BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**🙞 🕮 🙜**



**BÁO CÁO**

**MÔN MÁY HỌC NÂNG CAO**

**Đề tài**

**ỨNG DỤNG HỌC SÂU VÀO PHÂN LOẠI**

**HÌNH ẢNH**

**Sinh viên thực hiện:**

**Mã Chí Tài B17009563**

**Nguyễn Việt Hào B1812338**

**Phan Trịnh Minh Vinh B1812321**

*Cần Thơ, 05/2022*

BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**🙞 🕮 🙜**



**BÁO CÁO**

**MÔN MÁY HỌC NÂNG CAO**

**Đề tài**

**ỨNG DỤNG HỌC SÂU VÀO PHÂN LOẠI**

**HÌNH ẢNH**

**Giáo viên hướng dẫn:**

**PGS.TS. Phạm Nguyên Khang**

*Cần Thơ, 05/2022*

**NHẬN XÉT CỦA GIẢNG VIÊN**

**--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**

**LỜI CẢM ƠN**

Để có được bài niên luận này, chúng em xin được bày tỏ lòng biết ơn chân thành và sâu sắc đến Thầy Phạm Nguyên Khang người đã trực tiếp tận tình hướng dẫn,giúp đỡ chúng em. Trong suốt quá trình thực hiện, nhờ những sự chỉ bảo kiến thức và hướng dẫn quý giá đó mà chúng em có thể thực hiện bài báo cáo để hoàn thành một cách tốt nhất.

Chúng em cũng xin gửi lời cám ơn chân thành đến các Thầy Cô Giảng viên Đại học Cần Thơ, đặc biệt là các Thầy Cô ở Khoa CNTT & TT, những người đã truyền đạt những kiến thức quý báu trong thời gian qua.

Tuy có nhiều cố gắng trong quá trình thực hiện bài báo cáo, nhưng không thể tránh khỏi những sai sót. Chúng em rất mong nhận được sự đóng góp ý kiến quý báu của quý Thầy Cô và các bạn để bài niên luận hoàn thiện hơn.

**MỤC LỤC**

[PHẦN GIỚI THIỆU 4](#_Toc102863391)

[1. Đặt vấn đề 4](#_Toc102863392)

[2. Lịch sử giải quyết vấn đề 4](#_Toc102863393)

[3. Mục tiêu đề tài 4](#_Toc102863394)

[4.Đối tượng và phạm vi nghiên cứu 5](#_Toc102863395)

[5. Phương pháp nghiên cứu 5](#_Toc102863396)

[6. Kết quả đạt được 5](#_Toc102863397)

[7. Bố cục 5](#_Toc102863398)

[PHẦN NỘI DUNG 6](#_Toc102863399)

[CHƯƠNG 1 6](#_Toc102863400)

[MÔ TẢ BÀI TOÁN 6](#_Toc102863401)

[1. Mô tả chi tiết bài toán 6](#_Toc102863402)

[2. Vấn đề và giải pháp liên quan đến bài toán 6](#_Toc102863403)

[2.1 Giải pháp 6](#_Toc102863404)

[2.1.1 Tìm hiểu tensorflow là gì? 6](#_Toc102863405)

[2.1.2 Keras là gì? 7](#_Toc102863406)

[2.1.3 Phân loại hình ảnh là gì? 7](#_Toc102863407)

[2.1.4 Cách thu thập dữ liệu. 7](#_Toc102863408)

[2.1.5 Xây dựng mô hình. 7](#_Toc102863409)

[2.1.6 Huấn luyện mô hình. 7](#_Toc102863410)

[2.1.7 Dự đoán trên web app 7](#_Toc102863411)

[CHƯƠNG 2 8](#_Toc102863412)

[THIẾT KẾ VÀ CÀI ĐẶT 8](#_Toc102863413)

[1. Thiết kế hệ thống 8](#_Toc102863414)

[1.1. Xác định đối tượng và thu thập hình ảnh 8](#_Toc102863415)

[1.2. Đọc dữ liệu: 8](#_Toc102863416)

[1.3. Nhãn của dữ liệu: 9](#_Toc102863417)

[1.4. Xây dựng mô hình 10](#_Toc102863418)

[1.5. Biên dịch mô hình 11](#_Toc102863419)

[1.6. Huấn luyện mô hình 12](#_Toc102863420)

[1.7. Xây dựng một ứng dụng web 12](#_Toc102863421)

[PHẦN KẾT LUẬN 14](#_Toc102863422)

[1. Kết quả đạt được 14](#_Toc102863423)

[2. Hướng phát triển 14](#_Toc102863424)

[TÀI LIỆU THAM KHẢO 15](#_Toc102863425)

**ABSTRACT**

# PHẦN GIỚI THIỆU

## Đặt vấn đề

Ngày nay, công nghê thông tin ngày càng phát triển trong nhiều lĩnh vực như nông nghiệp, giáo dục, y tế,….Giúp cho đời sống con người ngày càng phát triển hơn.

Trong doanh nghiệp quá trình phân loại, đánh giá sản phẩm như rau củ quả, các loại hoa, bàn, ghế, giường,..còn phải thực hiện bằng phương pháp thủ công. Đây là một công việc không quá khó, nhưng tốn nhiều thời gian, công sức của con người và là rào cản đối với sự mở rộng của công ty.

Trong thời gian gần đây, nhờ có sự phát triển mạnh mẽ về khả năng tính toán của hệ máy tính hiện đại cũng như sự bùng nổ về dữ liệu thông qua mạng lưới Internet, ta chứng kiến nhiều sự đội phá trong lĩnh vực Học máy, đặt biệt là trong lĩnh vực Deep Learning. Đề tài “Ứng dụng học sâu vào phân loại hình ảnh” giúp cho việc phân loại sản phẩm một cách nhanh nhất.

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 phân loại hình ảnh, thị giác máy tính và xử lý ngôn ngữ tự nhiên…

## 2. Lịch sử giải quyết vấn đề

TensorFlow thế hệ thứ hai của hệ thống học máy của Google Brain, với một bản cài đặt tham khảo đã phát hành dưới dạng phần mềm mã nguồn mở vào ngày 9/11/2015.

Trong khi bản cài đặt tham khảo chạy trên một thiết bị đơn, TensorFlow có thể chạy trên nhiều CPU và GPU (với nhiều mở rộng CUDA tùy chọn cho việc tính toán đa năng trên các GPU). Các tính toán của TensorFlow được thể hiện dưới dạng các biểu đồ dataflow chi tiết.

Ngày nay công nghệ phân loại hình ảnh đã không còn xa lạ gì với mọi người, có thể áp dụng rất nhiều thuật toán khác nhau để xây dựng mô hình phân loại, không chỉ trong một mà rất nhiều lĩnh vực khác nhau như hệ thống bảo mật và xe không người lái, phân loại các vật thể, dụng cụ học tập, phân loại các loài động vật,...

Từ những kiến thức đã có tiến hành phân tích và áp dụng sự hiểu biết vào việc xây dựng mô phân loại dữ liệu hình ảnh có 10 lớp, cụ thể là nhận diện hình ảnh của 10 người.

## 3. Mục tiêu đề tài

Đề tài nghiên cứu ứng dụng Deep learning vào bài toán phân loại hình ảnh đồng thời đề xuất cải tiến các tham số đầu vào của thuật toán để tăng hiệu quả phân loại so với các thuật toán hiện tại.

## 4.Đối tượng và phạm vi nghiên cứu

* Đối tượng nghiên cứu:
* Ngôn ngữ Python.
* Tìm hiểu về Deep learning.
* Nghiên cứu về máy học và thư viện Tensorflow, Keras.
* Transfer learning và fine tuning
* Flask framwork
* Phạm vi nghiên cứu:
* Xây dựng mô hình mạng noron phân loại hình ảnh trong deep learning.
* Sử dụng được mô hình pretrained.

## 5. Phương pháp nghiên cứu

* Tìm kiếm hình ảnh.
* Xử lý hình ảnh
* Thư viện Tensorflow.
* Huấn luyện mô hình.
* Kiểm tra mô hình và đánh giá độ chính xác.

## 6. Kết quả đạt được

Phân loại được các hình ảnh, xây dựng mô hình, biên dịch mô hình, huấn luyện mô hình, dự đoán được nhãn của hình ảnh.

## 7. Bố cục

**Phần giới thiệu**

Giới thiệu tổng quát về đề tài.

**Phần nội dung**

**Chương 1**: Mô tả bài toán .

**Chương 2**: Thiết kế, cài đặt giải thuật.

**Phần kết luận**

Trình bày kết quả đạt được và hướng phát triển.

# PHẦN NỘI DUNG

# CHƯƠNG 1

# MÔ TẢ BÀI TOÁN

## 1. Mô tả chi tiết bài toán

Phân loại ảnh là một bài toán quan trọng bậc nhất trong lĩnh vực Computer Vision. Chúng ta đã có rất nhiều nghiên cứu để giải quyết bài toán này bằng cách rút trích các đặc trưng rất phổ biến như SIFT, HOG rồi cho máy tính học nhưng những cách này tỏ ra không thực sự hiểu quả. Nhưng ngược lại, đối với con người, chúng ta lại có bản năng tuyệt vời để phân loại được những đối tượng trong khung cảnh xung quanh một cách dễ làm.

Dữ liệu đầu vào của bài toán là một bức ảnh của một đối tượng nào đó. Một ảnh được biểu ảnh bằng ma trận các giá trị. Mô hình phân lớp sẽ phải dự đoán được lớp của ảnh đó thược đối tượng nòa từ ma trận điểm ảnh này.

Trong bài toán này chúng ta cần thu thập dữ liệu là các hình ảnh của 40 đối tượng khác nhau gồm: 2 tập dữ liệu (train\_dataset và test\_dataset).

* Hai tập dữ liệu là:
* Tập dữ liệu Train gồm: 360 hình ảnh. Của 9 đối tượng, mỗi đối tượng có 9 ảnh ( nhiều góc nhìn, biểu cảm của khuôn mặt … )
* Tập dữ liệu Test gồm: 40 hình ảnh, tương đương 40 đối tượng (mỗi hình ảnh là một đối tượng)
* 40 đối tượng này tương ứng với 40 lớp bao gồm: s1,s2,s3,s4,s5,…s40 (

## 2. Vấn đề và giải pháp liên quan đến bài toán

### 2.1 Giải pháp

### Tìm hiểu tensorflow là gì?

* Tensorflow là một thư viện phần mềm mã nguồn mở hỗ trợ mạnh mẽ các phép toán để tính toán trong Machine learning và Deep learning.
* Được phát triển bởi Google.
* Tensorflow không tích hợp sẵn và phải cài đặt bằng lệnh pip install tensorflow.
* Sau khi cài đặt xong ta import bằng lệnh import tensorflow.

### Keras là gì?

* API mức cao dùng để xây dựng và huấn luyện các mô hình học sâu.
* Xây dựng nhanh.
* Thân thiện, dễ sử dụng.
* Mô-đun, dễ kết hợp.
* Dễ mở rộng.

### Phân loại hình ảnh là gì?

Phân loại hình ảnh là nhiệm vụ gán nhãn cho hình ảnh từ một định nghĩa. Thực tế, điều này có nghĩa là nhiệm vụ của chúng ta là phân tích một hình ảnh đầu vào và trả lại một nhãn phân loại hình ảnh. Nhãn luôn từ một tập hợp các danh mục có thể được xác định trước.

### Cách thu thập dữ liệu.

Dữ liệu hình ảnh từ tập dữ liệu AT&T đã được thu thập chính xác khung hình vừa với khuôn mặt (phát hiện khuôn mặt)

### Xây dựng mô hình.

Sử dụng VGG16 để chuyển giao học tập (transfer learning) và tinh

chỉnh mô hình (fine tuning) là một phương thức để tạo mô hình lớn trên (big model) trên một tập dữ liệu nhỏ.

Transfer learning: Chuyển tiếp những kiến thức thu thập từ việc học

trên tập dữ liệu vào 1 tập dữ liệu khác.

Fine tuning: là 1 mothod sử dụng mạng đã được đào tạo và huấn luyện lại trên tập dữ liệu mới.

Thường thì (deep learning) sẽ cần một lượng lớn dữ liệu để train mà không phải lúc nào cũng có dữ liệu để xây dựng mô hình này.

Không để 1 epoch chạy lâu thì trong 1 số trường hợp cần tinh chỉnh (fine tuning) mô hình để giải quyết việc này.

### Huấn luyện mô hình.

Sử dụng model.fit huấn luyện mô hình.

### Dự đoán trên web app

Xây dựng một ứng dụng web bằng flask frame work để dự đoán hình ảnh được truyền vào.

Dự đoán ảnh mới, đầu tiên ta update hình ảnh lên, sau đó xử lý các hình ảnh đầu vào đó rồi đưa vào mô hình mạng đã được xây dụng để dự đoán nhãn của hình ảnh đó.

# CHƯƠNG 2

# THIẾT KẾ VÀ CÀI ĐẶT

## Thiết kế hệ thống

### Xác định đối tượng và thu thập hình ảnh

Bộ cơ sở dữ liệu ảnh là một trong các thành phần quan trọng hàng đầu trong các phương pháp Học máy nói chung, được sử dụng để phục vụ cho quá trình tính toán tham số và huấn luyện, tinh chỉnh các mô hình. Thông thường, bộ dữ liệu càng lớn và càng được chọn lọc tỉ mỉ cần thận thì độ chính xác của mô hình càng được cải thiện, nhưng trong phạm vi báo cáo này sẽ được hạn chế về số lượng ảnh. Cụ thể:

* Số lượng hình ảnh phân loại gồm 40 loại (tương đương với 4 đối tượng)
* S1
* S2
* S3
* S4
* S5
* …
* S40
* Số lượng ảnh cho mỗi loại:
* Tập train gồm 360 hình ảnh, mỗi lớp tương đương với 1 đối tượng (hình mặt người) và mỗi đối tượng sẽ có 9 hình ảnh trong tập train
* Tập test gồm: 40 hình ảnh mặt của 40 đối tượng khác nhau (40 lớp)

### Đọc dữ liệu:

Đưa giá trị mức xám về khoảng [0, 1].

Mọi hình ảnh kỹ thuật số được tạo thành bởi pixel có giá trị trong phạm vi 0 ~ 255. 0 là màu đen và 255 là màu trắng. 255 là giá trị pixel tối đa. 1/255 là chuyển đổi mọi giá trị pixel từ phạm vi [0,255] -> [0,1].

Text

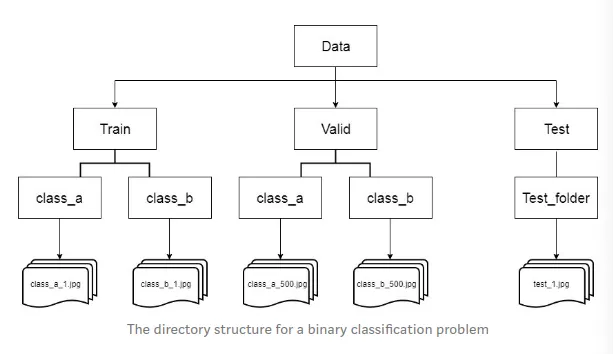
Description automatically generated

Hình 1: *Xử lý dữ liệu.*

Sau đó sử dụng ImageGenerator cũng là một dạng data generator được xây dựng trên framework keras và dành riêng cho dữ liệu ảnh.

Tiếp theo chúng ta sẽ truyền dữ liệu vào mô hình thông qua một hàm là flow\_from\_directory().

Hàm flow\_from\_directory() sẽ có tác dụng đọc các ảnh từ đường dẫn các thư mục hình ảnh và lấy ra những thông tin bao gồm ma trận ảnh sau biến đổi và nhãn tương ứng. Hàm này sẽ đọc dữ liệu theo cây thư mục :



Text

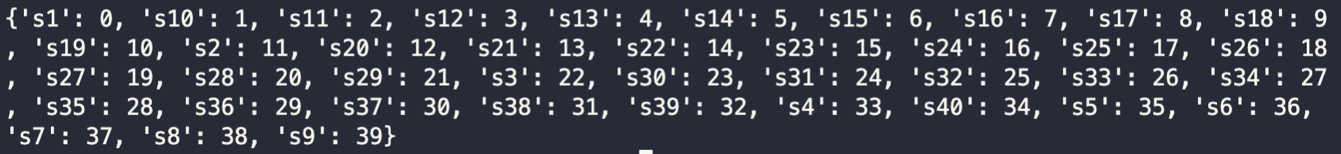
Description automatically generated

Dữ liệu ban đầu có kích thước 92x112, chúng ta sẽ chuyển đó về 90x90 bằng tham số target\_size để có thể đưa vào mô hình VGG16.

Dữ liệu có số lớp lớn hơn 2 nên sẽ sử dụng categorical cho tham số class\_mode

### Nhãn của dữ liệu:

Là các lớp từ s1 đến s40, tương ứng với mỗi lớp sẽ có một chỉ số riêng. Ví dụ s1 sẽ có chỉ số là 0 và s10 có chỉ số là 1…



### Xây dựng mô hình

Trong bài toán này ta sử dụng VCC16 một pre trained model

**Bước 1:** Ở layer đầu tiên cần chỉ rõ input\_shape của ảnh, input\_shape = (W, H, D), ta dùng ảnh kích thước (90,90) nên input\_shape = (90, 90, 3)

**Bước 2:** Có include\_top=False vì trong các tầng ẩn có các tầng convolution layers và mong muốn tầng Dense output ra số nơ ron tương ứng với tập dữ liệu. ở đay sẽ là 40.

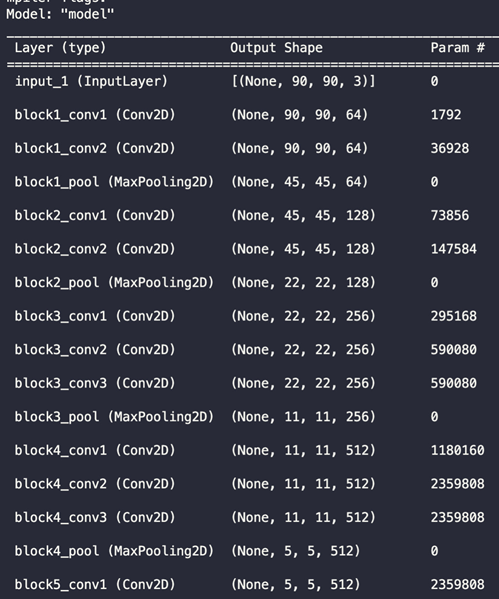
**Bước 3:** Kết nối thêm 1 Tầng Flatten vào VGG16 model. (transfer learning)

Diagram

Description automatically generated

**Bước 4:** Thêm tầng kết nối đầy đủ Dense cũng là output của model, cấu trúc: model.add(Dense(40,activation=”softmax”))

Hàm kích hoạt là softmax để tạo ra 1 vector xác suất.



Graphical user interface, text

Description automatically generated

Hình 2: *Mô hình mạng CNN.*

Hình 2 này có 3 cột , cột thứ nhất là tên các layers theo thứ tự từ trên xuống dưới là từ input => output, cột thứ hai là shape cho mỗi output của từng layers, cột thứ ba là số lượng param(hay weights) cho từng layers. Hàng cuối cùng là tổng hợp số lượng params cho cả model. Như mô hình trên có tổng là hơn 14 tham số, trong đó có 81960 para khả dụng để huấn luyện.

### Biên dịch mô hình

Biên dịch mô hình (compile).

Sau khi build model xong thì compile nó có tác dụng biên dịch mô hình lại toàn bộ mô hình của chúng ta đã xây dựng. Ở đây chúng ta có thể chọn các tham số để training model như : thuật toán training thông qua tham số optimizer, function loss của model chúng ta có thể sử dụng mặc định hoặc tự xây dựng thông qua tham số loss, chọn metrics hiện thị khi model được training.

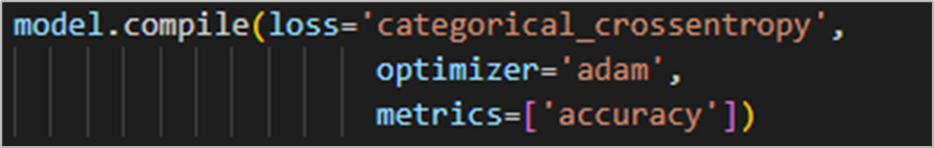
History dùng để xem accuracy, loss qua từng epochs.

Gắn kết mô hình với bộ tối ưu, hàm loss, và tiêu chí đánh giá trong quá trình huấn luyện:

• Optimizer: adam (một dạng của Gradient descent).

• Loss: CategoricalCrossentropy.

• Metrics: accuracy.



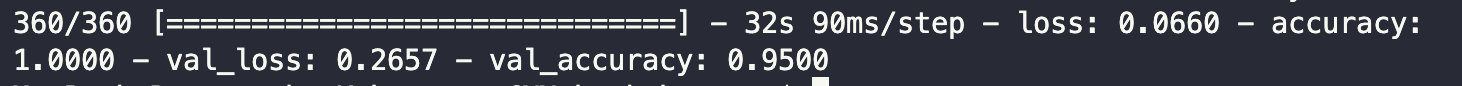
Hình 3*:* *Biên dịch mô hình.*

### Huấn luyện mô hình

* Đưa dữ liệu huấn luyện gồm ảnh và nhãn vào mô hình.
* Mô hình sẽ học cách liên kết ảnh thành nhãn.
* Dự đoán nhãn của ảnh trong tập test.



Hình 4: Huấn luyện mô hình.



Chỉ sau 5 lần epoch thì chúng ta có thể model đã đc accuracy của train là 1 và loss là 0.066; accuracy\_val là 0.95 và loss là 0.2657, ta tiếp tục với các epoch nhiều hơn, bằng cách tăng số lượng epoch lên đến lúc cảm thấy loss đã đc bão hòa.

Lưu dữ model và trọng số của model đã được huấn luyện để tiện cho việc load model và load weight để dự đoán cho dữ liệu mới tói.

Graphical user interface

Description automatically generated

Hình 2: lưu model

### Xây dựng một ứng dụng web

Dựa trên fask framework chúng ta xây dụng một ứng dụng web có thể nhận vào hình ảnh ở đây là sẽ khuôn mặt của một người và trả về đó là nhân vận nào (đối tượng)

Graphical user interface, application

Description automatically generated

Hình 3: ứng dụng web

Graphical user interface, application

Description automatically generated

Hình 4: kết quả trả về là đối tượng nào.

Chúng ta sử dụng mô hình đã được huấn luyện bên trên để dự đoán cho một ảnh bất kì bằng cahs tải lại model và bộ trọng số đã được huấn luyện. Sau đó, sử dụng mô hình này để dự đoán cho các ảnh mới.

# PHẦN KẾT LUẬN

## Kết quả đạt được

* Ứng dụng được pre trained model
* Xử lý dữ liệu.
* Tinh chỉnh được mô hình pretrained model.
* Biên dịch mô hình.
* Huấn luyện mô hình.
* Phân loại được hình ảnh với độ chính xác cao.

## Hướng phát triển

* Xây dựng mô hình mới với kiến trúc mạng khác giúp cho bài toán tốt hơn.
* Thử nghiệm thêm nhiều pretrained model khác nữa.
* Xây dụng giao diện web tốt hơn.

# TÀI LIỆU THAM KHẢO

1. [TensorFlow Open Source machine learning](https://www.youtube.com/watch?v=oZikw5k_2FM)
2. Ertam, Fatih, and Galip Aydın. "Data classification with deep learning using Tensorflow." *2017 international conference on computer science and engineering (UBMK)*. IEEE, 2017.
3. Zeng, Zixian, Qingge Gong, and Jun Zhang. "CNN model design of gesture recognition based on tensorflow framework." *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE, 2019.
4. <https://www.tensorflow.org/guide/effective_tf2>
5. *TensorFlow.org*
6. <https://www.tensorflow.org/tutorials/keras/classification>
7. [Google Just Open Sourced TensorFlow, Its Artificial Intelligence Engine](http://www.wired.com/2015/11/google-open-sources-its-artificial-intelligence-engine/).