

Mì “Python”

Mì AI Training

Bài số 05



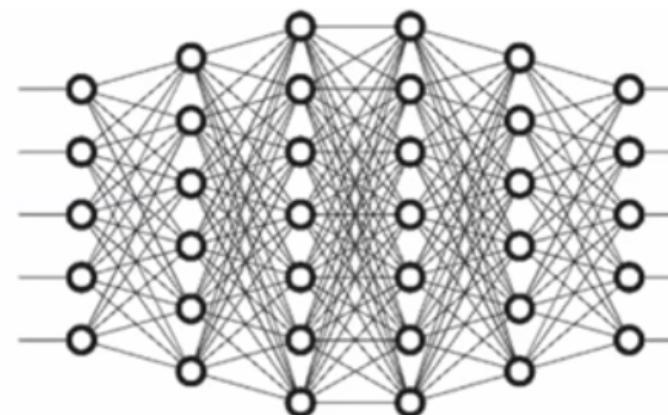
Nội dung khóa học

- Bài 1. Python cơ bản A
- Bài 2. Python cơ bản B
- Bài 3. Python với OpenCV
- Bài 4. Python với Keras
- Bài 5. Python với Keras 2
- Bài 6. Python với Pandas
- Bài 7. Xây dựng Backend Server với Python

Tại sao cần CNN?

Mạng NN thông thường sẽ gặp vấn đề về số lượng weights khi xử lý ảnh.

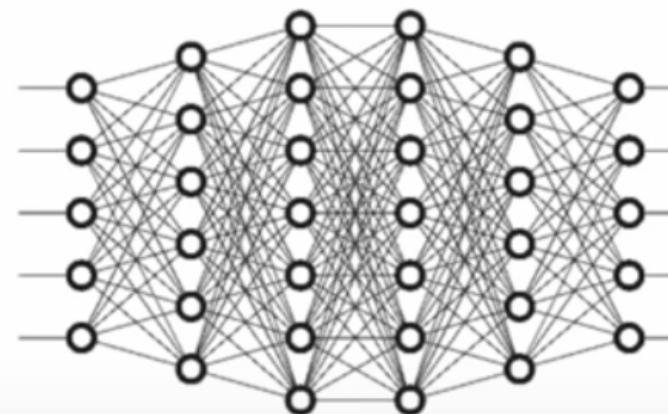
Image with
 $28 \times 28 \times 3$
pixels



*Number of weights in
the first hidden layer
will be 2352*



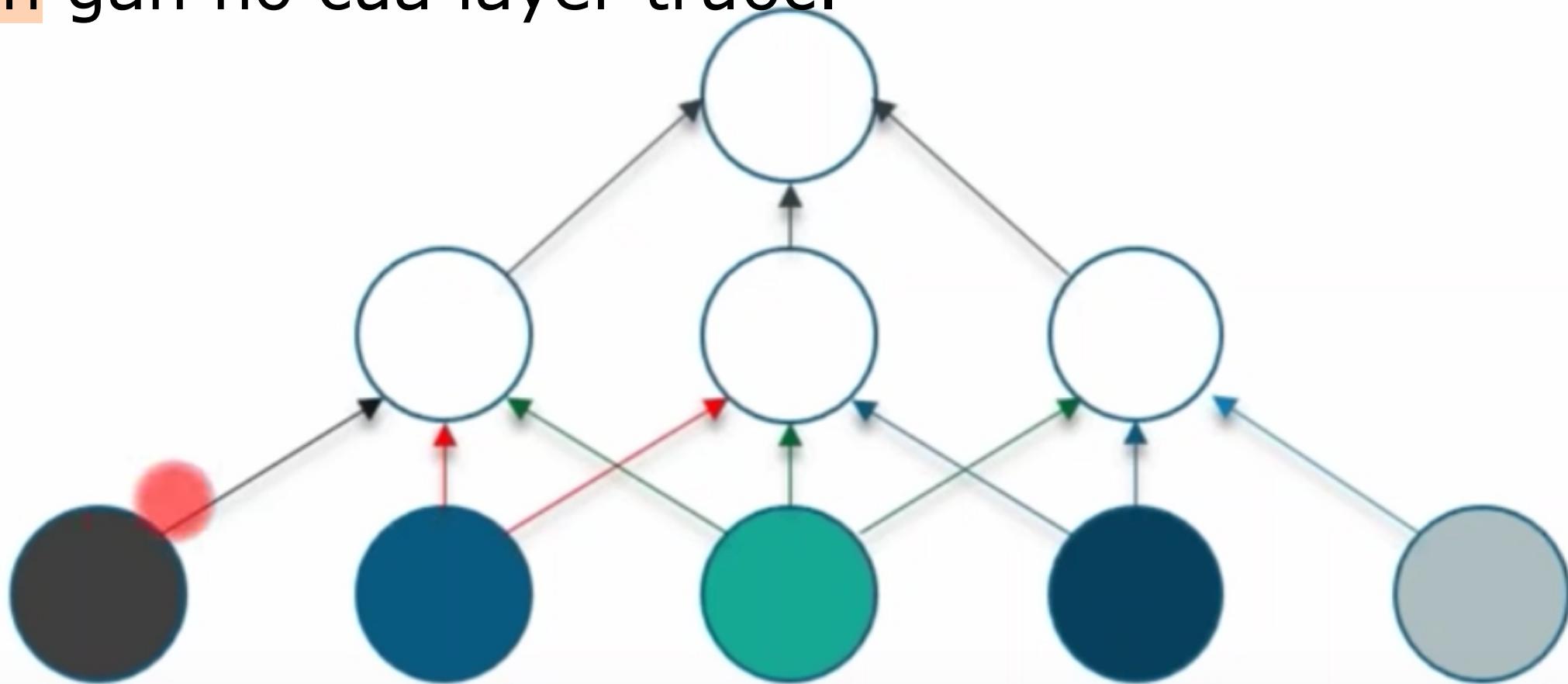
Image with
 $200 \times 200 \times 3$
pixels



*Number of weights in
the first hidden layer
will be 120,000*

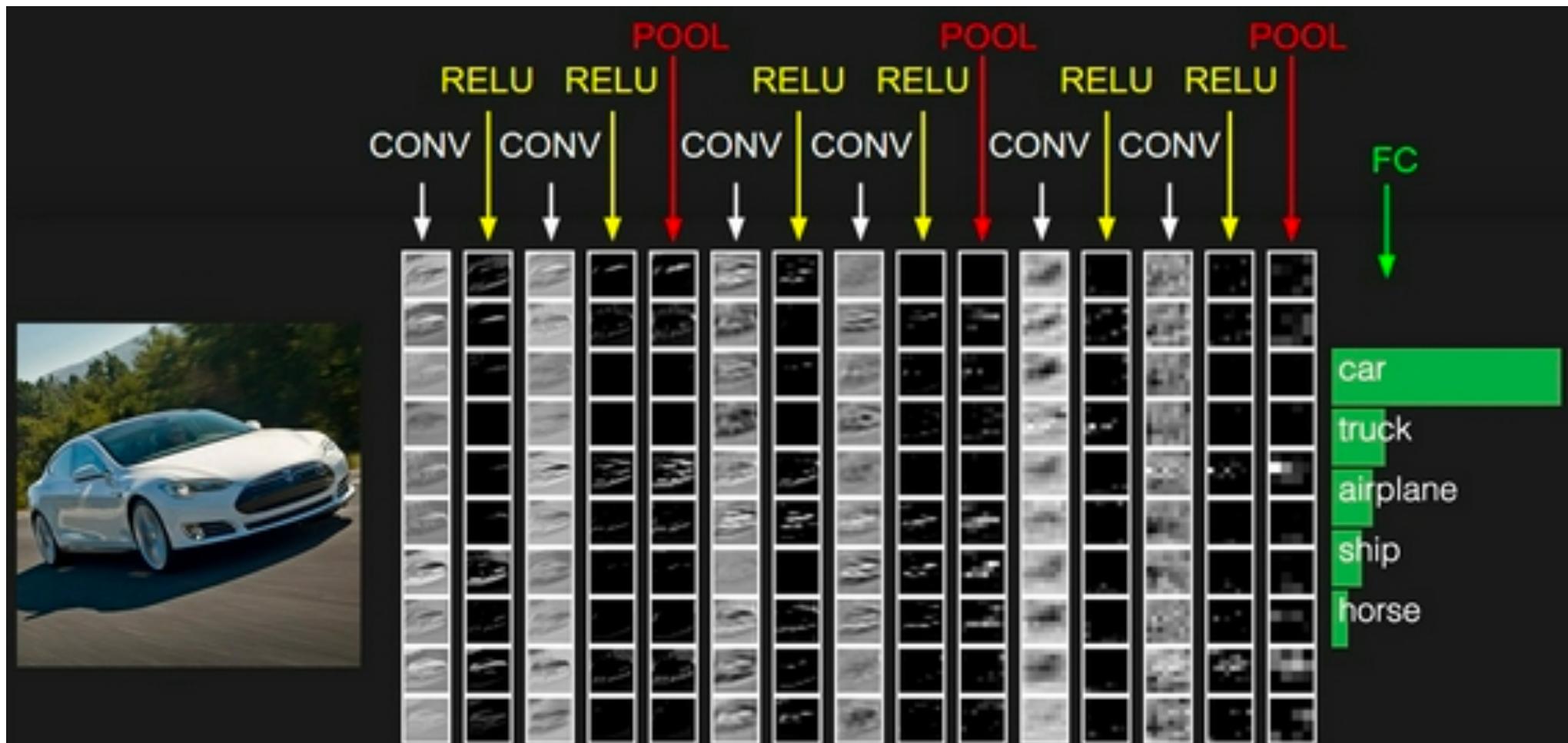
Tại sao cần CNN?

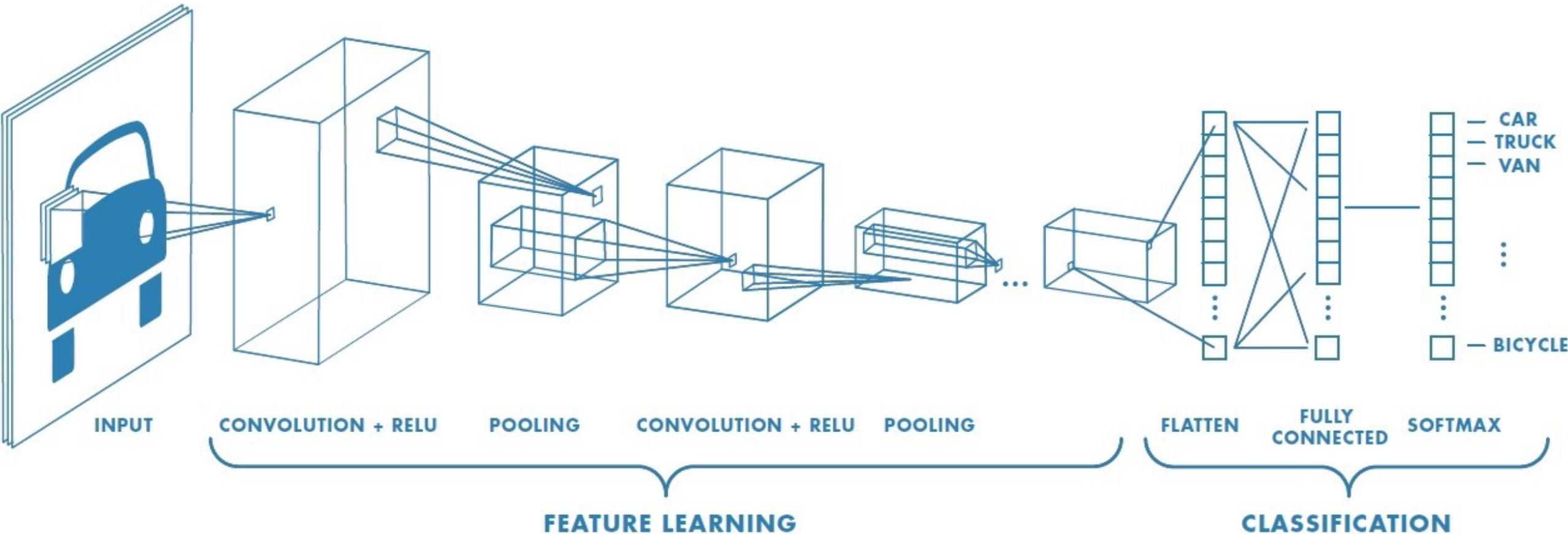
Với CNN, một Nơ ron trong một lớp chỉ kết nối tới vài nơ ron gần nó của layer trước.



Cấu trúc mạng CNN

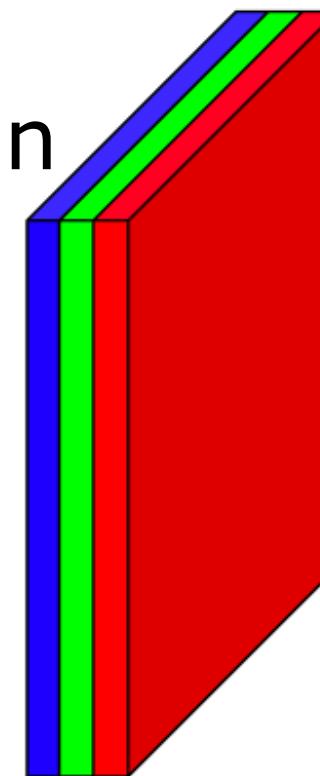
CNN có 02 phần chính: **Lớp trích lọc đặc trưng của ảnh** (Conv, Relu và Pool) và **Lớp phân loại** (FC và softmax).



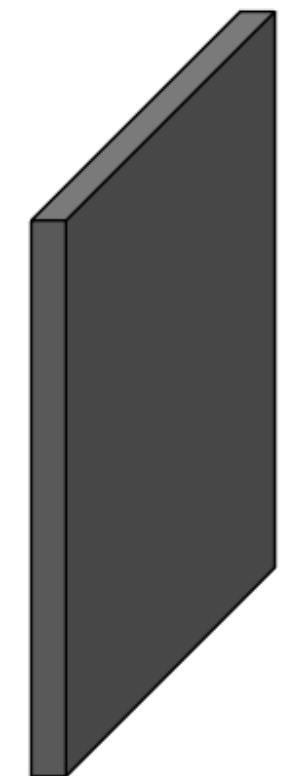


Convolution Layer

- Sử dụng để trích chọn các đặc trưng của ảnh đầu vào
- Input là ảnh biểu diễn bởi ma trận $[M \times N \times D]$. Trong đó
 - M: Chiều rộng
 - N: Chiều cao
 - D: Số chiều



RGB
Image
 $M \times N \times 3$



Grayscale/Binary
Image
 $M \times N \times 1$

Convolution Layer

- Sử dụng các cửa sổ trượt để trích xuất đặc trưng từng vùng ảnh
- Output là các đặc trưng của ảnh biểu diễn dưới dạng ma trận.
- Link tham khảo thêm [tại đây](#).

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved
Feature

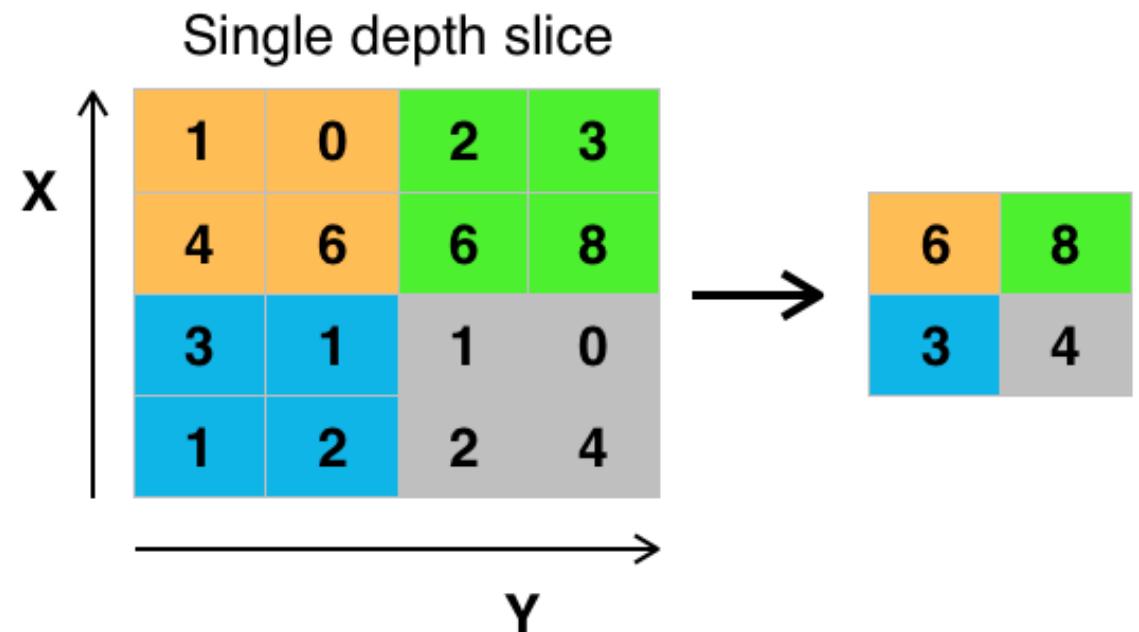
ReLU Layer

- ReLU layer áp dụng các kích hoạt (activation function) **max(0,x)** lên đầu ra của Conv Layer, có tác dụng đưa các giá trị âm về thành 0 tránh ảnh hưởng cho việc tính toán ở các layer sau đó.
- Layer này không thay đổi kích thước của ảnh và không có thêm bất kì tham số nào.
- Layer này đảm bảo cho các giá trị là non-linear (không tuyến tính) qua từng layer

Pooling Layer

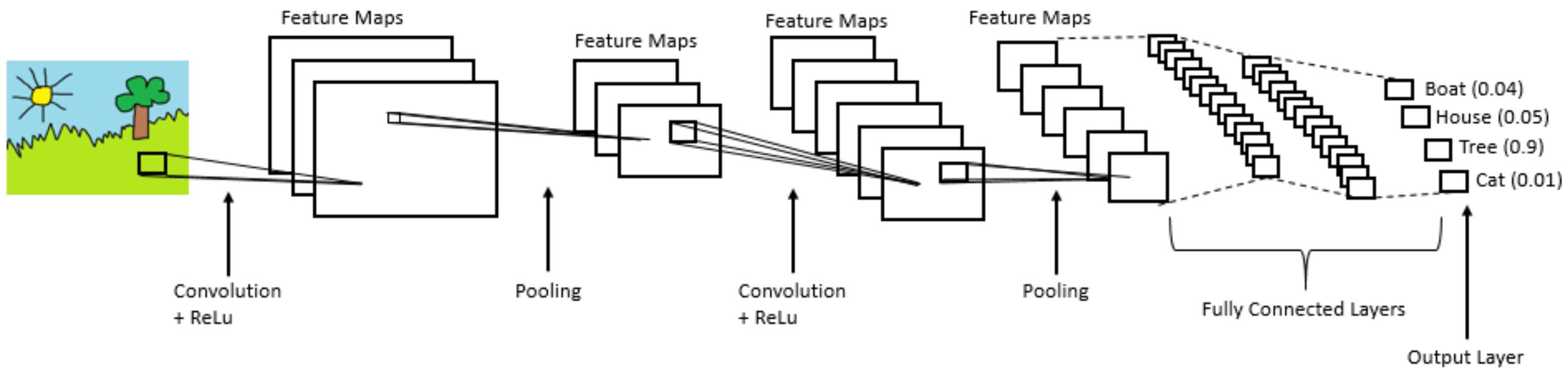
- Pool Layer thực hiện chức năng làm giảm chiều không gian của đầu và giảm độ phức tạp tính toán
- Max Pooling là được sử dụng nhiều vào phổ biến hơn cả với ý tưởng cũng rất sát với thực tế con người đó là: **Giữ lại chi tiết quan trọng** hay giữ lại **pixel có giá trị lớn nhất**.

Ví dụ: Max pooling với bộ lọc 2×2 và stride = 2. Bộ lọc sẽ chạy dọc ảnh. Và với mỗi vùng ảnh được chọn, sẽ chọn ra 1 giá trị lớn nhất và giữ lại.



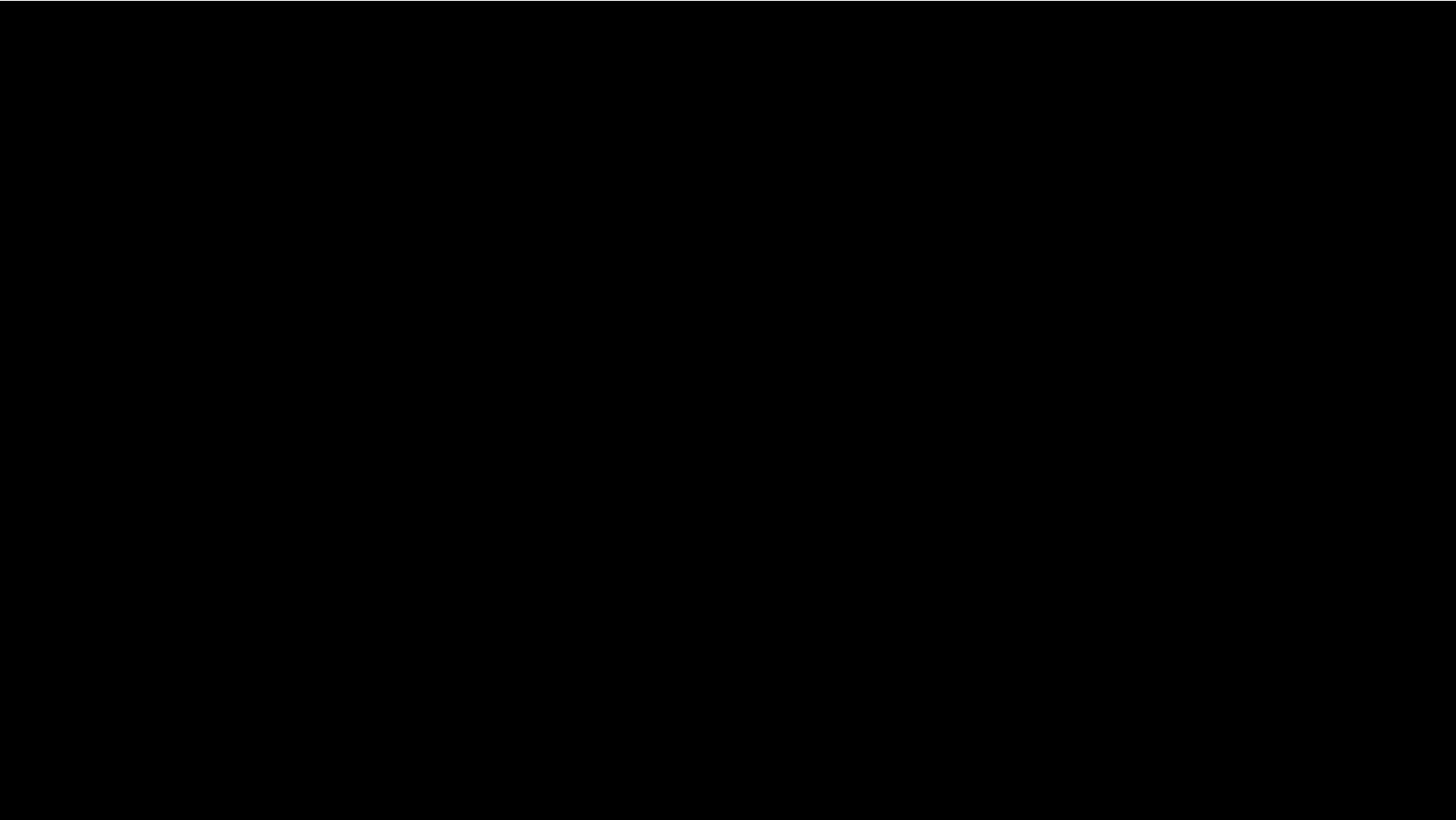
FC Layer ~ Fully Connected Layer

- Là các lớp kết nối đầy đủ như các mạng NN thông thường
- Input: Để đưa ảnh từ các layer trước vào mạng này, buộc phải dàn phẳng bức ảnh ra thành 1 vector thay vì là mảng nhiều chiều như trước.
- Tại layer cuối cùng sẽ sử dụng 1 hàm softmax để phân loại đối tượng dựa vào vector đặc trưng đã được tính toán của các lớp trước đó.



Complete CNN architecture

CNN vs Multilayer Perceptron



- Một nơ ron trong các layer kết nối với tất cả nơ ron trong layer trước đó
- Số lượng weights khi xử lý ảnh thấp, tiết kiệm tài nguyên
- Có các layer đặc biệt: Conv Layer và Pooling Layer

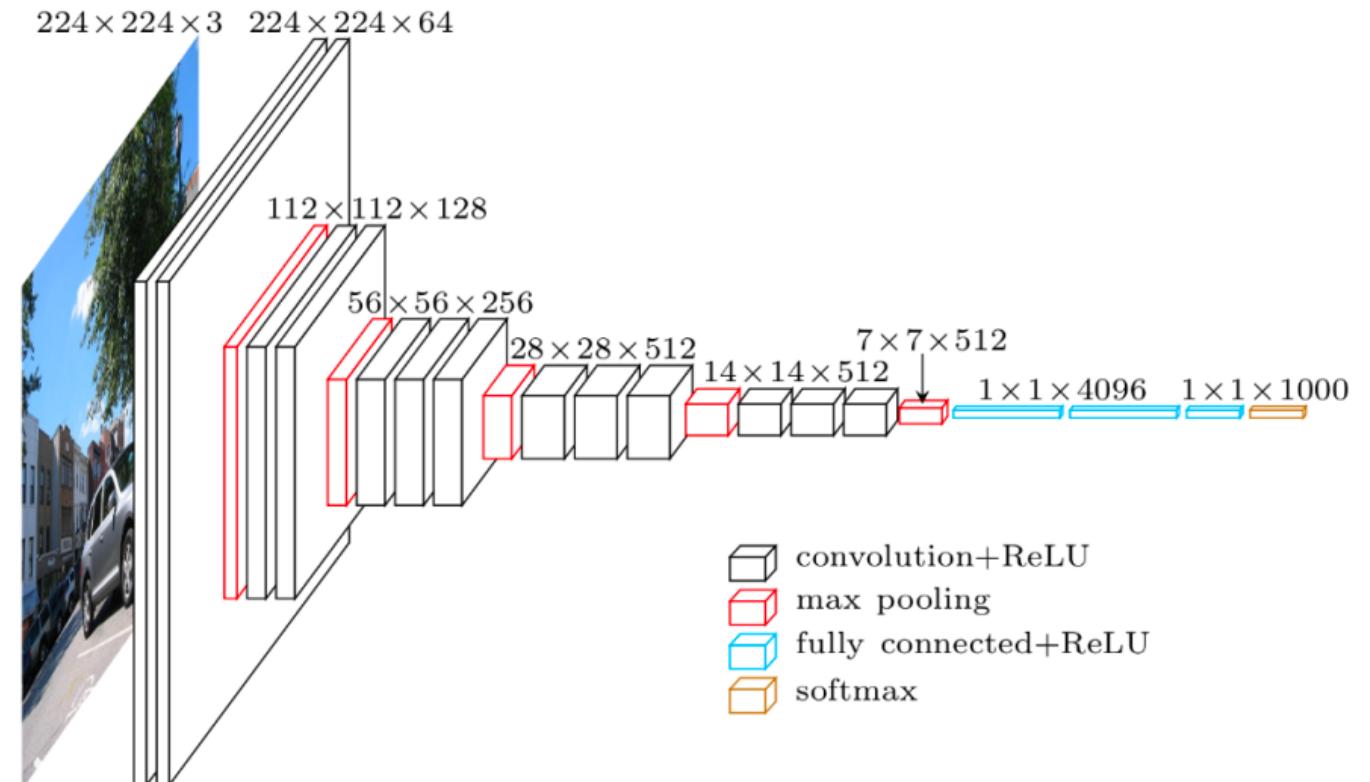
- Một nơ ron trong một layer chỉ kết nối tới một vùng nhỏ của layer trước đó
- Số lượng weights khi xử lý ảnh nhiều, tốn kém tài nguyên
- Không có các layer đặc biệt của CNN

Một số mạng CNN nổi tiếng

VGG-16 (hiện đã có VGG-19)

Được phát triển năm 2014, là một biến thể sâu hơn nhưng lại đơn giản hơn so với kiến trúc convolution (từ gốc: convolutional structure) thường thấy ở CNN. Kiến trúc như hình dưới, có thể thấy số mắc dù các tần cao hơn được đơn giản hóa so với LeNet, AlexNet ... thu gọn hơn.

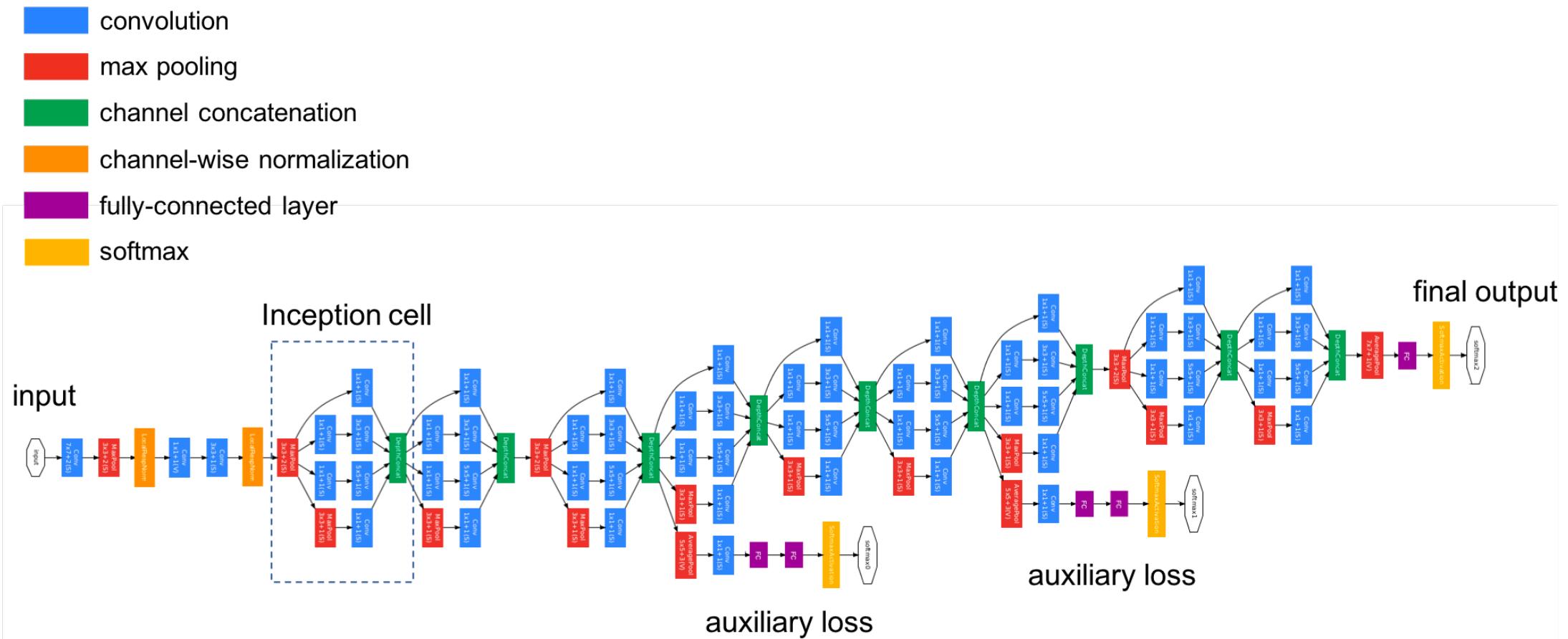
Số lượng tham số: 138.000.000



Một số mạng CNN nổi tiếng

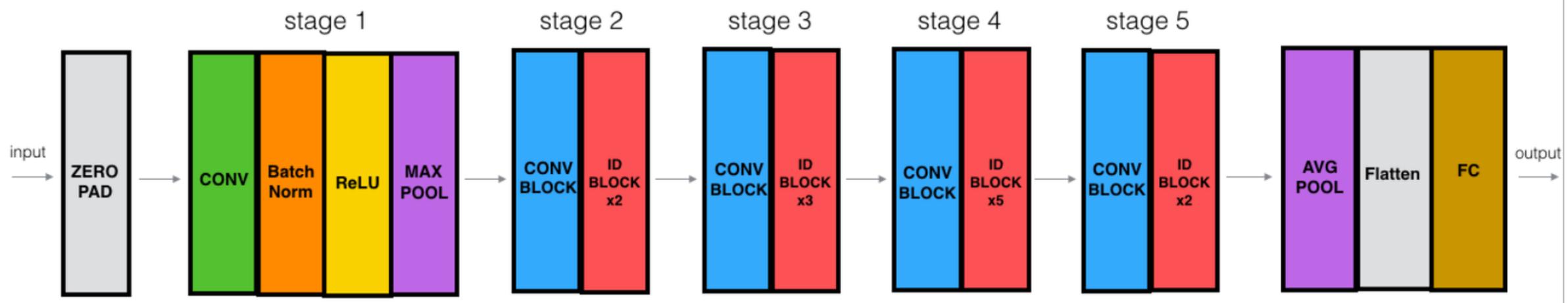
Inception (GoogleLeNet)

Vào năm 2014, các nhà nghiên cứu của google đã đưa ra mạng Inception -> tham dự cuộc thi ImageNet 2014.



Một số mạng CNN nổi tiếng

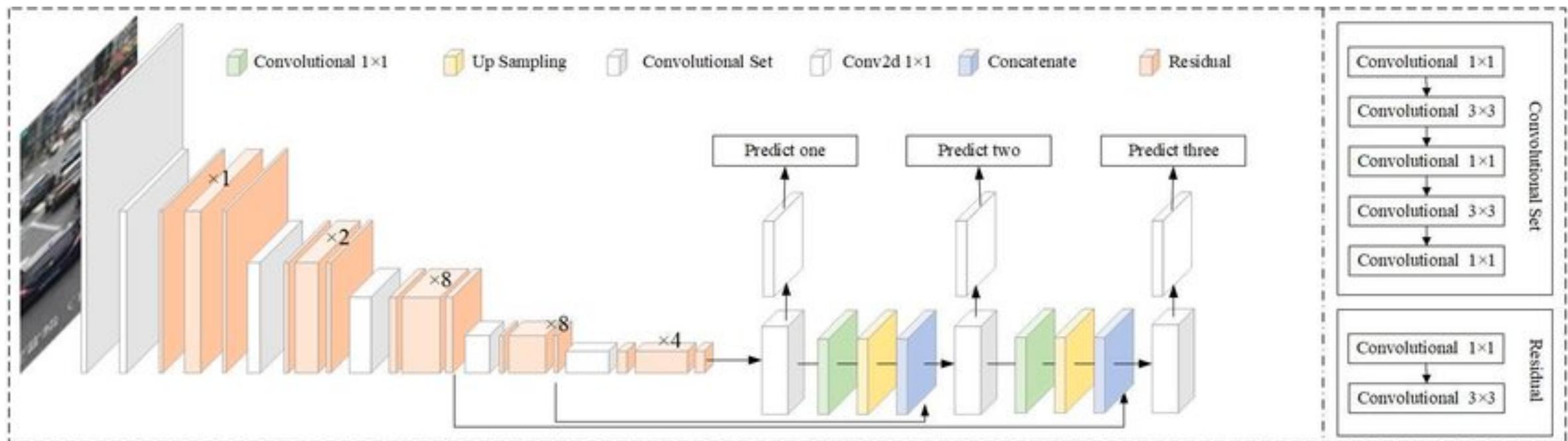
ResNet (viết tắt của residual network), là mạng học sâu nhận được quan tâm từ những năm 2012 sau cuộc thi [LSVRC2012](#) và trở nên phổ biến trong lĩnh vực thị giác máy. ResNet khiến cho việc huấn luyện hàng trăm thậm chí hàng nghìn lớp của mạng [nơ ron](#) trở nên khả thi và hiệu quả



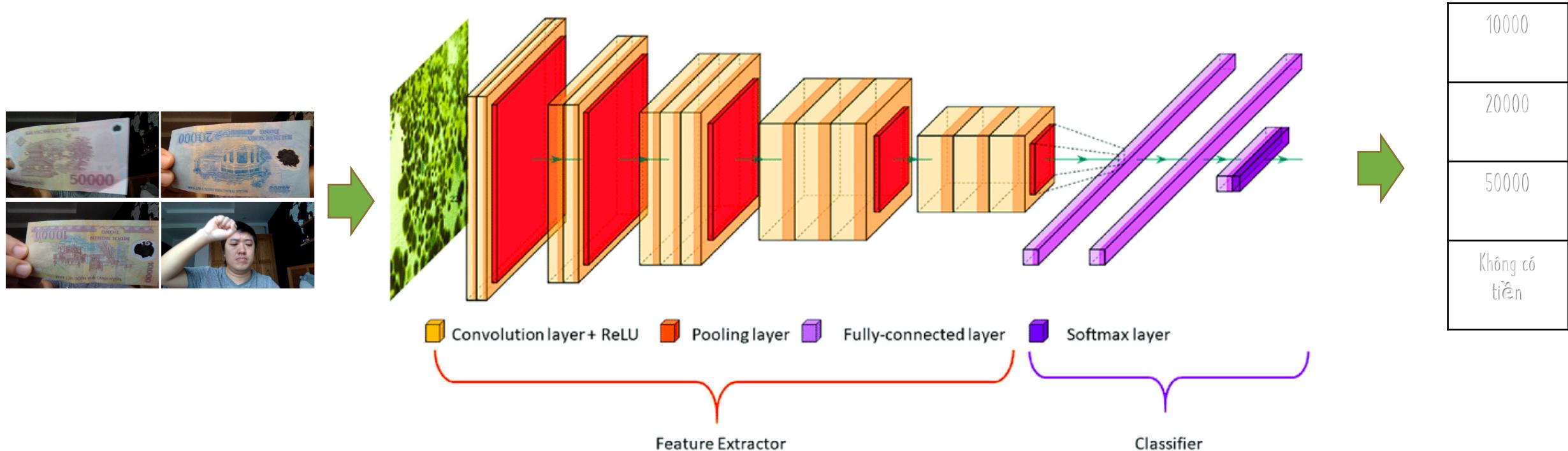
Một số mạng CNN nổi tiếng

Darknet-53

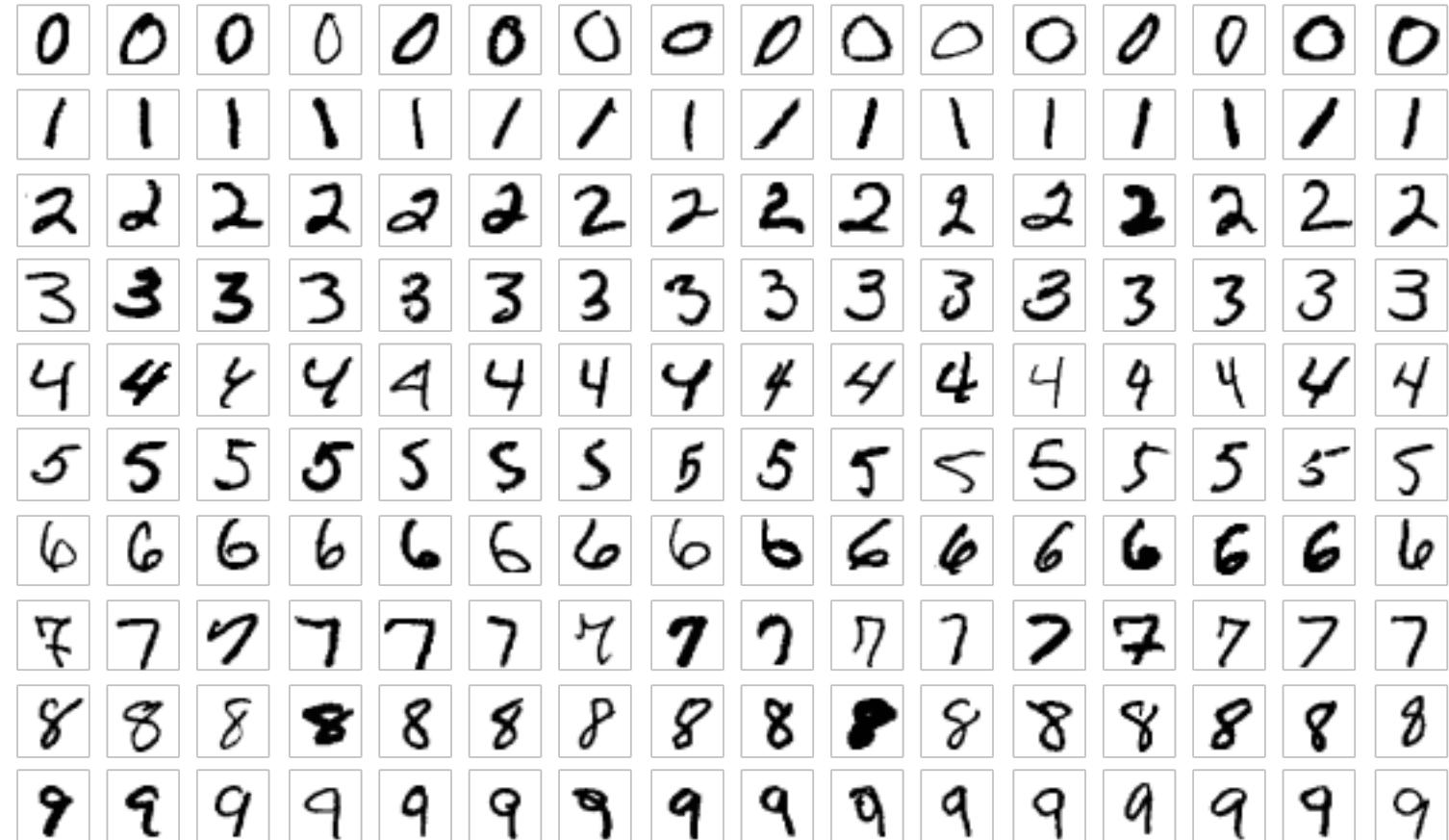
Là mạng backbone của YOLO v3



Bài toán nhận diện tiền



Bài toán nhận diện chữ số trong bộ MNIST



Bài toán nhận diện chữ số trong bộ MINST

