TRƯỜNG ĐẠI HỌC AN GIANG

**KHOA CÔNG NGHỆ THÔNG TIN**

**ĐỀ CƯƠNG KHOÁ LUẬN TỐT NGHIỆP**

**NGÀNH KỸ THUẬT PHẦN MỀM**

# NGHIÊN CỨU VÀ XÂY DỰNG HỆ THỐNG CAMERA GIÁM SÁT THÔNG MINH PHỤC VỤ CHĂM SÓC SỨC KHỎE TẠI NHÀ

Họ tên sinh viên : **Thái Trường An**

Mã số sv : **DPM175008**

Khóa : **18**

Giảng viên hướng dẫn: **TS. Đoàn Thanh Nghị**

## **AN GIANG, THÁNG 11 NĂM 2020**

**ĐỀ CƯƠNG KHÓA LUẬN TỐT NGHIỆP**

Họ và tên sinh viên: Thái Trường An Mã số sinh viên: DPM175008

Điện thoại: 0914518169 Email: thaitruonganlxag91@gmail.com

Là sinh viên lớp: DH18PM ngành Kỹ thuật phần mềm

thuộc Khoa Công nghệ Thông tin.

Tôi bắt đầu nhận đề tài khóa luận tốt nghiệp từ ngày 23 tháng 11 năm 2020 đến ngày 23 tháng 4 năm 2020

**Tên đề tài:**  Nghiên cứu và xây dựng hệ thống camera giám sát thông minh phục vụ chăm sóc sức khỏe tại nhà.

dưới sự hướng dẫn của Thầy/Cô: TS. Đoàn Thanh Nghị

**CHƯƠNG I**

**TỔNG QUAN**

* 1. **TÍNH CẤP THIẾT CỦA ĐỀ TÀI**

Ngày nay, khi kinh tế ngày càng phát triển cuộc sống con người cũng vì thế mà được nâng cao. Nhu cầu về chăm sóc sức khỏe trở nên là ưu tiên hàng đầu. Cùng với đó, là sự phát triển của khoa học và công nghệ trong cuộc cách mạng công nghiệp 4.0, lĩnh vực về chăm sóc sức khỏe lại được nâng lên tầm cao mới. Việc chăm sóc sức khỏe thông qua việc phát hiện các triệu chứng bất thường như: té xỉu, đau bụng, nhức đầu,… và cảnh báo các triệu chứng cho người thân trong gia đình là vô cùng cần thiết. Từ những thực tiễn đó là một sinh viên chuyên ngành Kỹ thuật phần mềm (KTPM) tôi mong muốn tìm hiểu, nghiên cứu và ứng dụng thị giác máy tính cũng như kỹ thuật học sâu vào tự động phát hiện các triệu chứng bất thường thông qua hành động con người.

Lịch sử giải quyết vần đề: nhận diện hành động con người đã từng được ứng dụng trong tóm tắt, truy vấn video và phân tích thể thao. Nhưng trên lĩnh vực camera giám sát chăm sóc sức khỏe tại nhà thì rất ít nghiên cứu trước đó. Vì vậy, nghiên cứu về camera giám sát thông minh có ý nghĩa khoa học và thực tiễn rất cao.

* 1. **LÍ DO CHỌN ĐỀ TÀI**

Cùng với sự phát triển của khoa học công nghệ trong cuộc cách mạng công nghiệp 4.0, các ứng dụng về trí tuệ nhân tạo (TTNT) và học máy ngày càng phát triển và là lĩnh vực được quan tâm nhiều nhất hiện nay, trong đó lĩnh vực học sâu (Deep Learning) là một loại phổ biến của máy học đã có những thành tựu phát triển vượt bậc. Mở ra một bước ngoặc mới trong việc giải quyết các bài toán về TTNT trước đây đã gặp khó khăn như nhận diện sự vật (Object Detection), nhận diện hình ảnh, hành động, … Trong đó phải kể đến sự phát triển của lĩnh vực thị giác máy tính (Computer Vision), đặt nền tảng cho nhiều ứng dụng thực tiển như các hệ thống giám sát thông minh, hệ thống xe tự lái, rôbôt thông minh,… Lĩnh vực học sâu là một kỹ thuật trong máy học sử dụng mạng nơ ron nhiều lớp để giải quyết các bài toán phức tạp dựa trên phương pháp tích chập và trích chọn các đặt trưng từ các tập dữ liệu lớn và đem lại kết quả chính xác cao trong giải quyết các bài toán TTNT trong thực tế. Cụ thể là nó tập trung giải quyết các vần đề về mạng nơ ron về thị giác máy tính, xử lý ngôn ngữ tự nhiên, xử lý giọng nói,… Hiện nay, cũng có nhiều nghiên cứu về nhận diện hành động con người thông qua video để xác định hành vi của con người. Tuy nghiên đây là một lĩnh vực tương đối rộng và hiện có nhiều kỹ thuật khác nhau để giải quyết bài toán này Vì vậy phạm vi luận văn này tập trung nghiên cứu lý thuyết và sử dụng các kỹ thuật mạng học sâu 3D CNN vào việc giải quyết bài toán nhận diện hành động con người đưa ra các dự đoán về các triệu chứng bất thường.

Trên cơ sở đó, tôi thực hiện khóa luận này để nghiên cứu ứng dụng kỹ thuật học sâu vào bài toán nhận diện hành động của con người trong video với đề tài “Nghiên cứu và xây dựng hệ thống camera giám sát thông minh phục vụ chăm sóc sức khỏe tại nhà”. Thông qua đề tài này góp phần phát triển hệ thống camera giám sát thông minh.

* 1. **MỤC TIÊU ĐỀ TÀI**

Sử dụng OpenCV, Python, Pytorch, Twilio để xây dựng mô hình camera giám sát thông minh với độ chính xác cao trong thực tế giúp nâng cao hiệu quả và chất lượng của quá trình giám sát đảm bảo sức khỏe cho các thành viên trong gia đình.

Sử dụng trực tiếp thuật toán xử lý trên bộ mạch Raspberry Pi 4 và tự động gửi tin nhắn thông báo TXT/MMS, hình ảnh, video clip khi camera được kích hoạt và phát hiện các hành động, triệu chứng bất thường của các thành viên trong gia đình.

* 1. **MỤC TIÊU NGHIÊN CỨU**

Mục tiêu chính của đề tài là nghiên cứu nhận dạng hành động người, tập trung vào các hành động, triệu chứng bất thường dựa trên kỹ thuật học sâu mạng tích chập 3D – CNN trong lĩnh vực thị giác máy tính.

Thực nghiệm áp dụng kỹ thuật mạng tích chập 3D-CNN trong trích xuất đặt trưng hình ảnh từ video để dự đoán hành động con người phục vụ chăm sóc sức khỏe tại nhà. Từ đó phân tích, đánh giá hiệu quả và độ chính xác trên các thư viện có sẵn.

Xây dựng tập dữ liệu và thực nghiệm dự đoán một số hành động, triệu chứng bất thường trong hệ thống camera giám sát thông minh.

* 1. **PHƯƠNG PHÁP NGHIÊN CỨU**

Phương pháp nghiên cứu tài liệu: Cơ sở lý thuyết về các kỹ thuật nhận diện hành động, các nghiên cứu tương tự về đề tài.

Phương pháp thực nghiệm: Xây dựng mô hình camera và cài đặt chương trình chạy thử nghiệm với một số video sẳn có và đánh giá kết quả. Từ đó xây dựng mô hình camera xử lý video thời gian thực.

**CHƯƠNG 2**

**CƠ SỞ LÝ THUYẾT**

**2.1. MỘT SỐ KHÁI NIỆM**

**2.1.1. Hệ thống thông minh**

**2.1.1.1. Khái niệm**

Hệ thống thông minh (Intelligent Systems - IS) là một hệ thống thực thi một mục đích hoạt động xã hội nào đó, nó được xem như một máy tính kết nối với các máy tính khác qua hệ thống mạng internet, có khả năng thu thập và phân tích dữ liệu và giao tiếp các hệ thống khác. Các tiêu chí của hệ thống thông minh bao gồm khả năng học hỏi từ kinh nghiệm, bảo mật, kết nối, khả năng thích ứng theo dữ liệu hiện tại và khả năng giám sát, quản lý từ xa.

Trong nghành công nghệ thông tin, một hệ thống thông minh được định nghĩa là một tập hợp các phần tử (Elements) hoặc các thành phần (Components) tạo nên một hệ thống thực hiện một mục đích chung. Như vậy, ngoài việc nó cấu thành từ các thiết bị thông minh, nó còn bao gồm các hệ thống kết nối các thiết bị lại với nhau: bao gồm mạng và các hệ thống lớn hơn, các hệ thống phần mềm thông minh (AI) như chatbots, hệ thông chăm sóc sức khỏe… và những loại phần mềm khác.

Như vậy các hệ thống thông minh có các đặc điểm sau:

* Cung cấp cách tiếp cận về mặt phương pháp luận chuẩn để giải quyết các vấn đề quan trong, phức tạp và nhận được các kết quả nhất quán, đáng tin cậy qua thời gian.
* Có khả năng kết nối, giao tiếp với các hệ thống khác, quản lý từ xa.
* Có trí tuệ bao gồm khả năng lĩnh hội, khả năng hiểu, khả năng rút kinh nghiệm, khả năng thu lượm và duy trì tri thức, khả năng đáp ứng nhanh và thành công trong các tình huống mới.
* Trí tuệ nhân tạo là cơ sở cốt yếu để xây dựng các hệ thống thông minh. Các hệ thống thông minh hiện nay được ứng dụng rất rộng rãi trong tất cả các lĩnh vực như y tế (hệ thống chẫn đoán bệnh), giáo dục (trường học thông minh), sản xuất Công nghiệp (Công nghệ in 3D), hệ thống xe tự hành, thành phố thông minh… là các hệ thống thông minh phổ biến hiện nay ở các nước phát triển.

**2.1.1.2. Hệ thống hỗ trợ chăm sóc sức khỏe bệnh nhân, người già**

Cùng với sự phát triển của trí tuệ nhận tạo, hệ thống robot chăm sóc sức khỏe cho người già ra đời và phát triển mạnh ở các nước phát triển như Nhật Bản, rôbôt có thể thay thế con người trong một số nhiệm vụ chăm sóc người già, tuy nhiên nó vẫn mang tính hỗ trợ, giải phóng các lao động đơn giản của y tá chưa hoàn toàn thay thế con người trong các dịch vụ chăm sóc y tế cho người già.

**2.1.2. Thị giác máy tính**

Thị giác máy tính là một lĩnh vực gồm các phương pháp thu nhận, phân tích, xử  
lý, nhận dạng hình ảnh và dữ liệu đa chiều từ thực tế để cho ra thông tin số hoặc biểu  
tượng dùng cho các mục đích nhất định.

Ý tưởng được dựa trên thị giác của con người về khả năng nhận dạng và hiểu  
một hình ảnh. Để làm được điều đó, thị giác máy tính áp dụng các mô hình xây dựng  
dựa trên các ngành lý thuyết học, vật lý, thống kê và hình học.

Trong các ứng dụng thị giác máy tính thường được lập trình để giải quyết một  
vấn đề cụ thể nhưng trên cơ bản các phương pháp đều dựa trên sự học (learning) trở  
nên ngày càng phổ biến. Ví dụ các ứng dụng thị giác máy tính như:

* Điều khiển tiến trình (Controlling processes) như rô bốt công nghiệp.
* Nhận diện khuôn mặt (Recognition Face).
* Sự kiểm tra tự động (Automation inspection) như kiểm tra chất lượng sản phẩm trong các nhà máy xí nghiệp.

**2.1.3. Nhận diện hành động của con người**

Nhận diện hành động của con người là một lĩnh vực nghiên cứu rộng rãi trong thị giác máy tính. Các ứng dụng của nó bao gồm các hệ thống giám sát, phân tích video, robot và một loạt các hệ thống có liên quan đến sự tương tác giữa con người và các thiết bị điện tử như giao diện người - máy tính.

Nhận diện hành động của con người được phát triển bắt đầu vào năm 1980. Hiện nay, các nghiên cứu chủ yếu tập trung vào việc học và nhận biết các hành động từ chuỗi video.

Một hành động là một chuỗi các chuyễn động cơ thể đồng thời. Từ quan điểm của thị giác máy tính, việc nhận dạng hành động này là để phù hợp với các quan sát (ví dụ: video) với các mẫu được xác định trước đó và sau đó gán cho nó một nhãn là loại hành động. Tùy thuộc vào độ phức tạp, hoạt động của con người có thể được phân loại thành bốn cấp độ cử chỉ, hành động, tương tác và hoạt động của nhóm và nhiều nghiên cứu theo hướng một cấu trúc từ dưới lên về nhận dạng hoạt động của con người. Các phần chính của hệ thống như vậy bao gồm trích xuất đặc trưng, học tập hành động, phân loại, nhận dạng hành động và phân loại. Một quy trình đơn giản gồm ba bước, cụ thể là phát hiện của con người hoặc các bộ phận cơ thể, theo dõi, và sau đó nhận bằng cách sử dụng kết quả theo dõi. Ví dụ, để nhận ra hành động “bắt tay”, cánh tay và bàn tay của hai người được phát hiện trước tiên và theo dõi để tạo ra một mô tả không gian – thời gian của chuyển động của họ. Mô tả này được so sánh với các mẫu hiện có trong dữ liệu huấn luyện để xác định hành loại hành động. Mô hình này dựa rất nhiều vào tính chính xác của việc theo dõi, điều này không đáng tin cậy trong những cảnh lộn xộn.

Nhiều phương pháp đã được đề xuất, và có thể được phân loại theo nhiều tiêu chí khác nhau như trong bài luận điều tra hiện có. Poppe thảo luận nhận dạng hành động của con người từ biễu diễn hình ảnh và phân loại hành động riêng rẽ. Weinland khảo sát các phương pháp cho biểu diễn hành động, phân đoạn và nhận dạng. Turaga chia vấn đề nhận dạng hành động và hoạt động theo độ phức tạp, và các hướng tiếp cận phân loại theo khả năng của mình để xử lý các mức độ phức tạp khác nhau. Có nhiều tiêu chí phân loại khác nhau. Trong số đó, Aggarwal và Ryoo là một trong những tổng kết toàn diện mới nhất và sự so sánh của sự tiến bộ quan trọng nhất trong lĩnh vực này. Dựa vào hành động nhận dạng từ hình ảnh đầu vào trực tiếp, Aggarwal và Ryoo phân chia phương pháp nhận dạng thành hai loại chính: phương pháp tiếp cận đơn lớp và phương pháp tiếp cận phân cấp. Cả hai đều là thêm loại còn phụ thuộc vào các phương pháp biễu diễn đặc trưng và học.

**2.2. GIỚI THIỆU VỀ MẠNG TÍCH CHẬP (CNN)**

**2.2.1. Giới thiệu Convolutional Neural Networks**

**2.2.1.1. Tổng quan**

Convolutional Neural Network (CNNs – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến. Nó giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay.

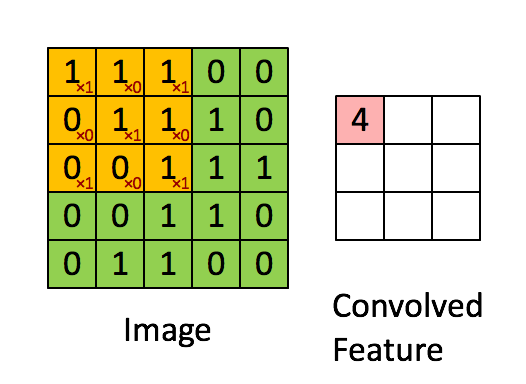
CNN được sử dụng nhiều trong các bài toán nhận dạng các object trong ảnh. Để tìm hiểu tại sao thuật toán này được sử dụng rộng rãi cho việc nhận dạng (detection), chúng ta hãy cùng tìm hiểu về thuật toán này.

Mạng CNN được có kiến trúc được cấu tạo bởi một số loại layer bao gồm:

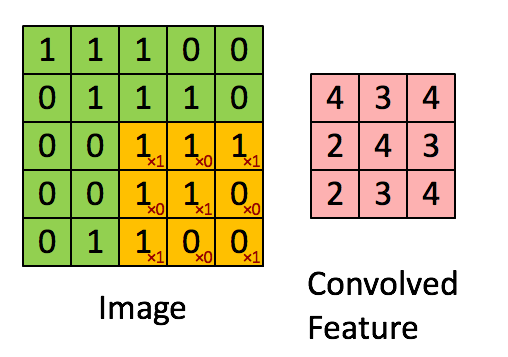
* Convolution layer
* Pooling layer
* Fully connected layer

**2.2.1.2. Convolutional layer**

Là một cửa sổ trượt (Sliding Windows) trên một ma trận như mô tả hình dưới:



Hình 1: Kernel quét qua phần tử đầu tiên của input

****

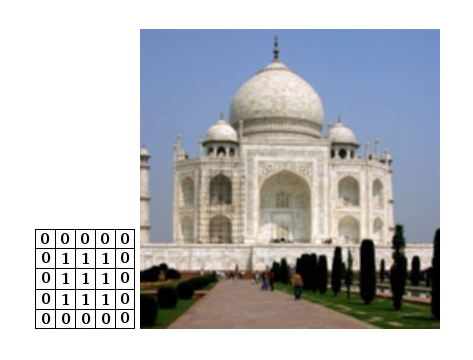
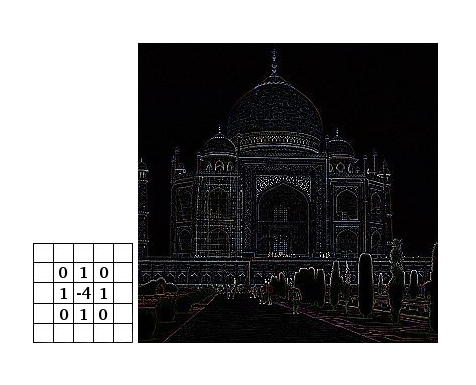
Hình 2: Kernel quét qua phần tử cuối cùng của input

Các convolutional layer có các parameter(kernel) đã được học để tự điều chỉnh lấy ra những thông tin chính xác nhất mà không cần chọn các feature.

Trong hình ảnh ví dụ dưới, ma trận bên trái là một hình ảnh trắng đen được số hóa. Ma trận có kích thước 5×5 và mỗi điểm ảnh có giá trị 1 hoặc 0 là giao điểm của dòng và cột.

Convolution hay tích chập là nhân từng phần tử trong ma trận 3. Sliding Window hay còn gọi là kernel, filter hoặc feature detect là một ma trận có kích thước nhỏ như trong ví dụ trên là 3×3.

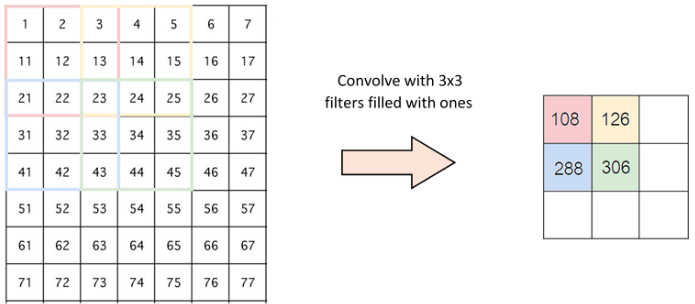
Convolution hay tích chập là nhân từng phần tử bên trong ma trận 3×3 với ma trận bên trái. Kết quả được một ma trận gọi là Convoled feature được sinh ra từ việc nhận ma trận Filter với ma trận ảnh 5×5 bên trái.



Hình 3: Ví dụ thực tế về input, kernel và feature map

**2.2.1.3. Bước nhảy trong CNN (Stride)**

Stride là số pixel thay đổi trên ma trận đầu vào. Khi stride là 1 thì ta di chuyển các kernel 1 pixel. Khi stride là 2 thì ta di chuyển các kernel đi 2 pixel và tiếp tục như vậy. Hình dưới là lớp tích chập hoạt động với stride là 2.



Hình 4: Minh họa Stride

**2.2.1.4. Pooling layer**

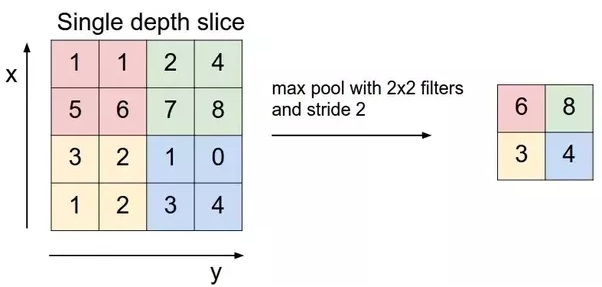
Lớp pooling sẽ giảm bớt số lượng tham số khi hình ảnh quá lớn. Không gian pooling còn được gọi là lấy mẫu con hoặc lấy mẫu xuống làm giảm kích thước của mỗi map nhưng vẫn giữ lại thông tin quan trọng. Các pooling có thể có nhiều loại khác nhau:

+ Max Pooling

+ Average Pooling

+ Sum Pooling

Max pooling lấy phần tử lớn nhất từ ma trận đối tượng, hoặc lấy tổng trung bình. Tổng tất cả các phần tử trong map gọi là sum pooling



Hình 5: Pooling kích thước 2x2

**2.2.1.5. Fully connected layer**

Sau khi ảnh được truyền qua nhiều convolutional layer và pooling layer thì model đã học được tương đối các đặc điểm của ảnh thì tensor của output của layer cuối cùng sẽ được là phẳng thành vector và đưa vào một lớp được kết nối như một mạng nơ-ron. Với FC layer được kết hợp với các tính năng lại với nhau để tạo ra một mô hình. Cuối cùng sử dụng softmax hoặc sigmoid để phân loại đầu ra.

Tôi hứa sẽ hoàn thành nội dung báo cáo khóa luận tốt nghiệp đúng theo đề cương đã viết và đúng thời gian theo kế hoạch thực hiện khóa luận mà Khoa Công nghệ thông tin đã phổ biến.

*An Giang, ngày … tháng ….. năm 202…*

**Xác nhận của Giảng viên hướng dẫn Sinh viên thực hiện**

*(Ký và ghi rõ họ tên) (Ký và ghi rõ họ tên)*