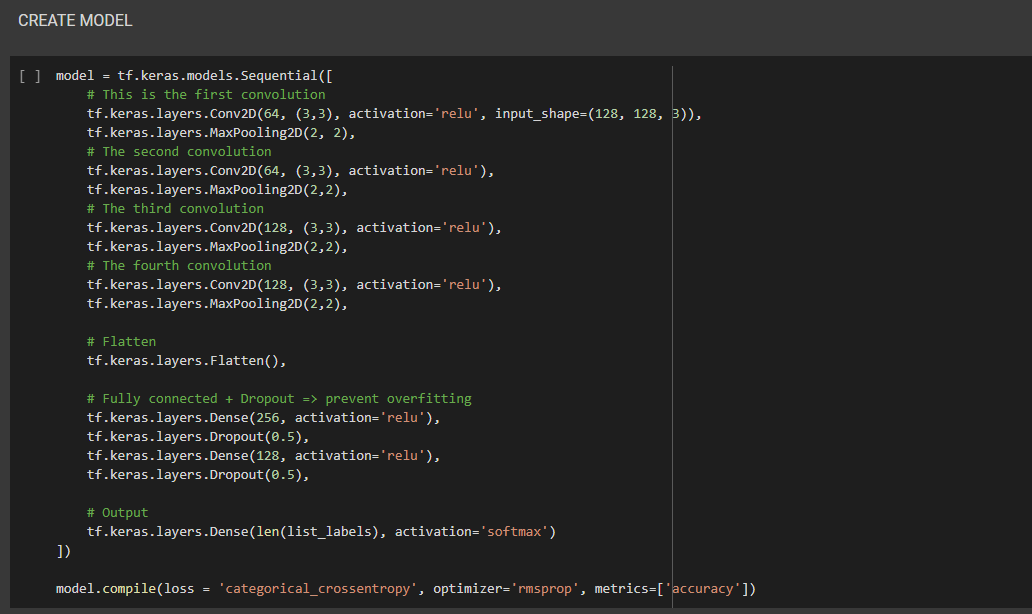
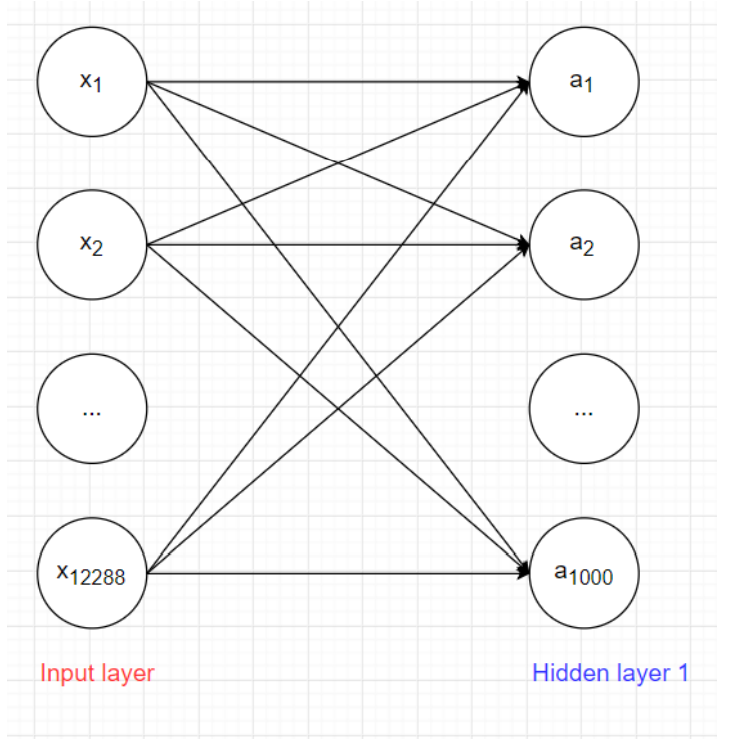
—------------------------------------------------------------------------------------------------------------------



**Lý do chọn CNN cho bài toán classification:**

* **Vấn đề của fully connected neural network với xử lý ảnh (Lớp DENSE trong model):**

Lấy ví dụ về bài toán xử lý ảnh, thì ảnh 64\*64 được biểu diễn dướ dạng 1 tensor 64\*64\*3 => để biểu thị hết nội dung của bức ảnh thì cần truyền vào input layer tất cả các pixel (64\*64\*3 = 12288). Nghĩa là input layer giờ có 12288 nodes.

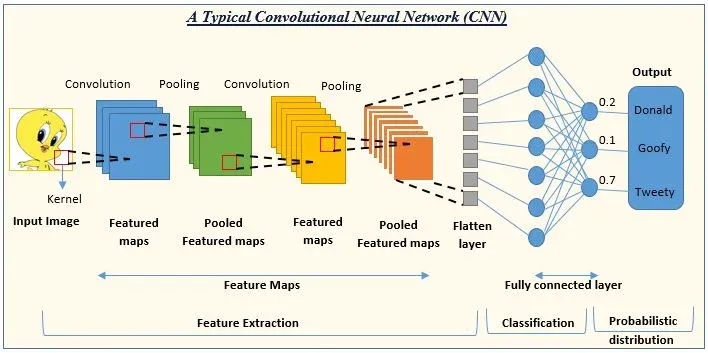


Với việc thể hiện lớp input layer và hidden layer1 như trên thì số lượng tham số đã lên tới 12288 \* 1000 +1000 . Trong khi model còn có thể có nhiều layer nữa => kích thước ảnh tăng => parameter tăng nhanh => Mạng nơ ron trở nên quá đồ sộ, gây khó khăn cho việc tính toán => CẦN GIẢI PHÁP. => CONVOLUTIONAL LAYER

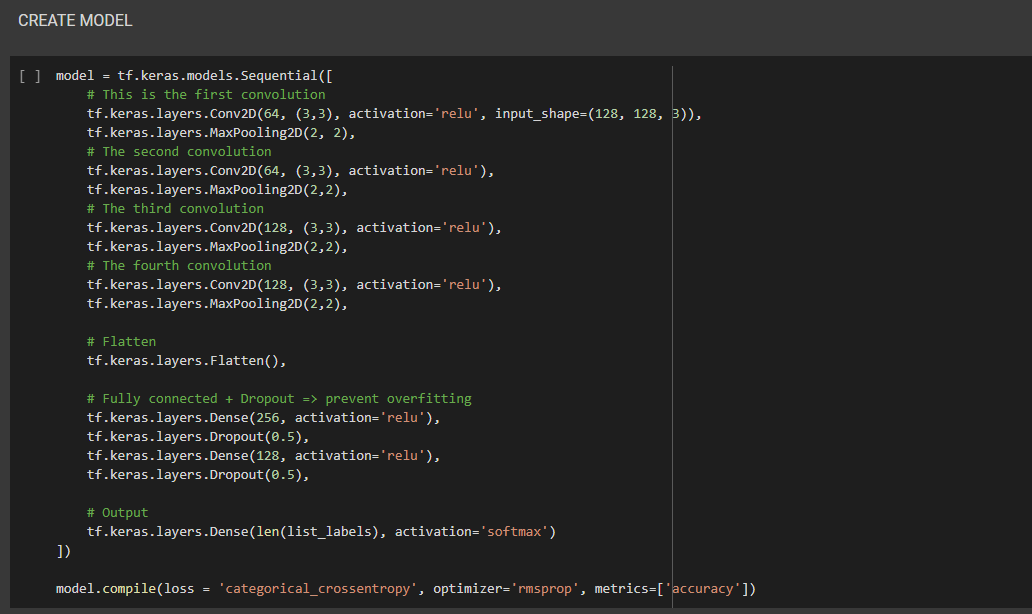
**CONVOLUTIONAL LAYER:**

Thông tin của các pixel thường chỉ chịu tác động bởi các pixel ngay gần nó, vậy nên việc bỏ qua một số nút ở tầng đầu vào trong mỗi lần huấn luyện sẽ không làm giảm độ chính xác của mô hình. Vậy nên người ta sử dụng cửa số tích chập nhằm giải quyết vấn đề số lượng tham số lớn mà vẫn trích xuất được đặc trưng của ảnh.

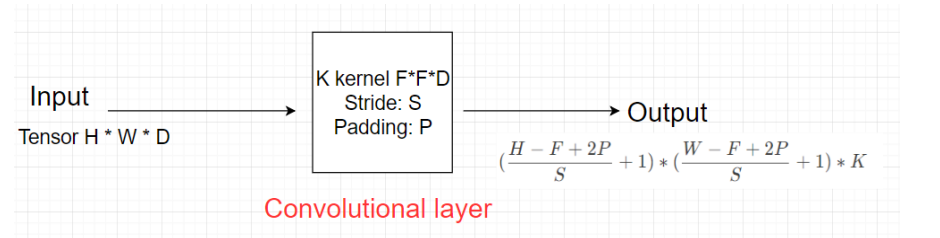
Về mặt kỹ thuật, trong mô hình học sâu CNN, mô hình ảnh đầu vào sẽ chuyển nó qua một loạt các lớp tích chập với các bộ lọc, sau đó đến lớp Pooling, rồi tiếp theo là các lớp được kết nối đầy đủ (FC — fully connected layers) và cuối cùng áp dụng hàm softmax để phân loại một đối tượng dựa trên giá trị xác suất trong khoản từ 0 đến 1.

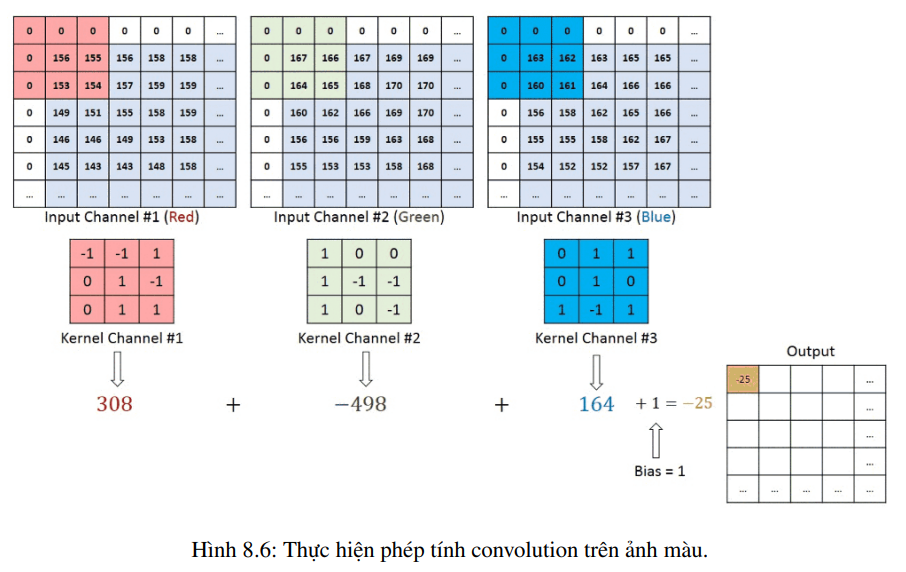


**Mô hình:**

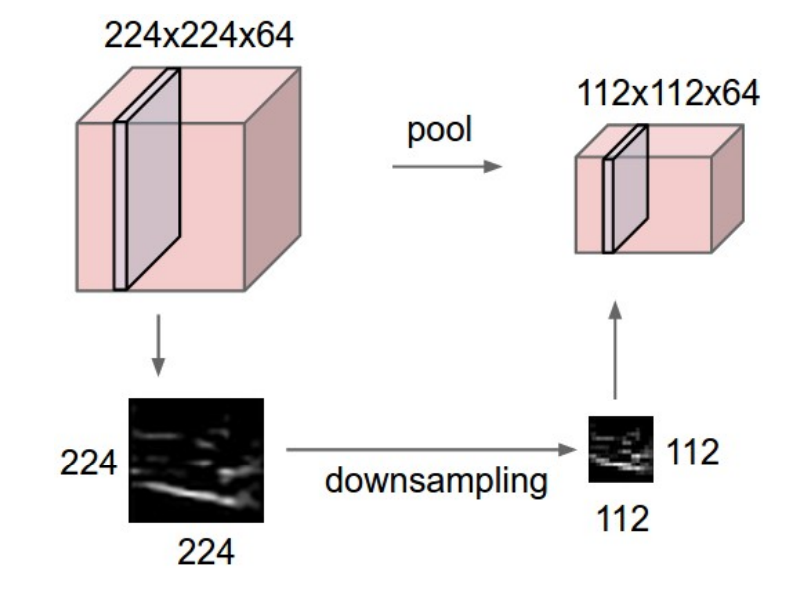


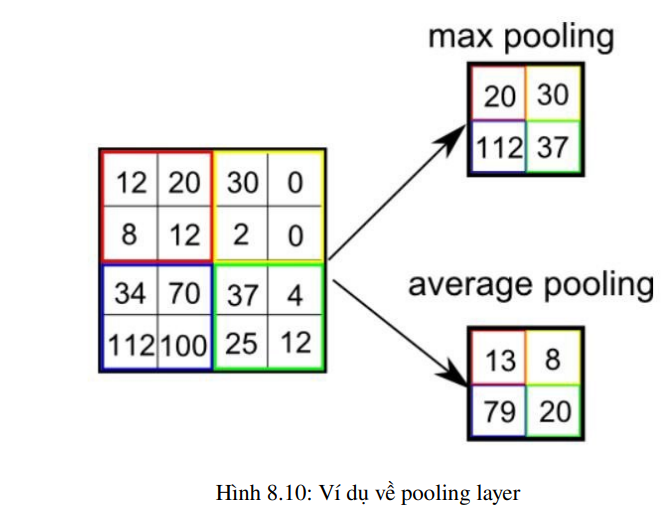
**Convolution layer:** kích thước kernel là 3\*3, mặc định stride = 1, padding = 1 => input từ các layer này sẽ cho width và height bằng với input





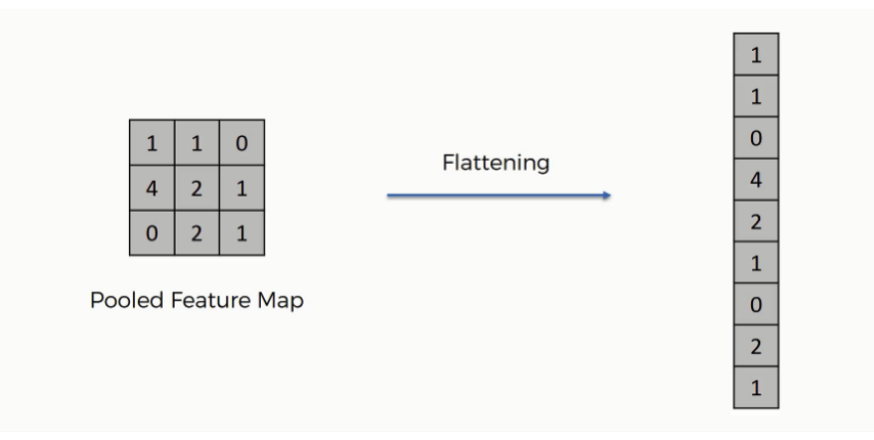
Xen kẽ giữa các Convolution layer là các **MAX Pooling laye**r với size 2\*2 => giảm kích thước dữ liệu nhưng vẫn giữ được các thuộc tính quan trọng. Việc giảm kích thước dữ liệu giúp giảm các phép tính toán trong model.





Số kernel trong mỗi Convolutional sau được thay đổi từ 64 lên 128 , kết hợp với việc giảm chiều dữ liệu bởi Max pooling layer thì output có height và width ngày càng giảm nhưng depth ngày càng tăng.

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 ( lớp Flatten) và đưa vào một lớp được kết nối như một mạng nơ-ron với các Fully connected layer (lớp DENSE)



Xen kẽ giữa các lớp DENSE là các lớp DROPOUT : Dropout là việc bỏ qua các đơn vị (tức là 1 nút mạng) trong quá trình đào tạo 1 cách ngẫu nhiên .<= Khi chúng ta sử dụng full connected layer, các neural sẽ phụ thuộc “mạnh” lẫn nhau trong suốt quá trình huấn luyện, điều này làm giảm sức mạnh cho mỗi neural và dẫn đến bị over-fitting tập train.

Cuối cùng là sử dụng softmax để phân loại đầu ra. (hàm softmax biến vector k chiều có các giá trị thực bất kỳ thành vector k chiều có giá trị thực có tổng bằng 1. Giá trị nhập có thể dương, âm, bằng 0 hoặc lớn hơn 1, nhưng hàm softmax sẽ luôn biến chúng thành một giá trị nằm trong khoảng (0:1].

Như vậy, chúng có thể được gọi là “xác suất)

Model sử dụng hàm kích hoạt (activation function) reLu. ReLU viết tắt của Rectified Linear Unit, là 1 hàm phi tuyến. Với đầu ra là: ƒ (x) = max (0, x).

* Tại sao reLu lại quan trọng: ReLU giới thiệu tính phi tuyến trong ConvNet. Vì dữ liệu trong thế giới mà chúng ta tìm hiểu (image) là các giá trị tuyến tính không âm.

