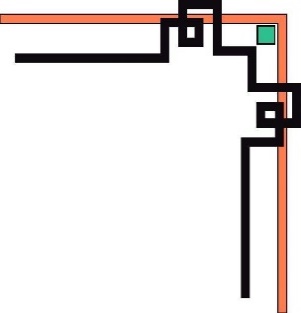
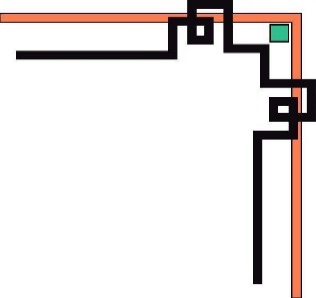
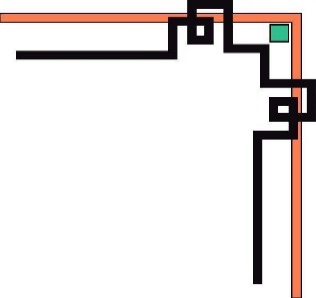
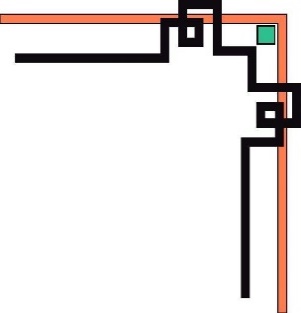
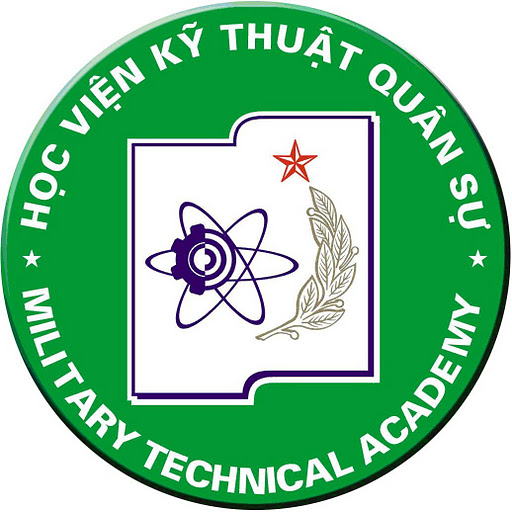
**HỌC VIỆN KỸ THUẬT QUÂN SỰ**



**Khoa Công Nghệ Thông Tin**



**Phương pháp nghiên cứu IT**

***Đề tài: Xây dựng chương trình nhận dạng hành động trong lớp học***

**Sinh viên thực hiện:** Bùi Vinh Hiếu

**Lớp:** CNGame&MP15

**Giáo viên hướng dẫn:** Th.S Phan Thị Hải Hồng

***Hà Nội, 2020***

# **MỤC LỤC**

[**MỤC LỤC** 2](#_Toc55465763)

[**Chương I. Đặt vấn đề** 3](#_Toc55465764)

[1. Đặt vấn đề 3](#_Toc55465765)

[2. Lý do chọn đề tài 3](#_Toc55465766)

[3. Mục tiêu của đề tài 4](#_Toc55465767)

[**Chương II. Tổng quan tài liệu và cơ sở lý thuyết** 4](#_Toc55465768)

[1. Tổng quan về Deep Learning và bài toán nhận diện hành động trong thị giác máy tính 4](#_Toc55465769)

[1.1. Deep Learning là gì? 4](#_Toc55465770)

[1.2 Một số ứng dụng của Deep Learning 4](#_Toc55465771)

[2. Giới thiệu Convolutional Neural Network 4](#_Toc55465772)

[2.1 Tổng quan 4](#_Toc55465773)

[2.2 Kiến trúc CNN 5](#_Toc55465774)

[3. Thư viện Tensorflow 5](#_Toc55465775)

[4. Thư viện Keras 7](#_Toc55465776)

[5. Phương pháp OpenPose 7](#_Toc55465777)

[5.1 Tổng quan về OpenPose 7](#_Toc55465778)

[5.2 Hoạt động 8](#_Toc55465779)

[**Chương III. Phương pháp nghiên cứu** 9](#_Toc55465780)

[1. Phương pháp nghiên cứu lý thuyết 9](#_Toc55465781)

[2. Phương pháp nghiên cứu thực nghiệm 10](#_Toc55465782)

[**Chương IV. Kết luận và kiến nghị** 10](#_Toc55465783)

[**Tài liệu tham khảo** 10](#_Toc55465784)

# **Chương I. Đặt vấn đề**

## 1. Đặt vấn đề

Ngày nay, dữ liệu video dễ dàng được tạo ra bởi các thiết bị như: điện thoại di động, máy tính xách tay, máy ảnh kỹ thuật số và các thiết bị hệ thống camera quan sát (CCTV)… Bên cạnh đó các trang web chia sẻ video cũng không ngừng tăng trưởng về số lượng lẫn chất lượng.

Bài toán nhận diện hành động của con người đóng góp một phần tự đông hóa khai thác tài nguyên dữ liệu nhiều thông tin này. Các ứng dụng liên quan đến bài toán nhận diện hành động như:

* An ninh và các hệ thống giám sát truyền thống gồm mạng lưới các camera và được giám sát bởi con người.
* Tương tác giữa người và máy vẫn còn nhiều thách thức, những dấu hiệu thị giác là phương thức quan trọng nhất trong giao tiếp phi ngôn ngữ. Khai thác hiệu quả phương thức giao tiếp này: Thông qua cử chỉ hành động, hoạt động hứa hẹn tạo ra thế hẹ máy tính tương tác chính xác và tự nhiên hơn với con người.
* Bên cạnh đó nhận diện hành động của con người trong video còn được ứng dụng tóm tắt, truy vấn video, phân tích thể thao.

Deep Learning là một chủ đề Trí tuệ nhân tạo, tập trung giải quyết các vấn đề liên quan đến mạng nơ-ron nhân tạo nhằm nâng cấp các công nghệ như nhận diện giọng nói, thị giác máy tính và xử lý ngôn ngữ tự nhiên.

## 2. Lý do chọn đề tài

Ngày nay, tại các trường Đại học và Phổ thông trên cả nước, việc quản lý học sinh – sinh viên gặp nhiều khó khăn, bất lợi do số lượng người trong một lớp đông, ít người quản lý, cơ sở vật chất chưa được cải thiện. Việc phát triển một công cụ giám sát học sinh trong lớp học là rất cần thiết và đem lại nhiều lợi ích cho Nhà trường và chính học sinh – sinh viên.

Vì lý do đó, em đã chọn đề tài:

**“Xây dựng chương trình nhận dạng hành động trong lớp học”**

Đồng thời đề xuất cải tiến các tham số đầu vào của thuật toán để tăng hiệu quả nhận diện so với cá thuật toán hiện tại.

## 3. Mục tiêu của đề tài

Thực hiện nghiên cứu, xây dựng chương trình nhận dạng hành động gồm 3 mục tiêu chính:

* Tìm hiểu tổng quan về Deep Learning, Machine Learning và các công cụ để thực hiện bài toán.
* Xây dựng chương trình nhận dạng hành động con người sử dụng Deep Learning và phương pháp OpenPose.
* Ứng dụng đề tài phục vụ cho việc quản lý và giám sát lớp học.

# **Chương II. Tổng quan tài liệu và cơ sở lý thuyết**

## Tổng quan về Deep Learning và bài toán nhận diện hành động trong thị giác máy tính

### 1.1. Deep Learning là gì?

Deep Learning là một chủ đề Trí truệ nhân tạo (AI) và là một phạm trù nhỏ của máy học. Deep Learning tập trung giải quyết các vấn đề liên quan đến mạng nơron nhân tạo nhằm nâng cấp các công nghệ như nhận diện giọng nói, thị giác máy tính và xử lý ngôn ngữ tự nhiên.

Deep Learning đang trở thành một trong những lĩnh vực quan trọng trong khoa học máy tính. Chỉ trong vài năm, Deep Learning đã thúc đẩy tiến bộ trong đa dạng các lĩnh vực như nhận thức sự vật (object perception), dịch tự động (machine translation), nhận diện giọng nói - đó là những vấn đề từng rất khó khăn với các nhà nghiên cứu trí tuệ nhân tạo.

### 1.2 Một số ứng dụng của Deep Learning

* Hệ thống xử lý trên các nền tảng.
* Nhận diện hình ảnh.
* Xử lý ngôn ngữ tự nhiên.

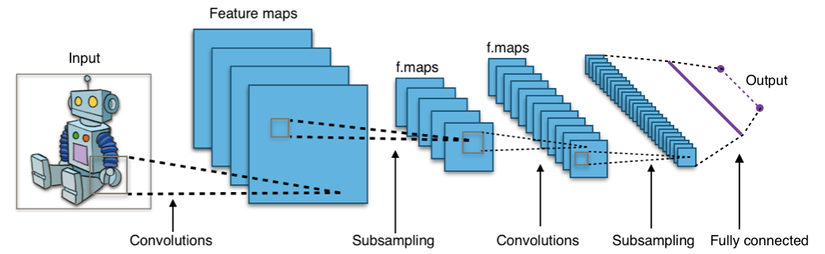
## Giới thiệu Convolutional Neural Network

### 2.1 Tổng quan

Convolutional Neural Network (CNN) bao gồm một hoặc nhiều lớp chập (thường với một bước lấy mẫu con) và sau đó theo sau bởi một hoặc nhiều hơn các lớp kết nối như trong một mạng nơron đa lớp chuẩn. Kiến trúc của một CNN được thiết kế để tận dụng lợi thế của cấu trúc 2 chiều của một hình ảnh đầu vào (hoặc đầu vào 2 chiều khác như một tín hiệu tiếng nói). Điều này đạt được với các kết nối cục bộ và trọng số ràng buộc theo một số hình thức tổng hợp mà kết quả là các đặc trưng không thay đổi. Một lợi ích khác của CNN là dễ dàng huấn luyện hơn và có ít thông số so với các mạng kết nối đầy đủ với cùng một số đơn vị ẩn.

### 2.2 Kiến trúc CNN

CNN gồm một số lớp chập và lớp lấy mẫu con tùy chọn, tiếp theo là các lớp | kết nối. Đầu vào cho một lớp chập là một hình ảnh m 3 m 3, với m là chiều cao và chiều rộng của hình ảnh và r là số kênh (ví dụ một ảnh RGB cố y = 3). Lớp chập sẽ có k bộ lọc có kích thước n < n < 4, trong đó n nhỏ hơn kích thước hình ảnh và q có thể bằng hoặc nhỏ hơn số kênh r. Kích thước của các bộ lọc tăng đến cấu trúc kết nổi cục bộ, trong đó mỗi bộ lọc xoắn với hình ảnh để tạo ra k bản đồ đặc trưng có kích thước m – n + 1, Mỗi bản đồ được lấy mẫu con đặc trưng với việc hợp nhất trung bình hoặc lớn nhất trên 2x2 khu vực lân cận, trong đó phạm vi của p là giữa 2 (cho các hình ảnh nhỏ) và 5 (chỉ các đầu vào lớn hơn). Trước hoặc sau khi lấy mẫu con, một bias bổ sung và xích ma phi tuyến được áp dụng cho mỗi bản đồ đặc trung.



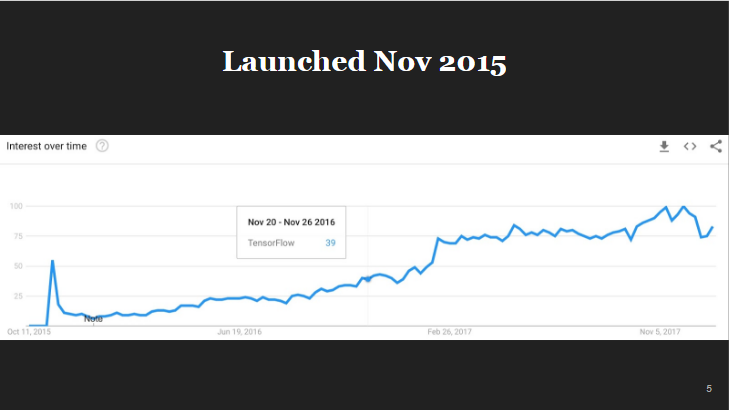
Hình 2.1 *Kiến trúc CNN trong nhận diện hành động con người*

Chập hình ảnh đầu vào với các bộ lọc huấn luyện khác nhau và các bias bổ sung, nhiều bản đồ đặc trưng được tạo trong lớp C1. Mỗi bản đồ đặc trưng trong S2 thu được bởi thao tác tổng hợp các bản đồ đặc trưng tương ứng trong lớp C1. Chập và tổng hợp cực đại trong lớp C3 và S4 thì giống trong lớp C1 và S2. Trong bước nhận diện cuối cùng, các đặc trưng thu được sau khi tổng hợp cực đại trong lớp S4 thì được mã hóa thành một vector 1 chiều.

## Thư viện Tensorflow

TensorFlow là một thư viện phần mềm mở cho tính toán số, sử dụng biểu đồ luồng dữ liệu. Các nút trong đồ thị biểu diễn cho hoạt động toán học, trong khi các cạnh đồ thị biểu diễn cho các mảng dữ liệu đa chiều (tensors) trao đổi giữa chúng. Kiến trúc linh hoạt cho phép chúng ta triển khai tính toán trên một hoặc nhiều CPU hoặc GPU trong một máy tính để bàn, máy chủ, hoặc thiết bị di động với một API đơn. TensorFlow ban đầu được phát triển bởi các nhà nghiên cứu và kỹ sư làm việc trong nhóm Google Brain cho các nghiên cứu máy học và deep neural network.

TensorFlow có các API với một số ngôn ngữ lập trình cho cả xây dựng và thực thi một đồ thị TensorFlow. Python API là hiện tại hoàn thiện nhất và dễ sử dụng nhất, nhưng API C++ có một vài ưu điểm về hiệu năng trong việc thực thi đồ thị, và hỗ trợ triển khai các thiết bị nhỏ như Android.



Hình 2.2 *Mức độ phổ biến của Tensorflow tính từ lúc được Opensource.*

Một số project nổi tiếng sử dụng thư viện Tensorflow:

* Phân loại ung thư da – Dermatologist-level classification of skin cancer with deep neural networks (Esteva et al., Nature 2017)
* WaveNet: Text to speech – Wavenet: A generative model for raw audio (Oord et al., 2016)
* Vẽ hình – Draw Together with a Neural Network (Ha et al., 2017)

Image Style Transfer Using Convolutional Neural Networks (Gatys et al., 2016) Tensorflow adaptation by Cameroon Smith (cysmith@github)

## Thư viện Keras

**Keras** là một thư viện nơ-ron mã nguồn mở được viết bằng ngôn ngữ lập trình Python. Nó có khả năng chạy trên đầu trang của Tensorflow, Microsoft Cognitive Toolkit, R, Theano hoặc StripeML. Được thiết kế để cho phép thử nghiệm nhanh với các mạng thần kinh sâu, nó tập trung vào việc thân thiện với người dùng, mô-đun và mở rộng.

**Keras** được coi là một thư viện ‘high-level’ với phần ‘low-level’ (còn được gọi là backend) có thể là TensorFlow, CNTK, hoặc Theano (sắp tới Theano sẽ không được duy trì nâng cấp nữa). Keras có cú pháp đơn giản hơn TensorFlow rất nhiều. Với mục đích giới thiệu về các mô hình nhiều hơn là các sử dụng các thư viện deep learning, tôi sẽ chọn Keras với TensorFlow là *‘backend’*.

Một số tính năng của Keras:

* Keras ưu tiên trải nghiệm của người lập trình
* Keras đã được sử dụng rộng rãi trong doanh nghiệp và cộng đồng nghiên cứu
* Keras giúp dễ dàng biến các thiết kế thành sản phẩm
* Keras hỗ trợ huấn luyện trên nhiều GPU phân tán

Keras hỗ trợ đa backend engines và không giới hạn bạn vào một hệ sinh thái.

## Phương pháp OpenPose

### 5.1 Tổng quan về OpenPose

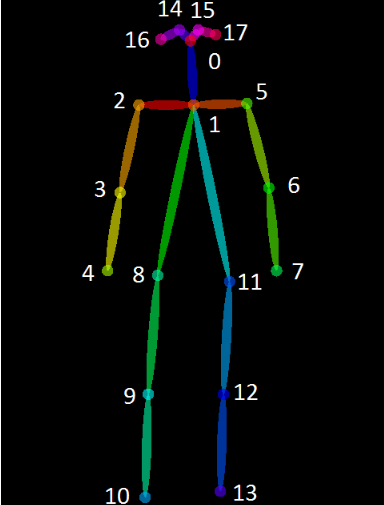
OpenPose là hệ thống đầu tiên sử dụng để phát hiện cơ thể con người, bàn tay và các điểm nhấn trên khuôn mặt (trong tổng số 130 điểm chính) trên các hình ảnh đơn lẻ.



Hình 2.3 *Openpose theo dõi chuyển động của cơ thể người*

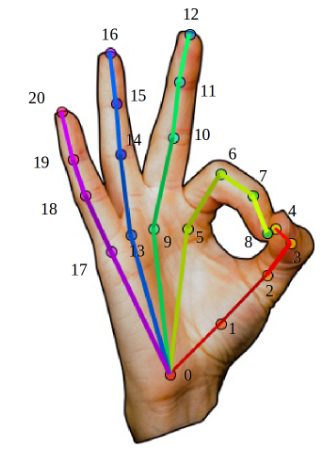
### 5.2 Hoạt động

* Openpose hoạt động dựa trên việc phát hiện các điểm được lưu lại trên cơ thể, trên khuôn mặt và bàn tay. Các điểm này sẽ được kết nối lại thành bộ khung chuyển động theo chuyển động của cơ thể.
* Các điểm được lưu lại trong Openpose.
* Cơ thể người: 18 điểm trên cơ thể người được Openpose đánh dấu và nhận diện. Các điểm này đại diện cho các khớp xương trên cơ thể và các điểm này có thể định hướng khuôn mặt của người được nhận diện.



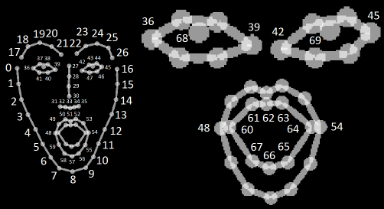
Hình 2.4 *18 điểm trên cơ thể được OpenPose lưu lại*

* Bàn tay: 21 điểm trên bàn tay được Openpose lưu lại và sử dụng để nhận diện chuyển động của bàn tay. Các điểm này đại diện cho các khớp xương của bàn tay.



Hình 2.5 *Điểm được OpenPose sử dụng để nhận dạng bàn tay*

* Khuôn mặt: 68 điểm được openpose lưu trữ và sử dụng để nhận chuyển động khuôn mặt.



Hình 2.6 *68 điểm được OpenPose sử dụng để nhận dạng chuyển động khuôn mặt*

# **Chương III. Phương pháp nghiên cứu**

## Phương pháp nghiên cứu lý thuyết

* Tìm hiểu tổng quan về Deep Learning và bài toán nhận diện hành động trong thị giác máy tính.
* Tìm hiểu tổng quan về Convolutional Neural Network (CNN).
* Tìm hiểu các thư viện Tensorflow, Keras, OpenCV và một số thư viện khác.
* Tìm hiểu phương pháp OpenPose và cách thực hiện.

## Phương pháp nghiên cứu thực nghiệm

* Tiến hành phân tích và cài đặt trên Python
* Tiếng hành training data cho máy học

Kết quả là chương trình nhận dạng hành động qua video hoặc camera sử dụng thời gian thực trên Hệ điều hành Windows.

# **Chương IV. Kết luận và kiến nghị**

# **Tài liệu tham khảo**

**[1]** Real-Time Action Detection – Paul Viola, Michal J. Jone.

**[2]** OpenPose – [Gines Hidalgo](https://www.gineshidalgo.com/)**,**[Zhe Cao](https://people.eecs.berkeley.edu/~zhecao)**,**[Tomas Simon](http://www.cs.cmu.edu/~tsimon)**,**[Shih-En Wei](https://scholar.google.com/citations?user=sFQD3k4AAAAJ&hl=en)**,**[Hanbyul Joo](https://jhugestar.github.io/)**, and**[Yaser Sheikh](http://www.cs.cmu.edu/~yaser).

**[3]** NHẬN DIỆN HÀNH ĐỘNG CỦA CON NGƯỜI TRONG VIDEO SỬ DỤNG DEEP LEARNING – Trần Thanh Tuấn.