

---

Pós-graduação em Ciência da Computação  
Centro de Informática  
Universidade Federal de Pernambuco

# Redes Convolucionais Aplicadas em Visão Computacional

Marcos José Canêjo Estevão de Azevêdo

[mjcea@cin.ufpe.br](mailto:mjcea@cin.ufpe.br)

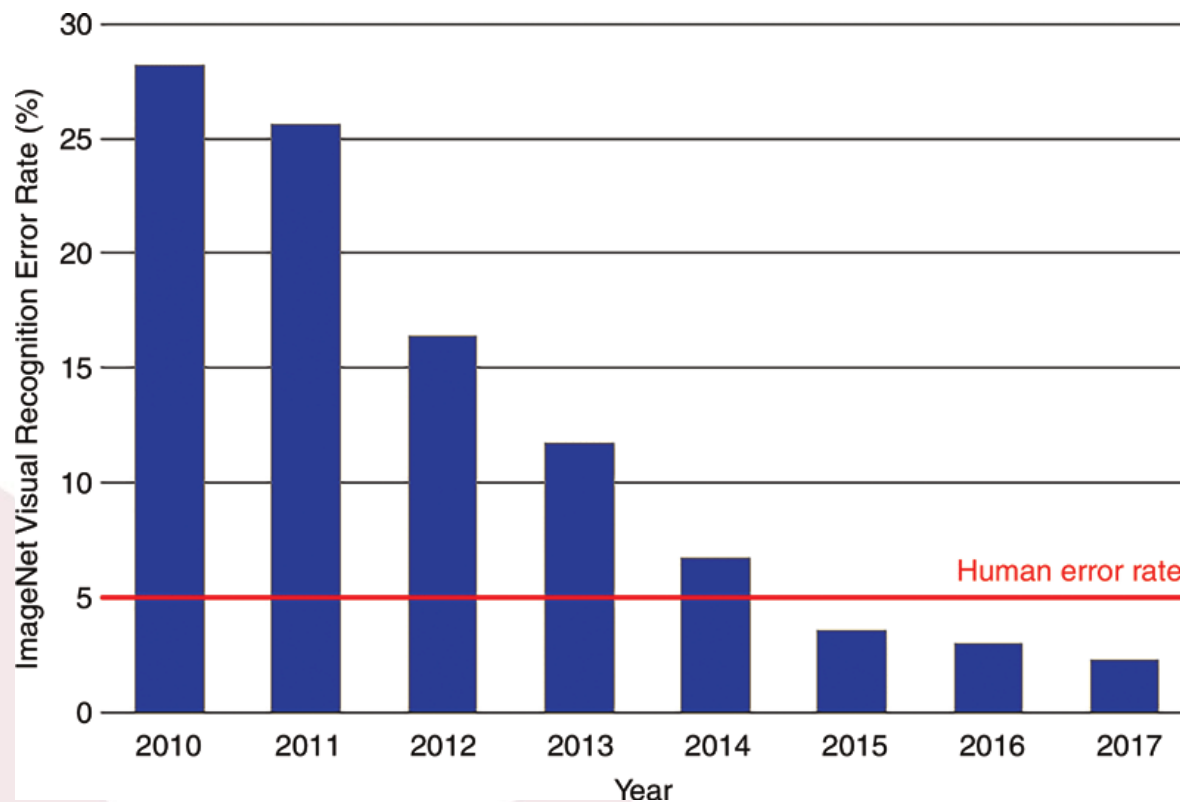
# Sumário

---

- História
- Evolução
- Tipos
- Aplicações
- Considerações Finais

# História

- Em 2012 ganharam destaque após ganhar a competição ImageNet Large-Scale Visual Recognition Challenge.



# História



Zebra (13.74%)



Dálmata (40.86%)



Banco de Parque (19.35%)



Zebra (26.31%)



Zebra (40.19%)

# Evolução



Corpo branco  
em forma oval



Cabeça arredondada  
e alongada com  
uma protuberância  
laranja

Pescoço branco e  
longo em forma  
retangular

# Evolução

Cabeça arredondada e  
alongada com protuberância  
preta

Pescoço branco e  
longo em forma retangular



Corpo branco  
em forma oval  
com ou sem asas  
simétricas  
grandes

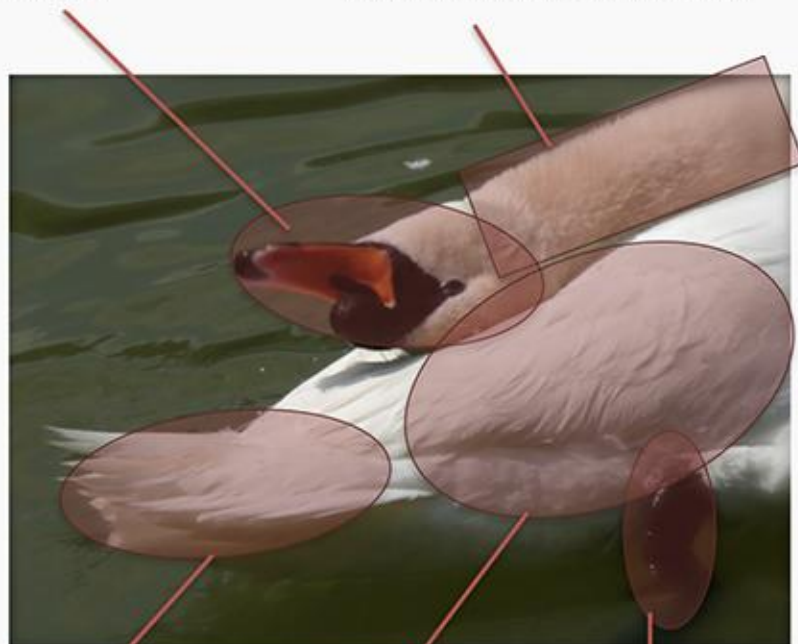
# Evolução

Cabeça arredondada e alongada com protuberância preta ou laranja **invertida**

Pescoço branco e longo em forma retangular que **pode se curvar**, não sendo necessariamente reto

Olhos pretos circulares que podem ou não aparecem na câmera **simultaneamente**

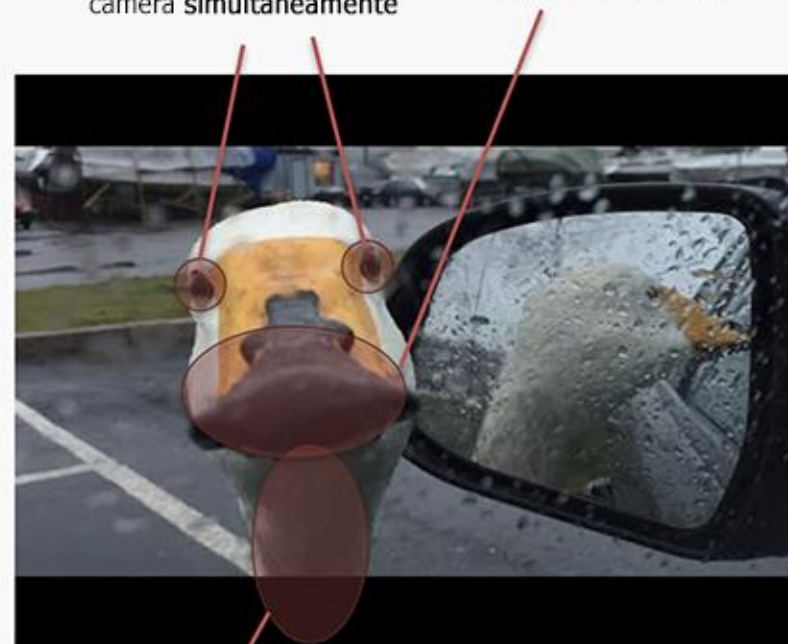
Bico em forma triangular totalmente preta com diferentes tamanhos



Cauda branca **geralmente** longe da cabeça

Asas visíveis

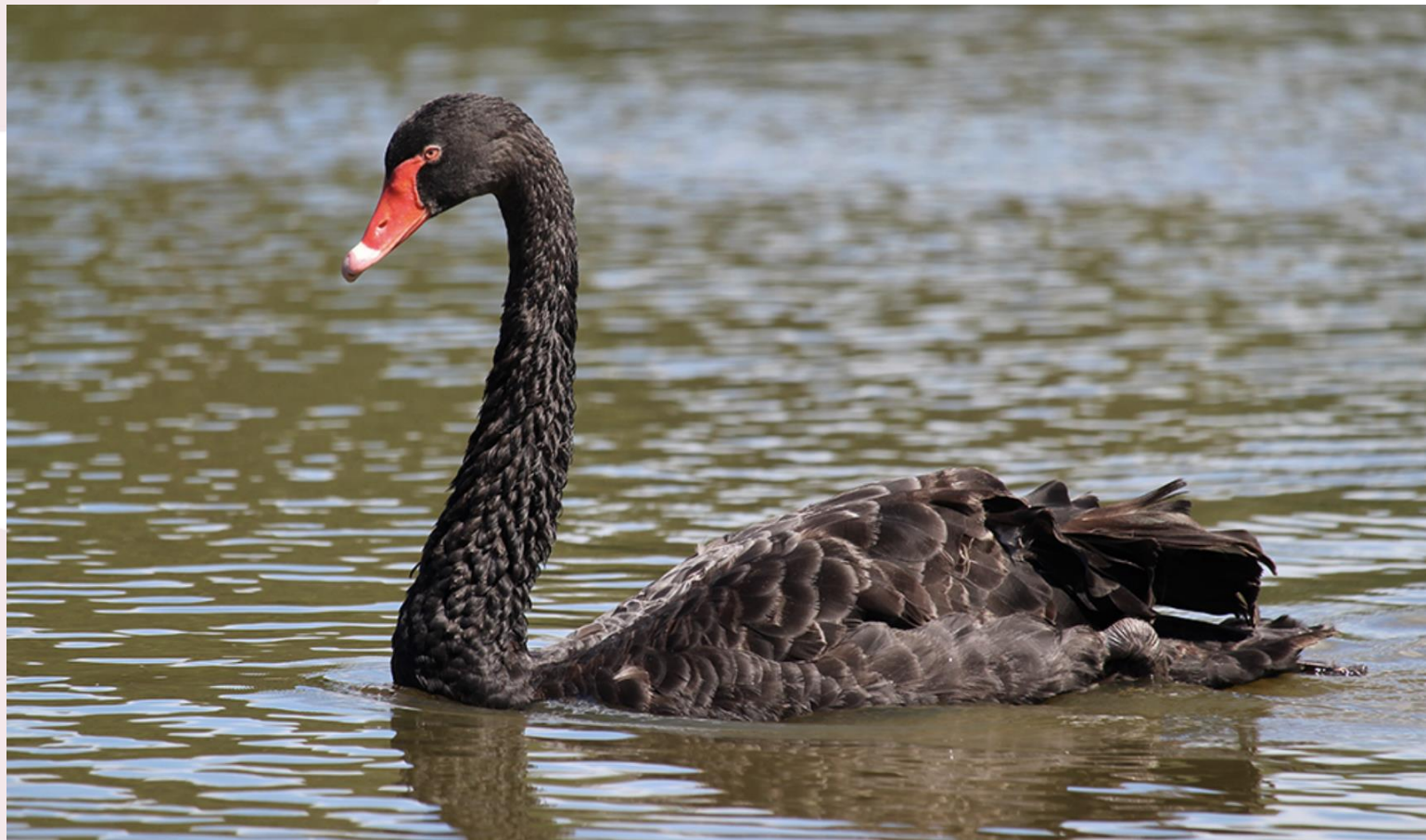
Pés **pretos** geralmente com formas diferentes



Pescoço com forma oval



# Evolução



# Evolução

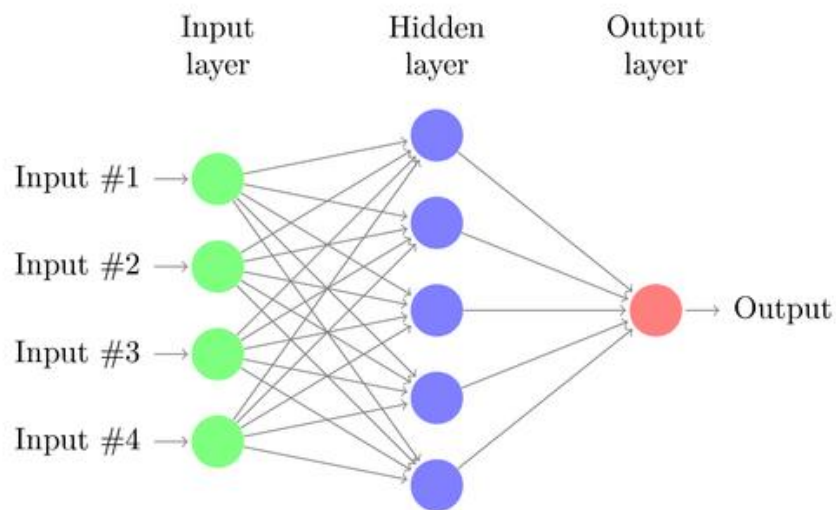


Homem em uma tenda de cisne tentando fotografar cisnes



## EXTRAÇÃO DE CARACTERÍSTICAS

→ SIFT, FAST, SURF,  
HOG, LBP e etc.



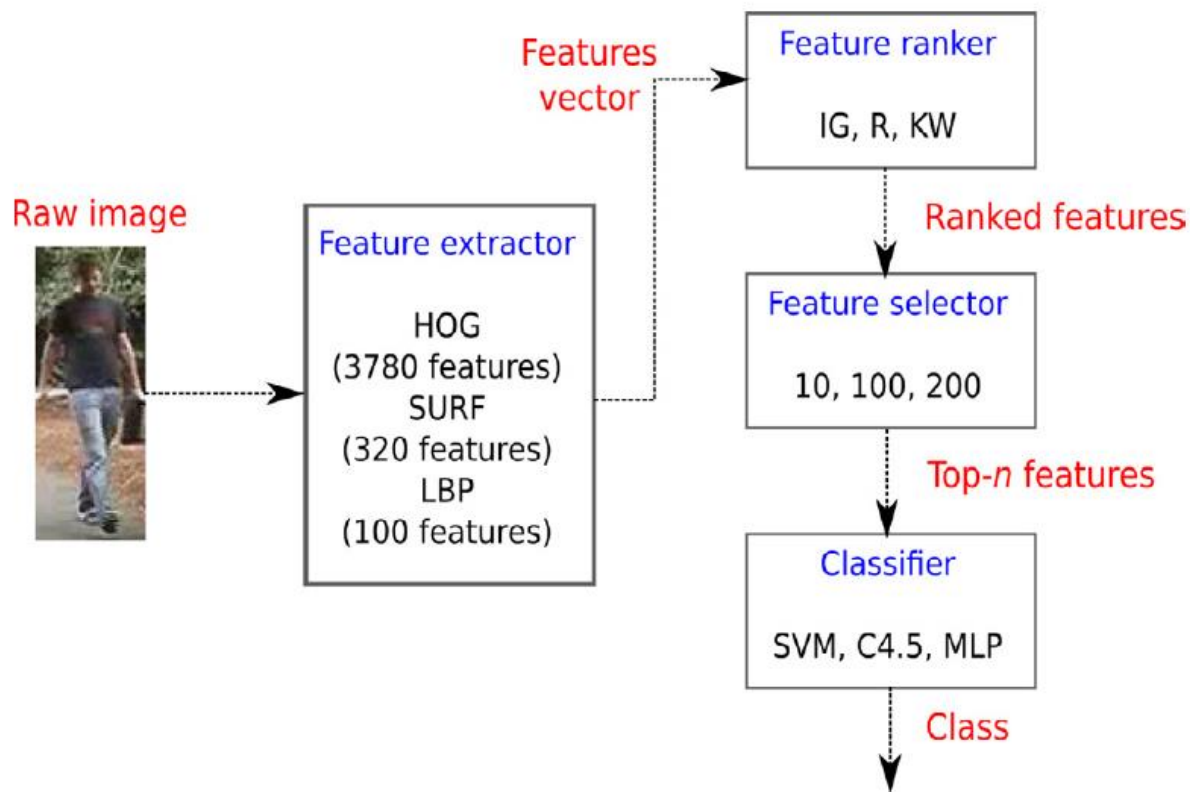
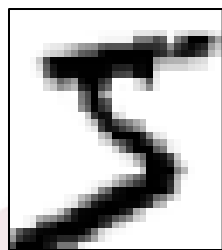


Fig. 2. Hand-designed feature extraction and classification process.

# Evolução



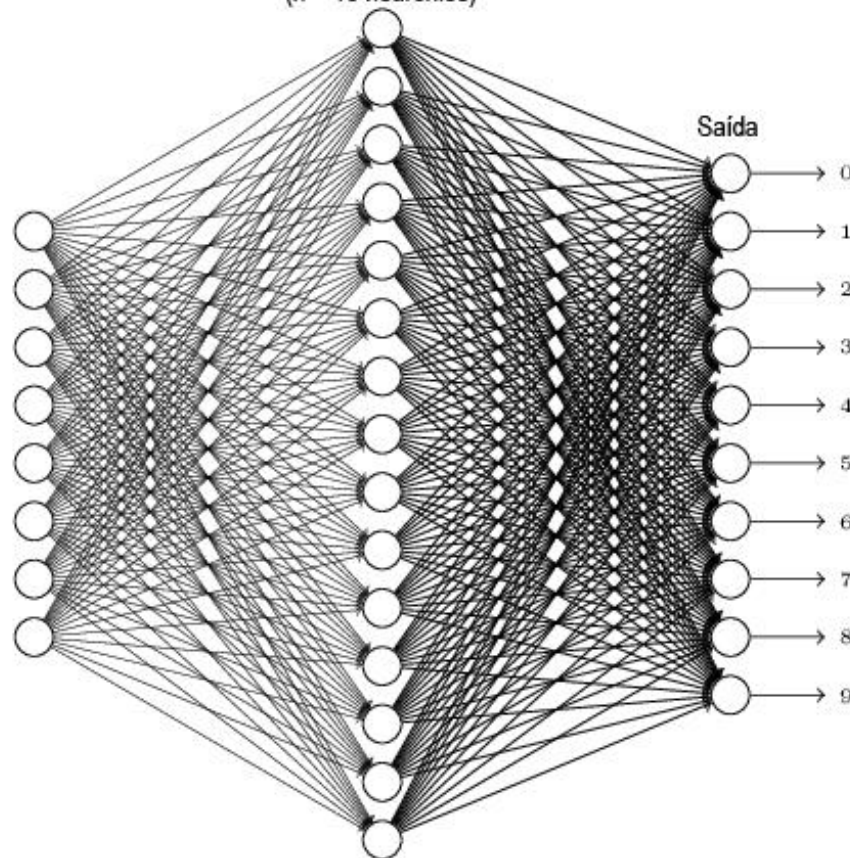
28x28 pixels



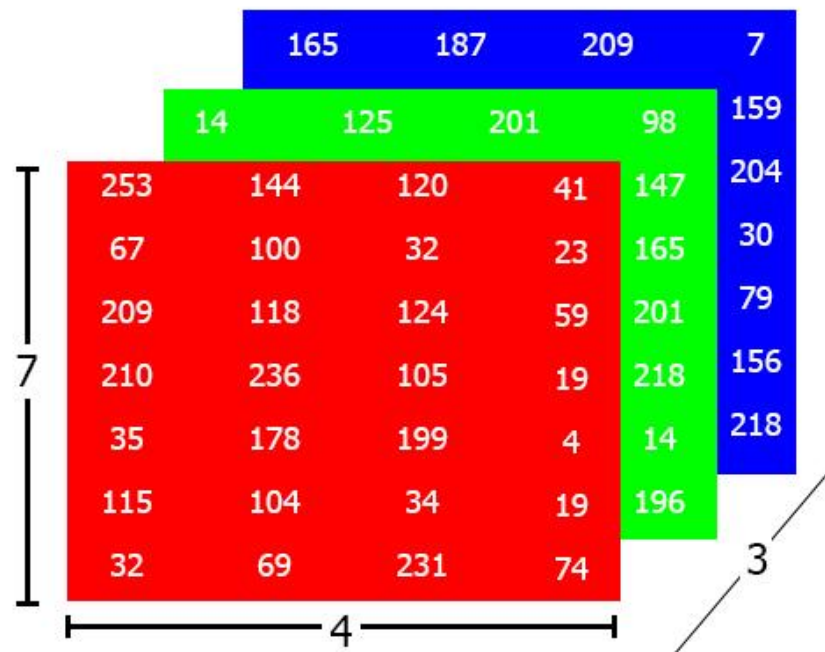
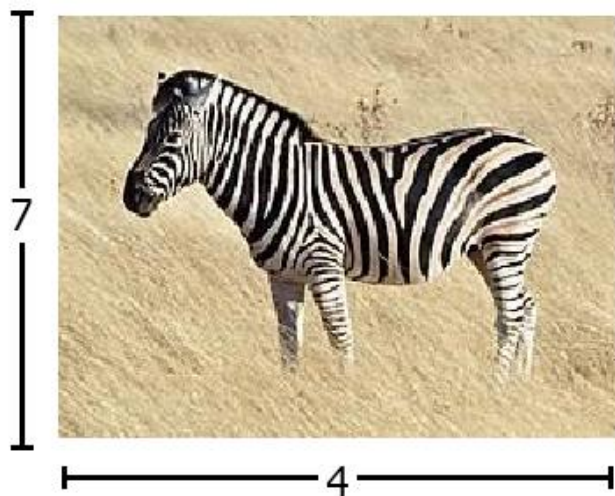
Entrada  
(784 neurónios)

Camada Escondida  
(n = 15 neurónios)

Saída



# Evolução

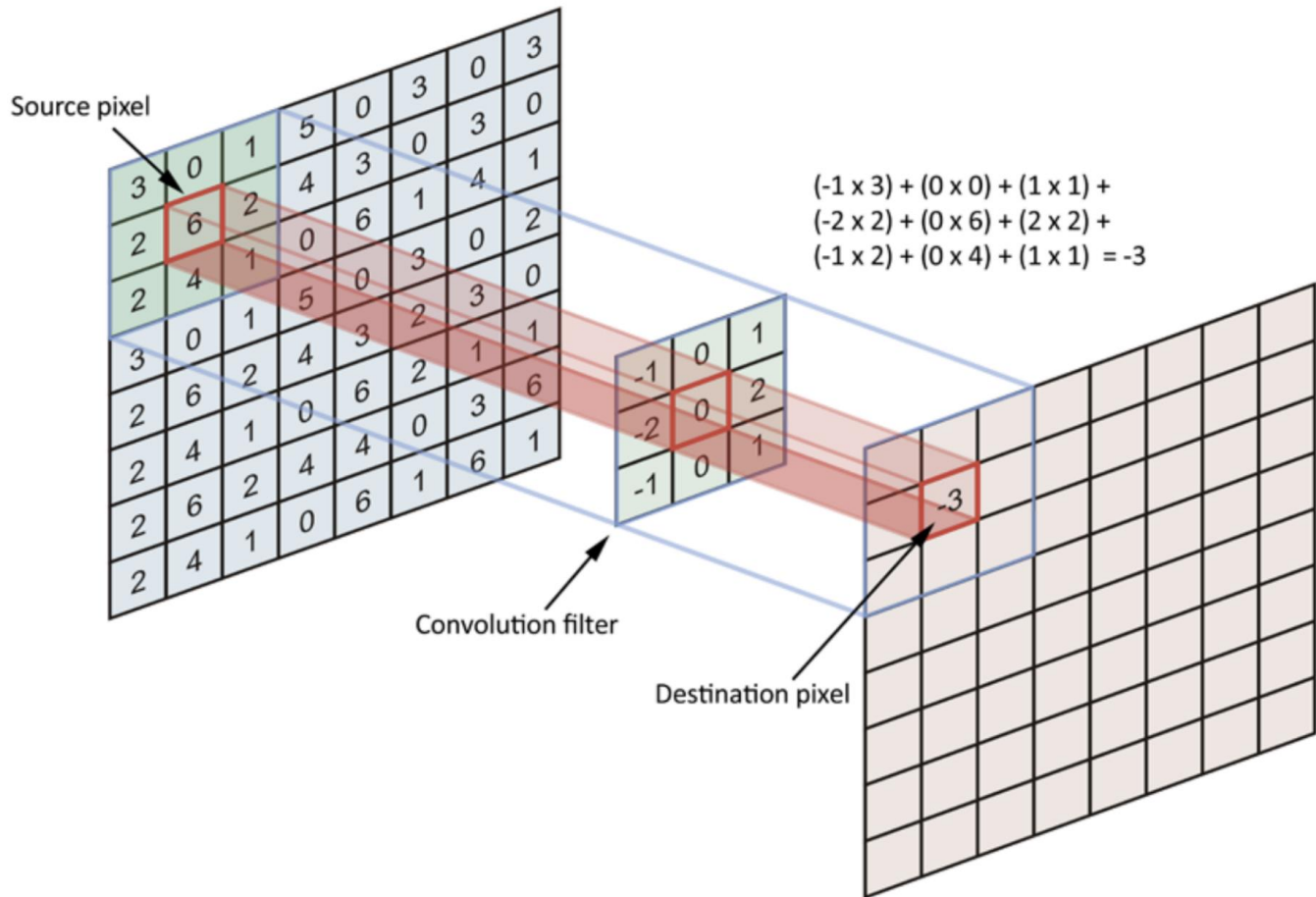




Os pixels próximos são mais fortemente relacionados que os distantes.



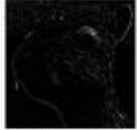



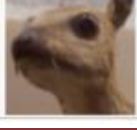
Objetos são construídos a partir de partes menores.

# Convolução

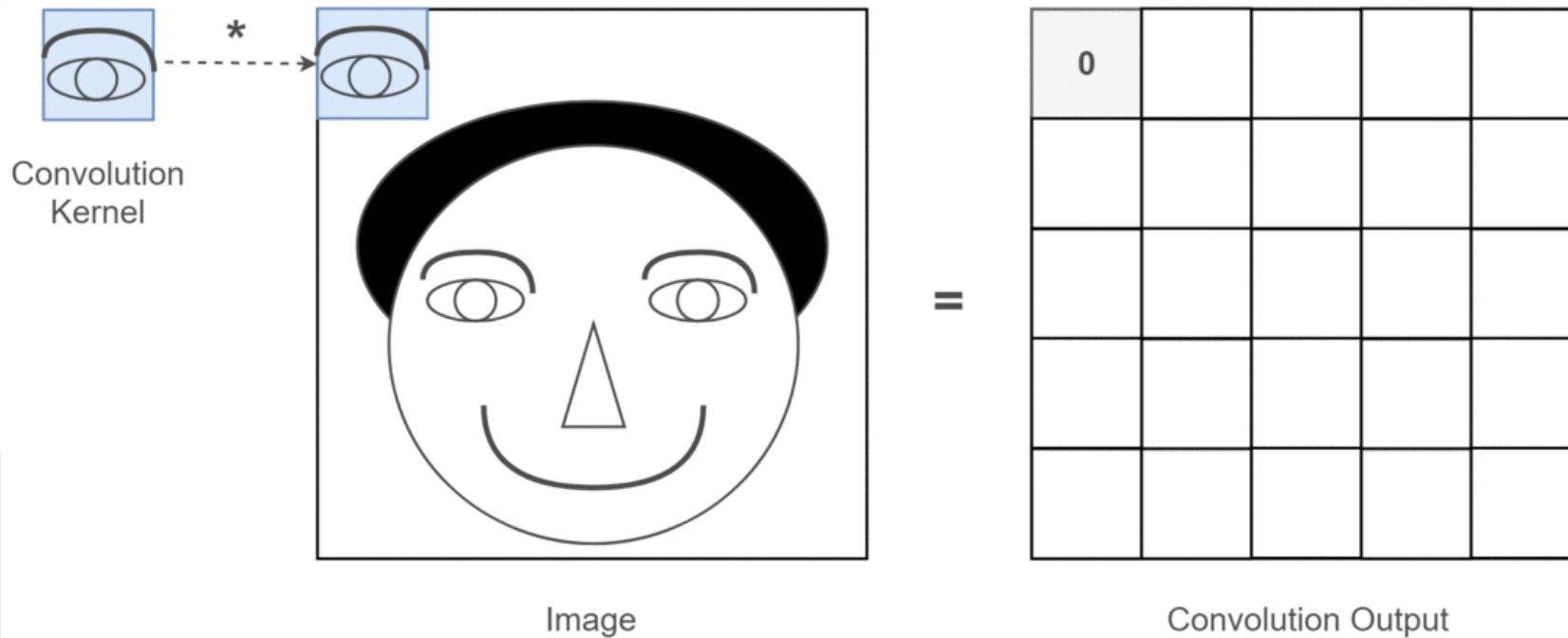




# Convolução

Operation	Filter	Convolved Image
<b>Identity</b>	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
<b>Edge detection</b>	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
<b>Sharpen</b>	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
<b>Box blur</b> (normalized)	$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
<b>Gaussian blur</b> (approximation)	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

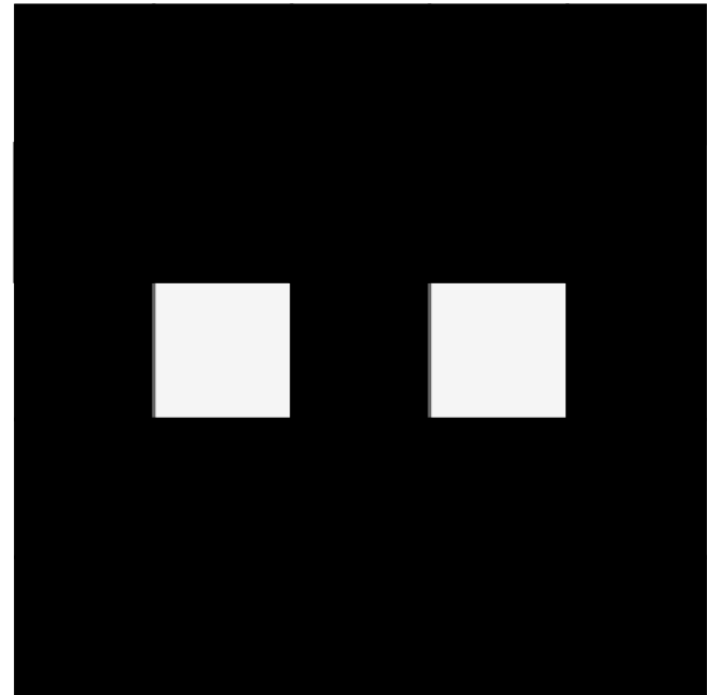
# Convolução



# Convolução

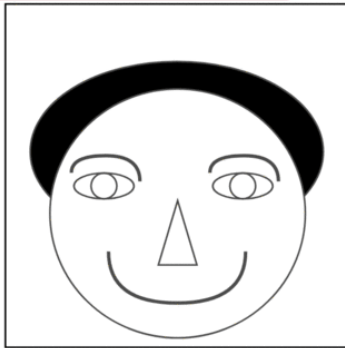
0	0	0	0	0
0	0	0	0	0
0	100	0	100	0
0	0	0	0	0
0	0	0	0	0

Convolution Output

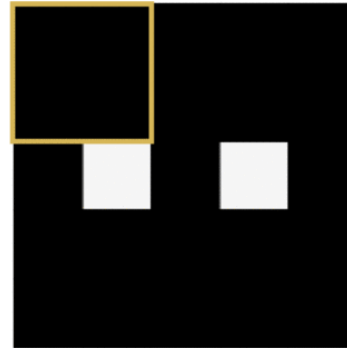


Convolution Output  
(as an Image)

# Pooling



Image

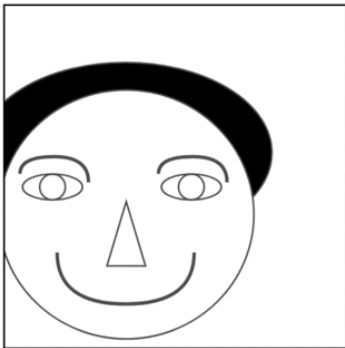


Convolution Output  
(as an Image)

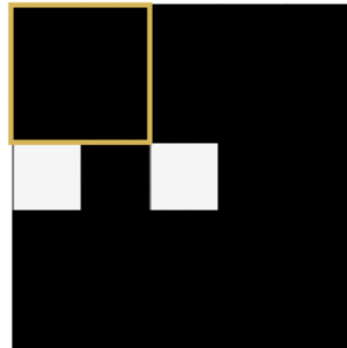


Pooling Output (as  
an Image)

Same image as above but  
translated on x-axis



Image



Convolution Output  
(as an Image)



Pooling Output (as  
an Image)

# Convolução

---

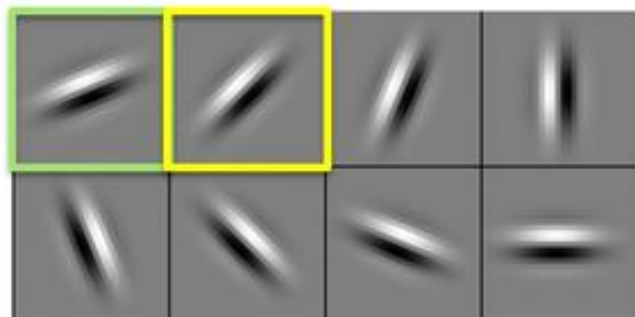


Input

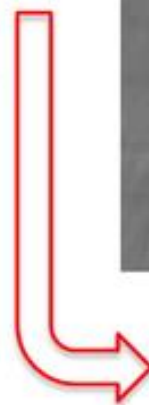
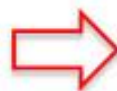
# Convolução



Imagem de entrada



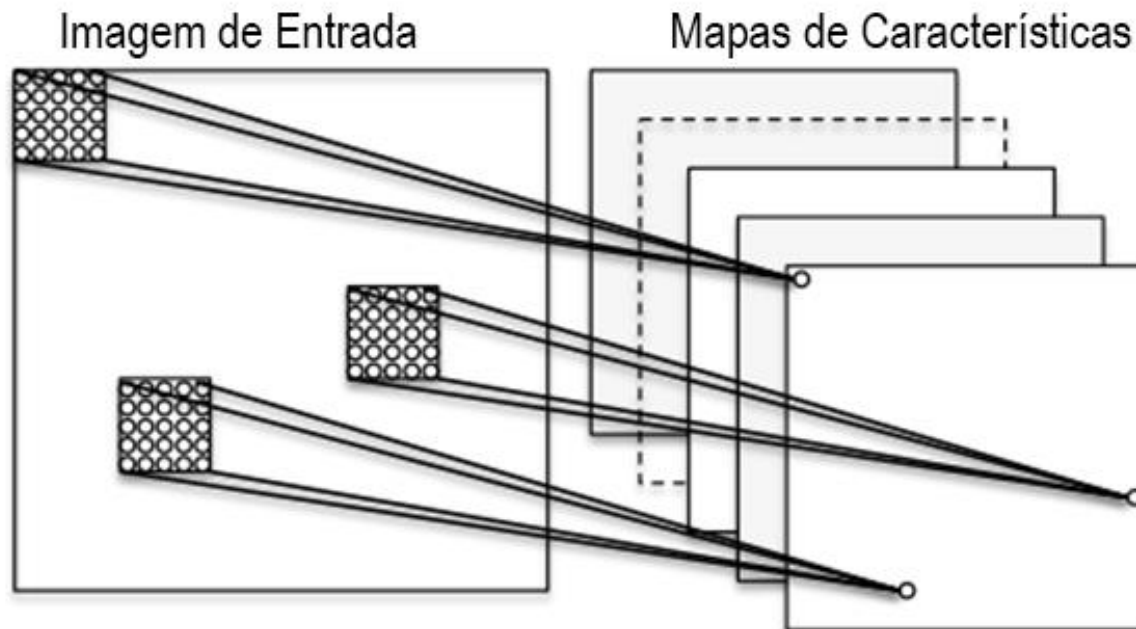
Filtros



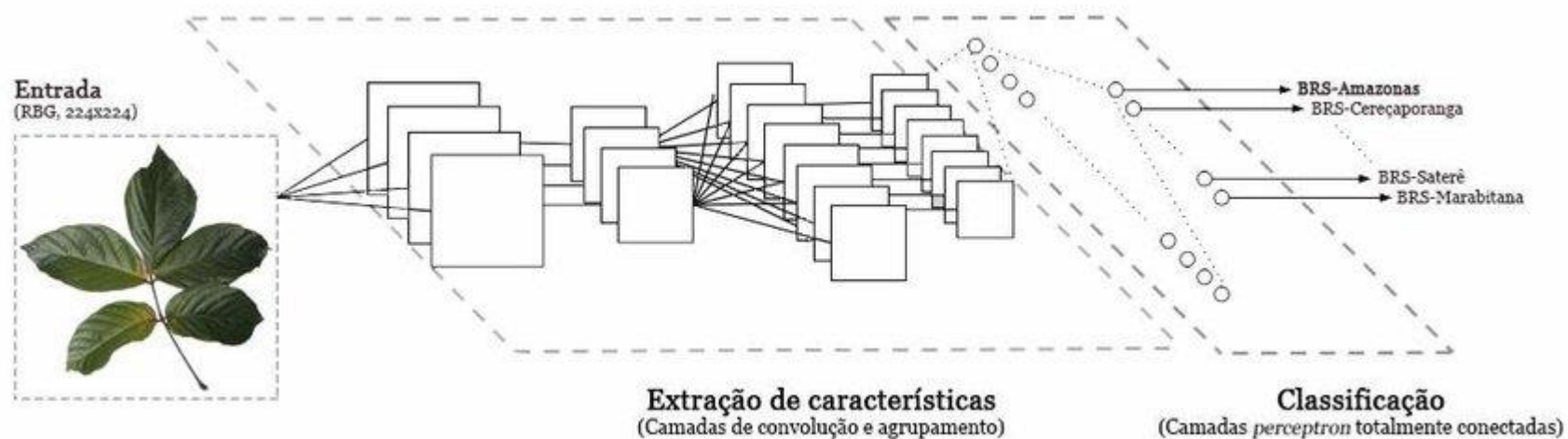
Mapas de Características (Feature Maps)

# Convolução

- O tamanho do **Mapa de Característica** é controlado por três parâmetros: **Profundidade** (*Depth*), **Passo** (*Stride*) e **Preenchimento** (*Padding*).

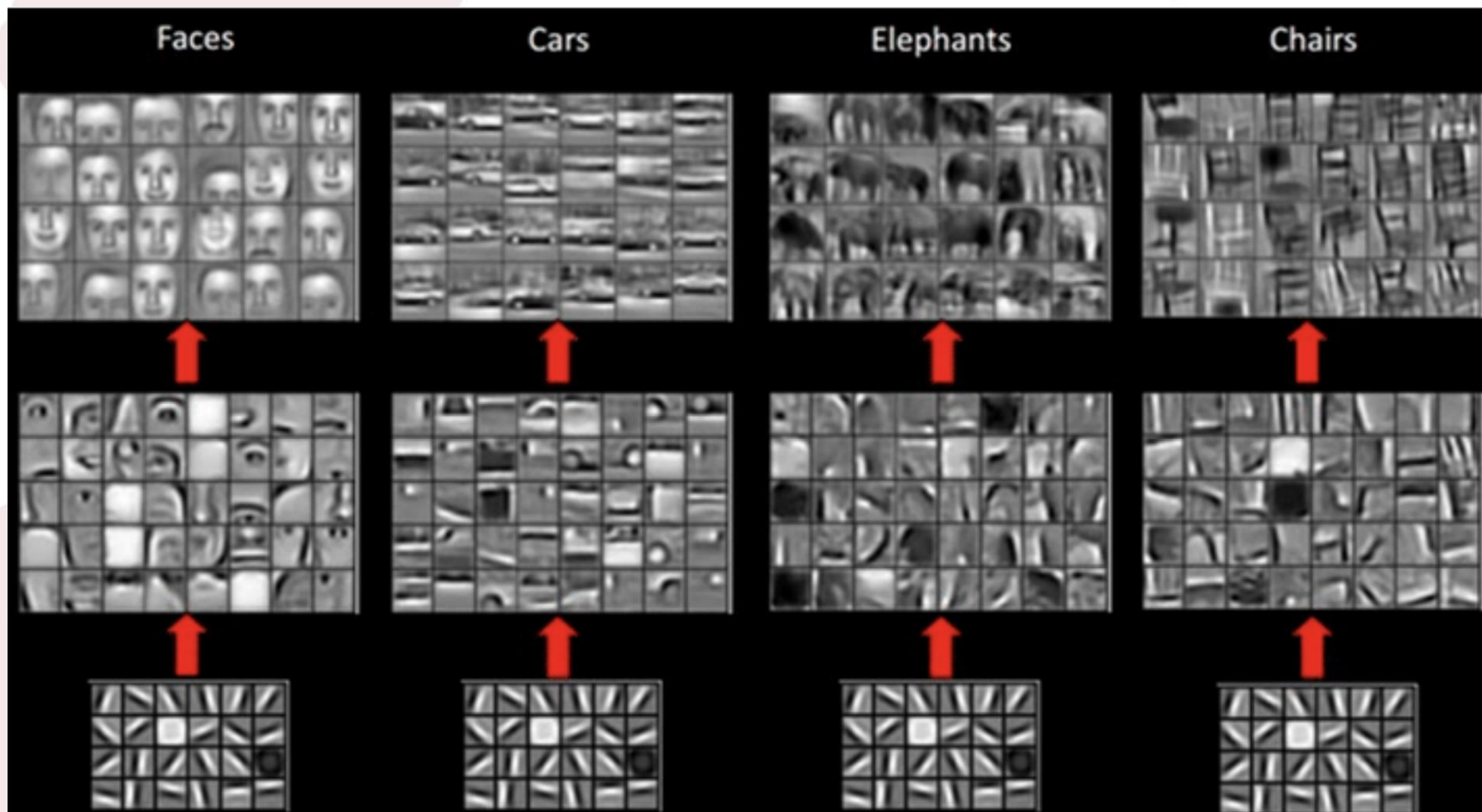


# Arquitetura

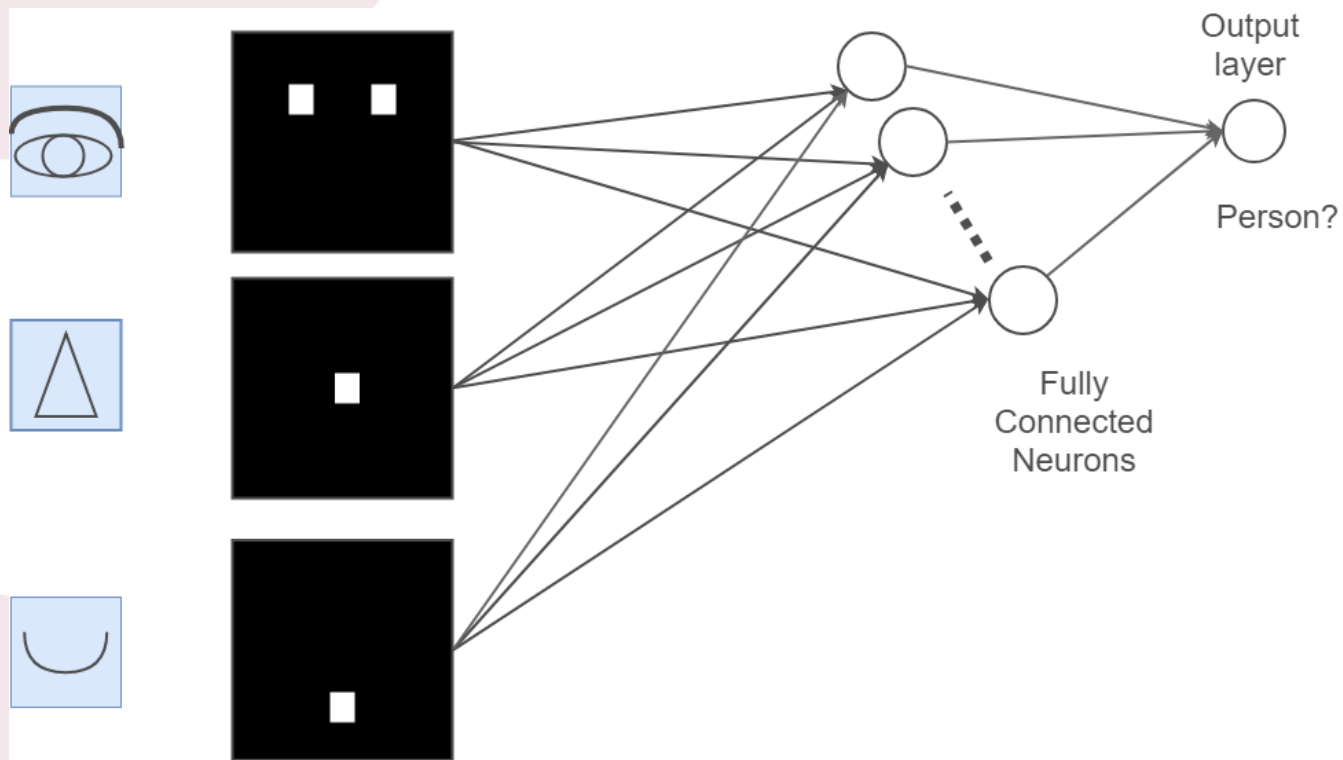




# Mapas de características gerados



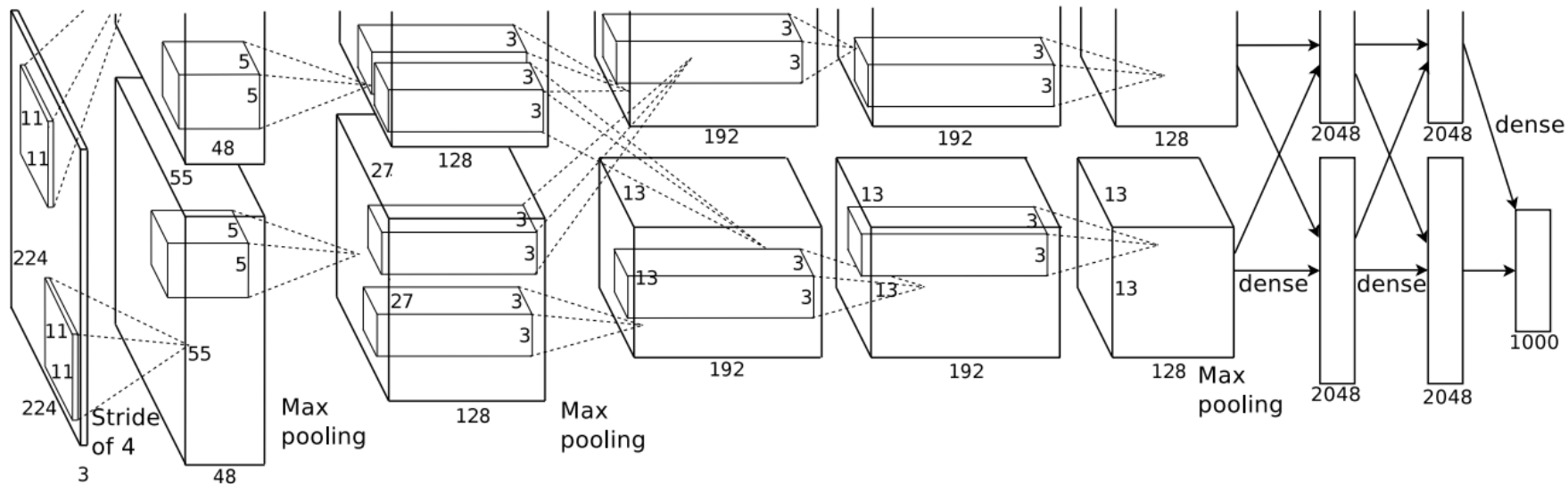
# Convolução



Convolution Outputs  
corresponding to the  
shown filters

- AlexNet (2012);
- ZFNet (2013);
- VGGNet (2014);
- GoogleNet/Inception (2015);
- ResNet(2015).

# AlexNet (2012)



Arquitetura da CNN conhecida por **AlexNet**. Uma GPU executa a parte superior da imagem, enquanto que uma outra executa a parte inferior.

# AlexNet (2012)

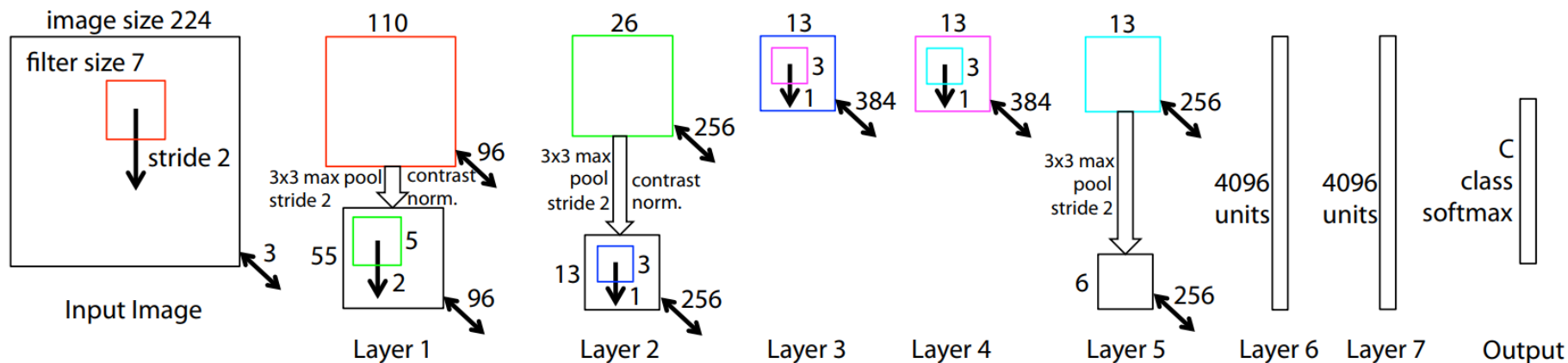
---



- Treinada em **duas GTX 580 (1,5 GB ram)** com os dados da ImageNet por **5/6 dias**, contendo mais de **15 milhões de imagens rotuladas** em mais de **22 mil categorias**;
- Uso da técnica de **Data Augmentation**;
- Primeira CNN a conseguir um bom resultado na ImageNet.



# ZFNet (2013)



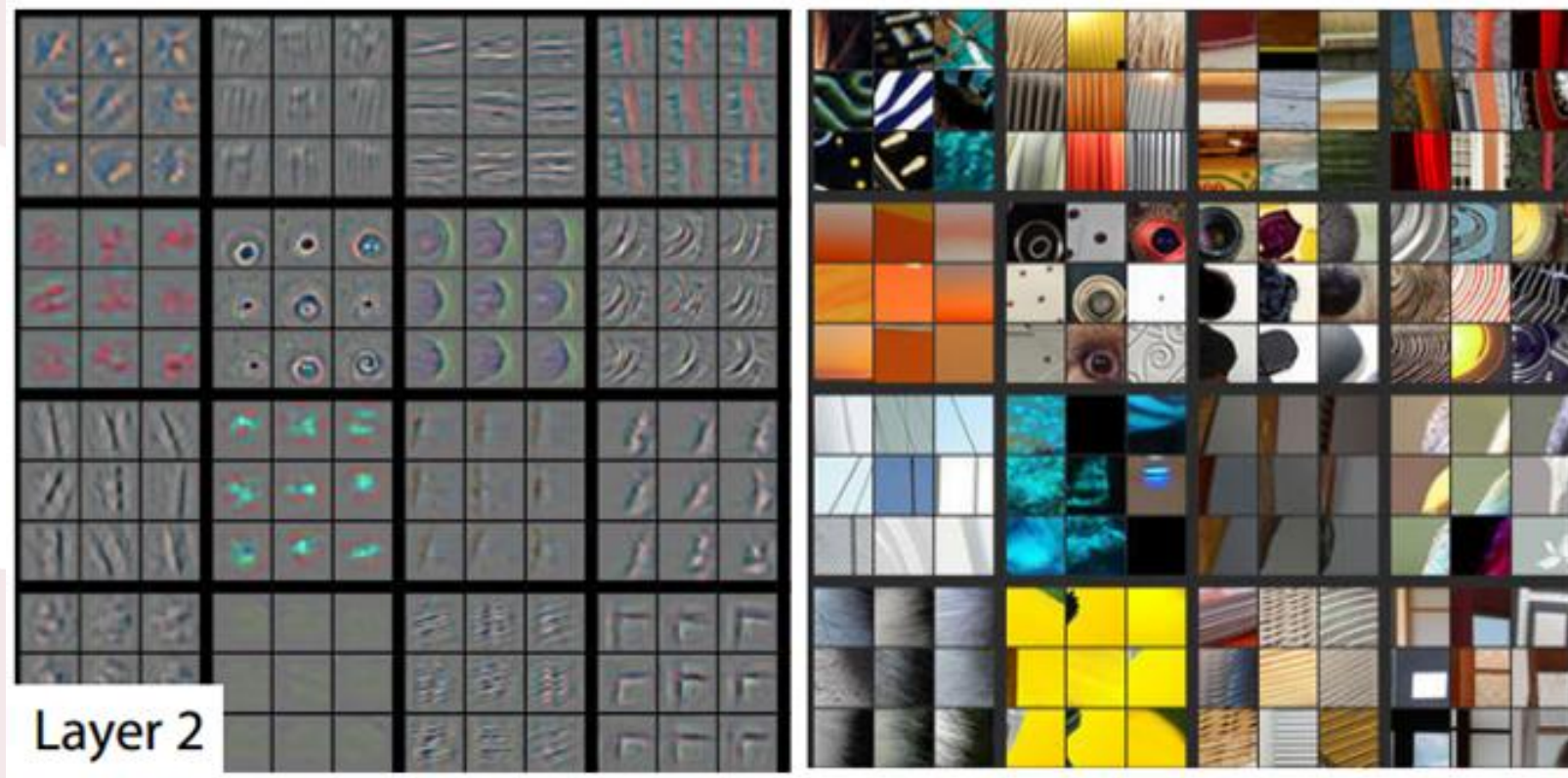
Arquitetura **ZFNet** bastante semelhante a AlexNet utilizando filtros 7x7 e um valor de *stride* menor na primeira camada com o objetivo de **reter mais informações da imagem de entrada**.

# ZFNet (2013)

---

- Treinada em uma **GTX 580 (1,5 GB ram)** por **12 dias** com 1.3 milhões de imagens;
- No mesmo trabalho foi desenvolvido uma técnica de visualização dos filtros chamado de **Deconvolutional Network (DeConvNet)**.

# ZFNet (2013)



Representação de 16 filtros diferentes para as imagens do lado direito.



# VGGNet (2014)

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

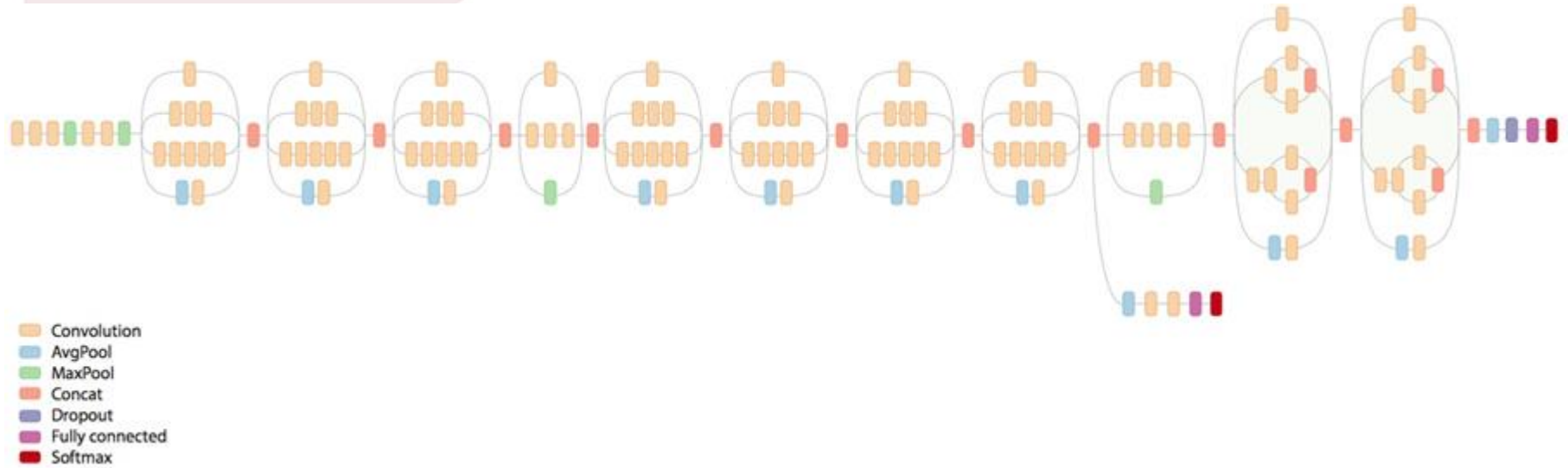
Arquitetura VGG com 6 configurações diferentes.

# VGGNet (2014)

---

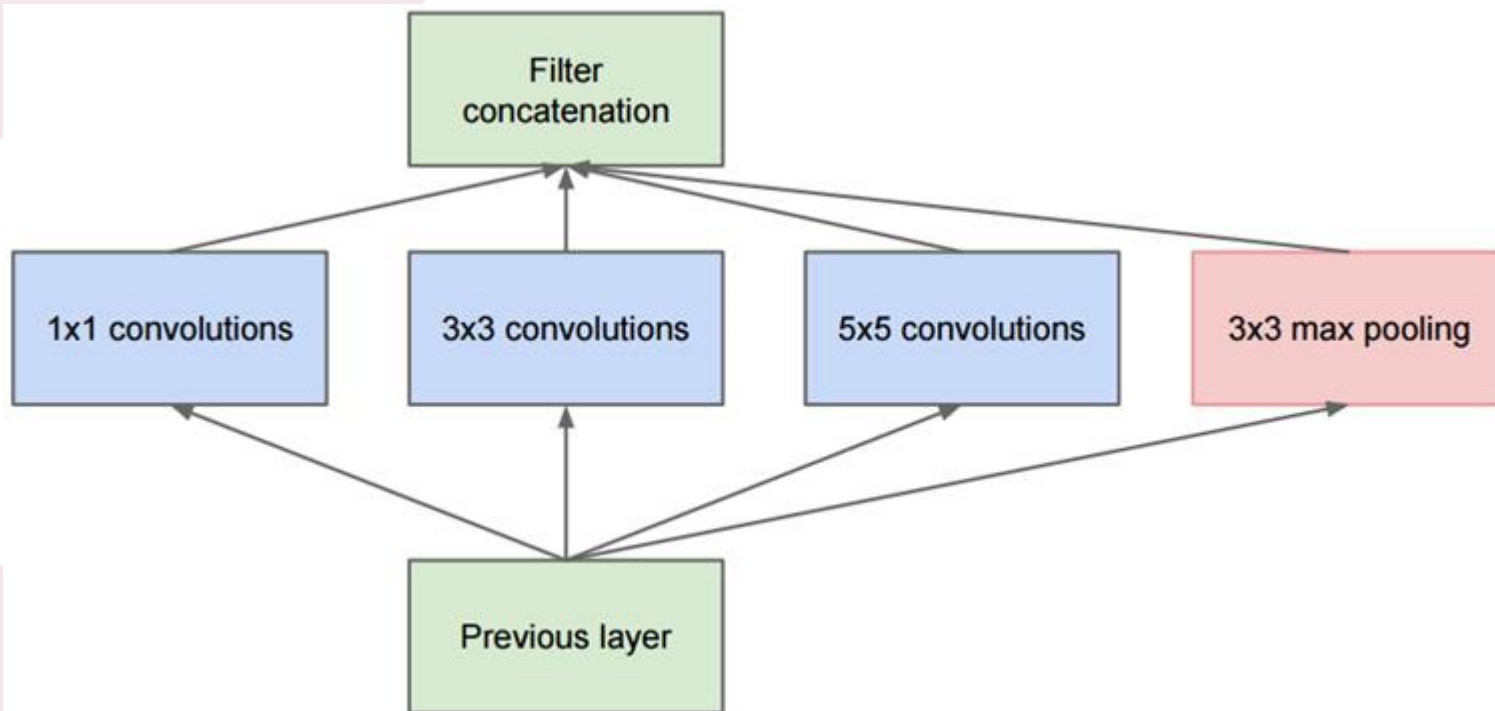
- Treinada em duas **Nvidia Titan Black (6GB ram)** por **2/3 semanas**;
- Aplicada para **classificação e localização**;
- Rede bastante profunda, mas simples em termos de tamanho de filtros.

# GoogleNet/Inception (2015)



Arquitetura **GoogleNet** e seus modulos **Inceptions**.

# GoogleNet/Inception (2015)



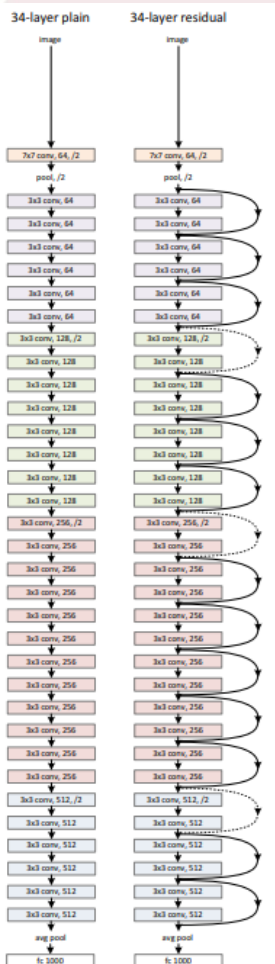
# GoogleNet/Inception (2015)

---

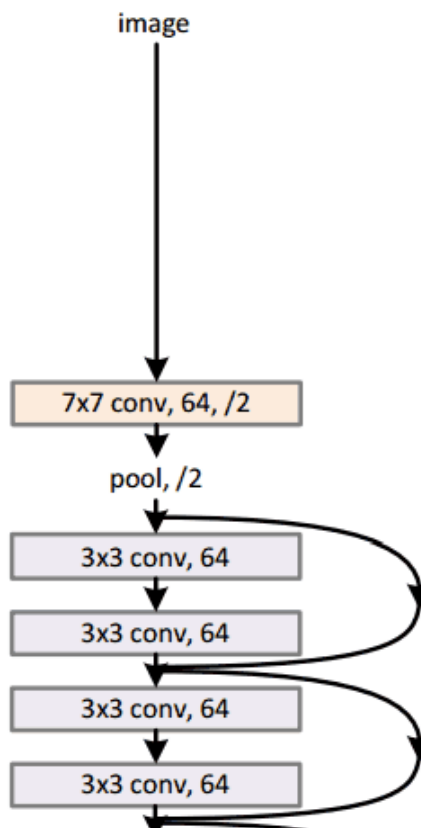


- Treinada com “algumas GPU de ponta” em **paralelo**;
- Utiliza 9 módulos Inception por toda a rede atingindo mais de **100 camadas**;
- Uma das primeiras redes a propor uma **arquitetura não sequencial**.

# ResNet (2015)



## 34-layer residual



Arquitetura **Microsoft ResNet** com 152 camadas atingiu bons resultados nos problemas de **classificação**, **detecção** e **localização**.

# ResNet (2015)

---



- Treinada em **8 GPUs** por 2/3 semanas;
- Testaram com **1202 camadas**, mas **não conseguiram bons resultados**;

# Aplicações

---



<https://lukeoakdenrayner.wordpress.com/2019/09/19/ai-competitions-dont-produce-useful-models/>

<http://scs.ryerson.ca/~aharley/vis/>

<https://paperswithcode.com/sota/image-classification-on-imagenet>





# Referências

---



- <http://scs.ryerson.ca/~aharley/vis/>
- <https://towardsdatascience.com/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac>
- <https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-convolution-neural-networks-e3f054dd5daa>
- <https://medium.com/datadriveninvestor/five-powerful-cnn-architectures-b939c9ddd57b>
- <https://adeshpande3.github.io/adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html>



---

Pós-graduação em Ciência da Computação  
Centro de Informática  
Universidade Federal de Pernambuco

# Redes Convolucionais Aplicadas em Visão Computacional

Marcos José Canêjo Estevão de Azevêdo

mjcea@cin.ufpe.br