**Rede neural convolucional**

Uma rede neural convolucional (CNN) é uma arquitetura de rede para aprendizado profundo que aprende diretamente dos dados. CNNs são particularmente úteis para encontrar padrões em imagens para reconhecer objetos. Eles também podem ser bastante eficazes para classificar dados que não sejam de imagem, como áudio, séries temporais e dados de sinal.

A arquitetura de uma rede neural convolucional é formada por uma sucessão de blocos de construção que extraem as características que discriminam a classe pertencente à imagem de outros. Um bloco de construção consiste em uma ou mais:

Diagrama

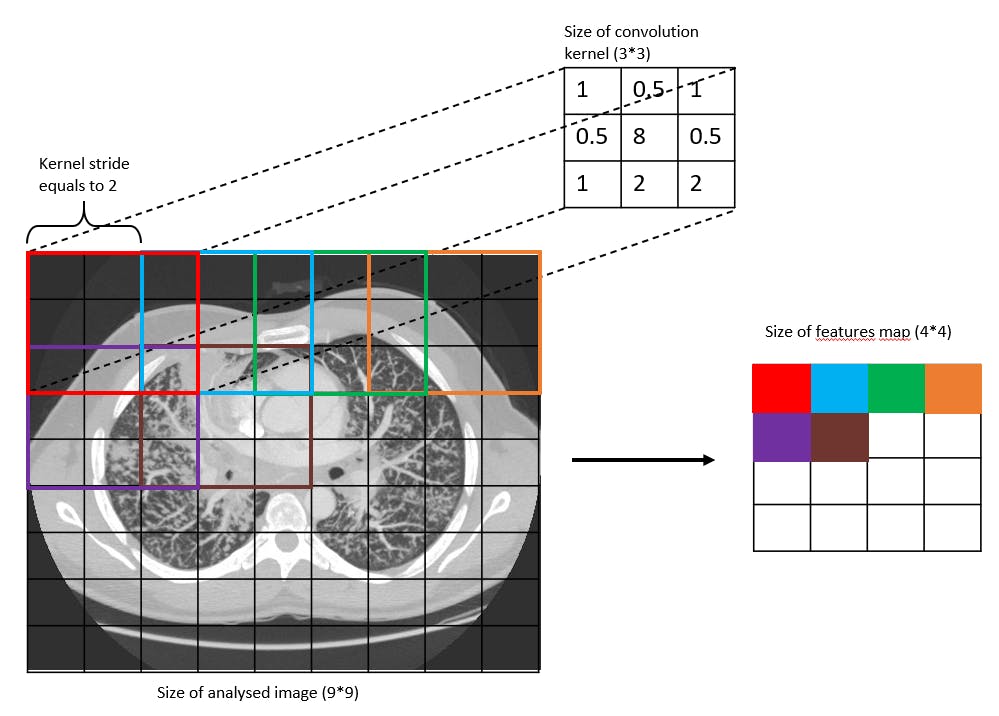
Descrição gerada automaticamente

**Camada convolucional**

Essa camada é a primeira camada usada para extrair os vários recursos das imagens de entrada, usamos filtro ou método Kernel para a extração. Cada kernel convolucional possui parâmetros específicos para as informações que são pesquisadas na imagem

**Kernel**

O kernel é o filtro de redes neurais que se move pela imagem, escaneando cada pixel e convertendo os dados em um formato menor ou, às vezes, maior.



Nota: Quando a imagem for RGB, o filtro deverá ter 3 canais (Vermelho , Verde e Azul ), sendo representada como uma matriz 3D (tensor).(height, width, 3). Isso ocorre porque uma imagem RGB possui 3 canais de cores e filtros de 3 canais são necessários para fazer os cálculos.

**Stride**

Em seguida, fazemos outro cálculo movendo o filtro na imagem horizontalmente um passo para a direita. O número de passos (pixels) que deslocamos o filtro sobre a imagem de entrada é chamado de **Stride**

Tela de celular

Descrição gerada automaticamente com confiança média

Se usarmos Stride=2, o tamanho será ainda mais reduzido. Se houver várias camadas convolucionais na CNN, o tamanho do mapa de recursos será reduzido ainda mais no final, de modo que não possamos fazer outras operações no mapa de recursos. Para evitar isso, usamos aplicar Padding à imagem de entrada.

**Padding**

É um termo relevante para redes neurais convolucionais, pois se refere ao número de pixels adicionados a uma imagem quando ela está sendo processada pelo kernel de uma CNN.

O preenchimento é simplesmente um processo de adicionar camadas de zeros às nossas imagens de entrada para evitar os problemas mencionados acima, isso evita o encolhimento, e aumenta a contribuição dos pixels na borda da imagem original, trazendo-os para o meio da imagem preenchida. Assim, as informações nas bordas são preservadas, bem como as informações no meio da imagem.

Diagrama

Descrição gerada automaticamente

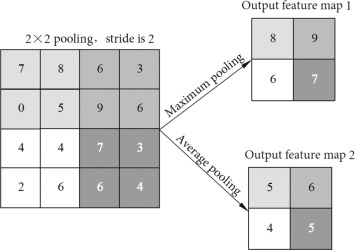
**Camada de agrupamento**

As camadas de agrupamento são o segundo tipo de camada usado em uma CNN. Pode haver várias camadas de agrupamento em uma CNN. O principal objetivo dessa camada é diminuir o tamanho do mapa de recursos convolvidos para reduzir os custos computacionais.   
Isso é realizado diminuindo as conexões entre as camadas e operando independentemente em cada mapa de recursos.

Dependendo do método usado, existem vários tipos de operações de pooling. Temos o agrupamento máximo e o agrupamento médio.

**Max pooling**: Obtenha o valor máximo na área onde o filtro é aplicado.

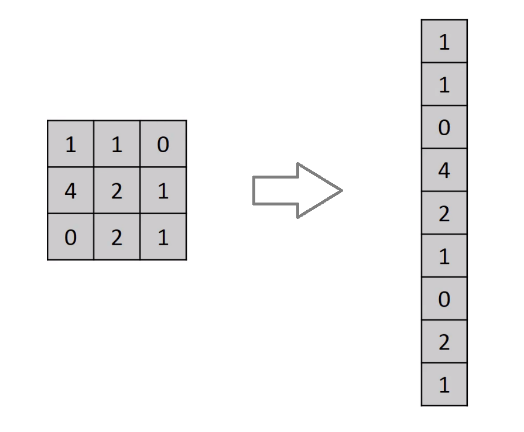
**Pooling médio**: Obtenha a média dos valores na área onde o filtro é aplicado



A operação de pooling mais comum é a máxima: **MaxPool**. É mais eficiente do que a média porque maximiza o peso de ativações fortes

**Flatten (nivelamento)**

O nivelamento é usado para converter todas as matrizes bidimensionais resultantes de mapas de recursos agrupados em um único vetor linear contínuo longo. A matriz achatada é alimentada como entrada para a camada totalmente conectada para classificar a imagem.



**Camada totalmente conectada (FC)**

A camada Fully Connected (FC) consiste nos pesos e bias junto com os neurônios e é usada para conectar os neurônios entre duas camadas diferentes.   
Essas camadas geralmente são colocadas antes da camada de saída e formam as últimas camadas de uma arquitetura CNN.

