

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Ứng dụng phát hiện người di chuyển
dùng mô hình CNN YOLO

NGÔ VĂN GIANG

giang.nv170065@sis.hust.edu.vn

Ngành Công nghệ thông tin

Giảng viên hướng dẫn: PGS. TS. Nguyễn Thị Hoàng Lan

Chữ kí GVHD

Khoa: Kỹ thuật máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 01/2024

LỜI CẢM ƠN

Trong những tháng năm học tập, trải nghiệm tại Đại học Bách Khoa, tôi đã trải qua những nốt thăng trầm, những khó khăn tưởng chừng như không vượt qua được và có cả những niềm vui trong cuộc sống. Hơn tất cả, tôi đã vượt qua và bước tới chặng cuối cùng này. Quyển báo cáo này như lời xin lỗi chân thành và lời cảm ơn sâu sắc đến những người đã đồng hành cùng tôi ở Bách Khoa này. Con xin cảm ơn gia đình, đặc biệt người bố đã mất trước khi nhìn thấy con tốt nghiệp. Em xin cảm ơn các thầy cô đã giảng dạy, giúp đỡ em trong suốt thời gian qua, đặc biệt cô Nguyễn Thị Hoàng Lan đã cho em cơ hội thứ hai này. Cháu xin cảm ơn cô Phạm Thanh Huyền đã giúp đỡ cháu vượt qua giai đoạn khó khăn nhất. Những kiến thức, kỹ năng tại nơi đây sẽ theo em suốt chặng đường sau này.

TÓM TẮT NỘI DUNG ĐỒ ÁN

Đồ án “*Ứng dụng phát hiện người di chuyển dùng mô hình CNN YOLO*” được thực hiện nhằm xây dựng một ứng dụng có thể thực thi cơ bản việc xử lý phát hiện người chuyển động từ video, cụ thể ở đây là dùng mạng YOLO. Mục đích của ứng dụng dùng để có thể áp dụng mạng YOLO để phát hiện người chuyển động, từ đó đưa ra những kết quả đánh giá mô hình đối với từng tập dữ liệu. Người dùng có thể thay đổi, điều chỉnh một số tham số, từ đó đưa ra số lượng người có trong khung hình, giá trị điểm đánh giá (accuracy) khi đã phát hiện được người, người chuyển động. Báo cáo đồ án này sẽ trình bày những nghiên cứu đã tìm hiểu được về bài toán phát hiện người chuyển động và các giải pháp đã được áp dụng. Qua những tìm hiểu chung đó để dẫn đến giải pháp dùng mạng YOLO để đưa ra phát hiện người chuyển động. Từ nền tảng đó, xây dựng một ứng dụng để thực thi giải pháp trên một cách trực quan, thân thiện, dễ sử dụng.

Sinh viên thực hiện
(Ký và ghi rõ họ tên)

Ngô Văn Giang

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và phạm vi đề tài.....	1
1.3 Định hướng giải pháp.....	2
1.4 Bố cục đồ án	2
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT	4
2.1 Tổng quan về mạng CNN	4
2.1.1 Giới thiệu chung về mạng CNN.....	4
2.1.2 Cấu trúc mạng CNN.....	5
2.2 Tổng quan về mạng YOLO	9
2.3 Vấn đề phát hiện người chuyển động	10
CHƯƠNG 3. CÔNG NGHỆ SỬ DỤNG.....	13
3.1 Mô hình YOLOv4.....	13
3.2 Tensorflow Framework	17
3.3 Qt Framework	20
3.4 Thư viện OpenCV	21
3.5 Google Colab	22
3.6 Công cụ thiết kế giao diện Figma	22
3.7 Các ngôn ngữ lập trình	23
CHƯƠNG 4. XÂY DỰNG ỨNG DỤNG VÀ KẾT QUẢ THỬ NGHIỆM ...	24
4.1 Thiết kế, xây dựng giải pháp của ứng dụng.....	24
4.2 Thiết kế luồng xử lý và các chức năng của ứng dụng	25
4.2.1 Biểu đồ hoạt động	25
4.2.2 Thiết kế các chức năng	27

4.3 Thiết kế chi tiết.....	28
4.3.1 Thiết kế giao diện	28
4.3.2 Cài đặt các chức năng.....	31
4.4 Xây dựng ứng dụng.....	32
4.4.1 Thư viện và công cụ sử dụng.....	32
4.4.2 Kết quả đạt được	32
4.5 Kịch bản thực nghiệm và đánh giá kết quả	35
4.5.1 Kịch bản thực nghiệm.....	35
4.5.2 Đánh giá kết quả	37
CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	39
5.1 Kết luận.....	39
5.2 Hướng phát triển.....	39
TÀI LIỆU THAM KHẢO.....	42

DANH MỤC HÌNH VẼ

Hình 2.1	Hình ảnh dưới góc nhìn của con người và máy tính	5
Hình 2.2	Đầu vào và đầu ra của mạng CNN	5
Hình 2.3	Ma trận cửa sổ 3x3 trượt trên hàng đầu tiên trong lớp tích chập	6
Hình 2.4	Ma trận cửa sổ 3x3 trượt trên hàng thứ 2 và kết quả của lớp tích chập	6
Hình 2.5	Kết quả khi áp dụng các bộ lọc khác nhau	7
Hình 2.6	Hàm kích hoạt ReLu	8
Hình 2.7	Max polling	8
Hình 2.8	Kiến trúc các lớp trong Yolo	9
Hình 2.9	Cách Yolo dự đoán	10
Hình 2.10	Tổng quan của phát hiện người chuyển động từ đối tượng chuyển động	11
Hình 2.11	Hệ thống phát hiện người chuyển động từ đối tượng chuyển động	12
Hình 2.12	Hệ thống phát hiện người chuyển động từ phát hiện người . . .	12
Hình 3.1	Cấu trúc tổng quan của YOLOv4	13
Hình 3.2	Sự khác biệt giữa Residual Block và CsPresBlock	13
Hình 3.3	Sự khác biệt giữa DarkNet53 với CSPDarkNet53	14
Hình 3.4	Sự khác biệt giữa DropOut và Drop Block	15
Hình 3.5	Kiến trúc của mạng PAN	16
Hình 3.6	Cách thức hoạt động của SSP với đầu ra là Fully-Connected Layer	16
Hình 3.7	Cách thức SPP hoạt động trong YOLOv4	17
Hình 3.8	Cấu tạo của Tensor	19
Hình 3.9	Các thức hoạt động của dataflow graph trong Tensorflow . . .	19
Hình 3.10	Mô hình cấu tạo của dataflow graph	20
Hình 3.11	OpenCV Module	22
Hình 4.1	Luồng hoạt động của ứng dụng	24
Hình 4.2	Biểu đồ use case tổng quan của ứng dụng	25
Hình 4.3	Biểu đồ hoạt động chính của ứng dụng	26
Hình 4.4	Thiết kế giao diện cơ bản	28
Hình 4.5	Thiết kế giao diện màn hình mở đầu	29
Hình 4.6	Thiết kế giao diện màn hình chính	30
Hình 4.7	Thiết kế giao diện màn hình hiển thị kết quả	31

Hình 4.8	Kết quả giao diện màn hình mở đầu	33
Hình 4.9	Kết quả giao diện màn hình chính	33
Hình 4.10	Kết quả giao diện màn hình chính khi xử lý video	34
Hình 4.11	Kết quả giao diện màn hình hiển thị kết quả với cả 2 chức năng	34
Hình 4.12	Kết quả giao diện màn hình hiển thị kết quả với 1 chức năng .	35
Hình 4.13	Kết quả chạy với kịch bản thử nghiệm	37

DANH MỤC BẢNG BIỂU

Bảng 4.1	Danh sách thư viện và công cụ sử dụng	32
Bảng 4.2	Kịch bản thực nghiệm	36
Bảng 4.3	Đánh giá kết quả theo từng video	37

DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

Thuật ngữ	Ý nghĩa
AI	Trí tuệ nhân tạo (Artificial Intelligence)
API	Giao diện lập trình ứng dụng (Application Programming Interface)
CNN	Mạng neural tích chập (Convolutional Neural Networks)
CPU	Bộ xử lý trung tâm (Central Processing Unit)
FPN	Feature Pyramid Networks
GPU	Bộ xử lý đồ họa (Graphics Processing Unit)
NN	Mạng thần kinh nhân tạo (Neural Network)
PAN	Mạng tổng hợp lũy tiến (Path Aggregation Network)
SOTA	state-of-the-art
SPP	Spatial Pyramid Pooling
TPU	Bộ xử lý Tensor (Tensor Processing Unit)
UI	Giao diện người dùng (User Interface)
UX	Trải nghiệm người dùng (User Experience)
WOCA	Viết một lần, biên dịch mọi nơi (Write Once, Compile Anywhere)
YOLO	You Only Look Once

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

1.1 Đặt vấn đề

Hiện nay, chúng ta đang sống trong thời đại mà xã hội đề cao tầm quan trọng của Khoa học, Kỹ thuật và Công nghệ. Và khi khoa học, công nghệ, kỹ thuật phát triển thì những trí tuệ nhân tạo như trí tuệ nhân tạo, học máy, học sâu,.. cũng dần trở nên phổ biến. Và khi cuộc cách mạng công nghiệp 4.0 vẫn đang len lỏi đến đời sống của chúng ta hàng ngày, thì những khái niệm đó cũng đã và đang dần dần trở nên quen thuộc, được ứng dụng nhiều hơn trong cuộc sống, hứa hẹn đây sẽ là một ngành công nghiệp đầy tiềm năng và phát triển vượt bậc của ngành khoa học kỹ thuật nói chung, và ngành công nghệ thông tin nói riêng.

Với việc trí tuệ nhân tạo càng được ứng dụng rộng rãi, thì các bài toán trước kia tưởng như là bất khả thi như nhận diện, phân loại, xử lý với các loại dữ liệu phức tạp như âm thanh, tập dữ liệu lớn... thì cũng có đã phương pháp giải quyết. Một trong những bài toán được quan tâm nhất trong lĩnh vực trí tuệ nhân tạo đó là "Phát hiện người chuyển động (Human motion detection)".

Phát hiện người chuyển động đã là bài toán không còn xa lạ trong cuộc sống hiện nay. Bài toán này có phạm vi ứng dụng rất rộng. Chúng ta có thể chia thành ba lĩnh vực chính là giám sát, kiểm soát và phân tích. Việc nghiên cứu, áp dụng phát hiện người chuyển động ngày càng trở nên phổ biến.

Với nhu cầu và tính thiết thực cao như vậy, em đã chọn đề tài: “*Ứng dụng phát hiện người di chuyển dùng mô hình CNN YOLO*” với mục đích nghiên cứu, tìm hiểu và hoàn thành một ứng dụng phát hiện chuyển động con người.

1.2 Mục tiêu và phạm vi đề tài

Hiện tại có khá nhiều bài toán phát hiện người và phát hiện chuyển động hoặc cả hai dùng mạng YOLO. Những năm gần đây đã có những nghiên cứu của các nhà khoa học cho phát hiện người chuyển động. Hơn nữa khi mức độ nghiên cứu cho ra một mô hình có khả năng phát hiện tốt nhất ngày càng tiệm cận tới mức tối đa ở hiện tại thì lại chuyển sang nghiên cứu có thể đáp ứng nhu cầu thời gian thực. Vậy nên các hệ thống camera hay kể cả camera cũng có tích hợp sẵn những module này. Ứng dụng của đề tài ra đời nhằm góp phần đóng góp phát triển các ứng dụng đáp ứng nhu cầu của xã hội ngày càng cao.

Với mục đích là xây dựng một ứng dụng phát hiện người chuyển động từ video, cụ thể là dùng mạng YOLO nên để giải quyết vấn đề đó có hai nhiệm vụ chính: Tìm hiểu mô hình mạng YOLO để phát hiện người chuyển động và xây dựng ứng dụng