



FIAP

Deep Learning - Redes Neurais Convolucionais – CNN

O que temos para hoje?

FIAP



E como trabalhar com Imagens?



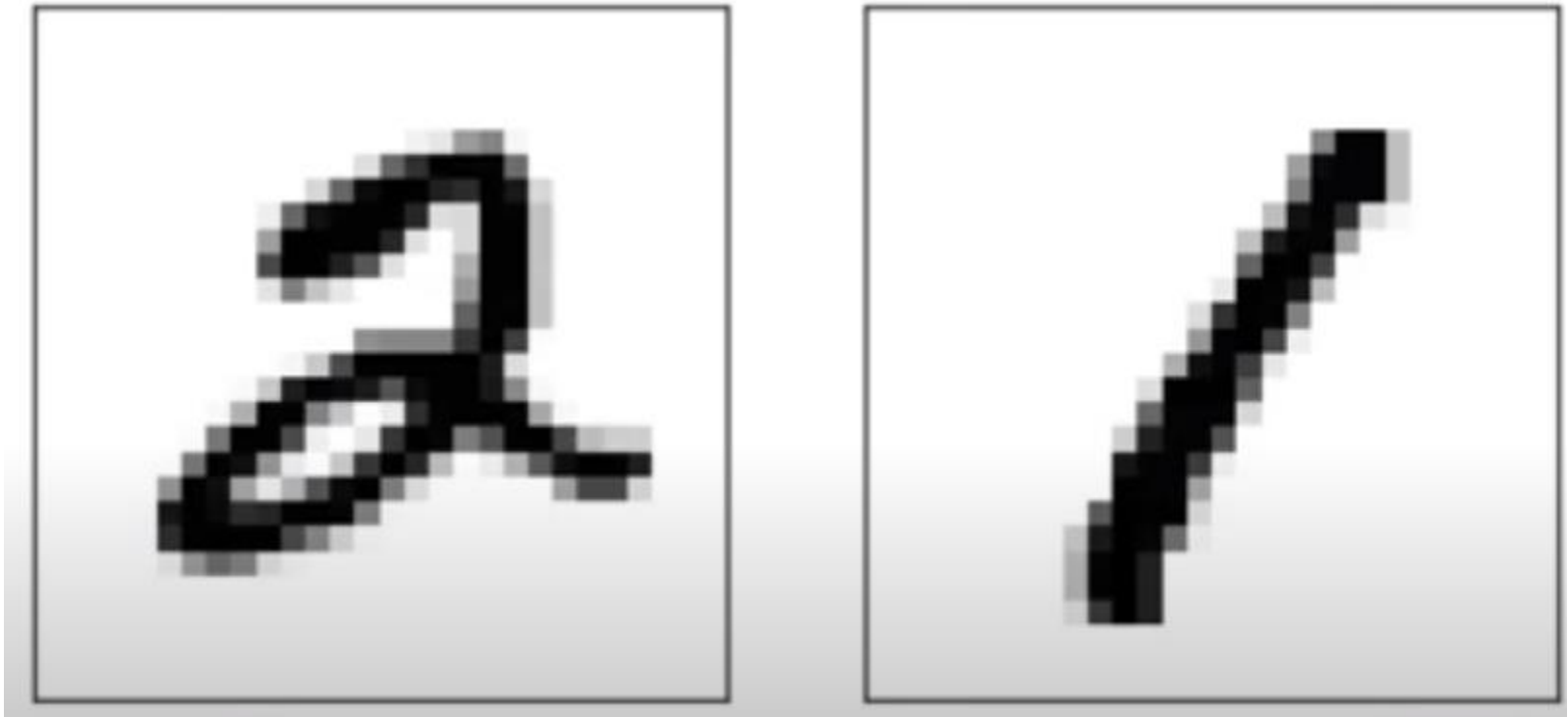
MNIST Dataset
Fonte: Wikipedia

Características do MNIST

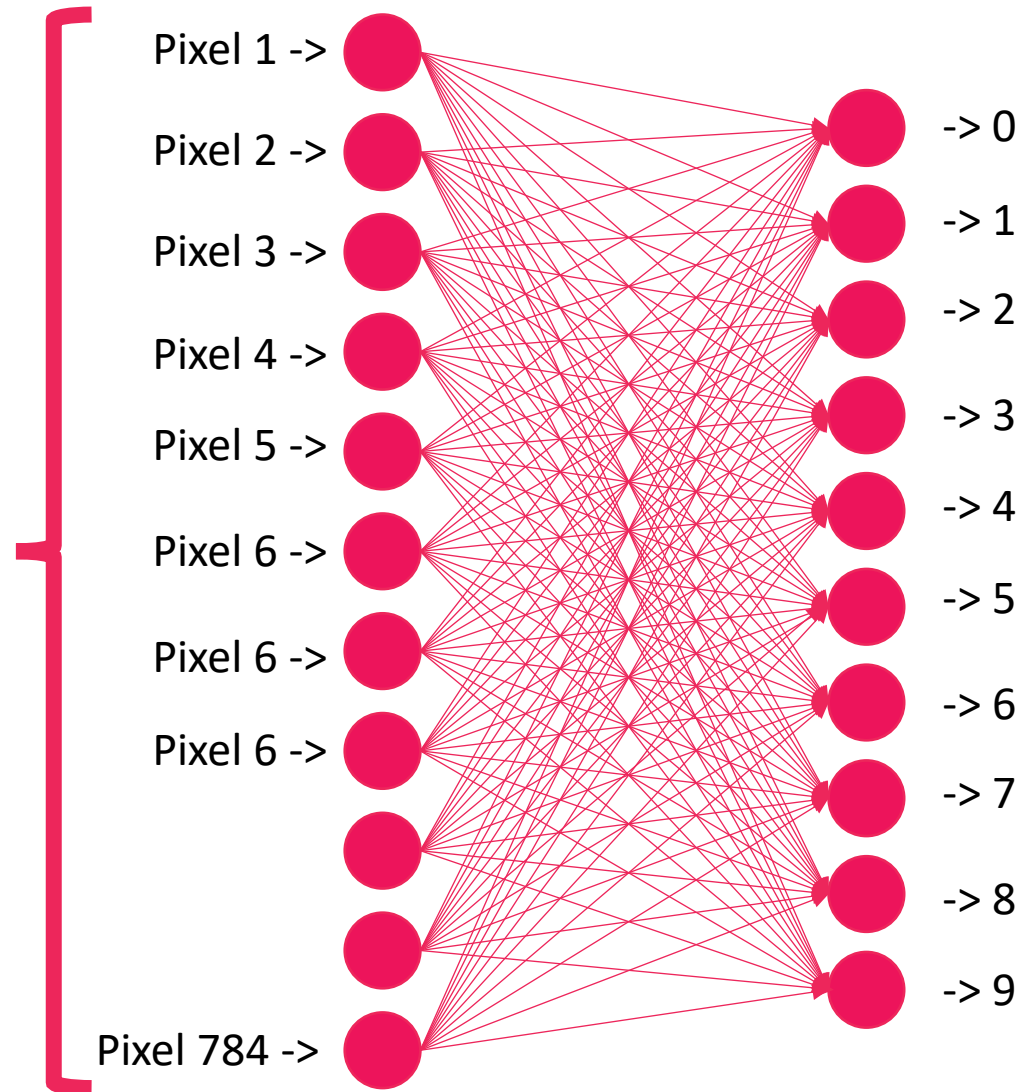
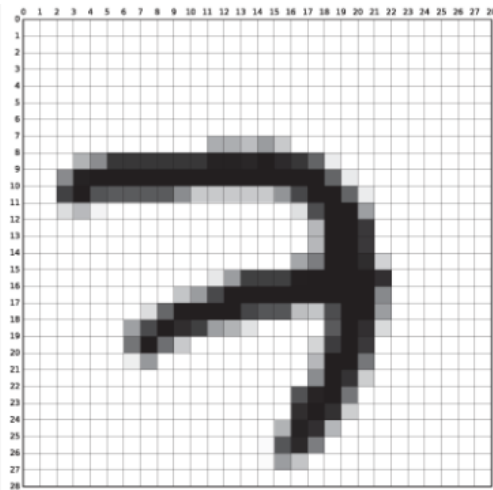


Como diferenciar os números?

FIA/P



E como aplicar numa RNA?



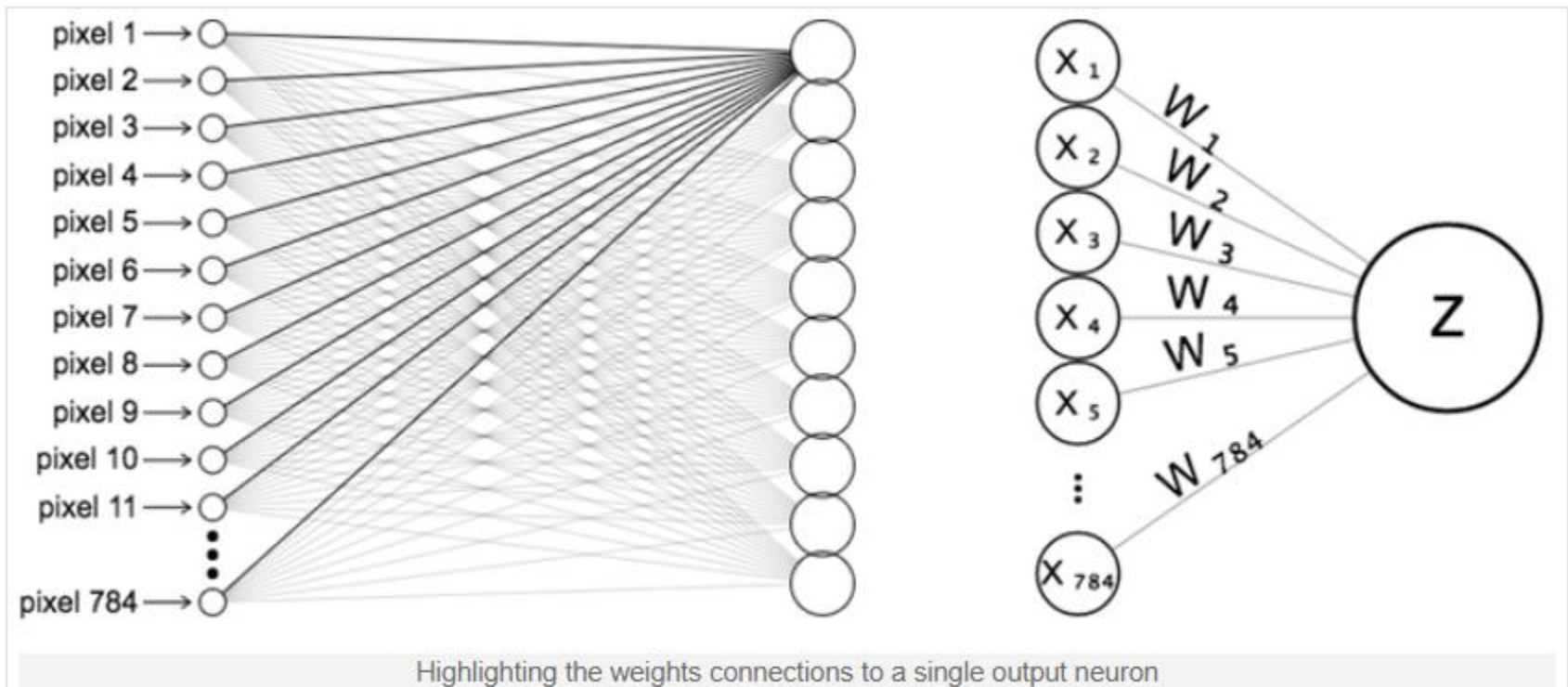
o MNIST, Modified National Institute of Standards and Technology database. É um dataset muito comum utilizado para introdução do assunto de visão computacional e processamento de imagens. Trata-se de um dataset com vários dígitos escritos a mão, de várias formas diferentes.

O objetivo deste laboratório é criar um modelo de **Rede Neural Multicamadas** para identificar os números

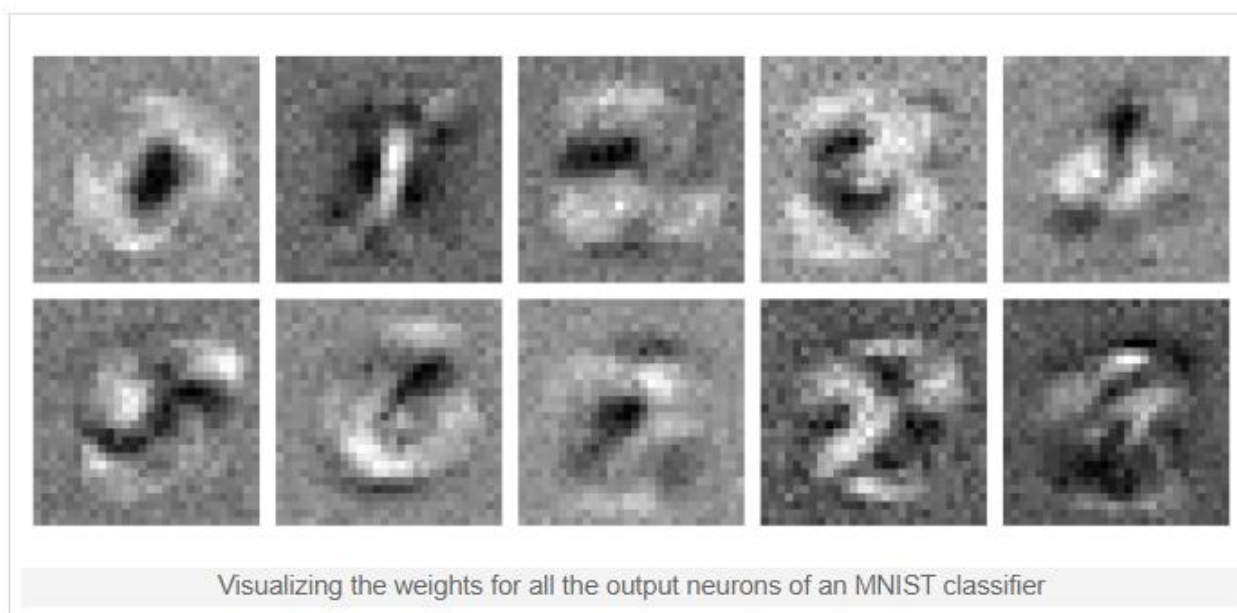
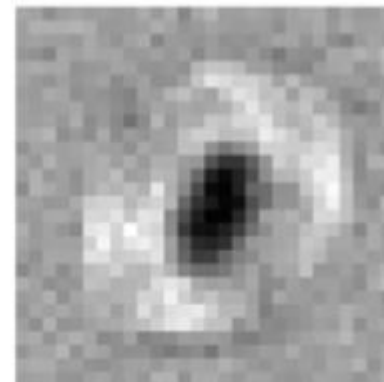
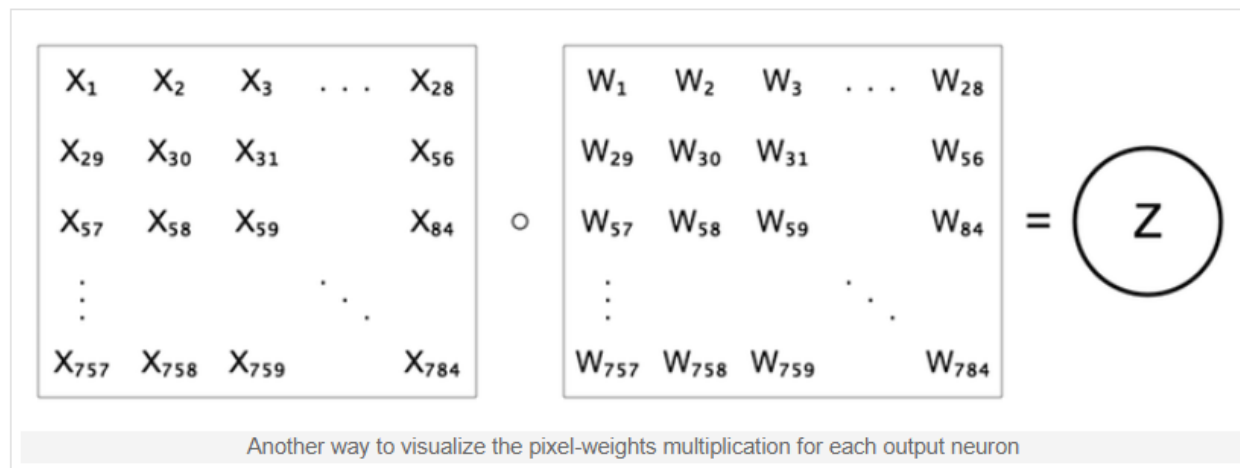


Spoiler: Como queremos classificar os números escritos em 10 algarismos (classes), trata-se de um problema de **Classificação!**

E o que aconteceu?

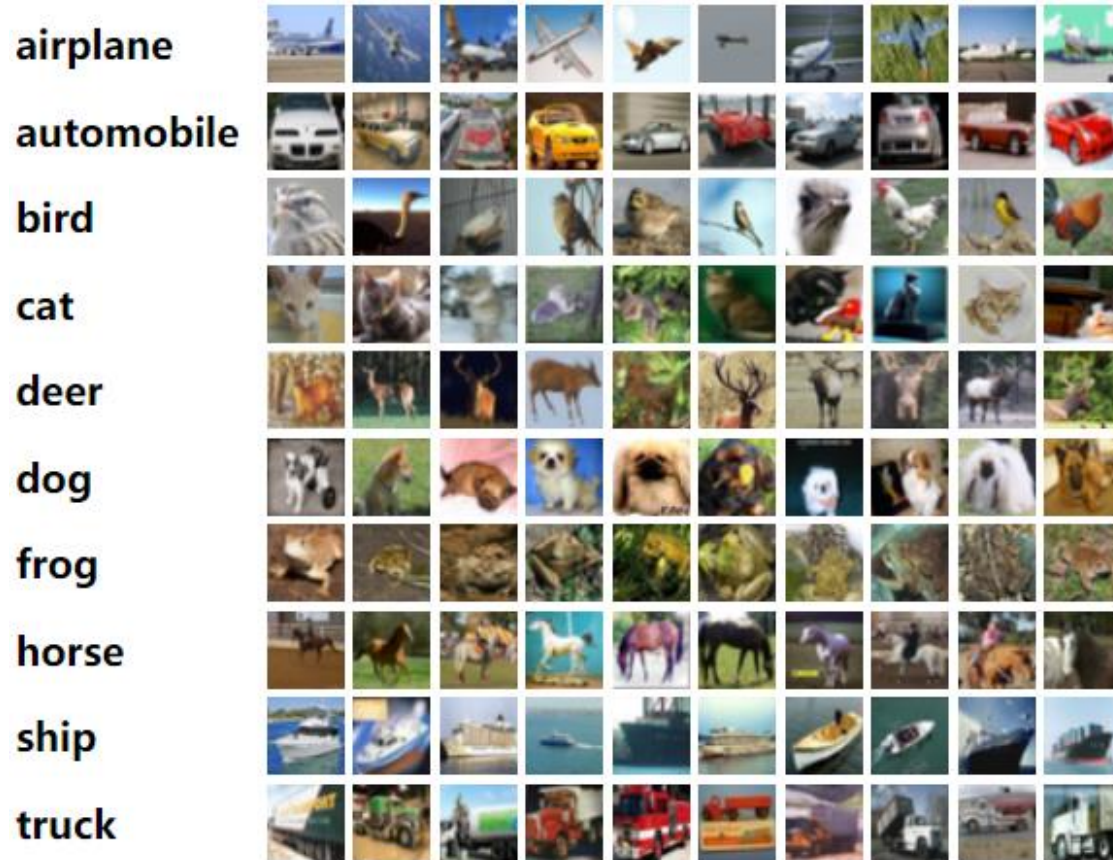


E o que aconteceu?



Visualizing the weights for all the output neurons of an MNIST classifier

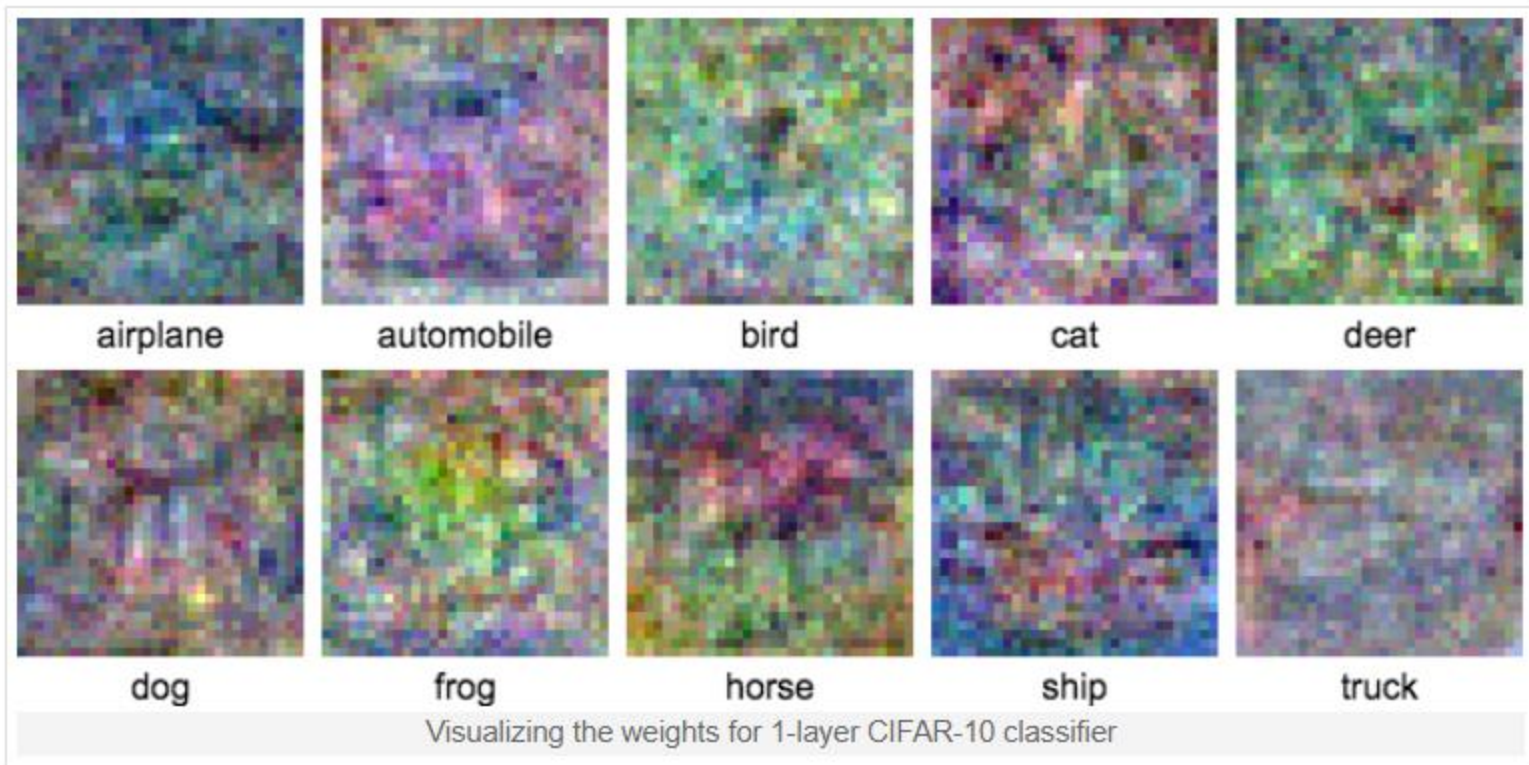
E se colocarmos cor nisso?



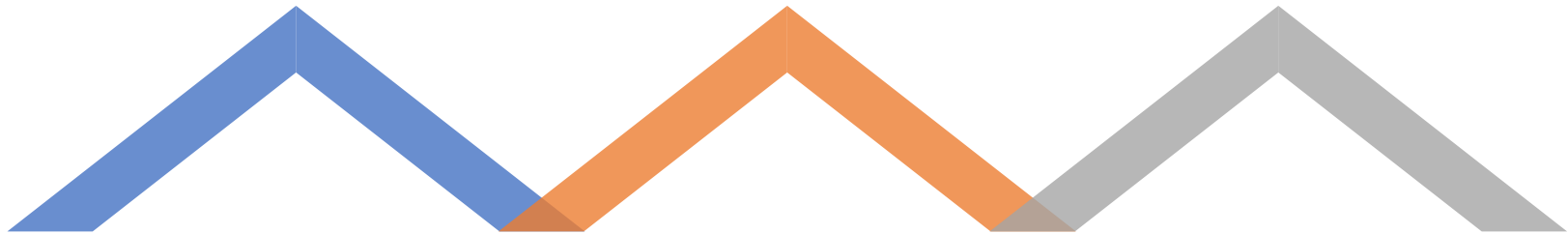
CIFAR-10

- 60.000 Imagens coloridas;
- 32 x 32 pixels;
- 10 classes independentes
- Classes balanceadas

Acontece isso...



Por que?



Fazer com que cada pixel da imagem seja um **atributo** ignora a dependência entre pixels vizinhos!

As características de contorno, profundidade e textura são percebidas por um **conjunto de pixels próximos** um do outro.

A inclusão de cores força a pensarmos em 3 Dimensões – **RGB!**

Existem duas alternativas para lidar com isso



Extrair as características de contorno, profundidade e textura através de técnicas de processamento de sinais... MUITO DIFÍCIL



Aprender a extrair as características através de parâmetros... MENOS DIFÍCIL

Redes Neurais Convolucionais (CNN) FIAP



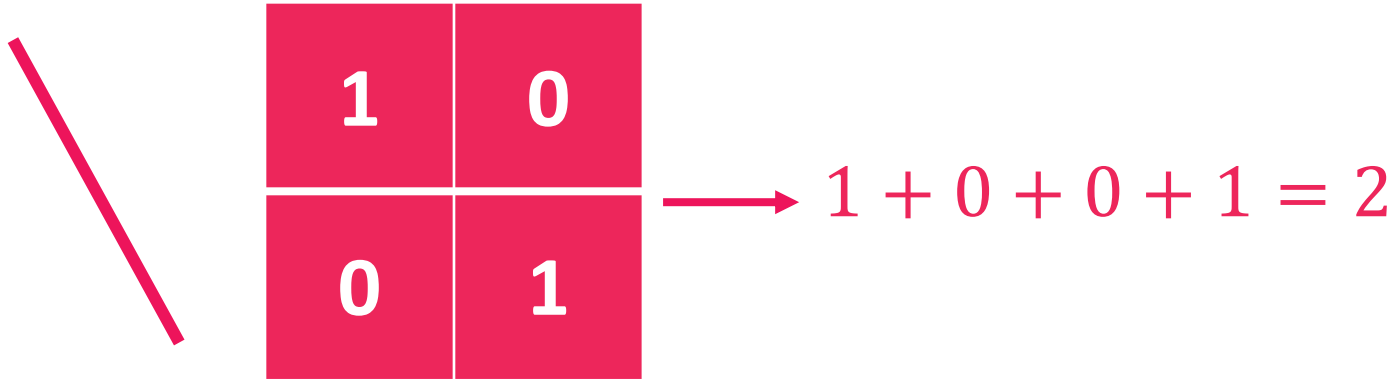
As CNN (Convolutional Neural Networks) são usadas para **reconhecimento visual**

Baseiam-se nos conceitos conhecidos sobre a **visão humana**;

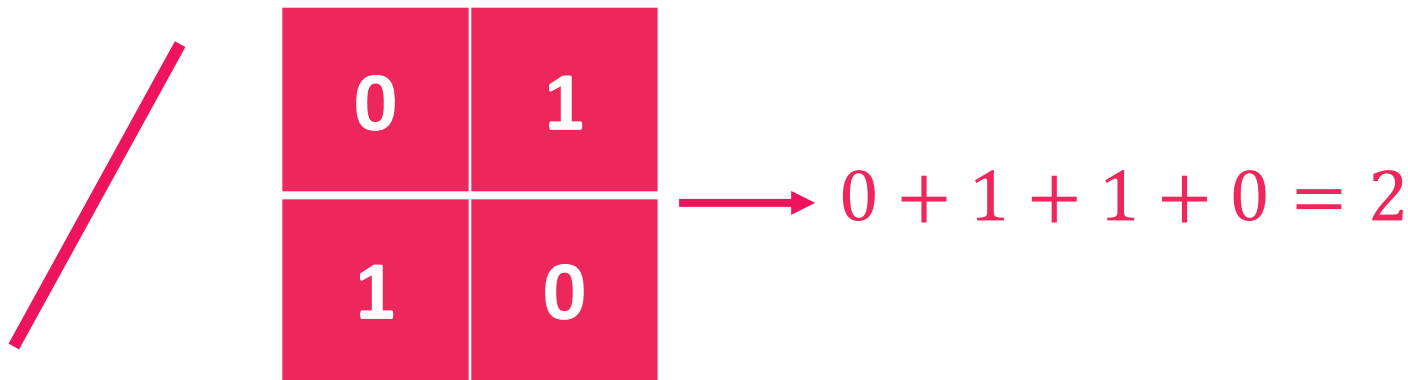
Considera a relação entre os pixels vizinhos para determinar contornos, texturas e profundidades;

A relação é estipulada através de **filtros**, que são aplicados através de um processo chamado **convolução**.

E o que são filtros?

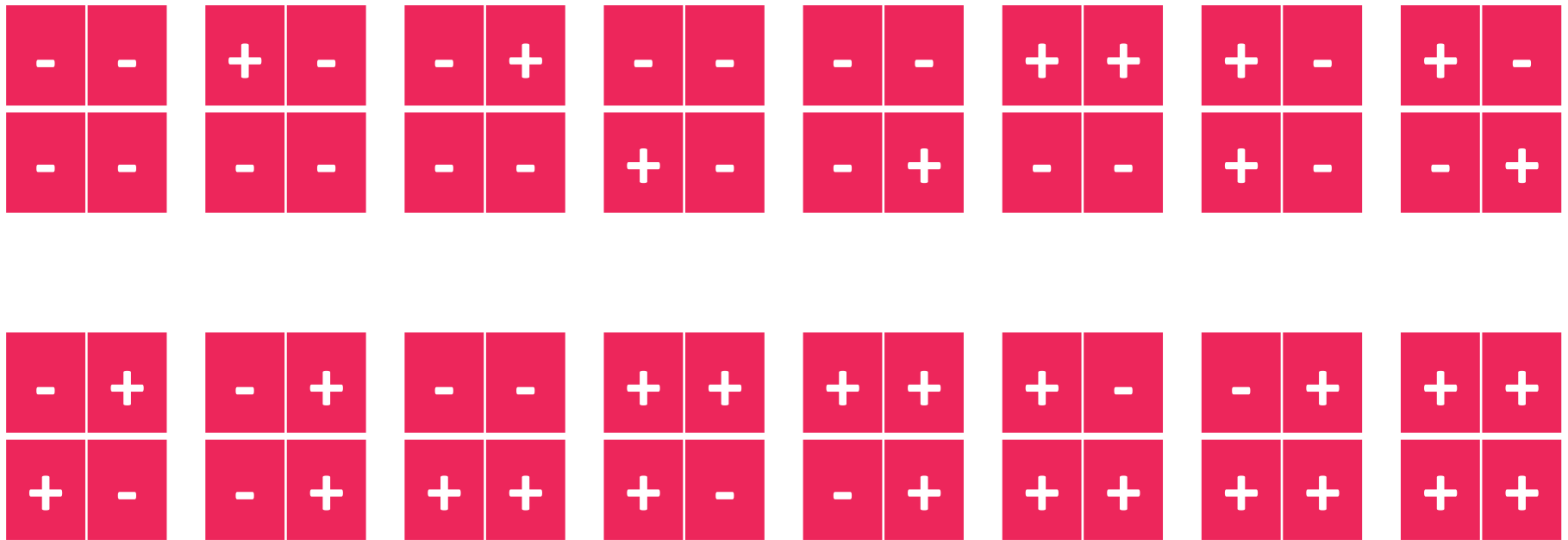


O que acham desse filtro?



E o que são filtros?

❖ Para uma imagem 2x2, apenas com “-” e “+” = $2^4 = 16$ possibilidades!



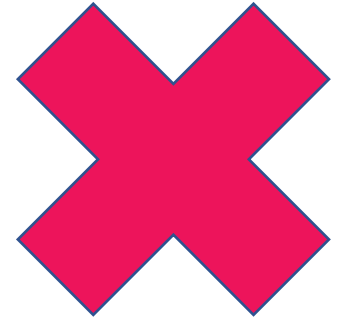
Voltando para CNN

FILAP



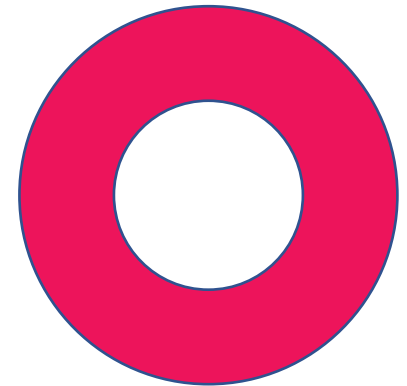
1	-1	-1
-1	1	-1
-1	-1	1

1	-1	1
-1	1	-1
1	-1	1

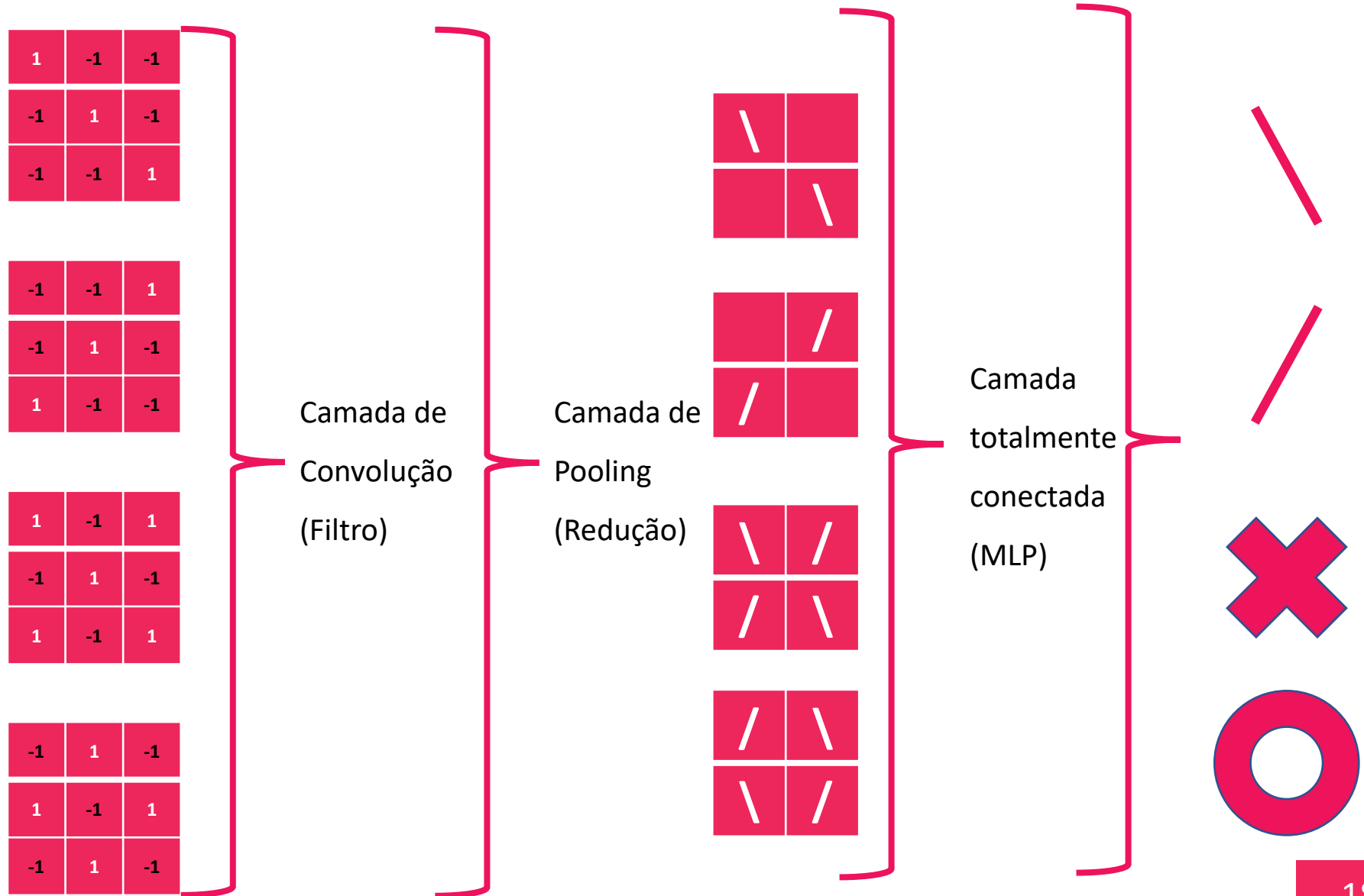


-1	-1	1
-1	1	-1
1	-1	-1

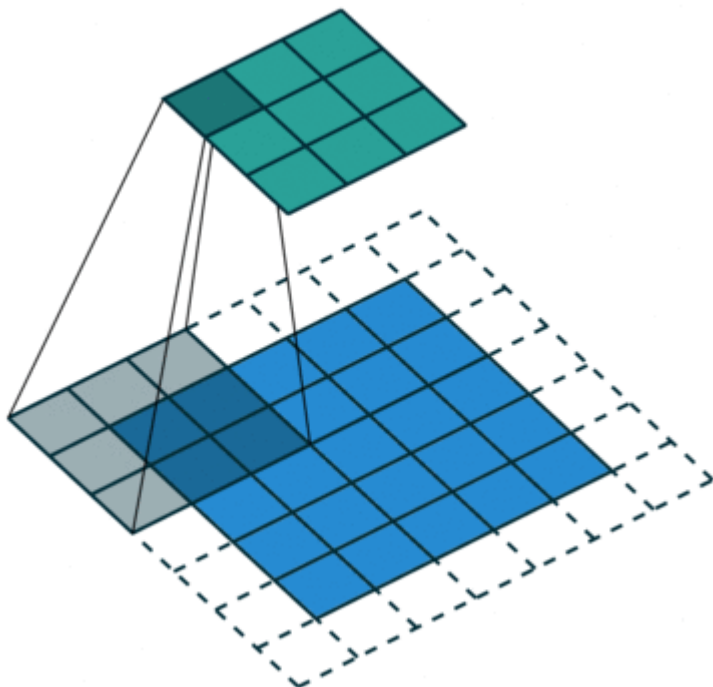
-1	1	-1
1	-1	1
-1	1	-1



Voltando para CNN



Camada de Convolução



Fonte: <https://giphy.com/gifs/blog-daniel-keypoints-i4NjAwytgIRDW>

1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

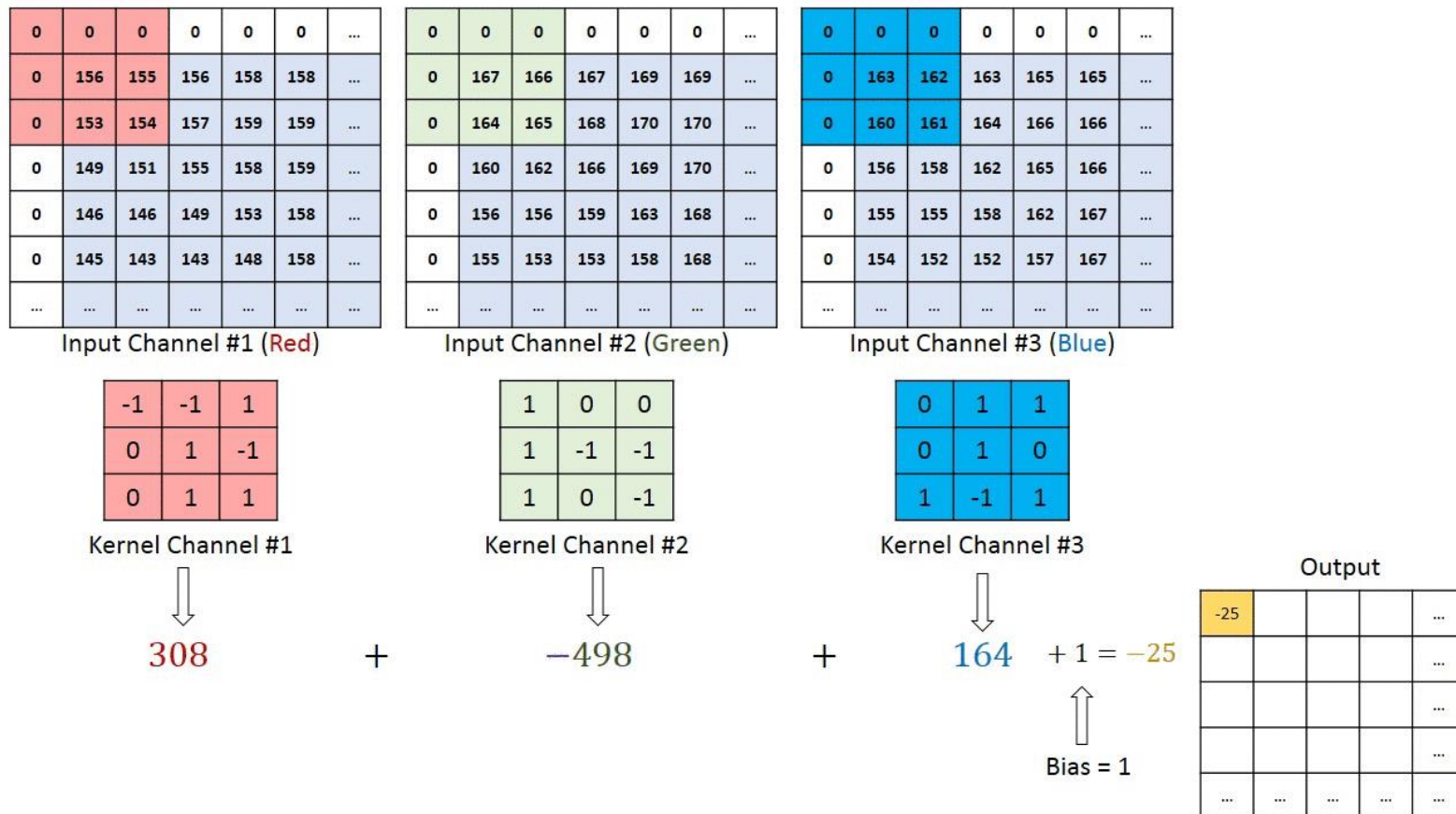
Image

4		

Convolved
Feature

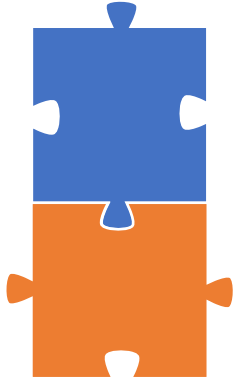
Fonte: https://commons.wikimedia.org/wiki/File:Convolution_arithmetic_-_Padding_strides.gif

Camada de Convolução em imagens $FILP$ coloridas

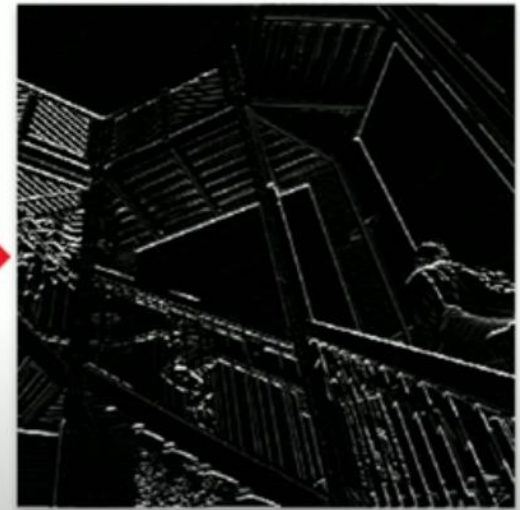
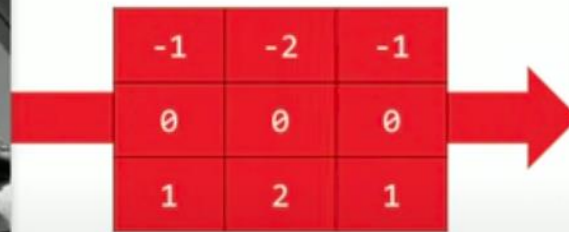


Fonte: https://miro.medium.com/v2/resize:fit:1280/1*ciDgQEjViWLnCbmX-EeSrA.gif

Mas para que serve isso?



- Serve para achar bordas nas imagens
- Serve para evidenciar alguma característica na imagem, aplicando filtros de Blur e Sharpen

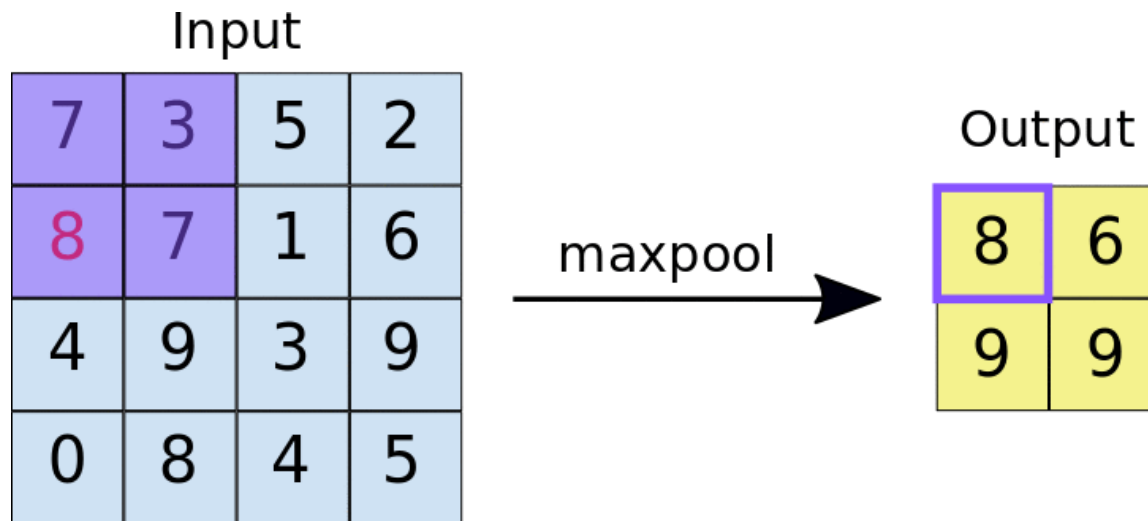


Camada de Pooling

❖ Usada para reduzir a dimensão da imagem e conseguir discrimina-la;

Exemplos:

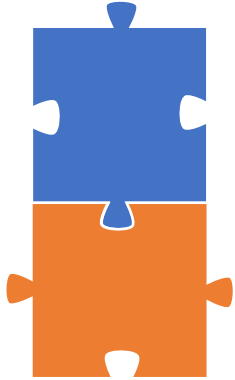
❖ Max Pooling



Fonte: <https://nico-curti.github.io/NumPyNet/NumPyNet/images/maxpool.gif>

❖ Outros exemplos: Min Pooling e Average Pooling

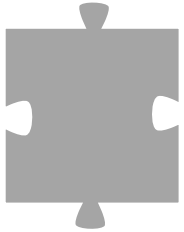
Mas para que serve isso?



- Diminui o custo computacional da operação
- Visa extrair as características principais da imagem

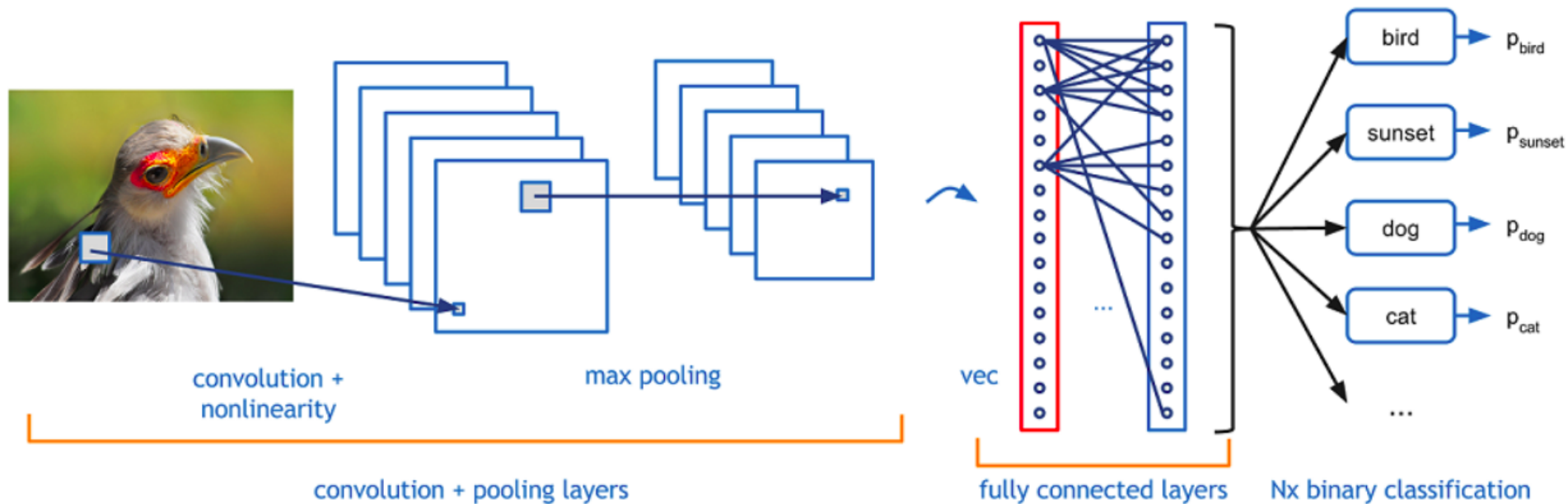


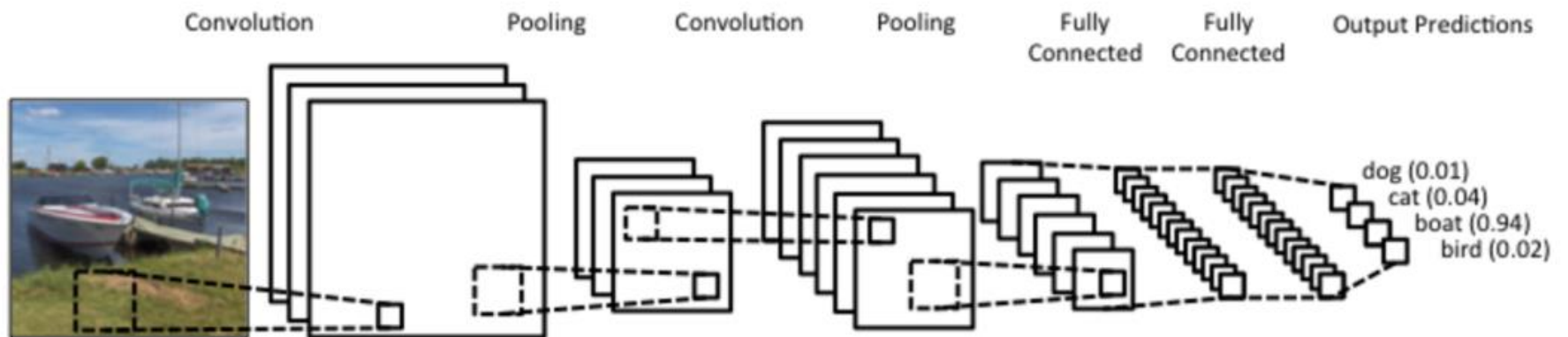
E o que isso tem a ver com CNN?



A extração de características é o processo pelo qual a CNN identifica padrões e características relevantes em uma imagem. As características são extraídas usando camadas convolucionais seguidas de camadas de pooling.

Após a extração de características é aplicado uma rede MLP para realizar a etapa de classificação da imagem.





O conjunto de dados Fashion MNIST é um grande banco de dados de imagens de moda disponível gratuitamente que é comumente usado para treinar e testar vários sistemas de aprendizado de máquina. Foi criado para substituir o banco de dados MNIST original para benchmarking de algoritmos de aprendizado de máquina, pois compartilha o mesmo tamanho de imagem, formato de dados e a estrutura das divisões de treinamento e teste.

O conjunto de dados contém 70.000 imagens em escala de cinza 28x28 de produtos de moda de 10 categorias de um conjunto de dados de imagens de artigos da Zalando, com 7.000 imagens por categoria. O conjunto de treinamento consiste em 60.000 imagens e o conjunto de teste consiste em 10.000 imagens. O conjunto de dados geralmente é incluído em bibliotecas de aprendizado de máquina padrão



O objetivo deste laboratório é criar um modelo de **Rede Neural Convolutiva** para identificar as classes das imagens