

*Conceitos Básicos de Deep
Learning e o seu uso através do*
NVIDIA® DIGITS™

Cristina Nader Vasconcelos
crisnv@ic.uff.br

Cronograma

Webinar I

- Motivação
- Abordagem Clássica para reconhecimento de padrões em sinais
- Exemplos: features, descriptors, classifiers
- Desafios existentes
- ML: modelos de aprendizado
- Redes neurais: formulação clássica
- CNNs
- Aplicações

Webinar II

- Treinamento
- Hierarquia de programação Deep Learning
- Caffe
- Digits
- Exemplos

Câmeras estão por toda parte!



Motivação

Em 2010 Eric Schmidt, Google CEO, estimou que a cada dois dias a humanidade produziria então 5 **exabytes** de dados, o que seria equivalente ao que foi produzido desde os primórdios da civilização até 2003.

1 EB = 1000^6 bytes = 10^{18} bytes = 1 milhão de terabytes = 1 bilhão de gigabytes.

Críticos dizem que ninguém sabe ao certo quanto de informação tem sido produzida ou quão veloz tem sido o aumento nessa produção, mas certamente estamos produzindo muita informação e a velocidades cada vez mais altas.

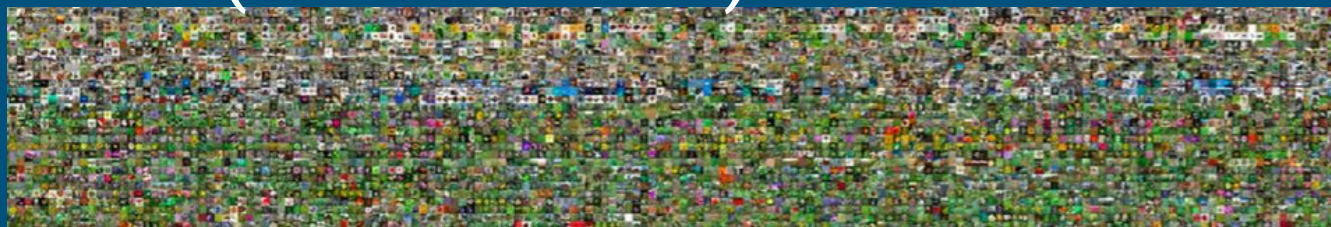
Grandes volumes de dados: desafios

Queremos acesso a informações relevante:

- consultas/organização
- análise/descoberta de padrões
- previsões
- entre outros desafios



Desenvolvimento de bancos padrões de referência (*benchmarks*)



IMGENET

www.image-net.org

22K categories and **14M** images

- Animals
 - Bird
 - Fish
 - Mammal
 - Invertebrate
- Plants
 - Tree
 - Flower
 - Food
 - Materials
- Structures
 - Artifact
 - Tools
 - Appliances
 - Structures
- Person
 - Scenes
 - Indoor
 - Geological Formations
 - Sport Activities

IMAGENET Large Scale Visual Recognition Challenge

Steel drum

The Image Classification Challenge:
1,000 object classes
1,431,167 images



Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle



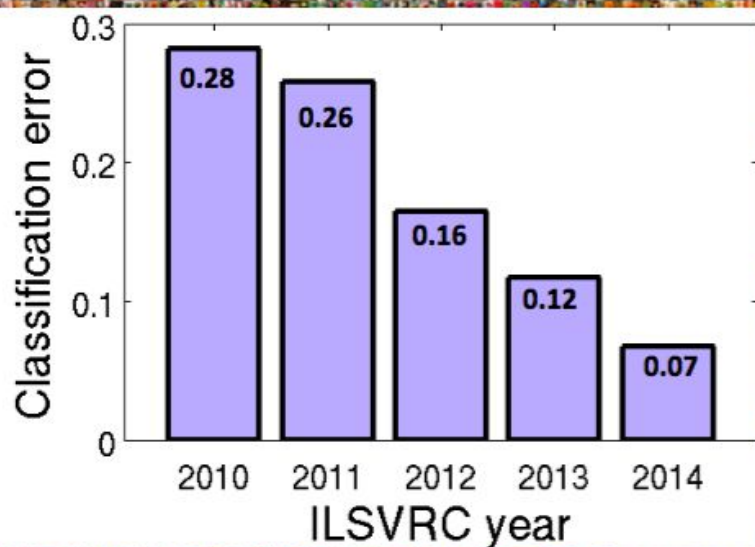
Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle



IMAGENET Large Scale Visual Recognition Challenge

Steel drum

The Image Classification Challenge:
1,000 object classes
1,431,167 images



abaixo
de
0.04
2015

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC



Dense grid descriptor:
HOG, LBP

Coding: local coordinate,
super-vector

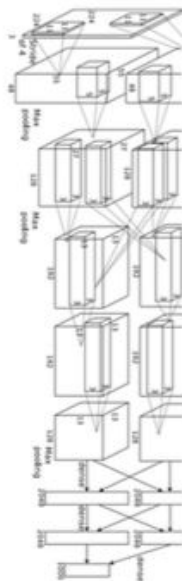
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

SuperVision



[Krizhevsky NIPS 2012]

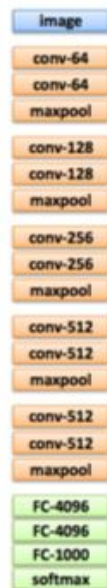
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

VGG



[Simonyan arxiv 2014]

MSRA



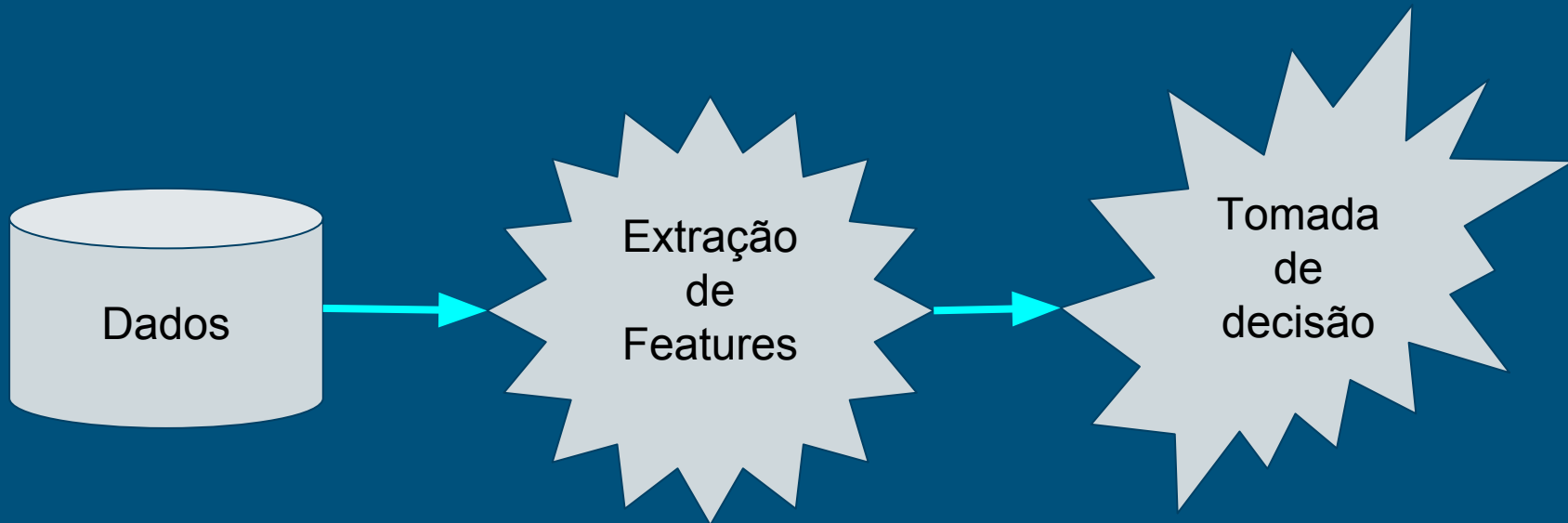
[He arxiv 2014]

Year 2015

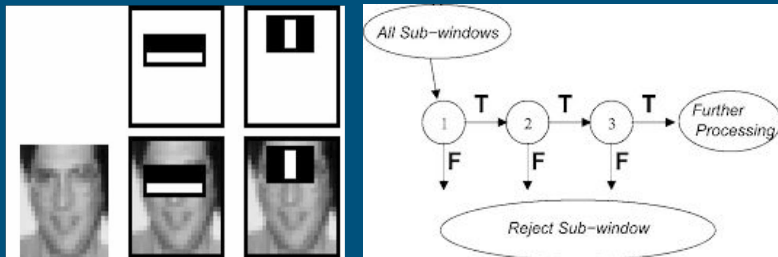
MSRA



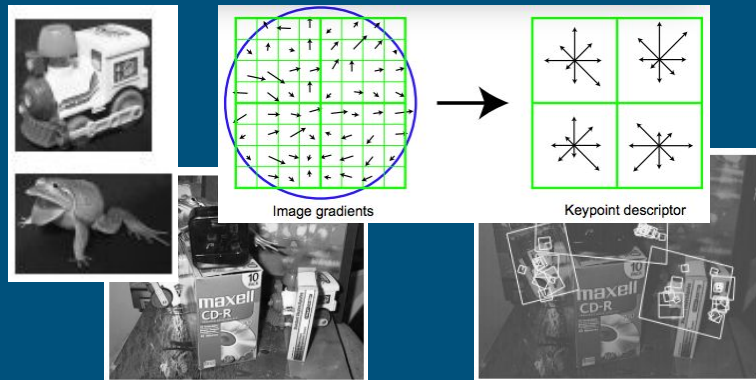
Abordagem clássica:



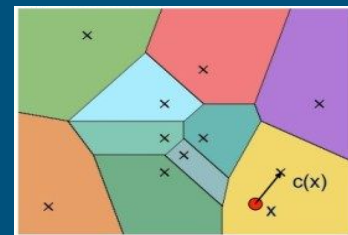
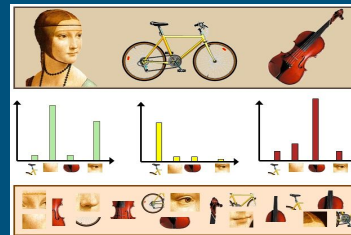
Diferentes arquiteturas...



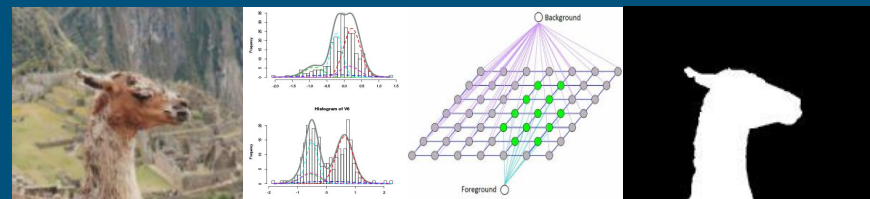
P. Viola and M. J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," CVPR 2001



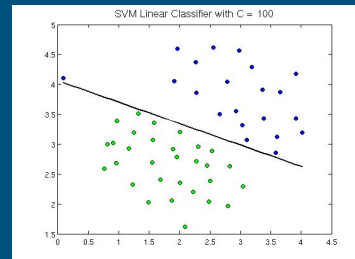
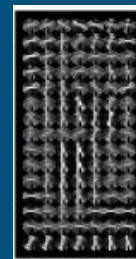
David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. IJCV 2004



J. Sivic and A. Zisserman, "Video google: a text retrieval approach to object matching in videos," ICCV '03



C. Rother, V. Kolmogorov, A. Blake. GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts. SIGGRAPH'04



N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," CVPR 2005.

Porque são problemas difíceis?

Desafios: ponto de vista, deformação, iluminação, ...



*[R. Tao, A. Smeulders, S. Chang] Attributes and Categories for Generic Instance Search from One Example. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp.177 - 186. 2015

**[K. Patil, S. Bojewar] A Survey on Face Recognition of Identical Twins. International Journal of Scientific & Engineering Research, Vol. 6, Issue 2, Feb.-2015 744.

Desafios: oclusão, distinção do fundo



Desafios: variação intra-classe



Aprendizado de máquina: abordagens

- Aprendizado Supervisionado (e semi): são fornecidos pares (x, y) . Busca-se aprender a produzir a saída y a partir de uma entrada x ;
- Aprendizado não-supervisionado: não é fornecida uma saída específica para a entrada x ;
- Aprendizado por reforço: busca-se tomar uma sequência de boas decisões para maximizar a recompensa em um ambiente influenciado por tais decisões;

Aprendizado Supervisionado

- mapeamento $x \rightarrow y$:

- problemas de classificação:
 - y representa uma classe;

ex: qual o seu animal preferido? {gato, cachorro, leão, cobra}

- problemas de regressão:
 - y representa um valor real;

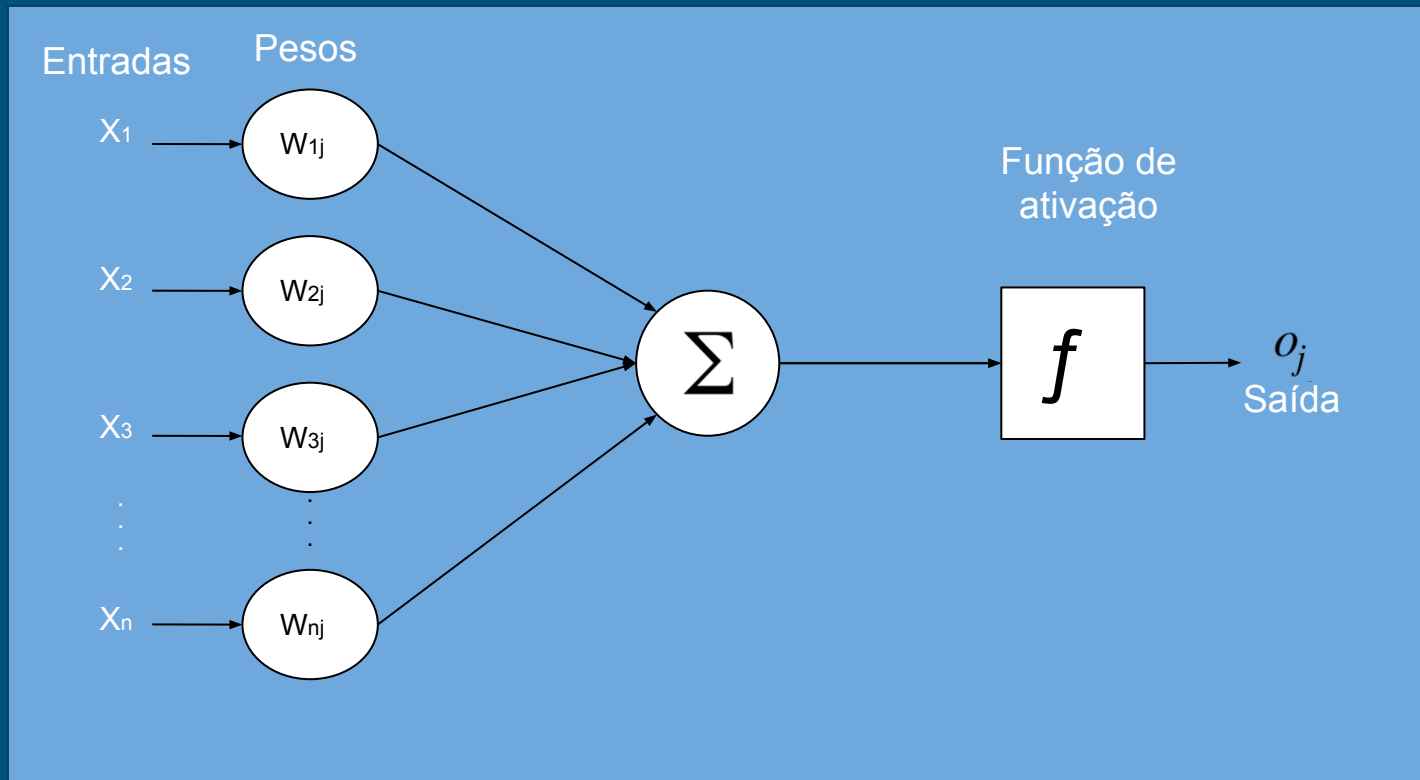
ex: qual é a sua altura? 1.71 metros

Ingrediente chave: dados

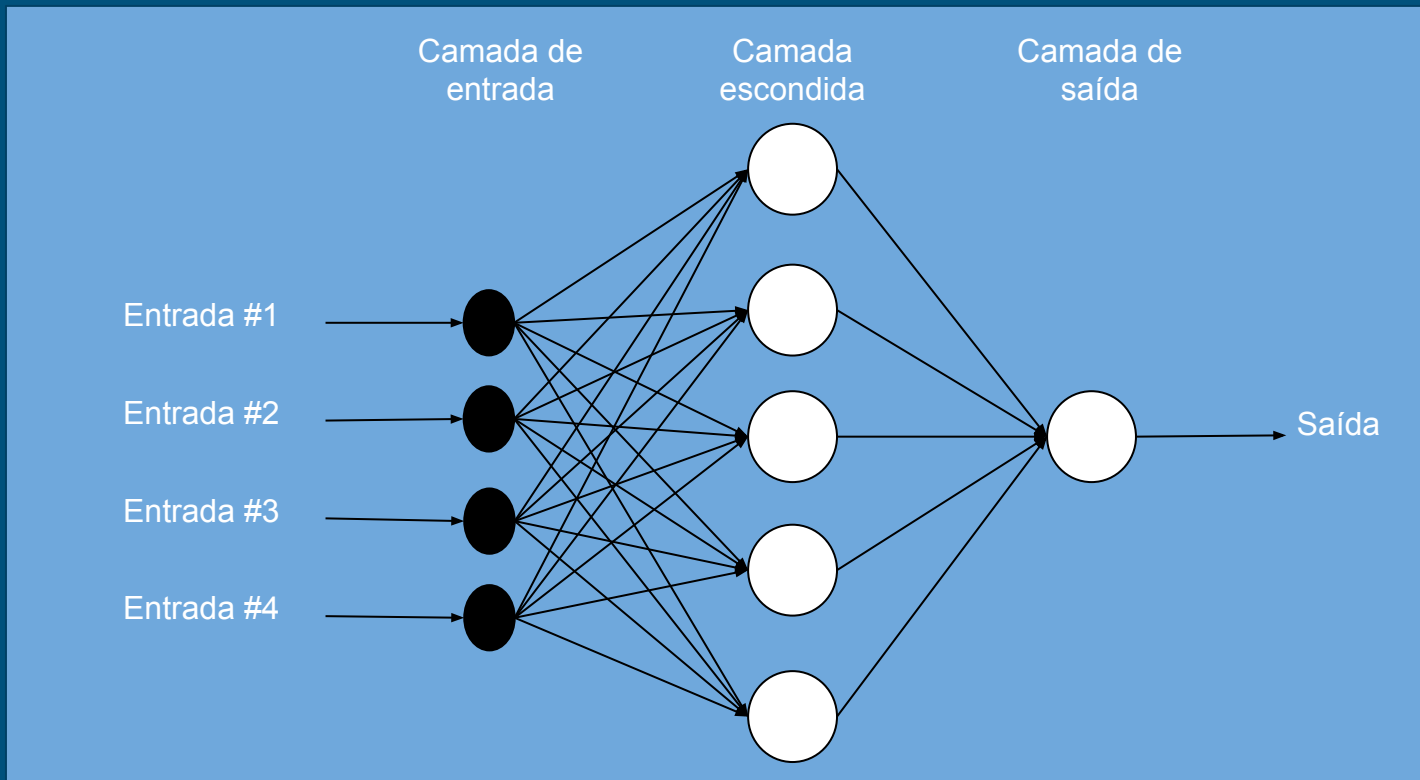
O aprendizado supervisionado se beneficia com o aumento do número de exemplos coletados para a formação da base de treinamento



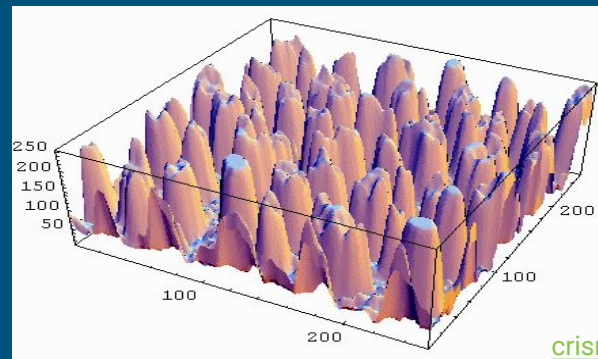
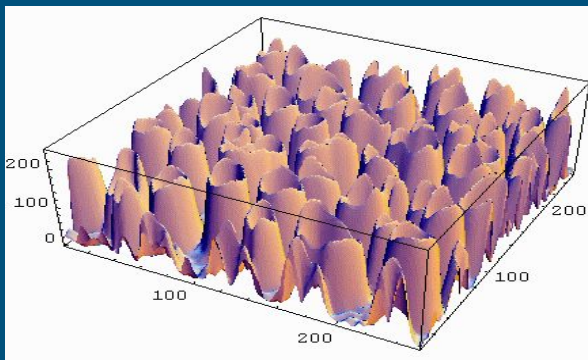
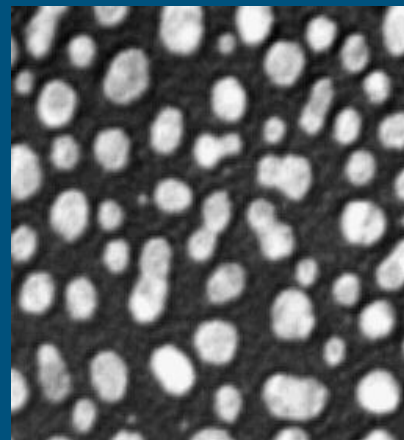
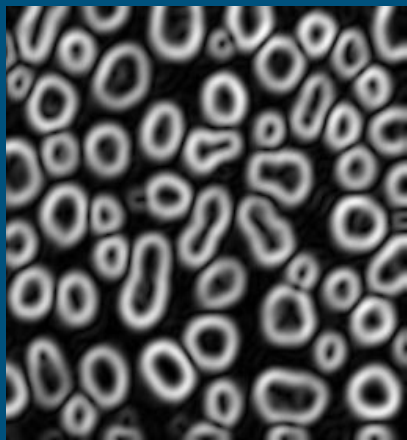
O perceptron



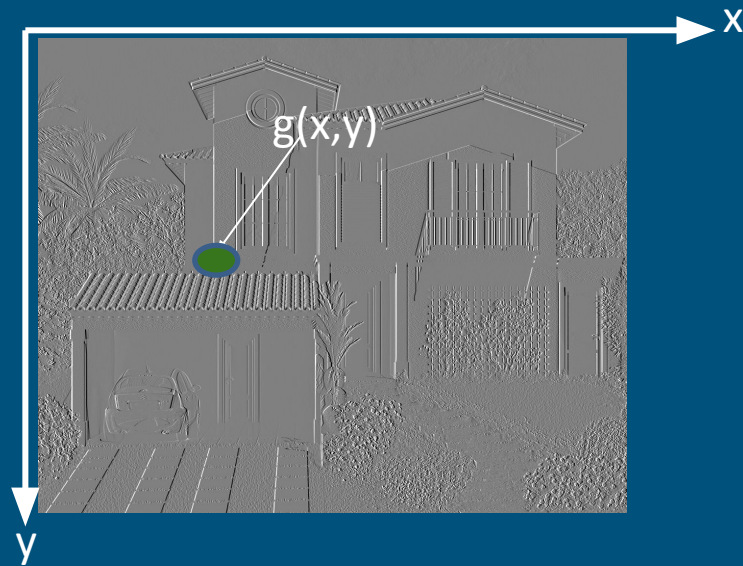
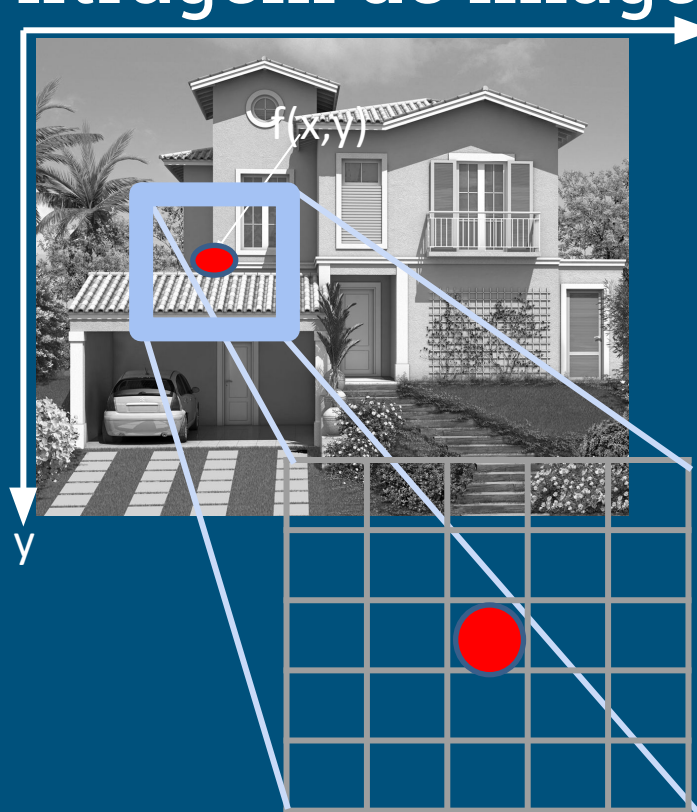
Redes Neurais: arquitetura de múltiplas camadas



Imagens como sinal 2D



Filtragem de Imagens

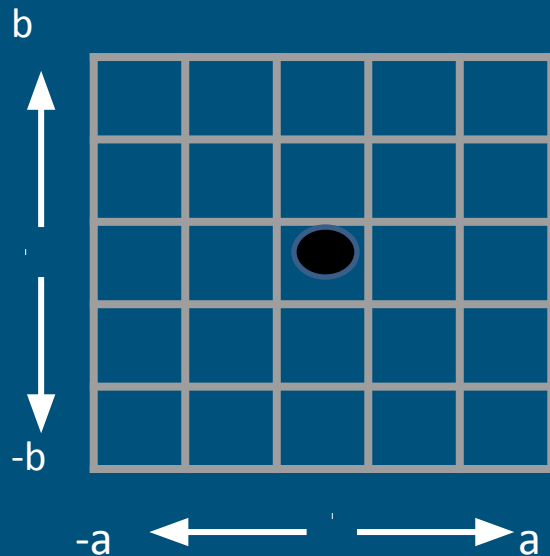


onde:

- $f(x,y)$: intensidade do pixel (x,y)
- $g(x,y)$: nova intensidade do pixel (x,y)
- N : vizinhança (ex: 5x5)
- T : operador definido na vizinhança N do pixel (x,y) . $g(x,y) = T_N(f(x,y))$

A convolução de uma imagem por um filtro

- Matriz de pesos:

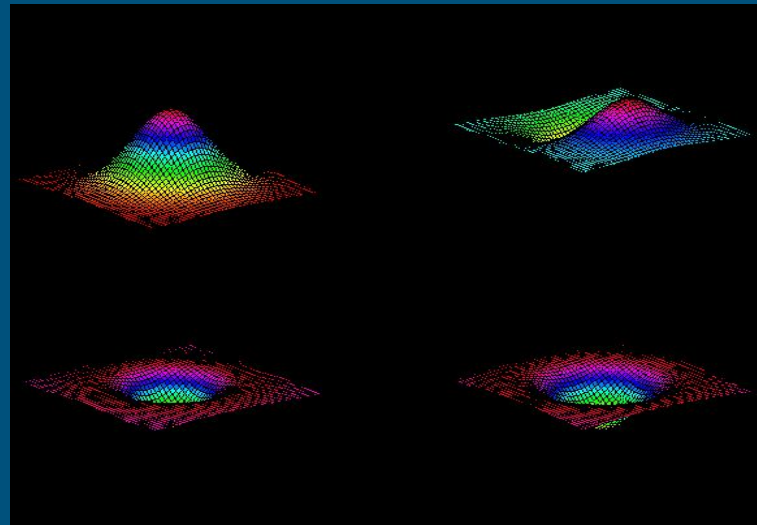
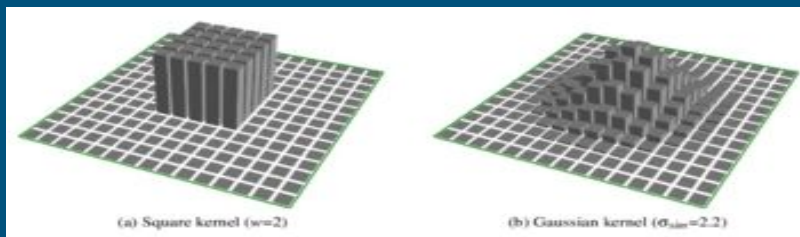
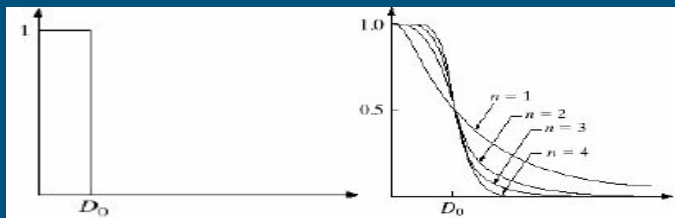


Convolução:

$$g(x, y) = \sum_{j=-a}^a \sum_{i=-b}^b w(i, j) f(x - i, y - j)$$

Visualização de Filtros

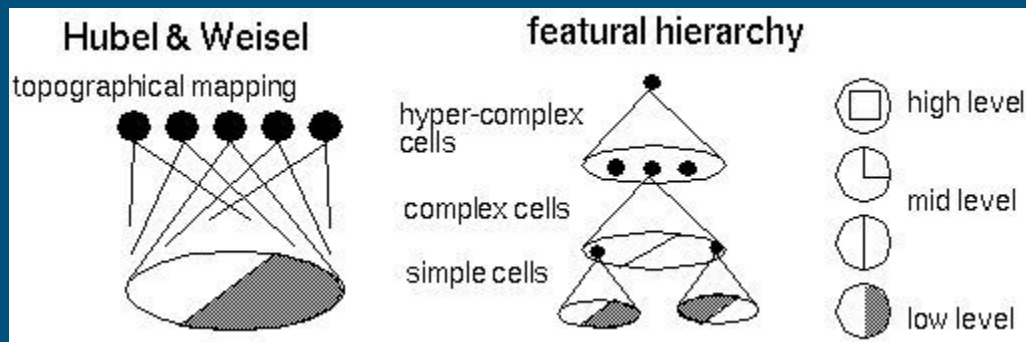
Diferentes formas de visualização



Convolutional Neural Networks: breve histórico

Uma rede neural de convolução (CNN, ou ConvNet) é uma variação de redes de múltiplas camadas de perceptrons.

Define padrões especiais de conexão entre seus neurônios inspirada pela organização do córtex visual de mamíferos.



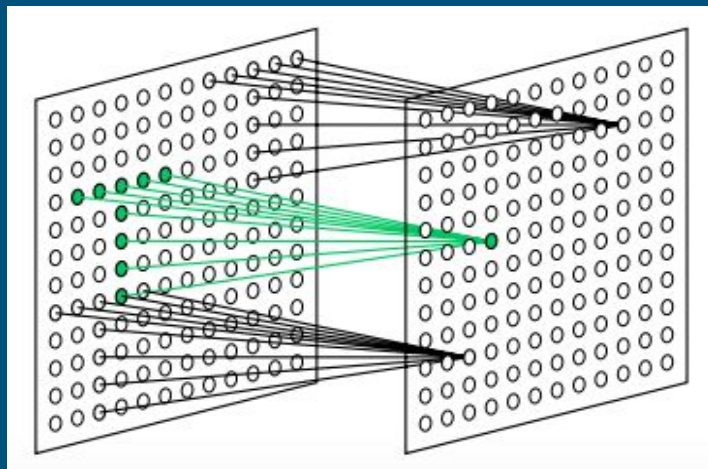
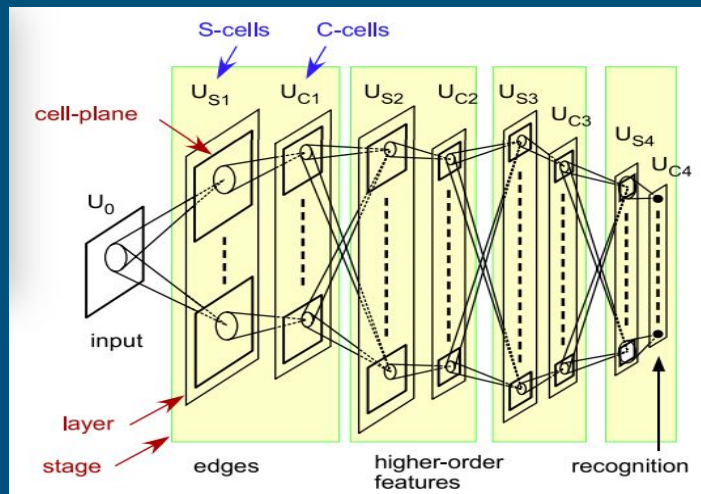
Hubel, D. H.; Wiesel, T. N. (1959). "Receptive fields of single neurones in the cat's striate cortex". The Journal of Physiology 148 (3): 574–591

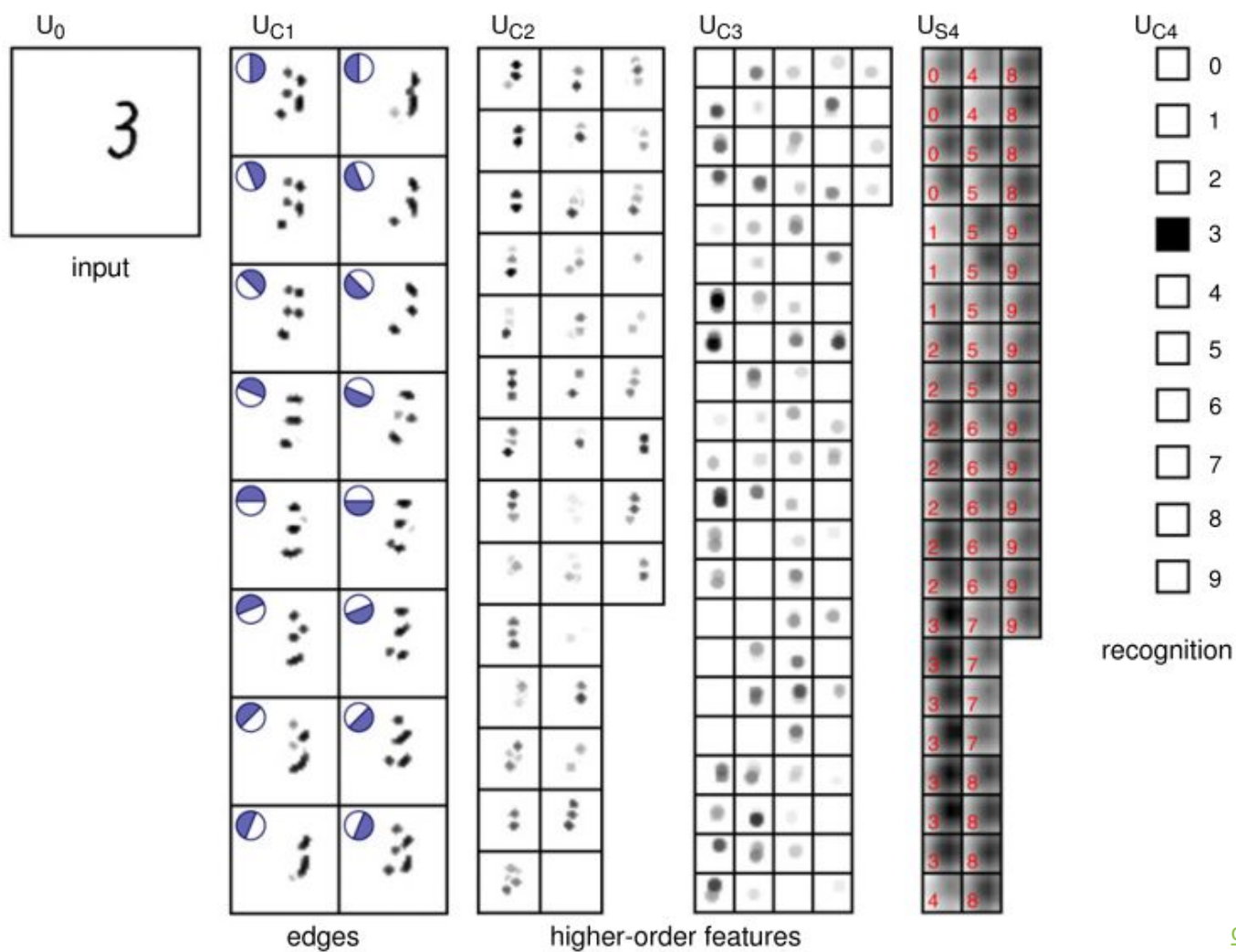
Convolutional Neural Networks: breve histórico

Nerocognition (1980)

Arquitetura alternada :

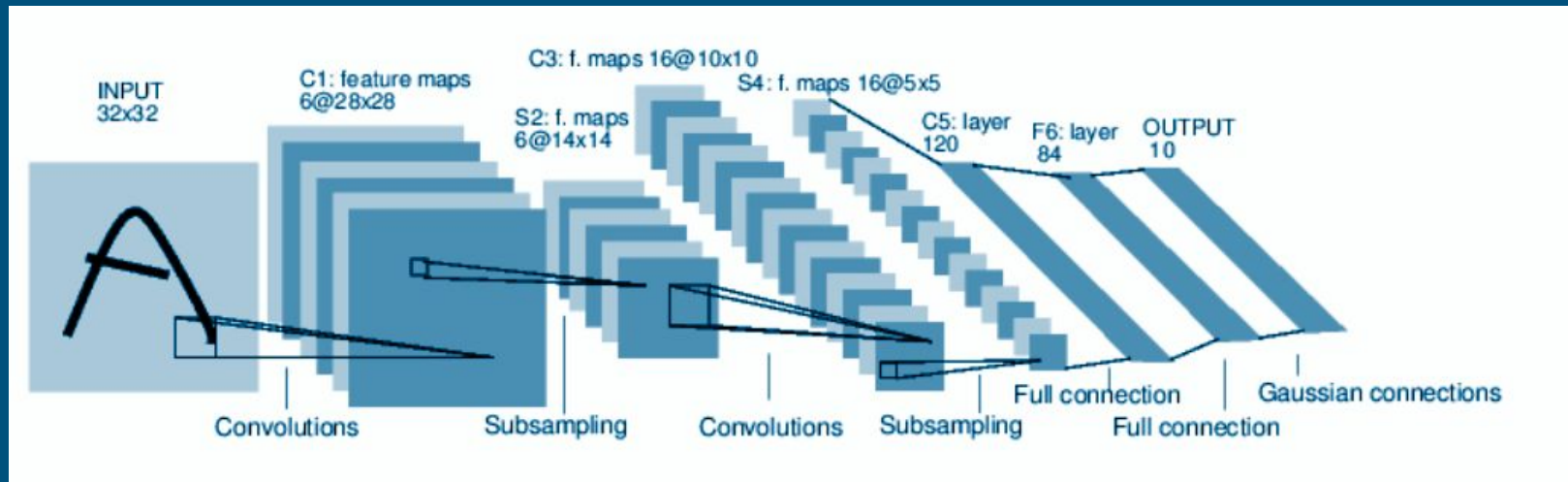
- células simples com parâmetros modificáveis (extratores de features);
- células complexas: realizando operações de *pooling* (agrupamento)





Convolutional Neural Networks: LeNet-5

Uso comercial: reconhecimento de anotações em milhares de cheques por dia



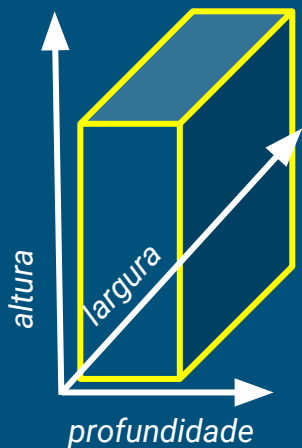
Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. Neural Computation 1989.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, november 1998

Componentes CNN:

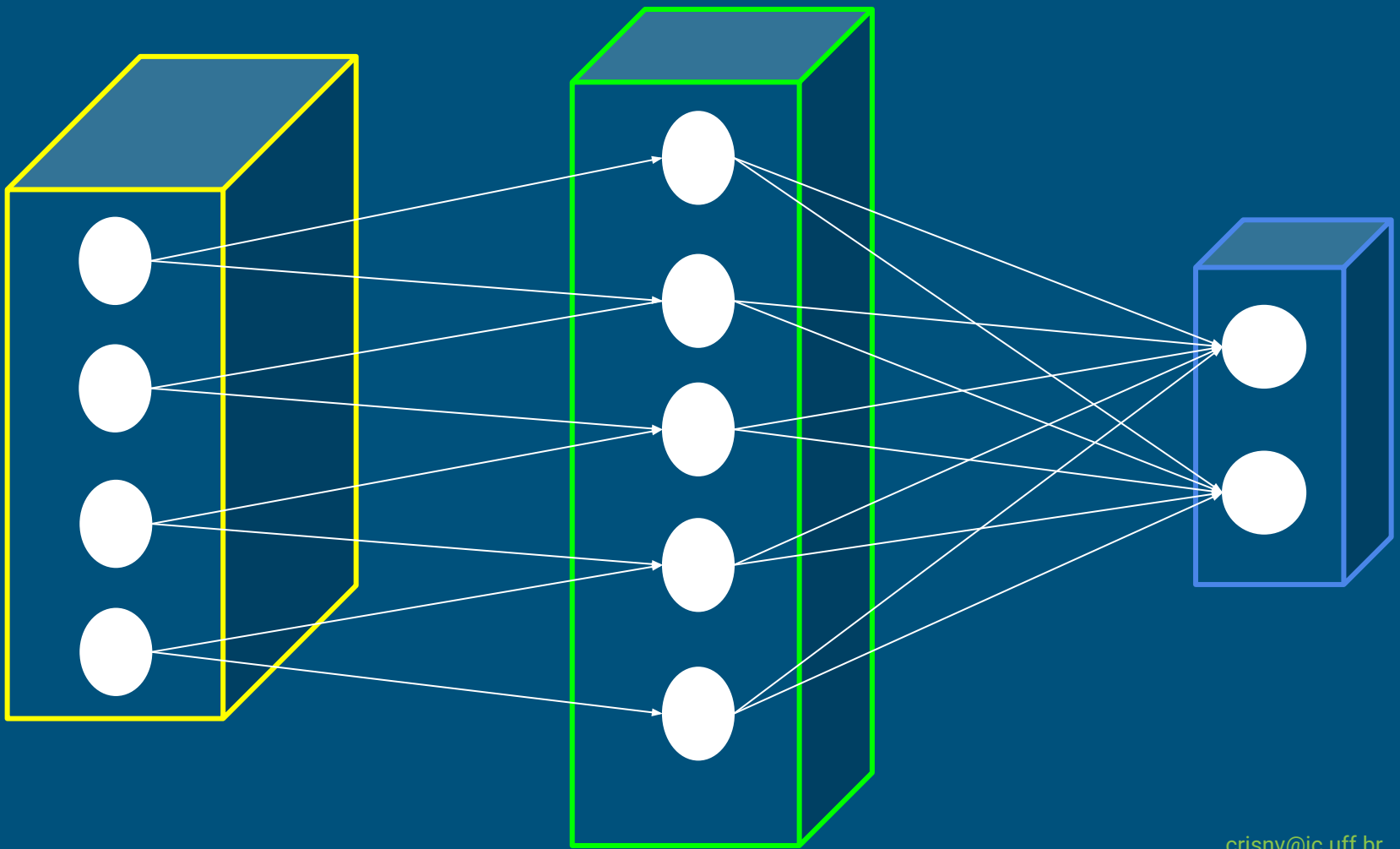
Ativações dispostas em 3 dimensões
ao longo da rede:

altura x largura X profundidade



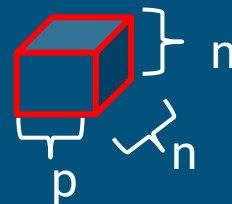
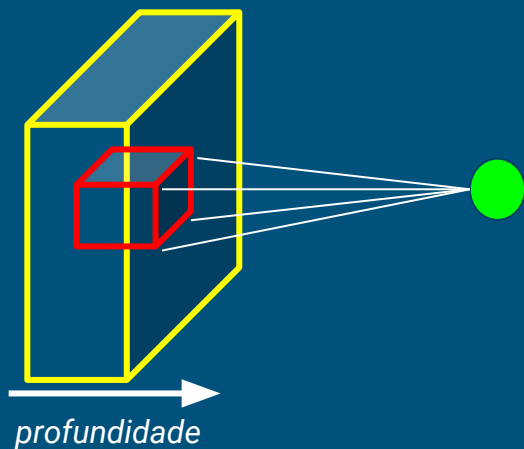
Ex: camada de entrada





Componentes CNN: campos receptivos locais

Conectividade Local: um neurônio em uma camada é ligado a um subconjunto de saídas da camada anterior

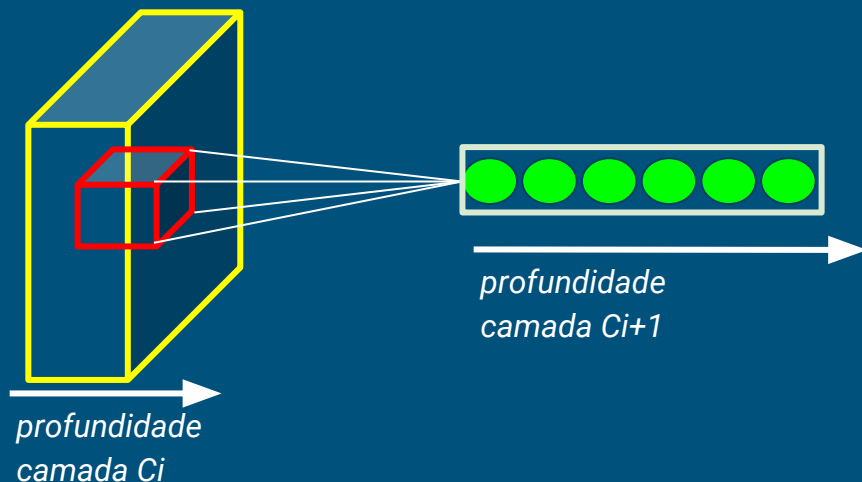


*onde: a conectividade é local em espaço,
mas completa em profundidade*

logo $n*n*p$ pesos

Componentes CNN:

- Conectividade Local

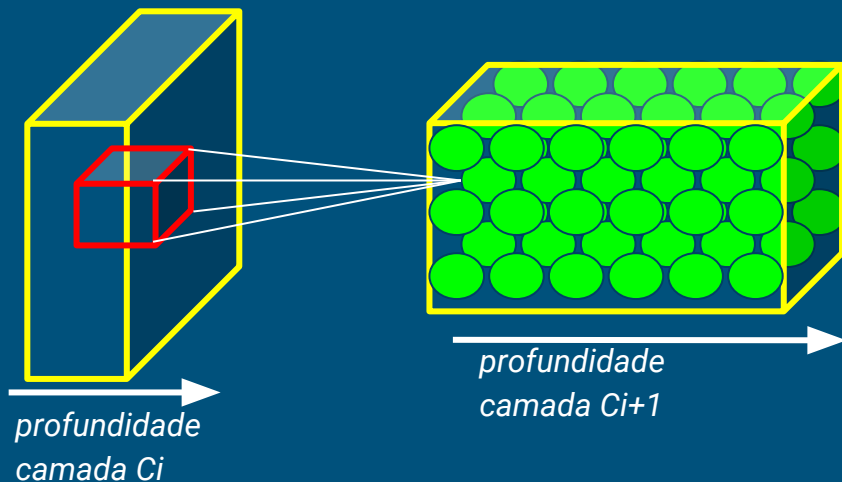


onde: múltiplos neurônios podem ser conectados a mesma região do volume de entrada para sua camada, organizados na dimensão profundidade

Ex: produz volume com profundidade 6

Componentes CNN:

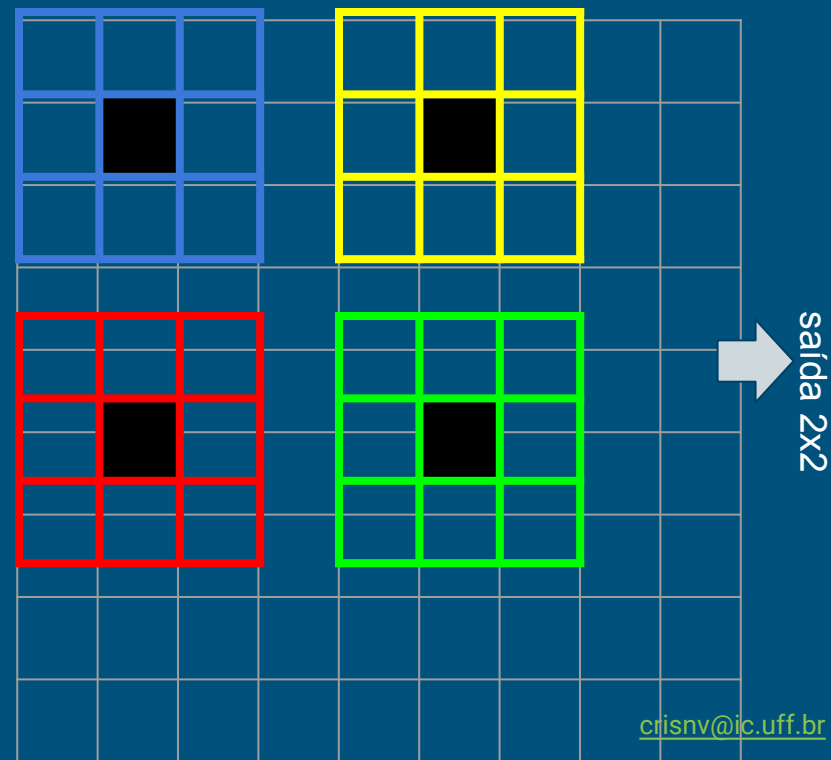
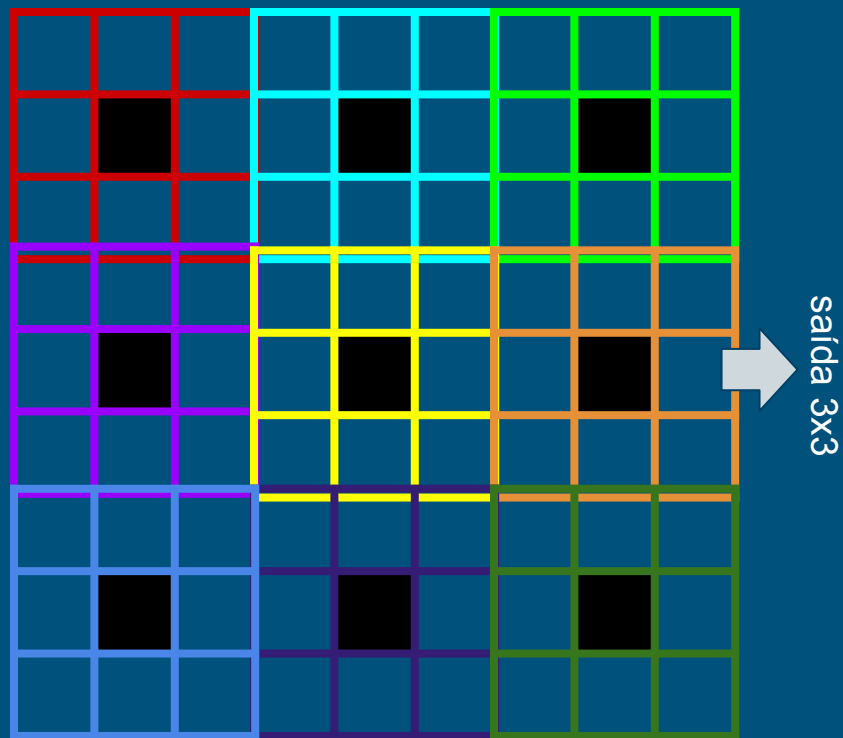
- Conectividade Local: replicamos o conjunto de neurônios ao longo das outras duas dimensões do espaço escolhendo um passo para essa replicação (stride)



*ex: ao escolher passo 1
criamos um volume com
mesma altura e largura da
camada anterior**

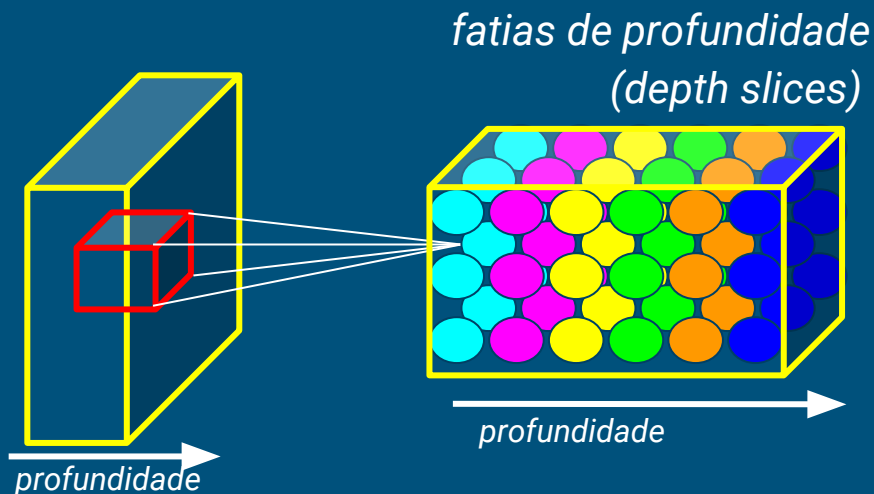
**tratamento de borda: preenchimento
(padding)*

Exemplos: dado 9x9, filtro 3x3, passo 3 e 4



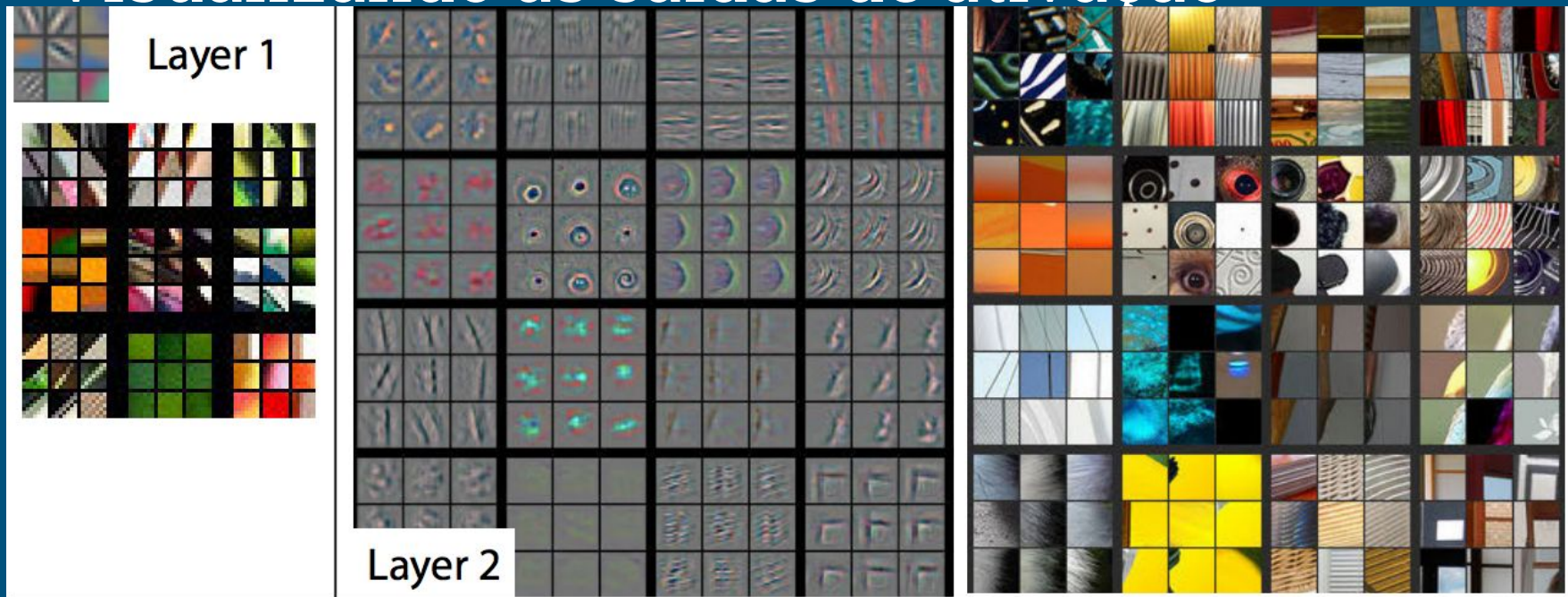
Componentes CNN:

- Compartilhamento de pesos (parâmetros) entre neurônios da mesma camada, na mesma profundidade;
- A saída produzida por uma fatia de profundidade é um mapa de ativação



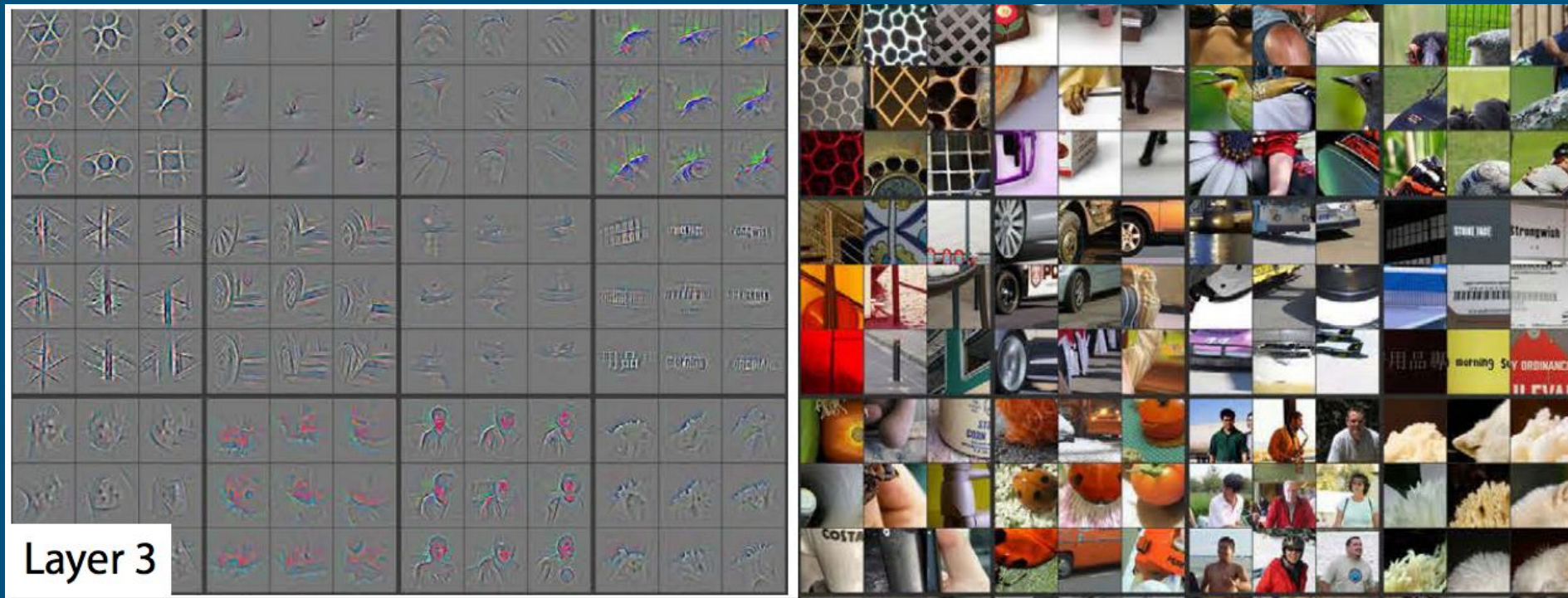
Repartição diminui drasticamente o número de parâmetros por fatia

Visualizando as saídas de ativação

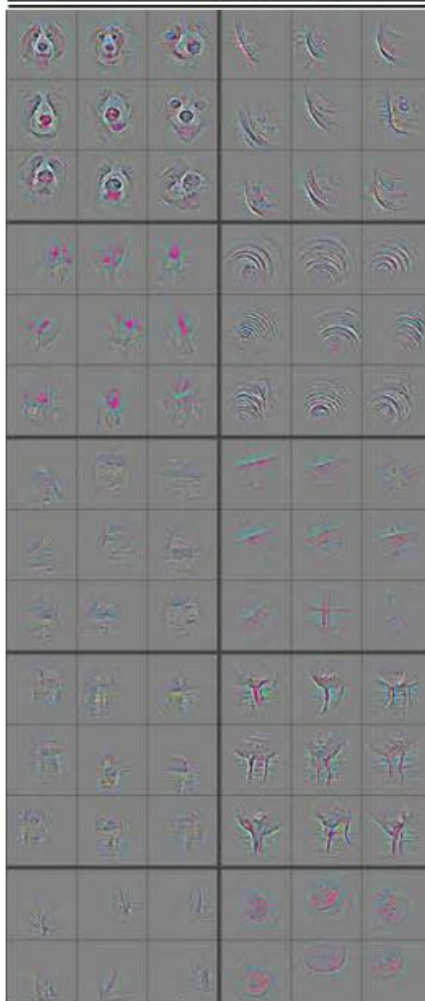


features de nível baixo

Visualizando as saídas de ativação



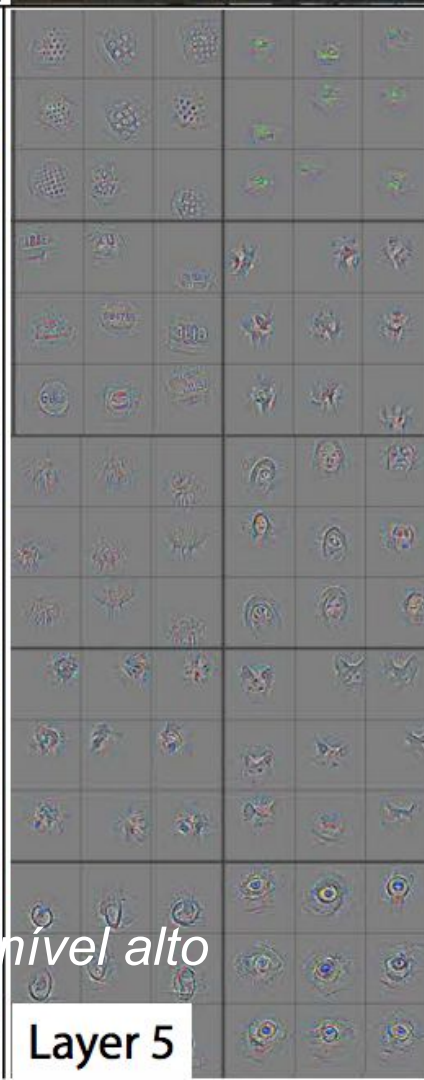
features de nível médio



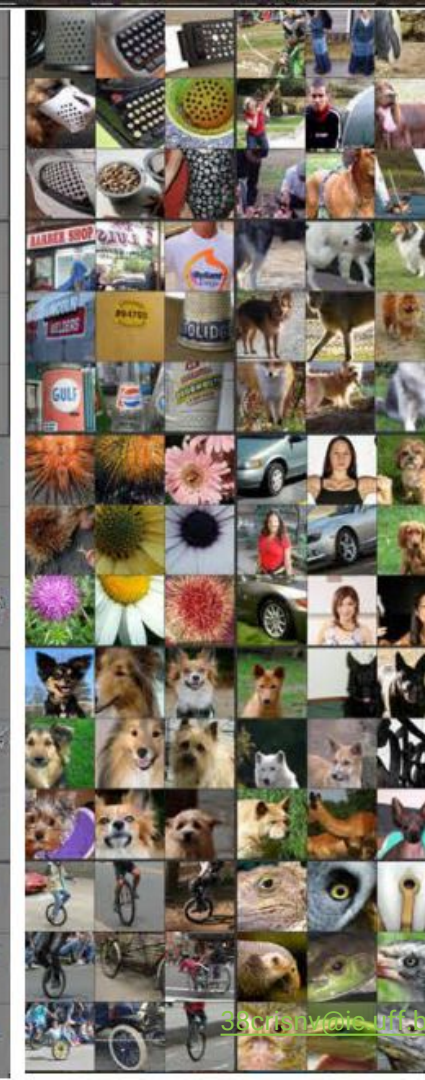
Layer 4



features de nível alto



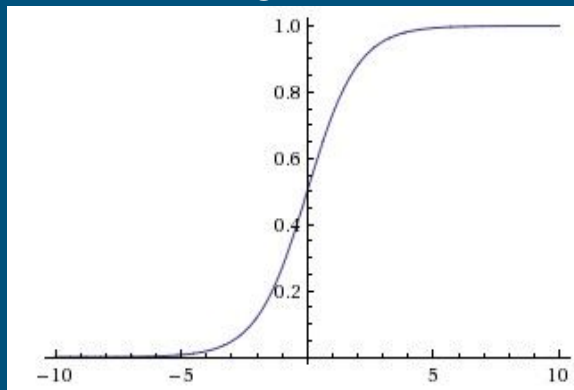
Layer 5



features de nível alto

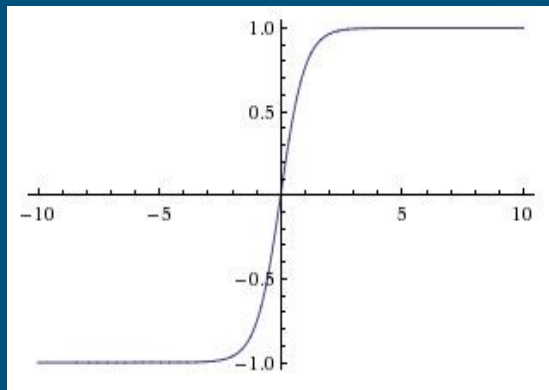
Funções de ativação: mais usadas

Sigmoid



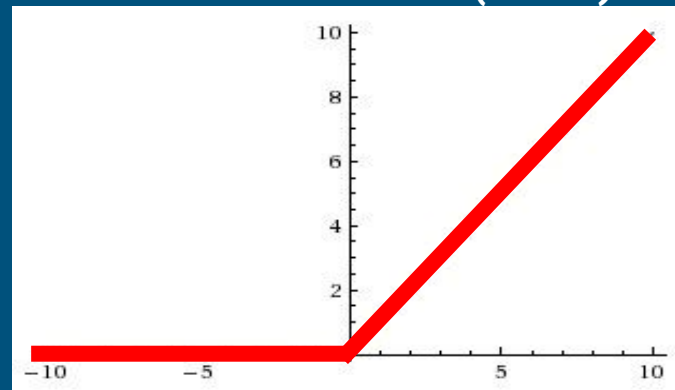
*mapeia números reais
para o intervalo [0,1]*

Tanh



*mapeia números reais para
o intervalo [-1,1].*

Rectified Linear Unit (ReLU)

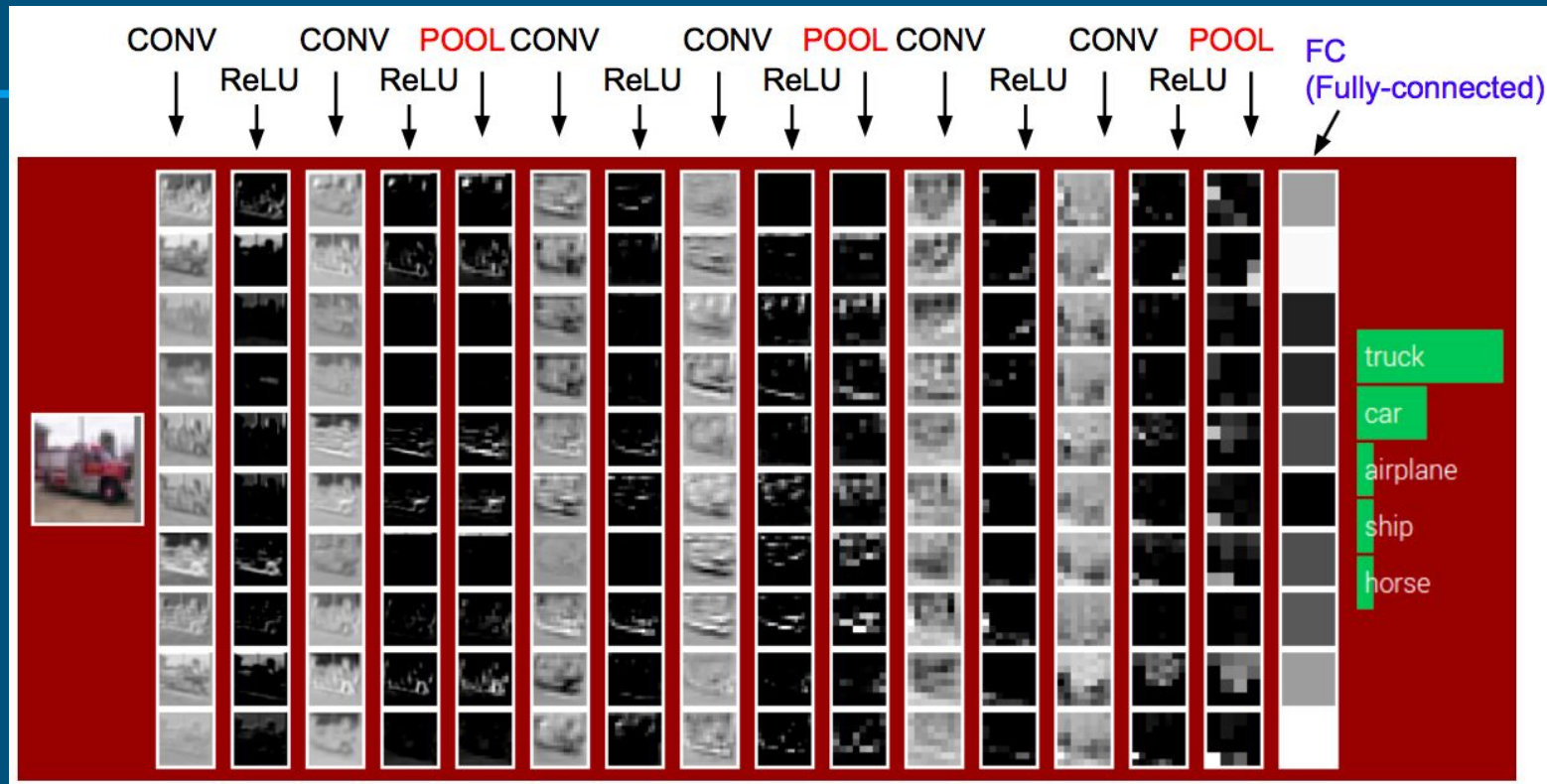


$f(x) = \max(0, x)$

Camadas de agrupamento (*Pooling layers*)



Ilustração de arquitetura completa



rede executada via browser em: <http://cs231n.stanford.edu/>

CNNs recentes

- adotam tamanhos de filtros pequenos (3x3, 2x2 e até 1x1)
- adotam tamanhos de pooling de 2x (ou menos em estratégias de pooling fracionais)
- tamanho de passo 1 (stride)
- muito profundas (MSRA 2015 - experimentos com 150 e 1200 camadas)

Arquiteturas pesadas podem ser repensadas aumentando fatores de agrupamento nas camadas de pooling ou aumentando o tamanho do passo das camadas (strides)

Fatores que impulsionaram o DL:

- desenvolvimento de algoritmos de aprendizado eficientes;
- bancos padrões de referência (benchmarks)
- função de ativação ReLU
- aumento da capacidade de processamento e poder de processamento de fins gerais utilizando hardware gráfico (GPU)

Conclusões

Vantagens de abordagens com aprendizado profundo

Robustez:

- uma hierarquia de features é aprendida de maneira ótima para a tarefa proposta
- mostra-se capaz de aprender as complexas variações existentes nos dados

Generalizável:

- a mesma arquitetura pode ser usada em diferentes tarefas e tipos de dados

Escalável:

- mais dados levam ao aumento de performance
- método massivamente paralelizável

Aplicações

Reconhecimento de números a partir de imagens do Google Street View



I. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, V. Shet. Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks. ICLR2014, ICLR2014

Reconhecimento de sinais de trânsito

IJCNN 2012 : Committes de CNNs 99.46 % X humanos 98.84 %



D. Ciresan, U. Meier, J. Masci, J. Schmidhuber. A Committee of Neural Networks for Traffic Sign Classification. IJCNN 2011

Caligrafia offline

ICDAR 2011, 2013



Places (205 categorias e 2.5 milhões de imagens)

categoryName (ImageNo), more images will be shown when click the category name.

a_abbey(46368)



a_airfield(10910)



a_airplane_cabin(5152)



a_airport_terminal(16174)

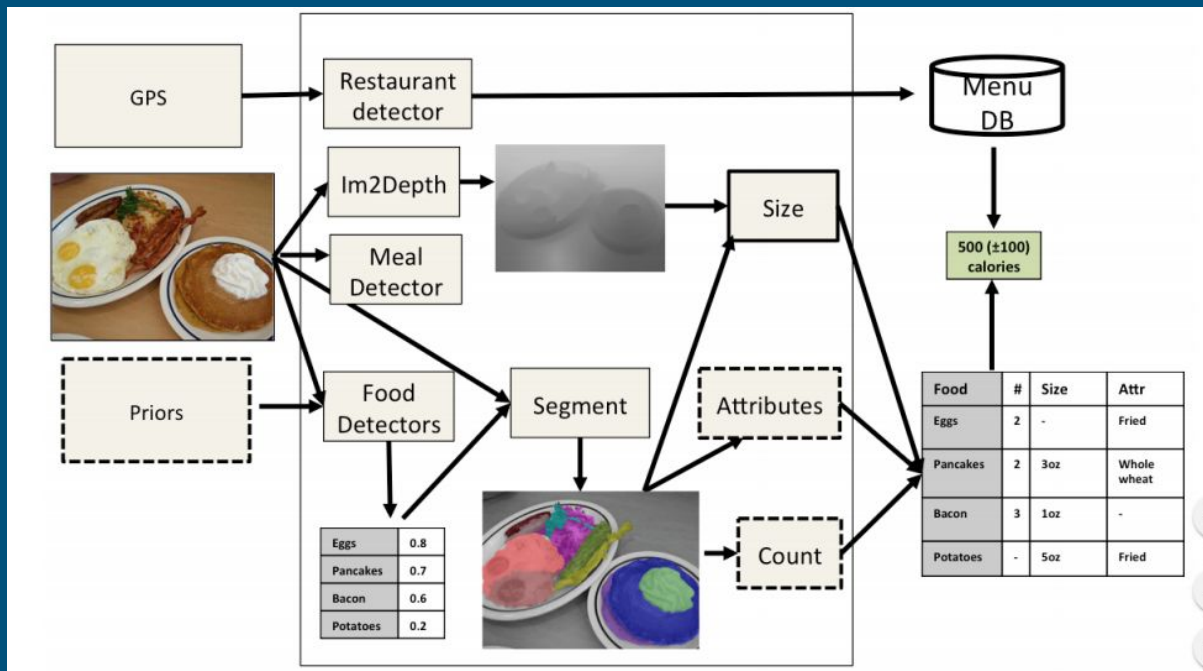


a_alcove(4966)



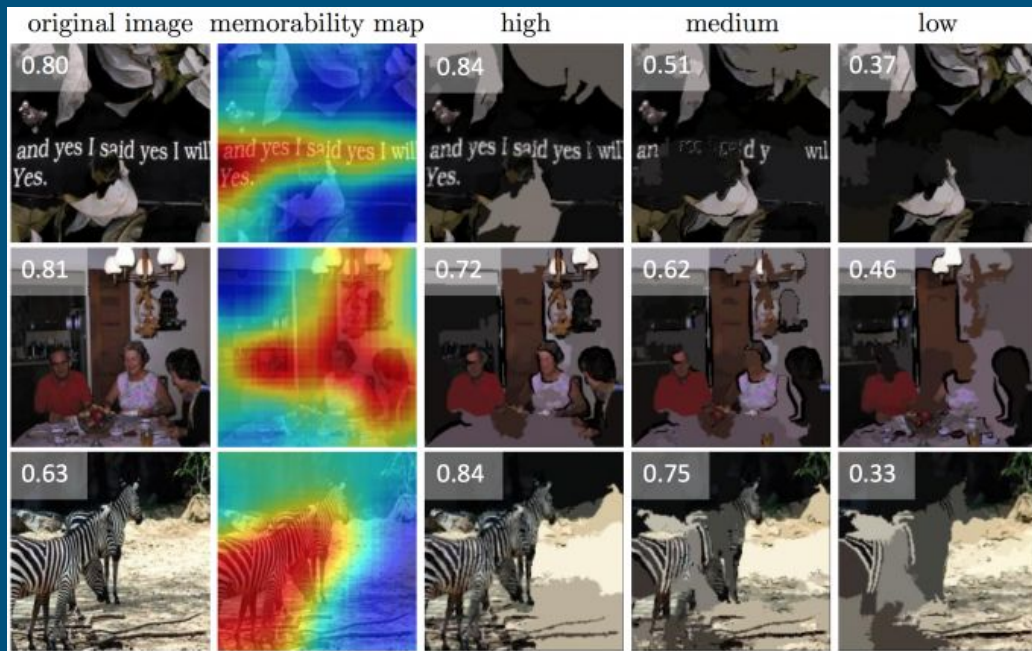
<http://places.csail.mit.edu/demo.html> B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. “Learning Deep Features for Scene Recognition using Places Database.” *Advances in Neural Information Processing Systems 27 (NIPS)*, 2014.

Im2calories



A. Myers, N. Johnston, et al. Im2Calories: towards an automated mobile vision food diary. ICCV 2015

Photo-memorability

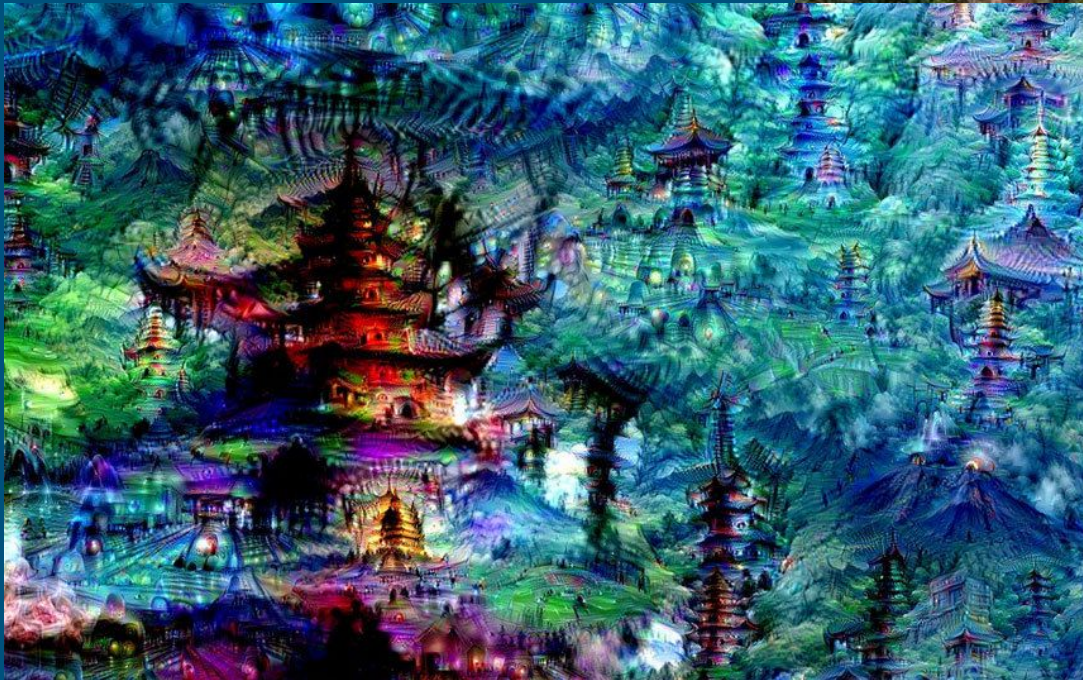


<http://memorability.csail.mit.edu/demo.html>

A. Khosla, A. S. Raju, A. Torralba and A. Oliva Understanding and Predicting Image Memorability at a Large Scale
ICCV 2015

Deep dreamer

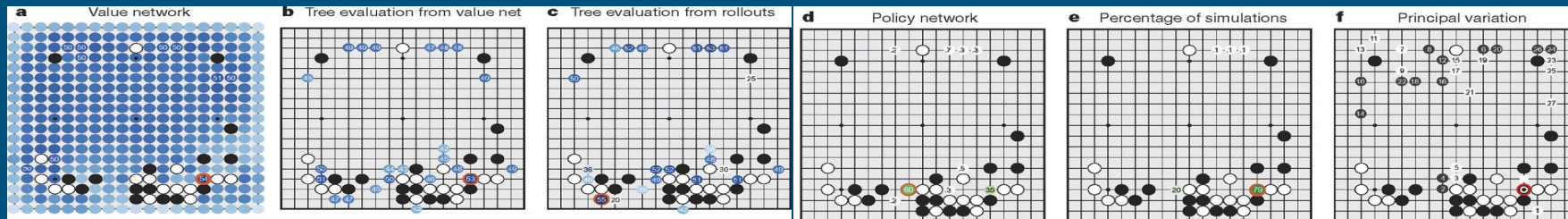
<http://deepdreamgenerator.com/>



Aprendizado Profundo por Reforço



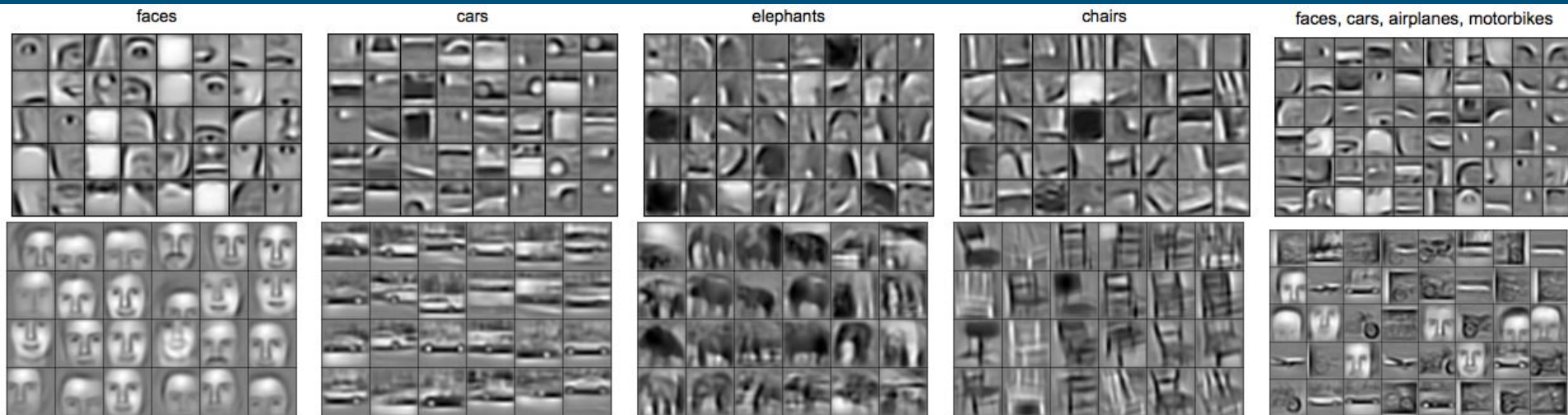
V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou D. Wierstra M. Riedmiller. Playing Atari with Deep Reinforcement Learning. NIPS Deep Learning Workshop, 2013



David Silver, Aja Huang et al. Mastering the game of Go with deep neural networks and tree search. Nature (Jan. 2016)

Aprendizado profundo não-supervisionado

Dados não anotados correspondem a maioria do que é produzido diariamente.



H. Lee, R. Grosse, R. Ranganath, A. Y. Ng. Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations. ICML 2009

obrigada!
crisnv@ic.uff.br

