

Hệ thống nhúng

Nhóm 9





Trần Quốc Nhơn - N19DCCN132

Lê Hoài Nhân - N19DCCN126

Lê Quang Phục - N19DCCN143

Huỳnh Lê Thanh Nhật - N19DCCN127

Phạm Minh Quang - N19DCCN151



Nội dung

1

Tổng quan

2

Cơ sở lý thuyết

3

Thiết bị phần cứng

4

Mô hình và huấn luyện dữ liệu

5

Hạ tầng, lưu trữ và xử lý dữ liệu

6

Flowchart và quá trình xây dựng hệ thống







Lý do chọn đề tài

Người khiếm thị là những người có khuyết tật về thị giác, bao gồm cả mù lòa hoặc mù màu.

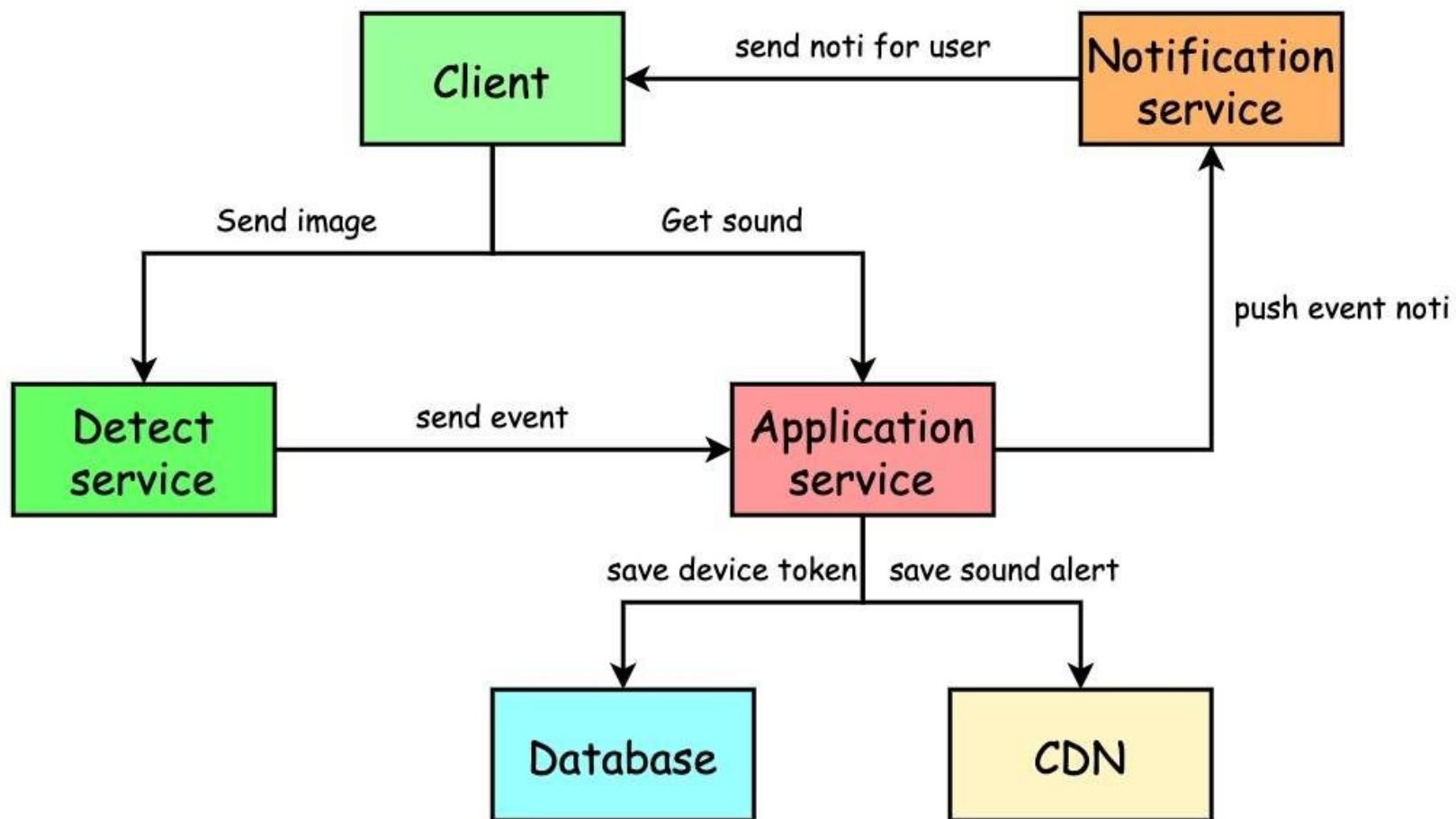
Do đó trong việc điều hướng và di chuyển thì người khiếm thị cần giúp đỡ trong việc điều hướng và di chuyển trong môi trường mới. Họ cần được hướng dẫn để tránh va chạm và đảm bảo an toàn. Giúp cho họ có thể có cuộc sống đầy đủ và được cảm nhận tình yêu thương và sự quan tâm của cộng đồng.



Tổng quan về đề tài

- là một sản phẩm công nghệ nhằm hỗ trợ người khiếm thị trong việc di chuyển và tìm đường đi.
- Đây là một giải pháp đáp ứng nhu cầu cần thiết cho người khiếm thị, giúp họ tự tin và độc lập hơn trong cuộc sống hàng ngày.
- Thiết bị được thiết kế với các tính năng như định vị, hướng dẫn đi lại, cảnh báo nguy hiểm.
- Thiết bị này sử dụng các công nghệ như Computer Vision, Machine Learning, IoT để thu thập thông tin và xử lý dữ liệu. Các thông tin được thông qua âm thanh để giúp người khiếm thị dễ dàng tiếp cận.
- Sản phẩm cũng có tính di động cao, thuận tiện cho người sử dụng mang theo bên mình.

LƯU ĐỒ HỆ THỐNG





Cơ sở lý thuyết



Học máy - Machine Learning

Machine Learning (Học máy) là một lĩnh vực của khoa học máy tính, là khả năng của chương trình máy tính sử dụng kinh nghiệm, quan sát hoặc dữ liệu trong quá khứ để cải thiện công việc của mình trong tương lai thay vì chỉ thực hiện theo đúng các quy tắc đã được lập trình sẵn. Chẳng hạn, máy tính có thể học cách dự đoán dựa trên các ví dụ, hay học cách tạo ra các hành vi phù hợp dựa trên quan sát trong quá khứ.



Nhận diện đối tượng(Object Detection)

- Object Detection là một trong những lĩnh vực của Computer Vision (thị giác máy tính) được sử dụng để nhận dạng và định vị các đối tượng trong ảnh hoặc video.
- Trong object detection, chúng ta cần phải tìm các đối tượng có trong ảnh, vị trí của chúng, và xác định lớp đối tượng đó là gì. Các kỹ thuật object detection thường được sử dụng hiện nay bao gồm các phương pháp cơ bản như Haar cascade, HOG, hoặc phương pháp deep learning như Faster R-CNN, SSD, YOLO.





Vi điều khiển

- là một bộ xử lý nhỏ gọn tích hợp một số phần mềm và phần cứng, được sử dụng để điều khiển và điều phối các hoạt động của các thiết bị điện tử. Vi điều khiển có thể tích hợp các cổng giao tiếp như I2C, SPI, UART, GPIO, v.v. để kết nối với các linh kiện khác nhau.
- Vi điều khiển thường được sử dụng trong các thiết bị điện tử như các hệ thống nhúng, điều khiển thiết bị, các thiết bị y tế, các thiết bị gia dụng thông minh, các thiết bị cảm biến, và các thiết bị IoT.



Thiết bị phần cứng

ESP32-CAM

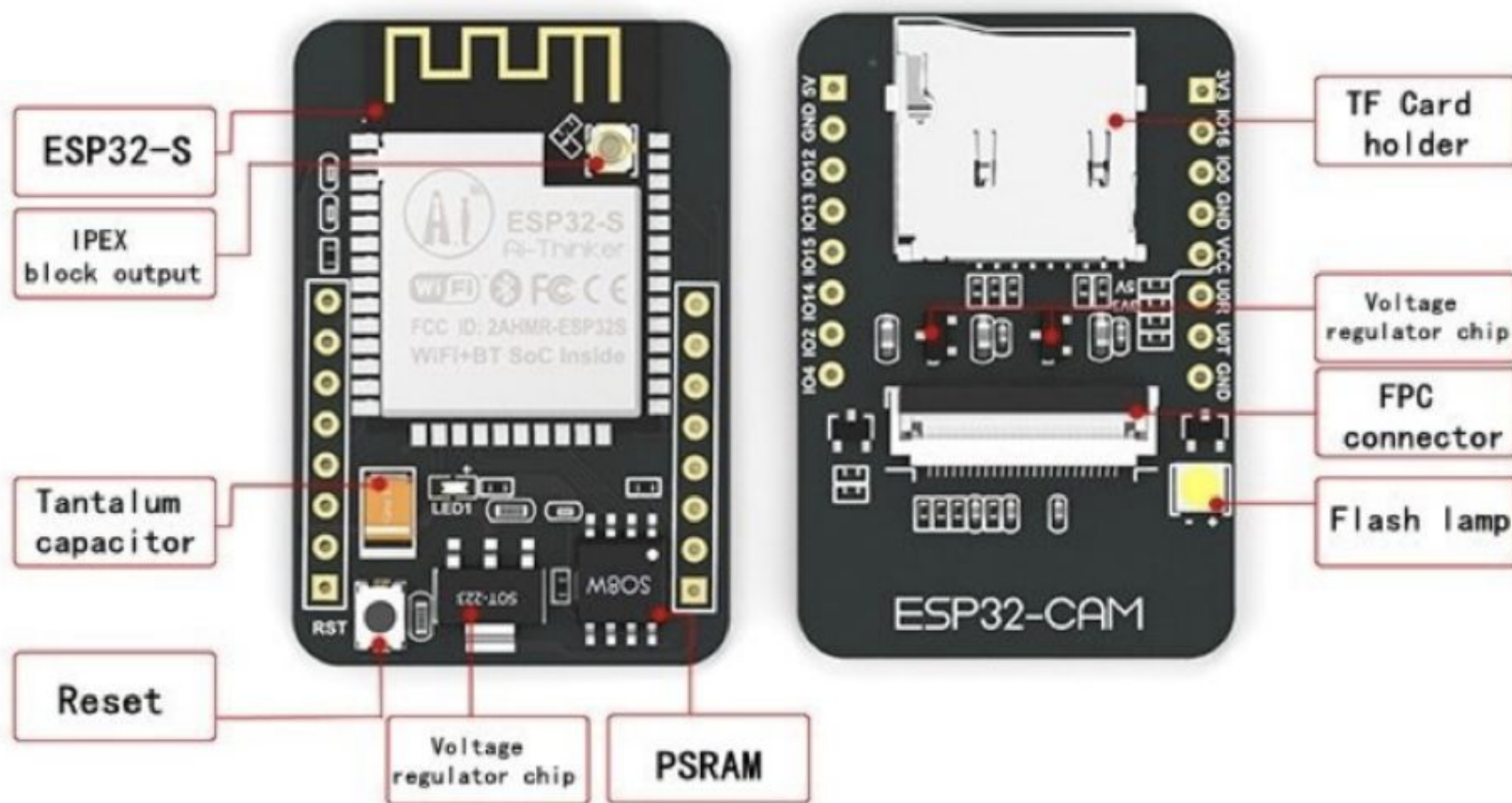
- ESP32-CAM là một module tích hợp WiFi và camera, được phát triển bởi Espressif Systems. Module này sử dụng chip ESP32 của Espressif Systems, bao gồm một vi xử lý hai nhân Tensilica LX6, kết nối WiFi băng tần kép và một loạt các tính năng khác như Bluetooth, GPIO, UART, SPI, I2C, ADC và DAC.
- Module ESP32-CAM có khả năng ghi hình video và chụp ảnh với độ phân giải cao lên đến 1600x1200 pixel. Module này được thiết kế với kích thước nhỏ gọn, có thể dễ dàng lắp đặt vào các thiết bị nhúng hoặc các ứng dụng IoT.



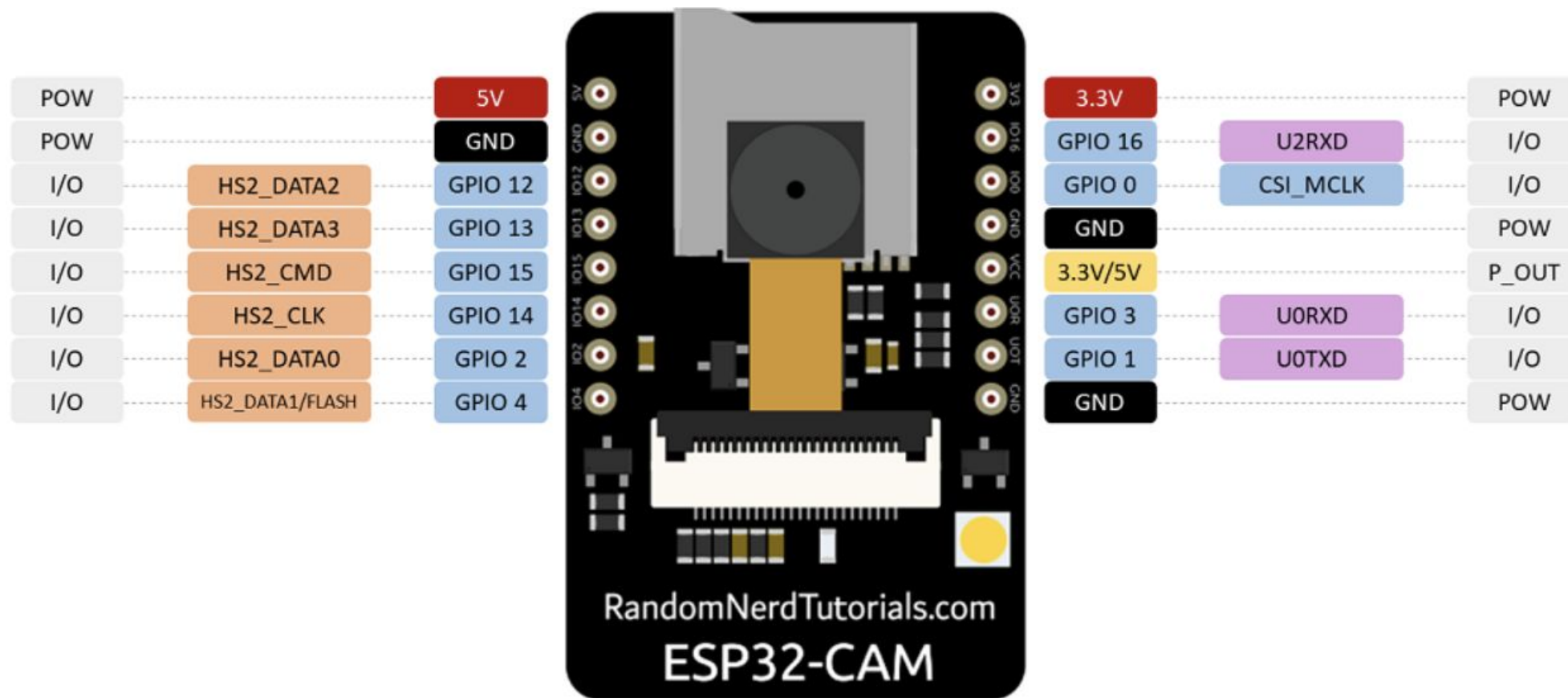
Thông số cơ bản ESP32-CAM

- IC chính: ESP32-S (AI-Thinker)
- Mô-đun Wi-Fi BT SoC 802.11 b/g/n/e/i
- CPU 32-bit công suất thấp, cũng có thể phục vụ bộ xử lý ứng dụng
- Tốc độ đồng hồ lên đến 160MHz, sức mạnh tính toán lên đến 600 DMIPS
- Tích hợp 520 KB SRAM, 4MPSRAM bên ngoài
- Dải tần số: 1421 ~ 2484 Mhz
- Bluetooth: 4.2 BR/EDR BLE
- Hỗ trợ UART / SPI / I2C / PWM / ADC / DAC
- Hỗ trợ máy ảnh OV2640 và OV7670, đèn flash tích hợp
- Hỗ trợ tải lên WiFi hình ảnh
- Hỗ trợ thẻ TF
- Hỗ trợ nhiều chế độ ngủ
- Nhúng Lwip và FreeRTOS
- Hỗ trợ chế độ hoạt động STA / AP / STA + AP
- Hỗ trợ cấu hình thông minh / công nghệ AirKiss
- Hỗ trợ nâng cấp cục bộ và từ xa cho cổng nối tiếp (FOTA)

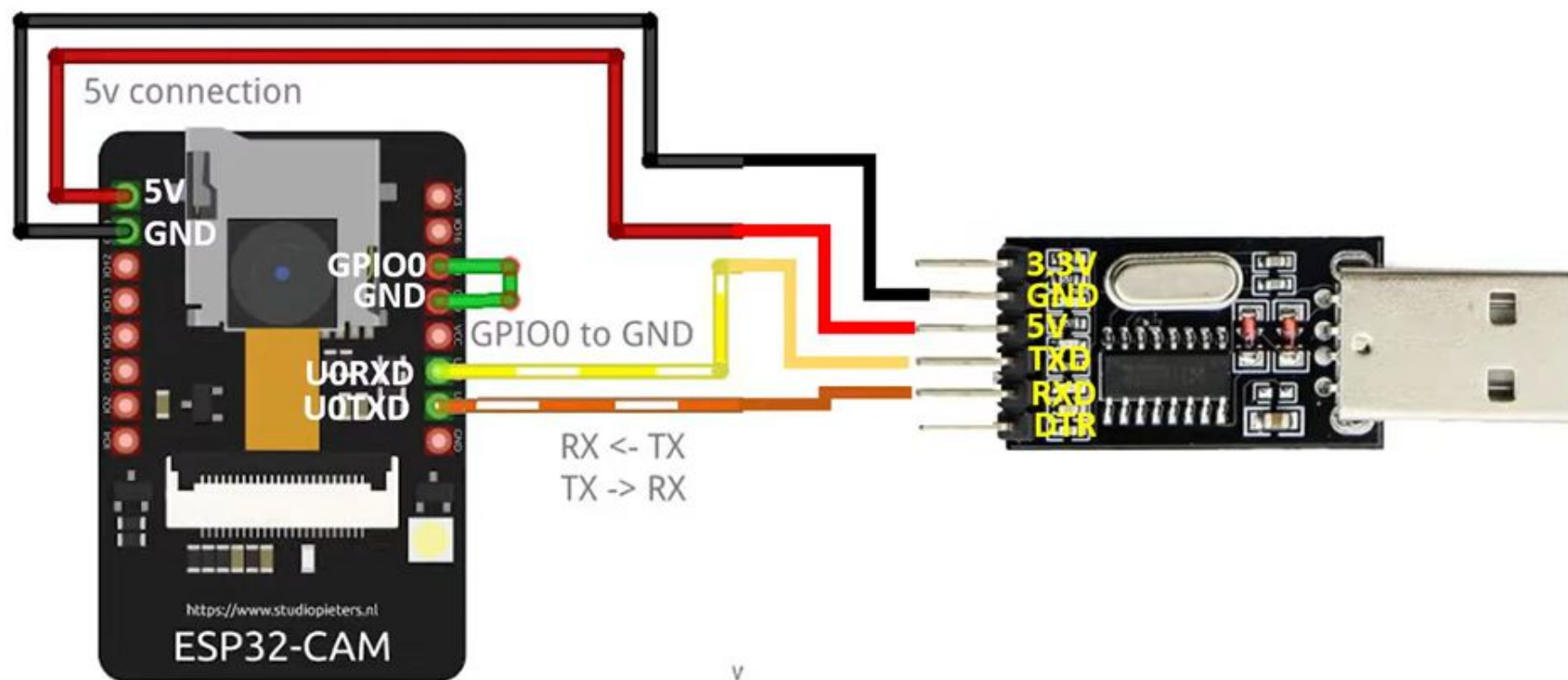
Sơ đồ nguyên lý mạch ESP32-CAM



Sơ đồ chân ESP32-CAM



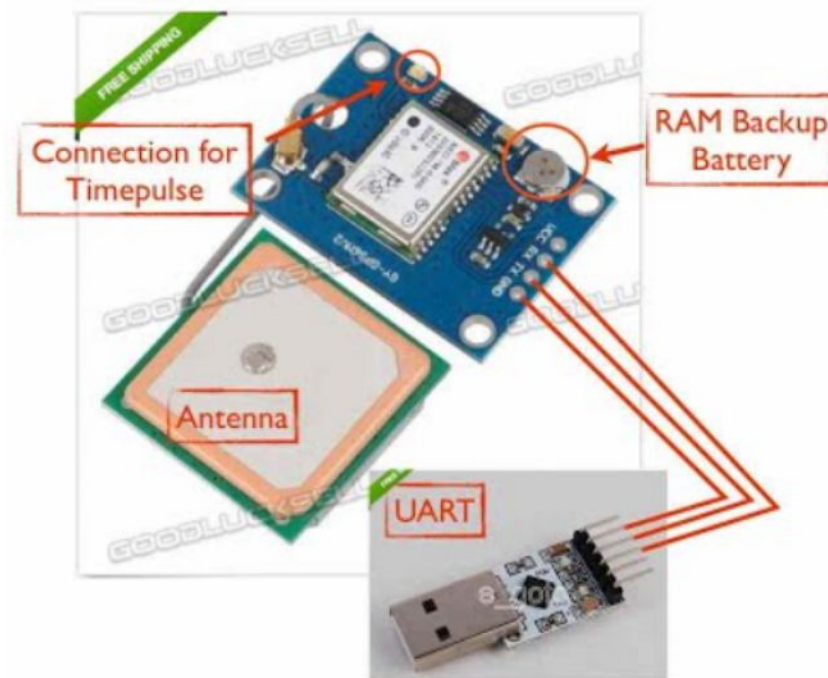
Cách nối mạch



Mạch định vị GPS NEO-6M V2

Module

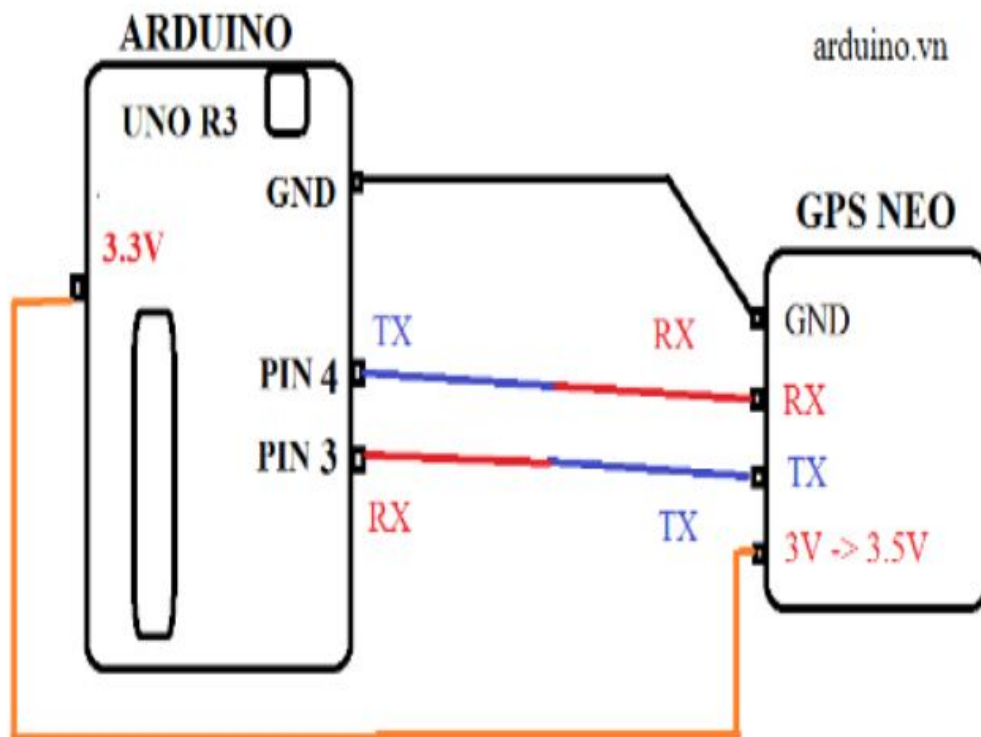
đi kèm theo là 1 ăng ten thu tín hiệu vệ tinh bằng sứ



Cách nối mạch

Nối mạch test thư viện GPS cho arduino.

PIN 4 : TX giả lập
PIN 3 : RX giả lập
PIN 1 : TX.
PIN 0 : RX.



Arduino - GPS NEO :

- Chân vcc - Nguồn 3v - 3.5v
- Chân GND - GND
- Chân RX - pin4TX
- Chân TX - pin3RX

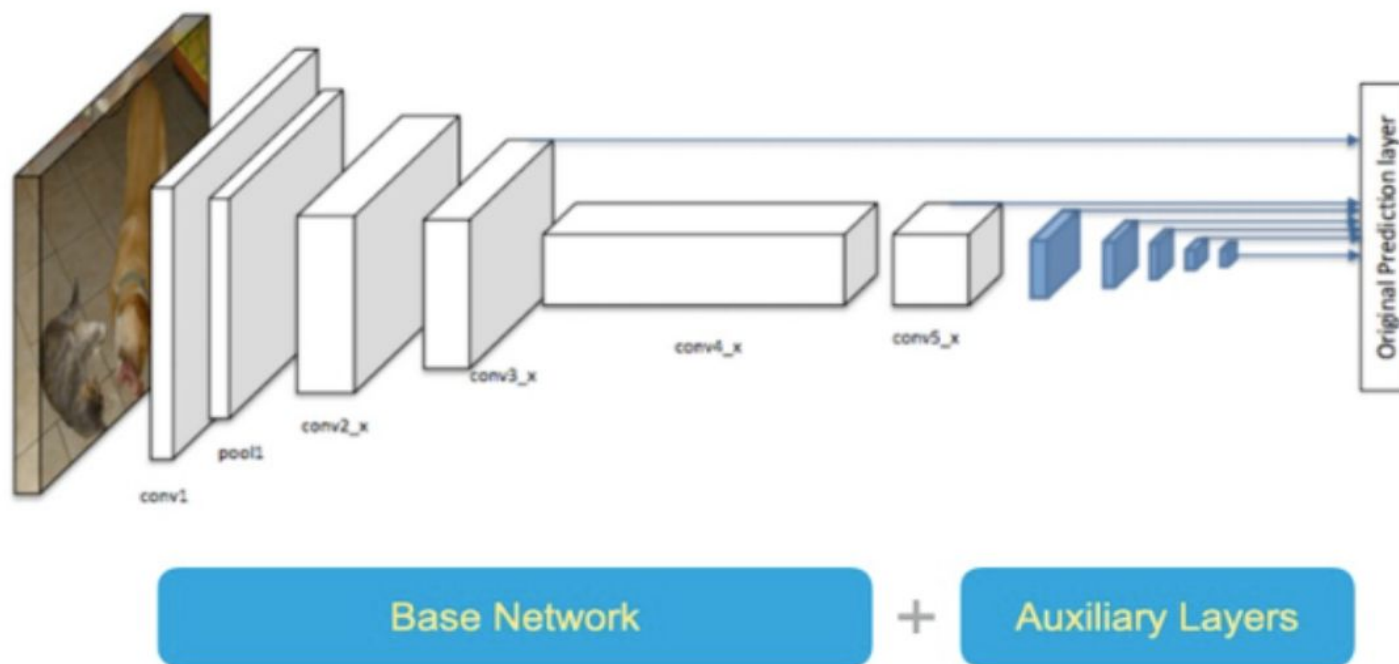




SSD

- SSD (Single Shot MultiBox Detector) model được giới thiệu vào năm 2016 là một kiểu model phát hiện và định vị các đối tượng trong lĩnh vực computer vision.
- Khác với các model trước đó như R-CNN hay cải tiến là Faster R-CNN, chúng nhận dạng vật thể qua 2 giai đoạn (2-stages) Region Proposals Network và Object Detection Network.
- Điểm làm nên thương hiệu của SSD là tốc độ nhanh, độ chính xác khá cao (cao hơn YOLO - một model 1-stage khác), khả năng phát hiện các vật thể với kích thước khác nhau.

SSD



Minh họa kiến trúc SSD



Kiến trúc SSD

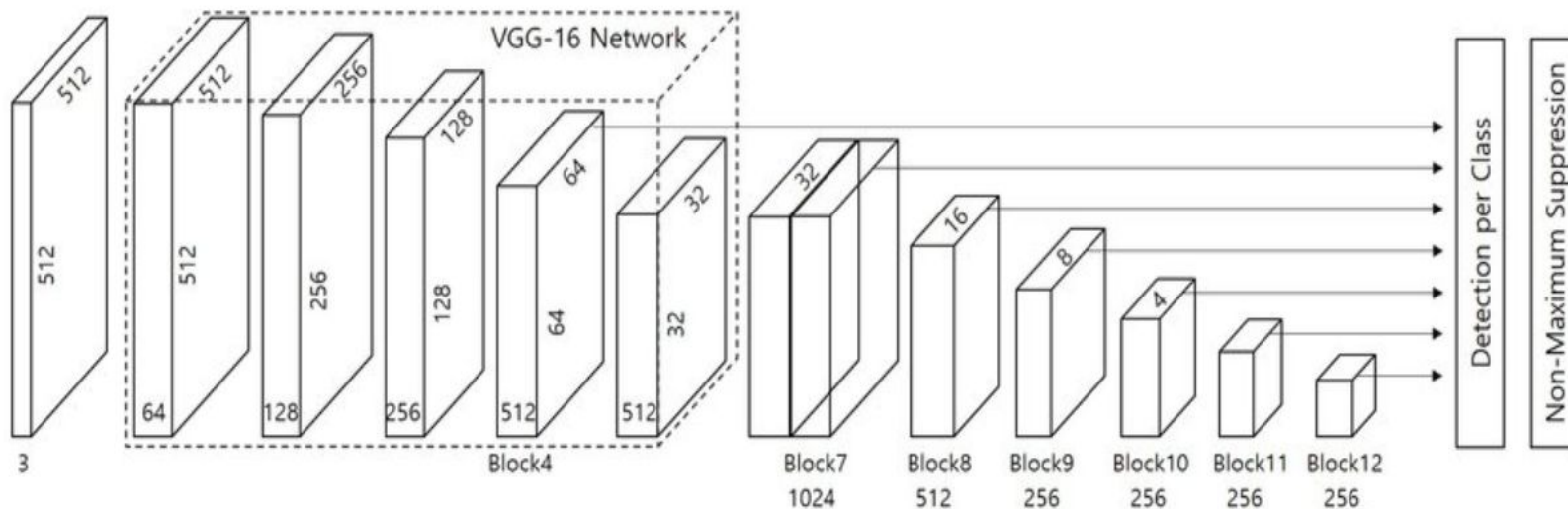
- Mô hình sử dụng một mạng neural duy nhất, bao gồm 2 thành phần: một mô hình mạng cơ sở (Base Network) và các lớp neuron phụ trợ (Auxiliary Layers).
- Trong đó, mạng cơ sở có nhiệm vụ trích xuất đặc trưng ảnh (là một mạng image classification), sau đó các đặc trưng này được sử dụng bởi các lớp neuron phụ trợ để dự đoán vị trí của các đối tượng trong ảnh.



Kiến trúc SSD

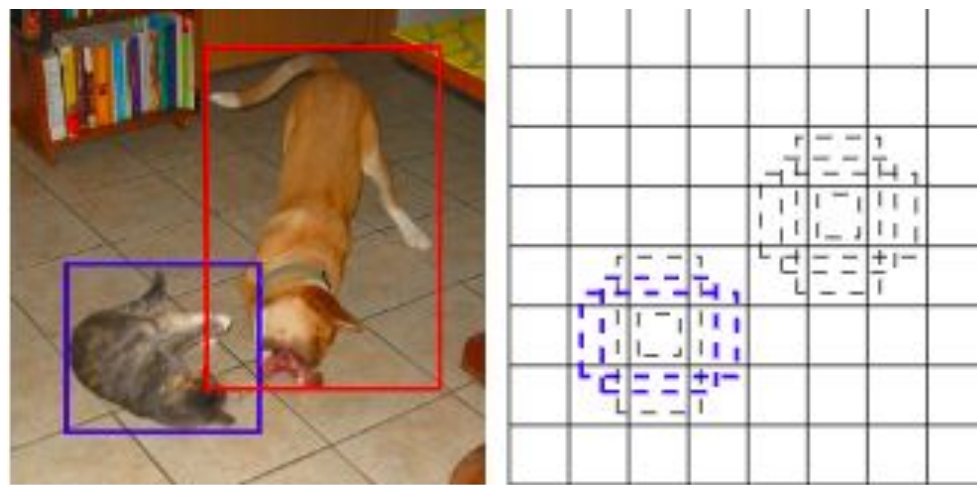
- Mô hình này có thể linh hoạt thay đổi phần base network để phù hợp với bài toán cần giải quyết.
- Sau khi thu được feature map từ base network, ta sẽ thêm vào phần sau đó các lớp neuron phụ trợ với kernel $3 \times 3 \times p$ (p là chanel ảnh) để đưa ra các bounding boxes và nhãn tương ứng.
- Sau cùng, các kết quả trên sẽ được đi qua một lớp gọi là Non-maximum Suppression layer nhằm khử các bounding boxes - chỉ đưa ra một bounding box duy nhất và chứa nhãn dự đoán vật thể.

Cách hoạt động SSD



- Trích xuất đặc trưng từ base network -> feature map
- Auxiliary layers sẽ dùng feature map này để tính toán tiếp, cụ thể:

- + Chia feature map thành một grid cell (vd: 32x32).
- + Mỗi cell sẽ tạo ra một số lượng bounding boxes (4 hoặc 6) kết hợp với aspect ratio và độ scale khác nhau.
- + Bên cạnh đó, trên mỗi feature map này còn dùng 1 kernel 3x3 để tính toán ra feature map ở layer kế tiếp.
- + Feature map này thực hiện các bước tương tự.
- + Lưu ý với mỗi feature map đều đưa ra bounding boxes tương ứng với class, vì thế để khử các bounding gần đúng, các output này sẽ được đi qua non-maximum suppression layer.



Ví dụ về bounding boxes

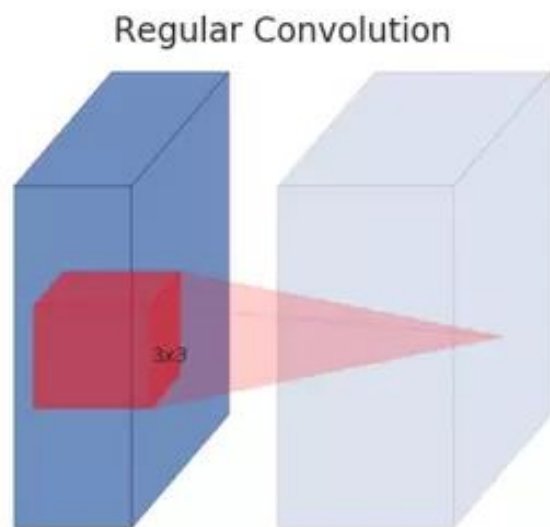
- IoU = $\frac{\text{Area of Overlap}}{\text{Area of Union}}$



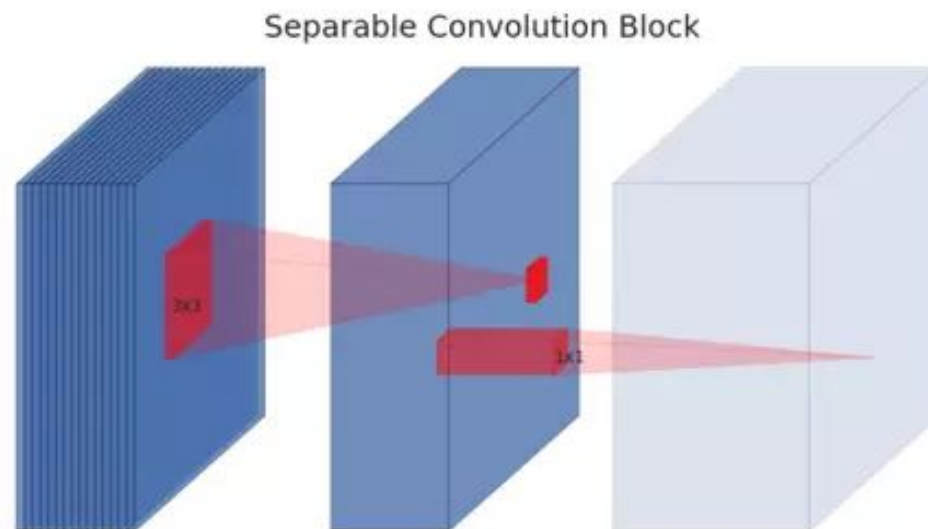
MobileNet

- MobileNet là một kiến trúc mạng nơ-ron tích chập (CNN) được thiết kế để sử dụng trên các thiết bị di động với các tài nguyên tính toán và bộ nhớ hạn chế.
- MobileNet được phát triển bởi Google và được giới thiệu trong bài báo khoa học "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications" vào năm 2017.

(a) Regular



(b) Separable



CNN vs Separable CNN

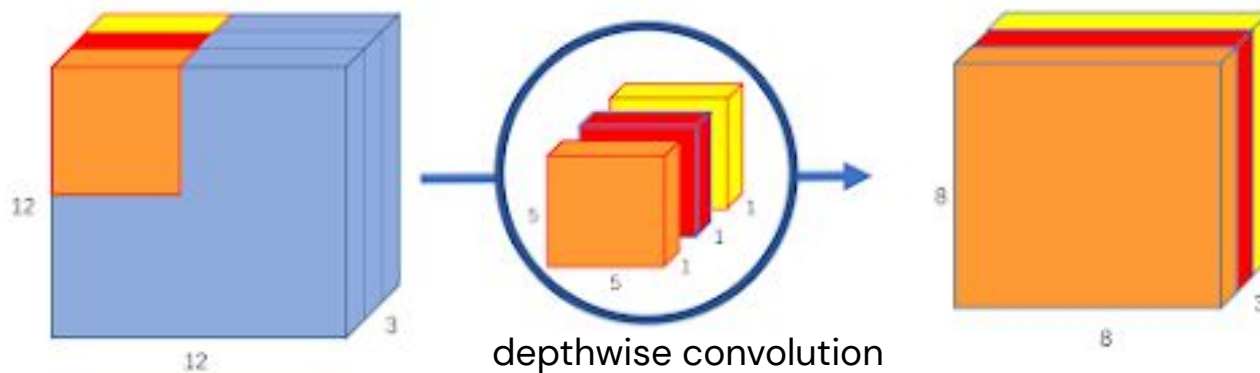


MobileNet

- MobileNet được thiết kế để giảm kích thước mô hình và tăng tốc độ tính toán bằng cách sử dụng hai kỹ thuật: Depthwise Convolution và Pointwise Convolution. Điều này giúp cho MobileNet có khả năng hoạt động nhanh và hiệu quả trên các thiết bị di động có tài nguyên hạn chế.
- Các ưu điểm của MobileNet so với các mô hình CNN khác là kích thước mô hình nhỏ hơn, tốc độ tính toán nhanh hơn và độ chính xác tương đối cao. Do đó, MobileNet được sử dụng rộng rãi trong các ứng dụng di động như phân loại hình ảnh, nhận dạng đối tượng và nhận dạng hành động.

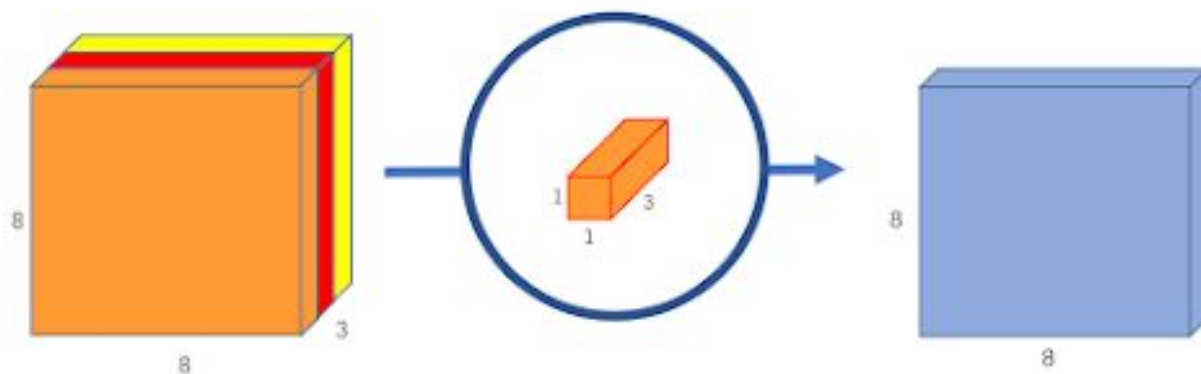
Depthwise convolution

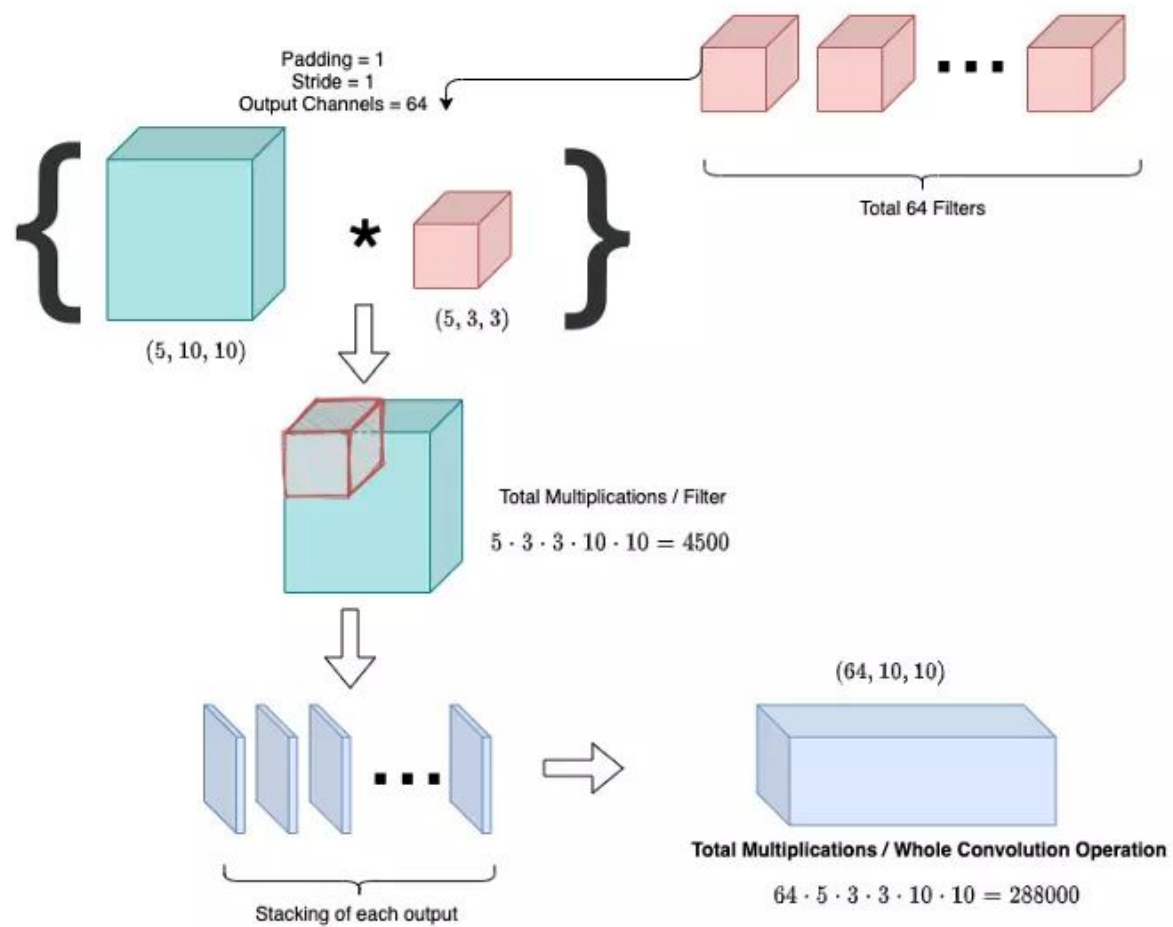
- Tách đầu vào thành các kênh riêng biệt.
- Chuyển đổi từng đầu vào với bộ lọc tương ứng.
- Xếp chồng các kết quả đầu ra được biến đổi với nhau.



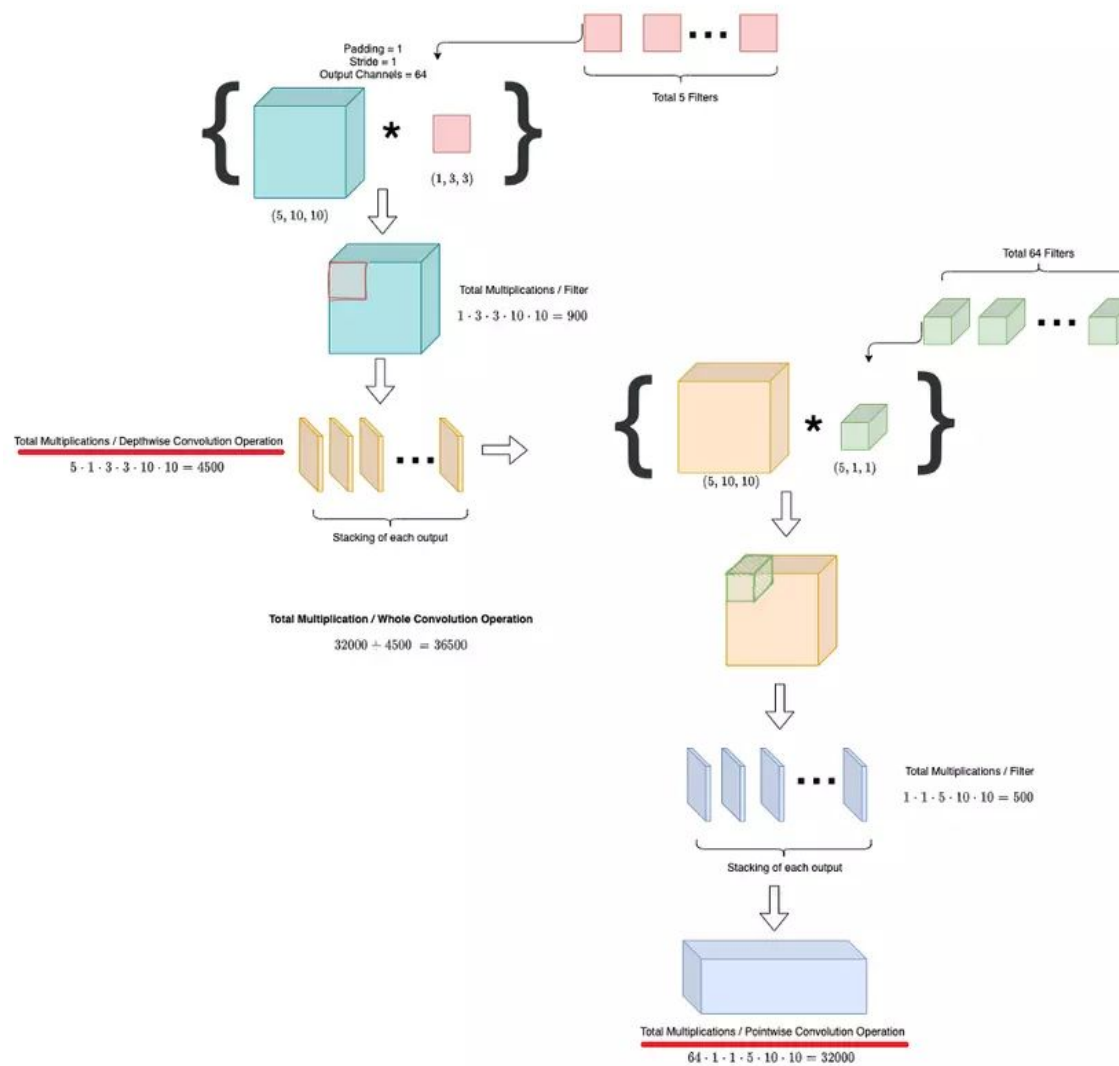
Pointwise convolution

- Sử dụng một nhân 1×1 (point) có độ sâu bằng số lượng kênh mà ảnh đầu vào có.





Standard CNN



Separable CNN

5

Hạ tầng, dữ liệu



Dataset

COCO (Common Objects in Context) là một tập dữ liệu lớn cho việc nhận dạng vật thể trong ảnh. COCO bao gồm hơn 330.000 hình ảnh được gán nhãn với hơn 1.5 triệu đối tượng thuộc 80 loại khác nhau. Bên cạnh đó, COCO còn có thông tin về mối liên hệ giữa các đối tượng và định vị vị trí của chúng trong không gian 2D.

COCO lưu trữ các dữ liệu annotation trong 1 file có định dạng JSON. Với một file dữ liệu JSON của COCO sẽ gồm những block chính như sau:



Dataset

- Info: trường dữ liệu này chứa thông tin về tập dataset.
- Licenses: trường dữ liệu này chứa thông tin về giấy phép của tập dataset.
- Categories: trường dữ liệu này chứa một danh sách tên các đối tượng.
- Images: trường dữ liệu này lưu trữ thông tin của các hình ảnh (đường dẫn, tên hình ảnh, id của ảnh).
- Annotations: trường dữ liệu này chứa thông tin nhãn của từng đối tượng trong ảnh.



TF record

- Là định dạng tệp binary được sử dụng trong TensorFlow để lưu trữ và truyền dữ liệu.
- Khi dữ liệu train quá lớn - không thể load vào RAM hết được; thì với TF Record, dữ liệu sẽ được format về binary giúp tăng tốc độ trong quá trình train , cũng như khả năng xử lý dữ liệu song song , tiết kiệm nhiều tài nguyên cho việc train.
- TF Record được lưu dưới dạng một chuỗi binary, vì vậy cần phải có một định dạng trước khi chuyển từ dữ liệu thô sang TF Record.

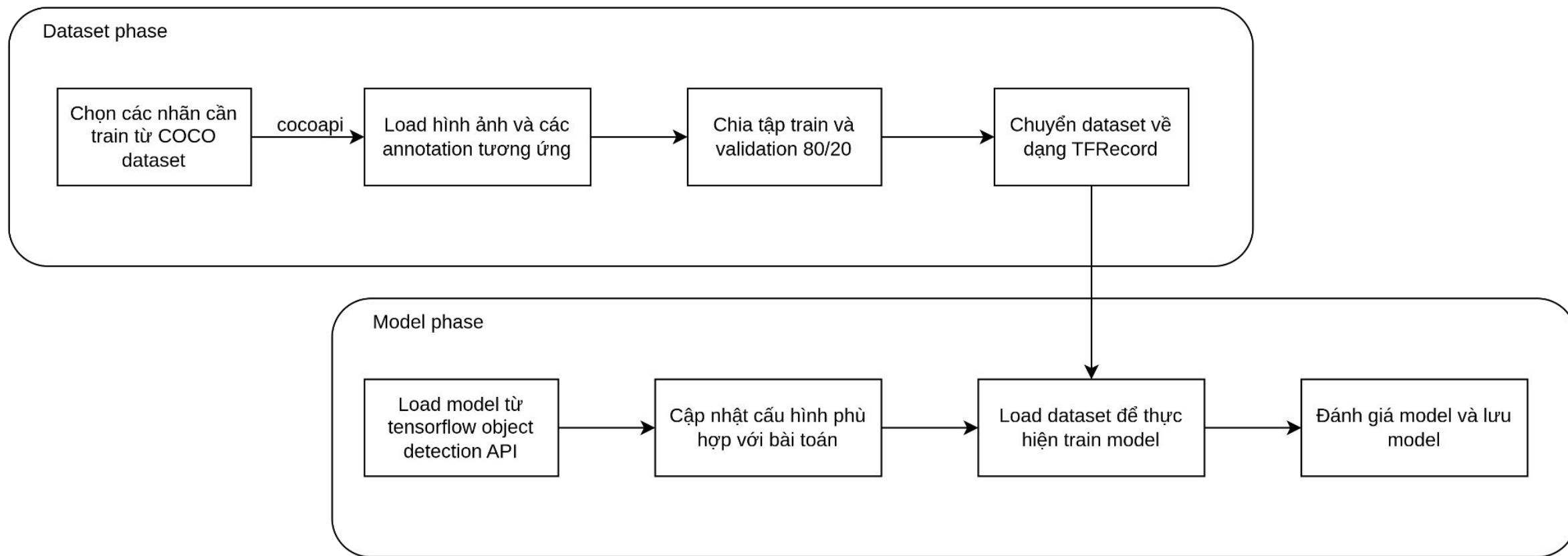


Checkpoint

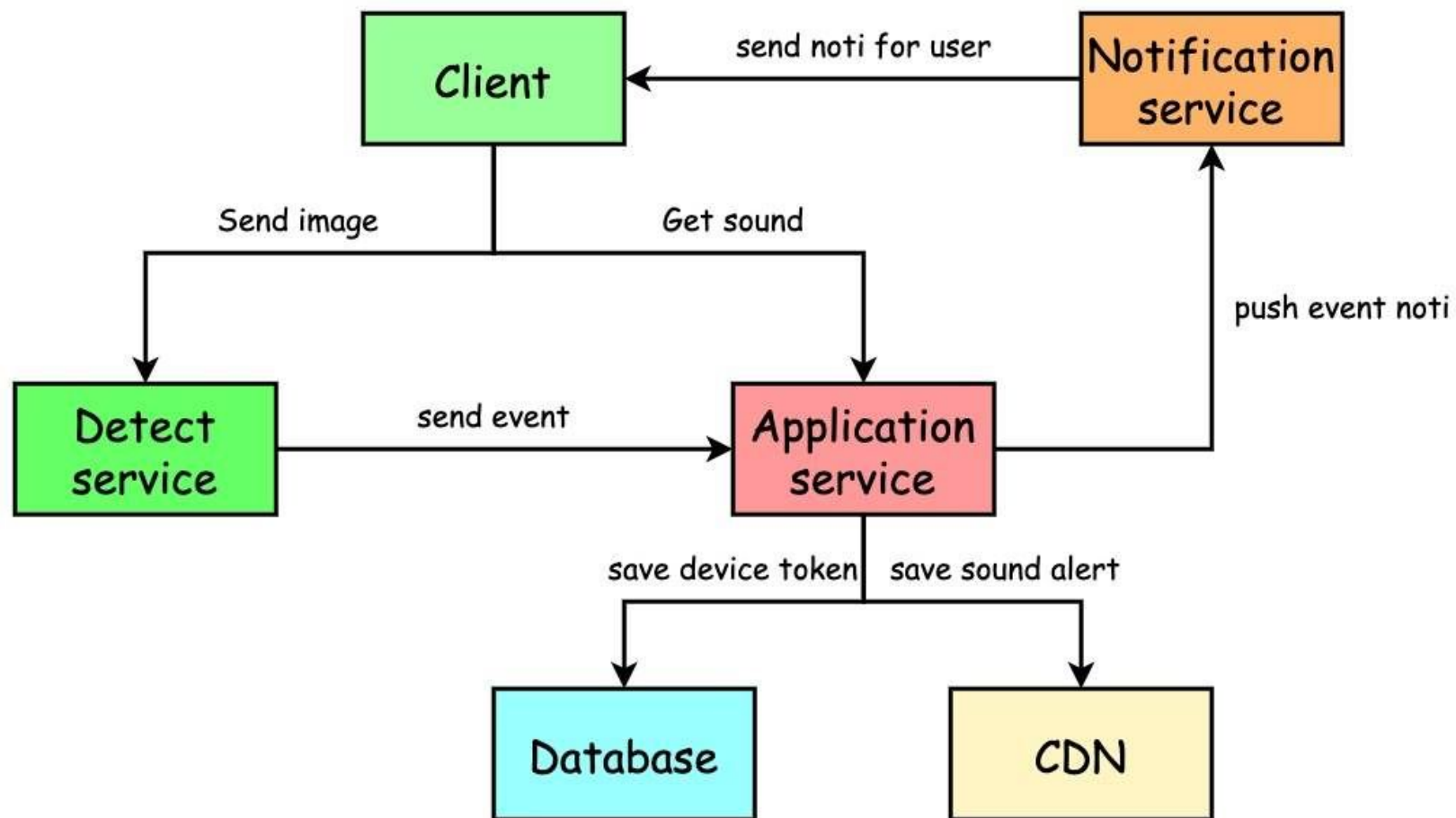
- Là một cơ chế để lưu trữ và quản lý các trọng số của model.
- Khi dữ liệu train lớn, việc lưu trữ checkpoint giúp ta tiết kiệm được thời gian train, vì ta không cần phải train lại từ đầu mỗi khi sử dụng model mà tiếp tục train tại checkpoint đã lưu trước đó.
- Checkpoint được sử dụng để theo dõi tiến trình huấn luyện và giám sát hiệu suất của model và điều chỉnh các tham số để cải thiện hiệu suất của mô hình
- Cho người dùng đánh giá độ chính xác của model từ checkpoint và khả năng triển khai model lên các môi trường thực tế

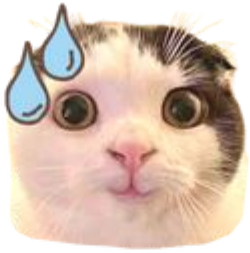


Quá trình build model



LƯU ĐỒ HỆ THỐNG





Thanks

