

Módulo 4 | Aula #3

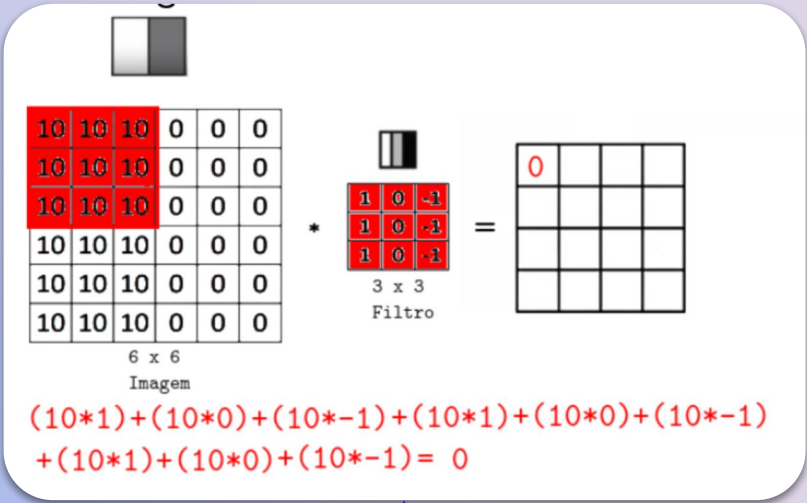
Detecção de objetos.
Selective search.
Análise da arquitetura R-CNN

Tempo estimado de leitura:
5 min

O que são as CNNs?

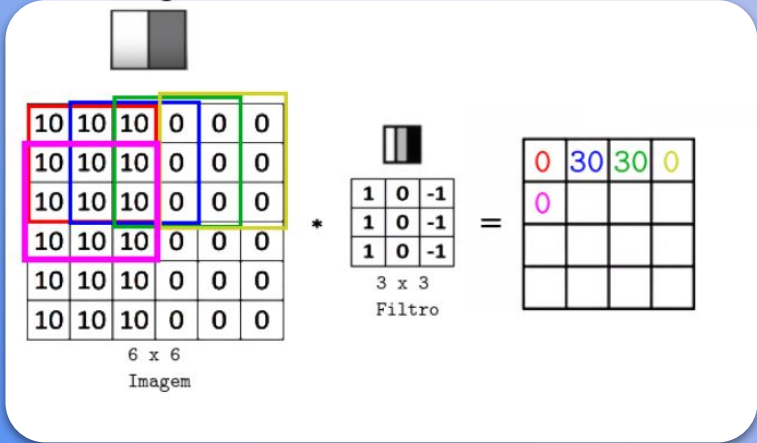
As **CNNs (Redes Neurais Convolucionais)** foram criadas por Yann LeCun, em 1998, e elas são amplamente utilizadas para tarefas que envolvem imagens e vídeos.

Estas redes são treinadas para capturar as características das imagens através da aplicação de filtros. Esta aplicação dos filtros acontece pelo processo de convolução, em que **sobreposomos os filtros nos pixels das imagens e aplicamos uma determinada função para extrair as características requeridas.**



Seguimos com esse processo até passarmos o filtro em todos os pixels da imagem.

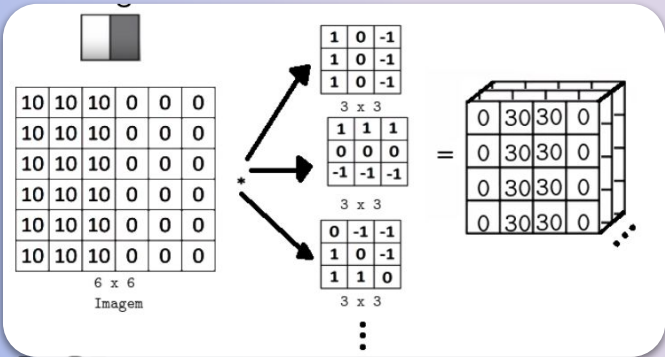
O resultado é a **imagem convoluída**.



As CNNs são compostas por diversos filtros, dessa forma, conseguimos capturar diferentes características de uma única imagem para que a rede aprenda sobre ela.

Na imagem que usamos como exemplo anteriormente, o filtro em questão verifica a existência e localização de bordas verticais na imagem.

A figura a seguir, ilustra o resultado da convolução após a aplicação de diversos filtros em uma imagem:

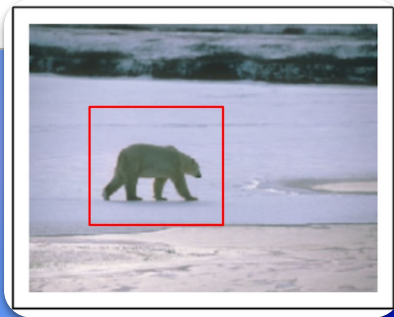


Uma das aplicações das CNNs é na **tarefa de detecção de objetos**. Utilizamos esta tarefa para identificar e localizar objetos de determinadas instâncias nas imagens, tais como: humanos, animais, carros, etc.

Os métodos de detecção de objetos realizam, basicamente, os seguintes passos:

- O método reconhece se há ou não o objeto que estamos procurando na imagem. Se existir, retorna a localização espacial do objeto.

Veja um exemplo do resultado do processo de detecção de ursos:



Muitos métodos de detecção utilizam classificação para avaliar a presença ou não dos objetos. Para isso, as redes extraem as características das imagens e estas características são utilizadas na comparação com características que estamos procurando.

Quer um exemplo para entender melhor este processo?

Precisamos identificar os rostos das pessoas em uma imagem. Inicialmente, o algoritmo precisa de um conjunto significativo de imagens contendo rostos para treinamento.

A partir de cada imagem são extraídas as suas características. Então, o classificador aprende as características dos rostos e depois, identifica em determinadas imagens se temos aquelas características aprendidas ou não.

Feito isso, se tiver rostos em uma determinada imagem o algoritmo irá realizar o processo de localização desses objetos para mostrar onde estão na imagem.