèo từ camera (góc nhìn từ trên xuống). Các ảnh dạng chụp bằng điện thoại thông minh hay tương tự có thể không mang lại hiệu quả cao.

Ưu tiên là, phạm vi phân tích để xác định các ảnh dễ bị nhận diện nhầm có thể được thu hẹp dựa trên phạm vi sử dụng của các camera, ví dụ các camera chỉ được áp dụng ở các khu vực trong nhà hoặc ngoài nhà thì các dữ liệu được thêm vào cũng chỉ cần nằm trong phạm vi tương ứng là trong nhà hoặc ngoài nhà.

Theo một phương án ưu tiên khác, tập dữ liệu huấn luyện đầu vào có thể được làm phong phú bằng cách thay đổi kích thước ảnh đầu vào, trộn lẫn các ảnh với nhau hoặc tương tự, và tốc độ học có thể nhỏ hơn hoặc bằng 0,0001 để nâng cao độ chính xác của mô hình sau khi được huấn luyện.

Các hình ảnh được thu thập trong vùng xâm nhập của camera bởi khối thu thập hình ảnh 200 có thể là các hình ảnh lấy trực tiếp từ camera hoặc thông qua các đầu thu được nối giữa camera với khối thu thập hình ảnh. Khối thu thập hình ảnh 200 có thể thu nhận các hình ảnh thông qua giao thức truyền phát thời gian thực đã biết nào đó. Theo một ví dụ cụ thể, hình ảnh hoặc chuỗi hình ảnh được

thu thập trong vùng xâm nhập của camera sẽ được thu về kích thước mặc định 512 x 288 để tối ưu tốc độ xử lý.

Bên cạnh đó, kích thước ảnh đầu có thể được thay đổi vào tùy thuộc vào số lượng camera, chất lượng camera của người dùng để tăng cường độ chính xác, ví dụ kích thước ảnh càng lớn thì độ chính xác càng cao. Số khung hình trên giây (FPS – Frames Per Second) có thể được thay đổi tùy thuộc vào số lượng camera tại một thời điểm sử dụng của người dùng để tăng hiệu suất của mô hình và hệ thống, ví dụ số khung hình trên giây càng lớn, thời gian phát hiện đối tượng xâm nhập và độ phản hồi của hệ thống càng nhanh.

Trong các điều kiện thiếu sáng vào buổi tối, dữ liệu bổ sung có thể được thêm vào để nâng cao độ chính xác.

Ưu tiên là, tỉ lệ duy trì cảnh báo được xác định trước nêu trên là 60%.

Ưu tiên là, hai thời điểm liên tiếp nhau nêu trên, mà dùng để tính toán sự sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện, cách nhau khoảng 1 giây. Việc xác định có hay không có sự di chuyển của đối tượng được nhận diện, dựa trên mức độ sự sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện ưu tiên là được xác định ở ba thời điểm liên tiếp nhau.

Theo một ví dụ cụ thể, việc xác định có hay không có sự di chuyển của đối tượng được nhận diện nhờ sử dụng thuật toán theo dõi IoU (Intersect over union Tracking) để xác minh trong ba giây liên tiếp có xuất hiện chuyển động hay không. Về thuật toán IoU, như đã trình bày ở trên, mỗi giây sẽ có các khung hình của cùng một đối tượng được nhận diện, ví dụ người, đầu người, hoặc mặt người. Ở các giây tiếp theo, nếu các khung hình của cùng một đối tượng được trả ra cơ bản là “trùng” với các khung hình của cùng một đối tượng ở giây trước đó thì đối tượng được nhận diện được coi là đứng yên (“trùng” ở đây là khi hai khung hình có $\text{IoU} = 0,9$). Cần hiểu rằng, việc xác định chuyển động không nhất

thiết phải được thực hiện cho tất cả các đối tượng được nhận diện mà có thể chỉ thực hiện cho một đối tượng được nhận diện, ví dụ người.

Theo một phương án ưu tiên, một hoặc nhiều camera (ví dụ, sáu camera chẳng hạn) có thể cùng được kết nối với một khối xử lý trí tuệ nhân tạo chung để giảm yêu cầu tài nguyên tính toán, và mỗi camera có thể được thiết lập hai vùng xâm nhập để nâng cao khả năng phát hiện. Các vùng xâm nhập này có thể được thiết đặt khi sử dụng tùy theo yêu cầu của người dùng hoặc trong hoàn cảnh thực tế.

Theo một phương án ưu tiên khác, việc phát cảnh báo có thể bao gồm bật đèn, bật còi hú, thông báo cho người dùng thông qua điện thoại, tin nhắn, thư điện tử, hoặc tương tự khi có người di chuyển tới trong vùng xâm nhập.

Việc gửi thông báo cho người dùng có thể được thực hiện thông qua ứng dụng phần mềm bất kỳ, ứng dụng này ưu tiên là sẽ gửi ảnh chụp nhanh tại thời điểm phát hiện cho chủ nhà để chủ nhà có thể dễ dàng xác nhận đây là cảnh báo chính xác hay cảnh báo nhầm. Với những trường hợp cảnh báo nhầm, ảnh chụp nhanh này có thể sẽ được gửi về máy chủ để xử lý lỗi và cập nhật mô hình mới nhất cho nhủ nhà, điều này có thể cần khách hàng đồng ý cho phép dữ liệu được chuyển lên máy chủ để làm điều kiện quyết định thực thi hay không.

Ưu tiên là, đèn ở các khu vực hoặc phạm vi xung quang camera có thể được tự động tắt khi không phát hiện thấy có người trong vùng xâm nhập.

Đã biết, trong quá trình huấn luyện mô hình YOLO, một hoặc một số giải thuật hoặc phương án khác nhau có thể được áp dụng để tối ưu hoá bộ tham số của mô hình YOLO. Dưới đây sẽ mô tả một số giải thuật hoặc phương án để làm ví dụ minh hoạ. Tuy nhiên, cần hiểu rằng, sáng chế không bị giới hạn ở đó.

Xử lý tập dữ liệu đầu vào

Một nhóm ảnh bao gồm số lượng ảnh xác định, ví dụ mười ảnh khác nhau có thể được trộn theo cách ngẫu nhiên và đưa vào huấn luyện. Việc này làm phong phú dữ liệu đầu vào, và so với việc chỉ dùng một ảnh đơn thì việc kết hợp

các ảnh sẽ giúp kết quả thu được có độ chính xác được nâng cao, theo nhận định của các tác giả có thể lên tới 1%.

Ngoài ra, các bức ảnh đầu vào cũng sẽ được nhân, giảm với tỉ lệ nhất định để thu phóng kích thước ảnh (multi-scale). Tỷ lệ có thể nằm trong khoảng $\pm 50\%$ so với kích thước ảnh ban đầu. Điều này có thể làm tăng độ chính xác lên tới khoảng xấp xỉ 1,5% theo nhận định của các tác giả.

Thay đổi tốc độ học

Việc thay đổi tốc độ học có thể sử dụng giải thuật được gọi là OneCycle, có thể được tham khảo rõ hơn theo bài báo có liên kết tại: <https://towardsdatascience.com/finding-good-learning-rate-and-the-one-cycle-policy-7159fe1db5d6>. Toàn bộ nội dung của nó được đưa vào đây để viện dẫn.

Tốc độ học có thể được thay đổi xuống 0,0001, và kết hợp với việc xử lý dữ liệu đầu vào, ví dụ học chín ảnh đồng thời, trộn hai ảnh với nhau trong đó ảnh có người được làm nổi, ảnh ko có người được làm mờ, thu phóng bức ảnh để học ở các tỉ lệ khác nhau, tiến hành thu thập nhiều dữ liệu không có người đặc biệt ở các trường hợp hay phát hiện nhằm để nâng cao độ chính xác, loại bỏ các bức ảnh gây nhiễu, không phù hợp.

Lựa chọn các hàm cụ thể

Hàm kích hoạt (Activation Function) cho các lớp ẩn và lớp trung gian có thể sử dụng giải thuật có tên gọi là Leaky Relu, và sử dụng hàm Sigmoid cho lớp phát hiện.

Hàm tối ưu có thể sử dụng hàm SGD (Stochastic Gradient Descent) để tối ưu, có thể được tham khảo rõ hơn theo bài báo có liên kết tại: <https://arxiv.org/pdf/1609.04747.pdf>. Toàn bộ nội dung của nó được đưa vào đây để viện dẫn. Nói chung, hàm SGD có thể mang lại hiệu quả tốt hơn các phương pháp tối ưu khác khoảng xấp xỉ 1,5% theo nhận định của các tác giả.

Kết quả mô hình YOLO với bộ tham số được tối ưu hoá có thể sẽ được chuyển đổi thành mô hình có kích thước nhỏ hơn để có thể chạy được trên các thiết bị máy tính nhúng có khả năng tính toán giảm mà vẫn không làm giảm đi độ chính xác một cách đáng kể, việc chuyển đổi này có thể được thực hiện nhờ công cụ Jetson Nano và thông qua bộ chuyển đổi Tensorrt.

Trong trường hợp mô hình YOLO với bộ tham số được tối ưu hoá có thể sẽ được chuyển đổi thành mô hình có kích thước nhỏ hơn, có lợi là toàn bộ hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh theo sáng chế có thể được đóng gói nhỏ gọn, ví dụ ở dạng các thiết bị nhỏ gọn như các hộp xử lý trí tuệ nhân tạo (AI Box), trong đó mỗi AI Box này có thể xử lý chung cho số lượng các camera lên tới sáu camera chẳng hạn. Tuy nhiên, sáng chế không bị giới hạn ở đó, các phương án thực hiện hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh có thể là bất kỳ, bao gồm cả trường hợp mà trong đó các thành phần chính của hệ thống này có thể được bố trí ở các vị trí cách xa nhau, ví dụ khối thu thập hình ảnh có thể được bố trí ở các vị trí lân cận với các camera mà nó thu nhận hình ảnh từ đó, trong khi khối xử lý trí tuệ nhân tạo có thể được đặt tại máy chủ từ xa, chẳng hạn.

Về mặt lý thuyết, việc xử lý và nhận diện đối tượng xâm nhập bằng hình ảnh sử dụng mô hình YOLO v5 thông thường có thể tạo ra độ chính xác nhận diện người lên tới 97%. Giải pháp theo sáng chế nâng cao được độ chính xác của mô hình YOLO v5 trong ứng dụng nhận diện đối tượng xâm nhập được mô tả trên đây thông qua việc nhận diện người, đầu người, mặt người nêu trên được thực hiện một cách đồng thời, và kết hợp phát hiện chuyển động.

Dễ thấy là, khi có người xâm nhập được phát hiện, không chỉ đối tượng cần nhận diện là người (toàn phần người này được phát hiện), mà đồng thời các đối tượng cần nhận diện là đầu và mặt của người này cũng được phát hiện để bổ sung thêm cơ sở cho việc đưa ra kết quả. Tiếp đó, chuyển động của một trong số các yếu tố là người này, đầu của người này, hoặc mặt của người này cũng được xác định, và khi kết hợp sẽ nâng cao đáng kể độ chính xác khi kết luận là có

người xâm nhập, cụ thể hơn cần xác định được đồng thời có người xâm nhập bằng hình ảnh và người xâm nhập này phải có chuyển động thì mới đưa ra kết quả cuối cùng là có người xâm nhập.

Tiếp theo, quá trình hoạt động của hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh sẽ được theo dõi dựa trên các kết quả cảnh báo (chính xác hoặc không chính xác) trong thực tế và thu thập dữ liệu để có thể cập nhật mô hình. Đã biết, các mô hình trí tuệ nhân tạo có khả năng liên tục thu thập dữ liệu và học từ những trường hợp thực tế để làm phong phú thêm tập dữ liệu phục vụ việc nâng cao độ chính xác, có thể tổng quát hóa và xử lý toàn bộ các trường hợp trong thực tế với độ chính xác tốt hơn.

Nói chung, giải pháp theo sáng chế có thể được thực hiện sử dụng các camera có sẵn kết hợp với máy tính (hoặc các thiết bị tính toán tương tự) được cài đặt phần mềm/chương trình máy tính chứa các lệnh có thể được chạy trên máy tính để làm cho máy tính thực thi phương pháp phát hiện đối tượng xâm nhập bằng hình ảnh hoặc tạo ra các chức năng cho hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh. Theo đó, một phương án ưu tiên của sáng chế cũng đề cập đến vật ghi để lưu trữ chương trình máy tính nêu trên.

Tiếp theo, lưu đồ giải thuật để thực hiện sáng chế theo một ví dụ cụ thể sẽ được mô tả chi tiết có dựa trên Hình 3. Cần hiểu rằng, sáng chế không bị giới hạn ở đó, và sáng chế có thể được thực hiện theo nhiều cách thức và giải thuật khác nhau, miễn là tạo ra các chức năng hoặc đặc điểm như được mô tả trên đây.

Như được thể hiện trên hình vẽ, sau khi bắt đầu, giải thuật theo sáng chế sẽ thực hiện theo nhánh S1, xác định từ vùng người dùng chọn làm vùng xâm nhập cần phát hiện đối tượng xâm nhập, dịch ra theo biên của vùng này để lấy vùng có kích thước lớn hơn. Tiếp theo, sẽ lấy dữ liệu từ camera và/hoặc đầu ghi, dữ liệu này là các ảnh được thu nhận trong vùng xâm nhập được thiết đặt camera, và dữ liệu này được đưa vào để nhận diện bằng mô hình AI, cụ thể là mô hình YOLO v5 được đã mô tả chi tiết trên đây.

Tiếp theo, sẽ xác định trong một khoảng thời gian (block), tỷ lệ số khung hình nhận diện có đối tượng xâm nhập (ở đây đối tượng xâm nhập được xác định là người) so với tổng số khung hình được nhận diện hay không, trong đó mỗi khung hình được nhận diện bởi mô hình AI sẽ trả về kết quả là khung hình này có đối tượng xâm nhập cần phát hiện trong vùng quan sát hay không. Nếu tỷ lệ này nhỏ hơn 60%, thì không kích hoạt thiết bị, và không bật/phát cảnh báo an ninh. Nếu tỷ lệ này lớn hơn hoặc bằng 60%, thì sẽ thực hiện việc kiểm tra đối tượng nằm trong vùng to hay vùng nhỏ, ở đây vùng nhỏ là vùng người dùng chọn làm vùng xâm nhập và vùng to là vùng có kích thước lớn hơn nêu trên.

Nếu là vùng to, sẽ chuyển sang thực hiện theo nhánh S11, nếu là vùng nhỏ sẽ chuyển sang thực hiện theo nhánh S12.

Theo nhánh S11, sẽ thực hiện kiểm tra trước đó có phát sự kiện cảnh báo hay không? Nếu không sẽ chuyển sang thực hiện theo nhánh S111 và nếu có sẽ chuyển sang thực hiện theo nhánh S112. Ở nhánh S111, sẽ thực hiện việc phát hiện xem có sự xuất hiện của đầu người và mặt người bổ sung căn cứ khi xác định có đối tượng xâm nhập là người ở nhánh S1 nêu trên, nếu không thì không kích hoạt thiết bị (ở đây có thể được hiểu là thiết bị báo động hoặc các thiết bị tương tự, hoặc các thiết bị liên quan), nếu có sẽ gửi bản tin và kích hoạt thiết bị. Ở nhánh S111 cũng đồng thời phát hiện về trạng thái của vật thể (đối tượng xâm nhập), nếu trạng thái là tĩnh thì không kích hoạt thiết bị, nếu trạng thái là động thì gửi bản tin và kích hoạt thiết bị, có thể kết hợp còi hú xua đuổi trộm trong một số trường hợp.

Ở nhánh S112, sẽ phát hiện về trạng thái của vật thể, nếu trạng thái là động thì gửi bản tin và tiếp tục kích hoạt thiết bị, nếu trạng thái là tĩnh thì sẽ kiểm tra trong một khoảng thời gian, tỷ lệ số khung hình nhận diện được đầu người so với tổng số toàn bộ khung hình nhận diện đầu người có lớn hơn hoặc bằng 60% không. Nếu tỷ lệ này nhỏ hơn 60% thì sẽ tắt kích hoạt thiết bị, nếu tỷ lệ này lớn hơn hoặc bằng 60% sẽ gửi bản tin và tiếp tục kích hoạt thiết bị.

Theo nhánh S12, sẽ thực hiện kiểm tra trước đó có phát sự kiện cảnh báo hay không? Nếu không sẽ chuyển sang thực hiện theo nhánh S121 và nếu có sẽ chuyển sang thực hiện theo nhánh S122. Ở nhánh S121, sẽ thực hiện việc phát hiện xem có sự xuất hiện của đầu người và mặt người bổ sung căn cứ khi xác định có đối tượng xâm nhập là người ở nhánh S1 nêu trên, nếu không thì không kích hoạt thiết bị, nếu có sẽ gửi bản tin và kích hoạt thiết bị. Ở nhánh S121 cũng đồng thời phát hiện về trạng thái của vật thể (đối tượng xâm nhập), nếu trạng thái là tĩnh thì không kích hoạt thiết bị, nếu trạng thái là động thì gửi bản tin và kích hoạt thiết bị, có thể kết hợp gửi thông báo qua ứng dụng, thư điện tử, tin nhắn trong một số trường hợp.

Ở nhánh S122, sẽ phát hiện về trạng thái của vật thể, nếu trạng thái là động thì gửi bản tin và tiếp tục kích hoạt thiết bị, nếu trạng thái là tĩnh thì sẽ kiểm tra trong một khoảng thời gian, tỷ lệ số khung hình nhận diện được đầu người so với tổng số toàn bộ khung hình nhận diện đầu người có lớn hơn hoặc bằng 60% không. Nếu tỷ lệ này nhỏ hơn 60% thì sẽ tắt kích hoạt thiết bị, nếu tỷ lệ này lớn hơn hoặc bằng 60% sẽ gửi bản tin và tiếp tục kích hoạt thiết bị.

Hiệu quả có thể đạt được của sáng chế

Các giải pháp hiện nay đều chỉ tập trung vào nhận diện người riêng biệt nên khả năng phát hiện nhằm gây báo động giả. Hơn nữa bản chất trí tuệ nhân tạo trong xử lý hình ảnh là bất chước đôi mắt của con người nên sẽ có xu hướng lặp lại những lỗi đã từng xảy ra.

Giải pháp theo sáng chế xử lý thêm hai thuộc tính đặc trưng của con người là: nhận diện mặt, đầu và phát hiện chuyển động để bổ sung thêm cơ sở cho việc xác định nhằm tăng cường độ chính xác, giảm thiểu báo động giả.

Tính năng an ninh: khi phát hiện đối tượng ở gần vùng cấm xâm nhập, kết hợp với điều kiện là đối tượng đó di chuyển hoặc phát hiện đầu, mặt của đối tượng đó sẽ đủ khẳng định đó là người và tiến hành ghép nối đầu ra ở bước sau;

khi người vào vùng cấm xâm nhập, tương tự sẽ áp dụng logic trên để ghép nối đầu ra ở bước sau.

Tính năng cảm biến ánh sáng: khi thấy có người đi lân cận vùng, hệ thống tiến hành nhận diện, phát hiện đối tượng kết hợp phát hiện chuyển động và nhận diện đầu để khẳng định đó là người. Sau đó khi đối tượng vào vùng sẽ ngay lập tức bật đèn. Khi không có đối tượng trong vùng thì sẽ tắt đèn. Nếu đối tượng ở trong vùng nhưng đứng yên, thì hệ thống sẽ nhận dạng xem có đầu hay mặt của đối tượng đó không để quyết định đó có phải là người hay không hay là phát hiện nhầm.

Trên đây, sáng chế đã được mô tả chi tiết theo các phương án ưu tiên thực hiện. Hiển nhiên là, người có hiểu biết trung bình về lĩnh vực kỹ thuật này có thể dễ dàng tạo ra các thay đổi và cải biến dựa vào các phương án được mô tả. Do đó, các thay đổi và cải biến này được coi là không nằm ngoài phạm vi của sáng chế như được xác định trong yêu cầu bảo hộ kèm theo.

YÊU CẦU BẢO HỘ

1. Phương pháp phát hiện đối tượng xâm nhập bằng hình ảnh bao gồm các bước:

thu thập dữ liệu hình ảnh, bởi khối thu thập hình ảnh, để thu thập các hình ảnh trong vùng xâm nhập của camera;

xử lý, bởi khối xử lý trí tuệ nhân tạo, sử dụng dữ liệu hình ảnh nêu trên làm dữ liệu đầu vào, để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra liên quan đến đối tượng được nhận diện gồm có dữ liệu nhận diện người, dữ liệu nhận diện đầu người, và dữ liệu nhận diện mặt người, nhờ đó xác định có hay không sự có mặt của đối tượng xâm nhập, trong đó:

khối xử lý trí tuệ nhân tạo sử dụng mô hình YOLO v5 để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra nêu trên, mô hình này được huấn luyện để đưa ra các tham số mô hình dựa trên tập dữ liệu huấn luyện đầu vào gồm có các dữ liệu có người, đầu người, và mặt người được dán nhãn, được trộn cùng với các dữ liệu không có người, đầu người, và mặt người,

việc nhận diện người, đầu người, mặt người nêu trên được thực hiện một cách đồng thời, và

dữ liệu đầu ra có dạng khung hình bao quanh đối tượng được nhận diện và có chứa ít nhất là thông tin về tọa độ vị trí của khung hình này;

lưu trữ các dữ liệu đầu ra liên quan đến ít nhất là một đối tượng được nhận diện là các khung hình có thông tin về tọa độ vị trí theo mảng dữ liệu, và tính toán để xác định có hay không có sự di chuyển của đối tượng được nhận diện, dựa trên mức độ sự sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện ở hai hoặc nhiều hơn hai thời điểm liên tiếp nhau;

phát cảnh báo, bởi khối phát cảnh báo, để phát cảnh báo có đối tượng xâm nhập trong trường hợp phát hiện sự có mặt của đối tượng xâm nhập và phát hiện có sự di chuyển của đối tượng được nhận diện tương ứng,

trong đó, khối phát cảnh báo tiếp tục phát cảnh báo trong trường hợp tỉ lệ số lần phát hiện sự có mặt của đối tượng xâm nhập đối với số các lần phát hiện trong một khoảng thời gian nhất định lớn hơn hoặc bằng tỉ lệ duy trì cảnh báo được xác định trước, và khối phát cảnh báo không phát cảnh báo hoặc không duy trì phát cảnh báo trong các trường hợp khác với các trường hợp được nêu trên.

2. Phương pháp theo điểm 1, trong đó tập dữ liệu huấn luyện đầu vào bao gồm các ảnh dễ bị nhận diện nhằm để nâng cao độ chính xác, với số lượng ảnh có thể lên tới 50% so với các ảnh có người.

3. Phương pháp theo điểm 1 hoặc 2, trong đó tập dữ liệu huấn luyện đầu vào có thể được làm phong phú bằng cách thay đổi kích thước ảnh đầu vào, trộn lẫn các ảnh với nhau hoặc tương tự, và tốc độ học có thể nhỏ hơn hoặc bằng 0,0001 để nâng cao độ chính xác của mô hình sau khi được huấn luyện.

4. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó tỉ lệ duy trì cảnh báo được xác định trước nêu trên là 60%.

5. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó hai thời điểm liên tiếp nhau nêu trên, mà dùng để tính toán sự sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện, cách nhau khoảng 1 giây.

6. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó một hoặc nhiều camera để thu thập các hình ảnh có thể cùng được kết nối với một khối xử lý trí tuệ nhân tạo chung để giảm yêu cầu tài nguyên tính toán, và mỗi camera có thể được thiết lập hai vùng xâm nhập để nâng cao khả năng phát hiện.

7. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó việc phát cảnh báo có thể bao gồm bật đèn, bật còi hú, thông báo cho người dùng thông qua điện thoại, tin nhắn, thư điện tử, hoặc tương tự khi có người di chuyển tới trong vùng xâm nhập.

8. Phương pháp theo điểm 7, trong đó phương pháp này còn bao gồm bước tự động tắt đèn khi không phát hiện thấy có người trong vùng xâm nhập.

9. Hệ thống phát hiện đối tượng xâm nhập bằng hình ảnh bao gồm:

khối thu thập hình ảnh để thu thập dữ liệu hình ảnh là các hình ảnh trong vùng xâm nhập của camera;

khối xử lý trí tuệ nhân tạo, sử dụng dữ liệu hình ảnh nêu trên làm dữ liệu đầu vào, để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra liên quan đến đối tượng được nhận diện gồm có dữ liệu nhận diện người, dữ liệu nhận diện đầu người, và dữ liệu nhận diện mặt người, nhờ đó xác định có hay không sự có mặt của đối tượng xâm nhập, trong đó:

khối xử lý trí tuệ nhân tạo sử dụng mô hình YOLO v5 để xử lý dữ liệu đầu vào và đưa ra kết quả là các dữ liệu đầu ra nêu trên, mô hình này được huấn luyện để đưa ra các tham số mô hình dựa trên tập dữ liệu huấn luyện đầu vào gồm có các dữ liệu có người, đầu người, và mặt người được dán nhãn, được trộn cùng với các dữ liệu không có người, đầu người, và mặt người,

việc nhận diện người, đầu người, mặt người nêu trên được thực hiện một cách đồng thời, và

dữ liệu đầu ra có dạng khung hình bao quanh đối tượng được nhận diện và có chứa ít nhất là thông tin về tọa độ vị trí của khung hình này;

khối lưu trữ và tính toán để lưu trữ các dữ liệu đầu ra liên quan đến ít nhất là một đối tượng được nhận diện là các khung hình có thông tin về tọa độ vị trí theo mảng dữ liệu, và tính toán để xác định có hay không có sự di chuyển của đối tượng được nhận diện, dựa trên mức độ sự sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện ở hai hoặc nhiều hơn hai thời điểm liên tiếp nhau;

khởi phát cảnh báo để phát cảnh báo có đối tượng xâm nhập trong trường hợp phát hiện sự có mặt của đối tượng xâm nhập và phát hiện có sự di chuyển của đối tượng được nhận diện tương ứng,

trong đó, khởi phát cảnh báo tiếp tục phát cảnh báo trong trường hợp tỉ lệ số lần phát hiện sự có mặt của đối tượng xâm nhập đối với số các lần phát hiện trong một khoảng thời gian nhất định lớn hơn hoặc bằng tỉ lệ duy trì cảnh báo được xác định trước, và khởi phát cảnh báo không phát cảnh báo hoặc không duy trì phát cảnh báo trong các trường hợp khác với các trường hợp được nêu trên.

10. Hệ thống theo điểm 9, trong đó tập dữ liệu huấn luyện đầu vào bao gồm các ảnh dễ bị nhận diện nhằm để nâng cao độ chính xác, với số lượng ảnh có thể lên tới 50% so với các ảnh có người.

11. Hệ thống theo điểm 9 hoặc 10, trong đó tập dữ liệu huấn luyện đầu vào có thể được làm phong phú bằng cách thay đổi kích thước ảnh đầu vào, trộn lẫn các ảnh với nhau hoặc tương tự, và tốc độ học có thể nhỏ hơn hoặc bằng 0,0001 để nâng cao độ chính xác của mô hình sau khi được huấn luyện.

12. Hệ thống theo điểm bất kỳ trong số các điểm từ 9 đến 11, trong đó tỉ lệ duy trì cảnh báo được xác định trước nêu trên là 60%.

13. Hệ thống theo điểm bất kỳ trong số các điểm từ 9 đến 12, trong đó hai thời điểm liên tiếp nhau nêu trên, mà dùng để tính toán sự sai khác tọa độ vị trí giữa các khung hình của cùng một đối tượng được nhận diện, cách nhau khoảng 1 giây.

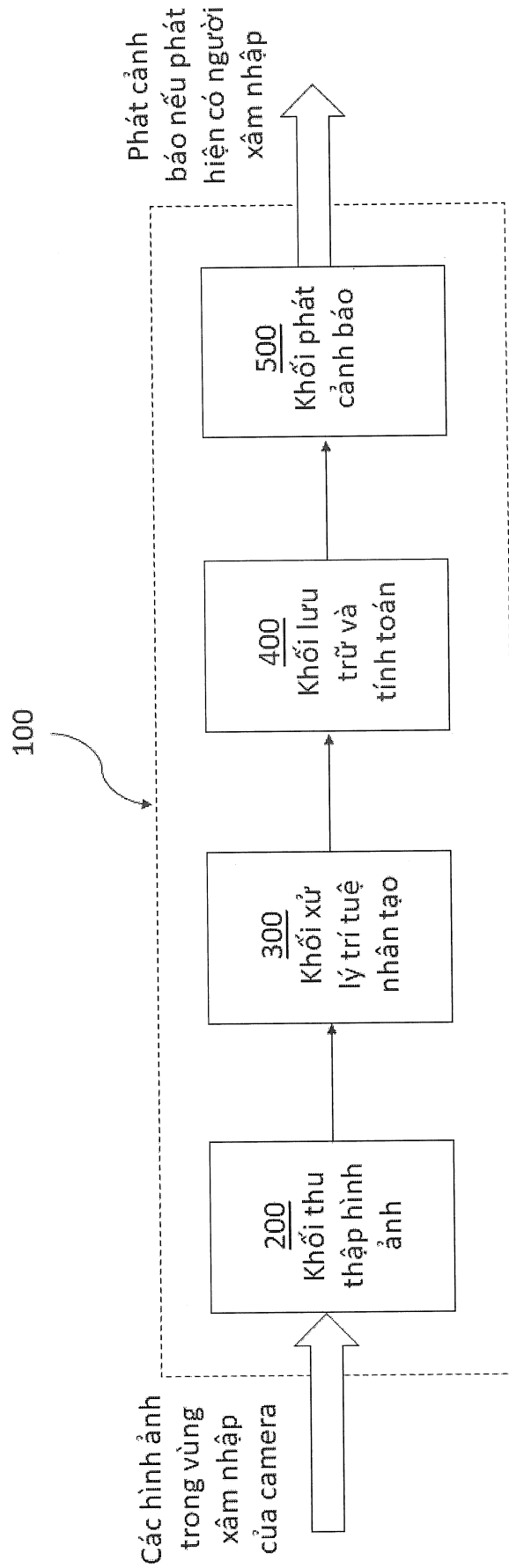
14. Hệ thống theo điểm bất kỳ trong số các điểm từ 9 đến 13, trong đó một hoặc nhiều camera để thu thập các hình ảnh có thể cùng được kết nối với một khối xử lý trí tuệ nhân tạo chung để giảm yêu cầu tài nguyên tính toán, và mỗi camera có thể được thiết lập hai vùng xâm nhập để nâng cao khả năng phát hiện.

15. Hệ thống theo điểm bất kỳ trong số các điểm từ 9 đến 14, trong đó việc phát cảnh báo có thể bao gồm bật đèn, bật còi hú, thông báo cho người dùng thông

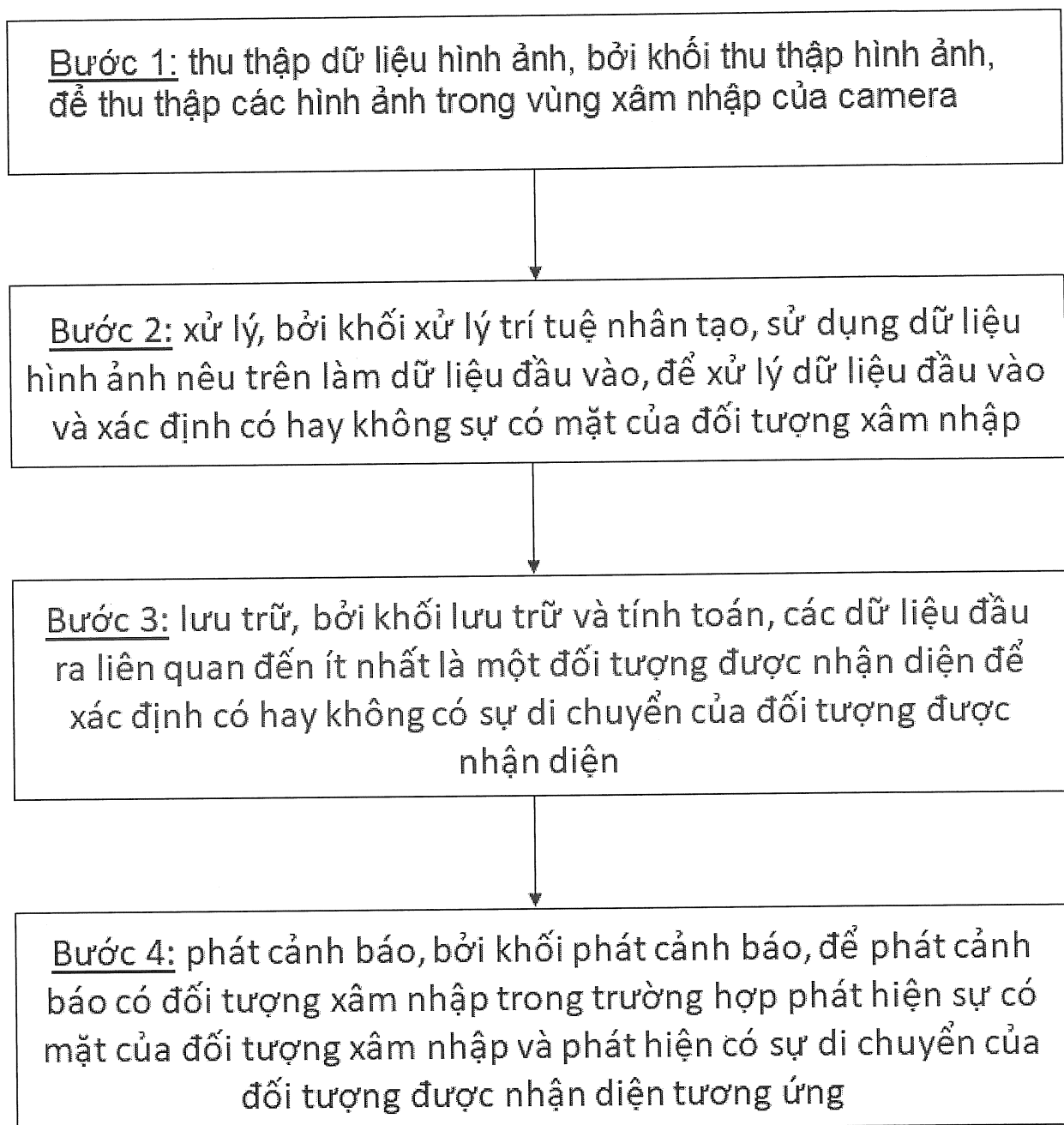
qua điện thoại, tin nhắn, thư điện tử, hoặc tương tự khi có người di chuyển tới trong vùng xâm nhập.

16. Hệ thống theo điểm 15, trong đó phương pháp này còn bao gồm bước tự động tắt đèn khi không phát hiện thấy có người trong vùng xâm nhập.

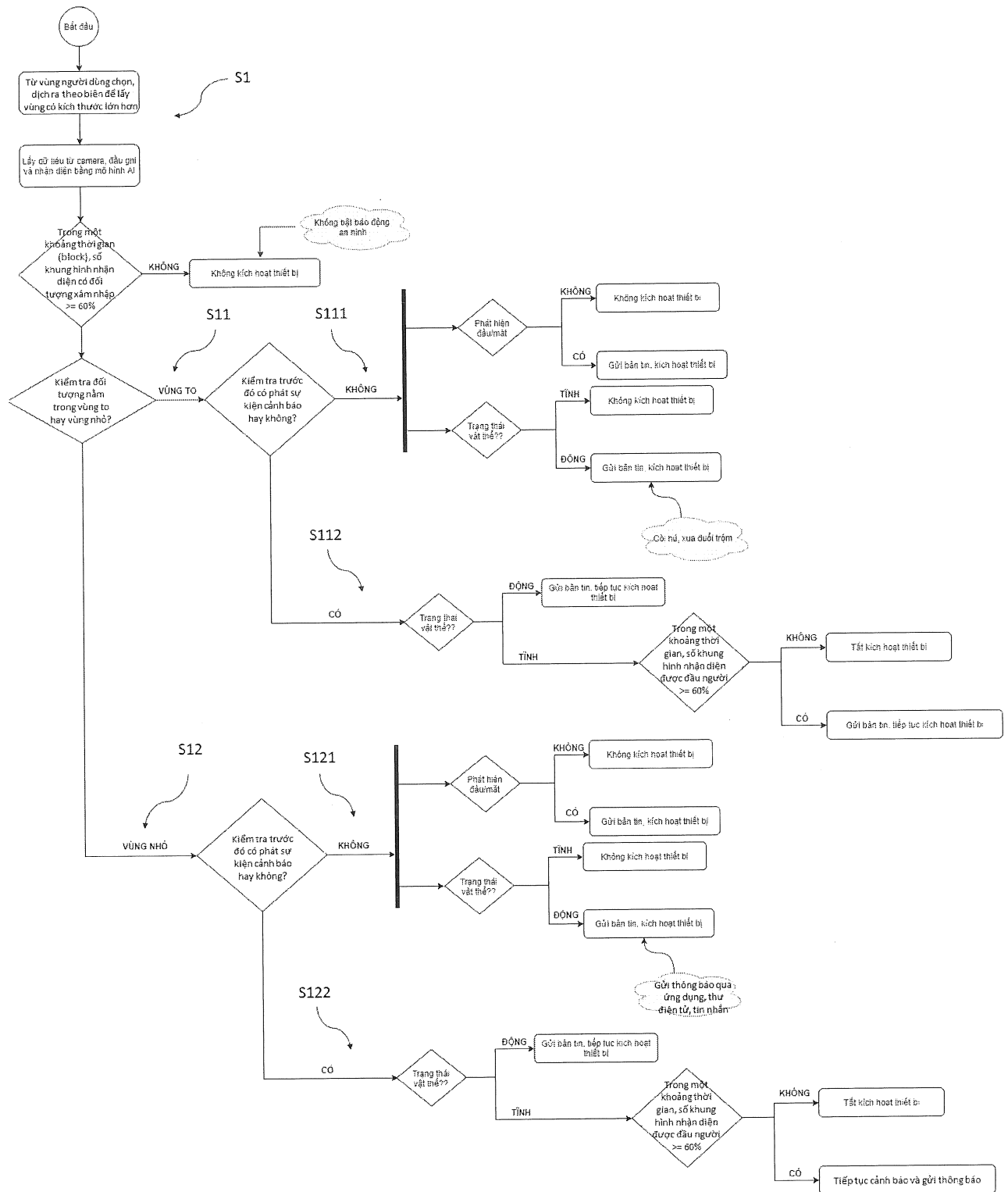
17. Vật ghi bất khả biến có chứa các lệnh máy tính mà khi được thực thi sẽ làm cho máy tính thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 8.



Hình 1



Hình 2



Hình 3