

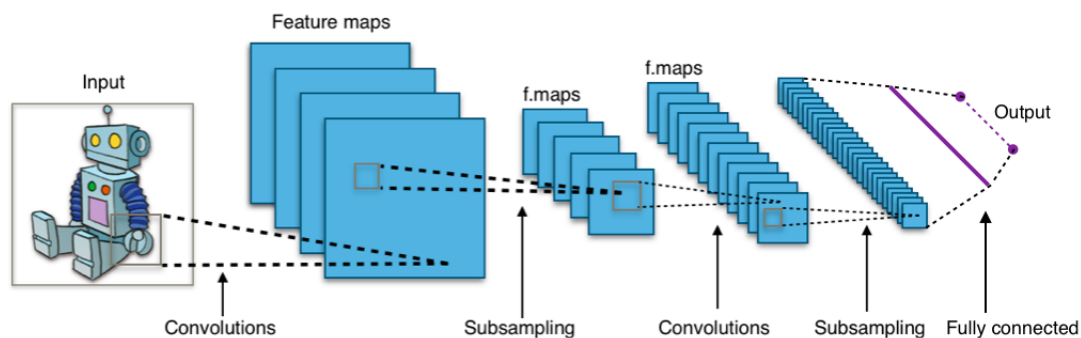
Tìm hiểu về CNN

1. Giới thiệu về CNN

Convolutional Neural Networks (CNN) là một trong những mô hình deep learning phổ biến nhất và có ảnh hưởng nhiều nhất trong cộng đồng Computer Vision. CNN được dùng trong nhiều bài toán như nhận dạng ảnh, phân tích video, ảnh MRI, hoặc cho các bài của lĩnh vực xử lý ngôn ngữ tự nhiên, và hầu hết đều giải quyết tốt các bài toán này

2. Cấu trúc mạng CNN

Mạng CNN là một tập hợp các lớp tích chập chồng lên nhau và sử dụng các hàm nonlinear activation như ReLU và tanh để kích hoạt các trọng số trong các node. Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo.



Convolution Layer

Đây là lớp quan trọng nhất của CNN, lớp này có nhiệm vụ thực hiện mọi tính toán. Những yếu tố quan trọng của một convolutional layer là: stride, padding, filter map, feature map.

- Convolution là một phép nhân element-wise. Sẽ quét một phần của hình ảnh, thường có kích thước là 3×3 và nhân nó với một filter.
-

-
- Stride là khoảng cách giữa 2 kernel khi quét.
 - Padding là các giá trị 0 được thêm vào với lớp input để giữ nguyên kích cỡ của feature map so với ban đầu.
 - Feature map: Nó thể hiện kết quả của mỗi lần filter map quét qua input. Sau mỗi lần quét sẽ xảy ra quá trình tính toán.

Relu Layer

Relu layer là hàm kích hoạt trong neural network và hàm này còn được gọi là activation function. Hàm kích hoạt có tác dụng mô phỏng các nơ ron có tỷ lệ truyền xung qua axon. Trong activation function thì nó còn có hàm nghĩa là: Relu, Leaky, Tanh, Sigmoid, Maxout,...Hiện nay, hàm relu được dùng phổ biến và vô cùng thông dụng.

Nó được sử dụng nhiều cho các nhu cầu huấn luyện mạng neuron thì relu mang lại rất nhiều ưu điểm nổi bật như: việc tính toán sẽ trở nên nhanh hơn,... Quá trình sử dụng relu, chúng ta cần lưu ý đến vấn đề tùy chỉnh các learning rate và theo dõi dead unit. Những lớp relu layer đã được sử dụng sau khi filter map được tính ra và áp dụng hàm relu lên những giá trị của filter map.

Pooling layer

Khi đầu vào quá lớn, những lớp pooling layer sẽ được xếp vào giữa giữa những lớp Convolutional layer để làm giảm parameter,từ đó giảm thời gian tính toán, tránh overfitting.

Fully Connected Layer

Sau khi ảnh được truyền qua nhiều convolutional layer và pooling layer thì model đã học được tương đối các đặc điểm của ảnh.Sau đó ta dùng các fully connected layer để kết hợp các đặc điểm của ảnh để ra được output của model.

3. Ứng dụng của CNN trong Computer Vision

- **Classification:** Đây là nhiệm vụ được biết đến nhiều nhất trong computer vision. Ý tưởng chính là phân loại nội dung chung của hình ảnh thành một tập hợp các danh mục, được gọi là nhãn.
-

-
- **Localization:** Mục đích chính của localization là tạo ra một hộp giới hạn mô tả vị trí của đối tượng trong hình ảnh. Đầu ra bao gồm một nhãn lớp và một hộp giới hạn.
 - **Detection:** Nhiệm vụ này bao gồm thực hiện localization trên tất cả các đối tượng trong ảnh. Các đầu ra bao gồm nhiều hộp giới hạn, cũng như nhiều nhãn lớp.
 - **Segmentation:** Nhiệm vụ ở đây là xuất ra cả nhãn lớp và đường viền của mỗi đối tượng hiện diện trong hình ảnh. Điều này chủ yếu được sử dụng để đánh dấu các đối tượng quan trọng của hình ảnh cho phân tích sâu hơn

4. Minh họa mạng CNN đơn giản

```
✓ class Net(nn.Module):  
✓     def __init__(self):  
        super().__init__()  
        self.conv1 = nn.Conv2d(3, 16, kernel_size=3, padding=1)  
        self.conv2 = nn.Conv2d(16, 8, kernel_size=3, padding=1)  
        self.fc1 = nn.Linear(8 * 8 * 8, 32)  
        self.fc2 = nn.Linear(32, 10)  
  
✓     def forward(self, x):  
        out = F.max_pool2d(torch.tanh(self.conv1(x)), 2)  
        out = F.max_pool2d(torch.tanh(self.conv2(out)), 2)  
        # flatten về dạng vector để cho vào neural network  
        out = out.view(-1, 8 * 8 * 8)  
        out = torch.tanh(self.fc1(out))  
        out = self.fc2(out)  
        return out
```