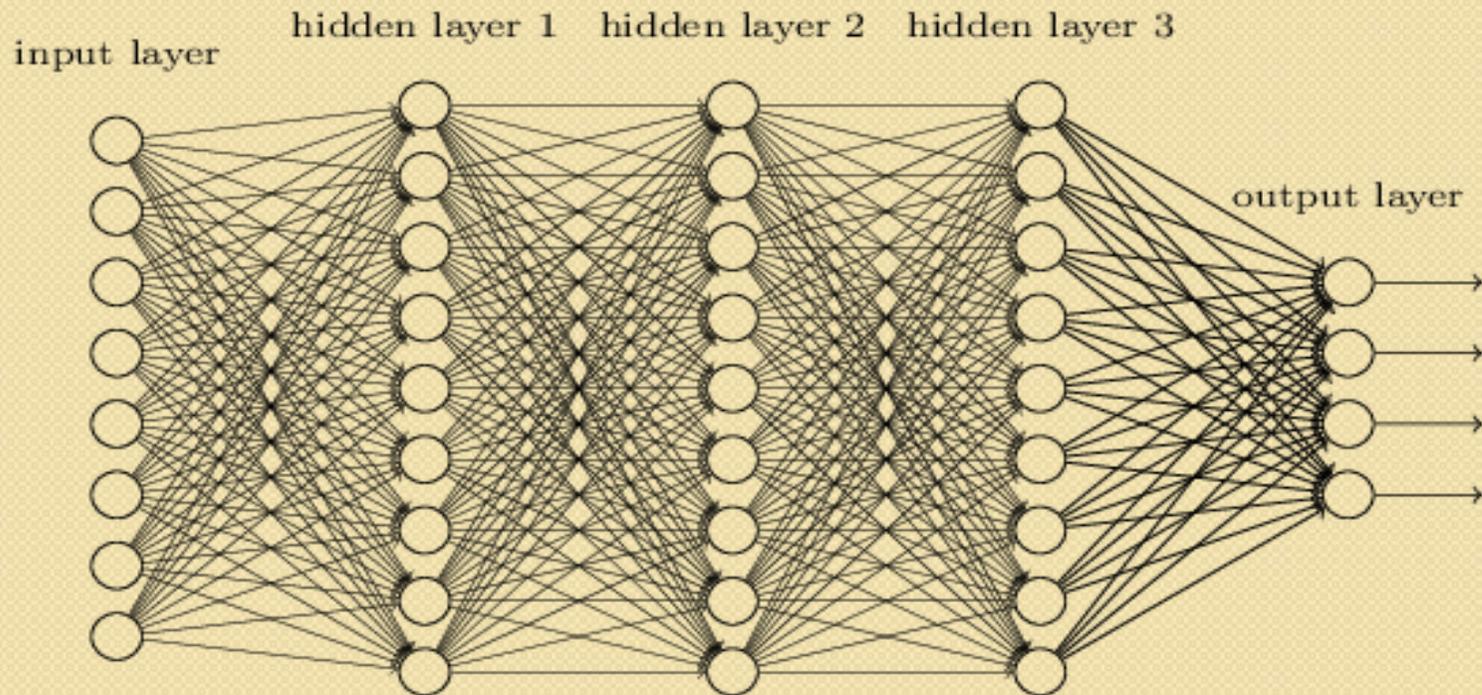


Redes Convolucionais e *Deep Learning*

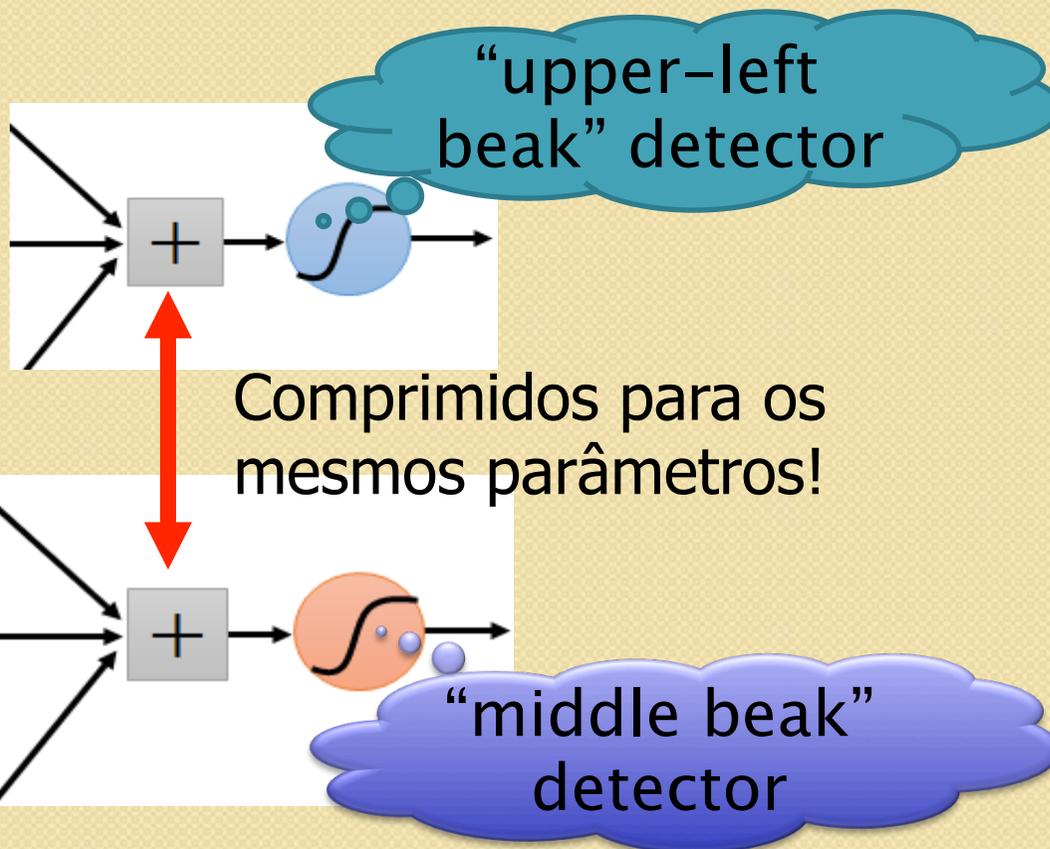
Germano C. Vasconcelos
Centro de Informática – UFPE

Estrutura Feedforward Tradicional

- Rede Totalmente Conectada (Fully Connected)

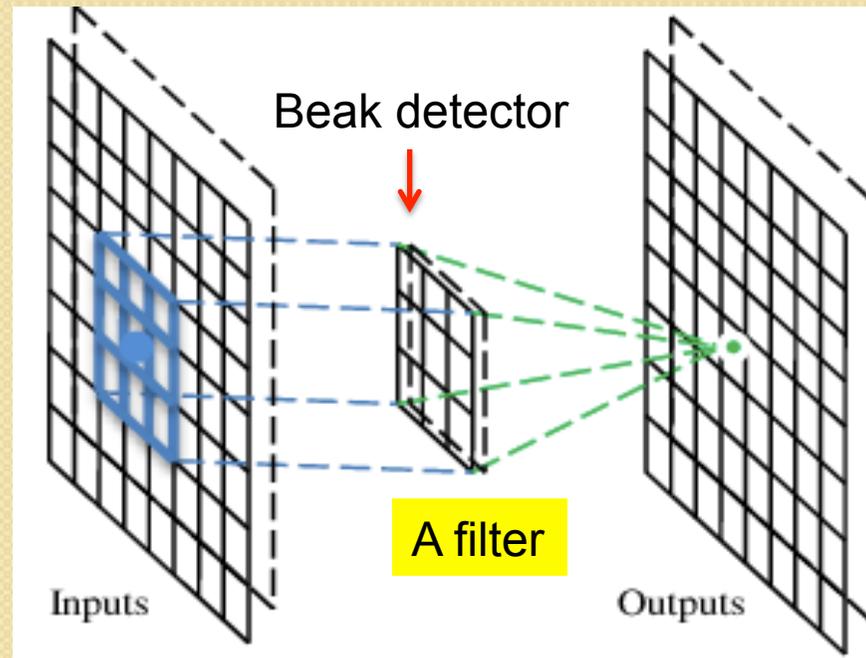


Como Lidar com Padrões em Áreas Diferentes? Que tal detectores em áreas diferentes?



1o Passo: Camada Convolutiva

- CNN: rede neural com camadas convolucionais e outras camadas
- Camada convolutiva: filtros que fazem operação de convolução



Convolução

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

Parâmetros a serem aprendidos!

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

Filtro 1

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

Filtro 2

⋮ ⋮

Cada filtro detecta
pequeno padrão(3 x 3)

Convolução

stride=1

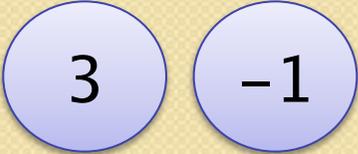
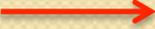
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

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

Filter 1

Dot product



Convolução

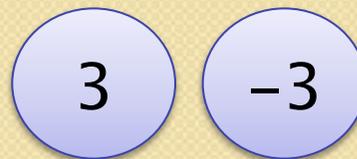
If stride=2

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

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

Filter 1



Convolução

stride=1

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

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

Filtro 1

3	-1	-3	-1
-3	1	0	-3
-3	-3	0	1
3	-2	-2	-1

Convolução

stride=1

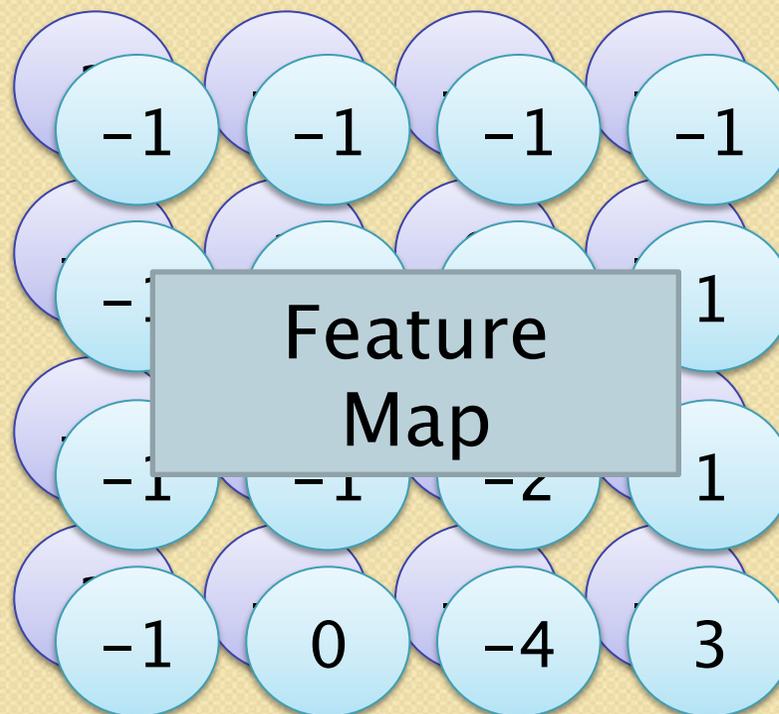
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

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

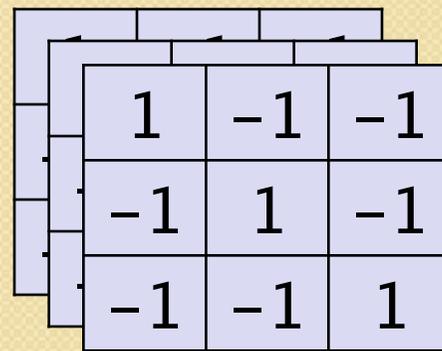
Filtro 2

Repita para cada filtro

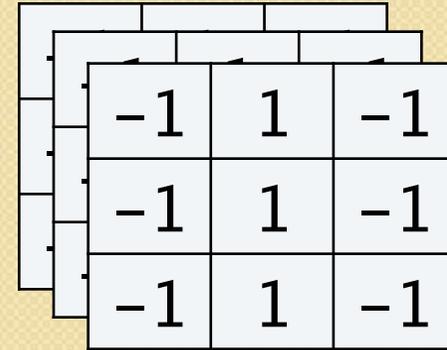


Duas imagens 4 x 4
Formando matriz 2 x 4 x 4

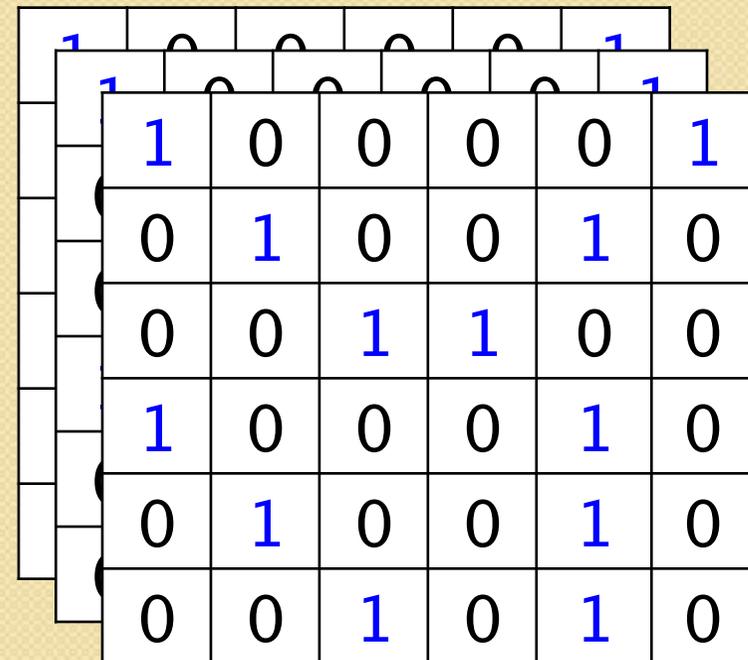
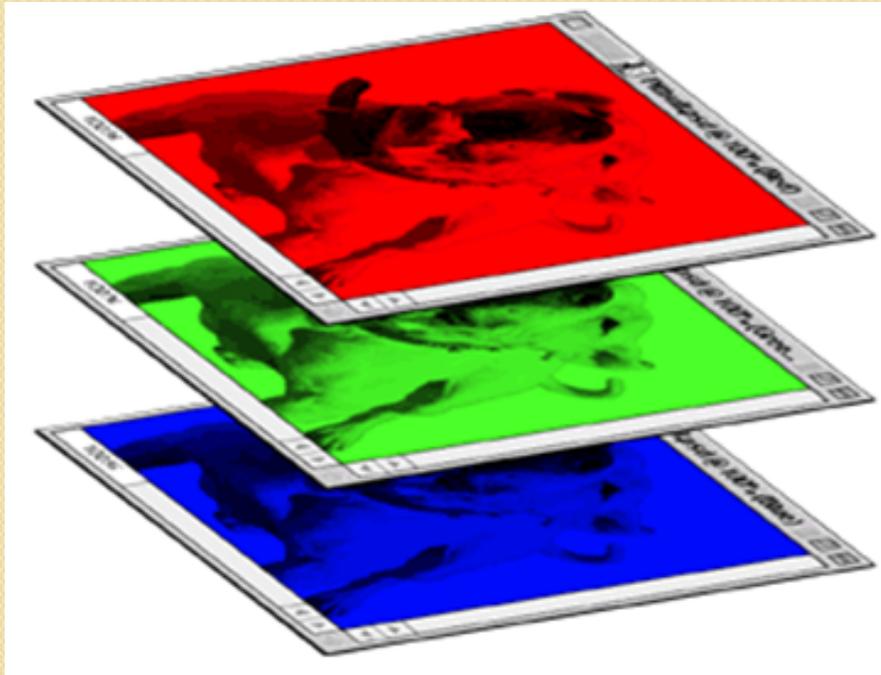
Imagem Colorida: 3 Canais RG



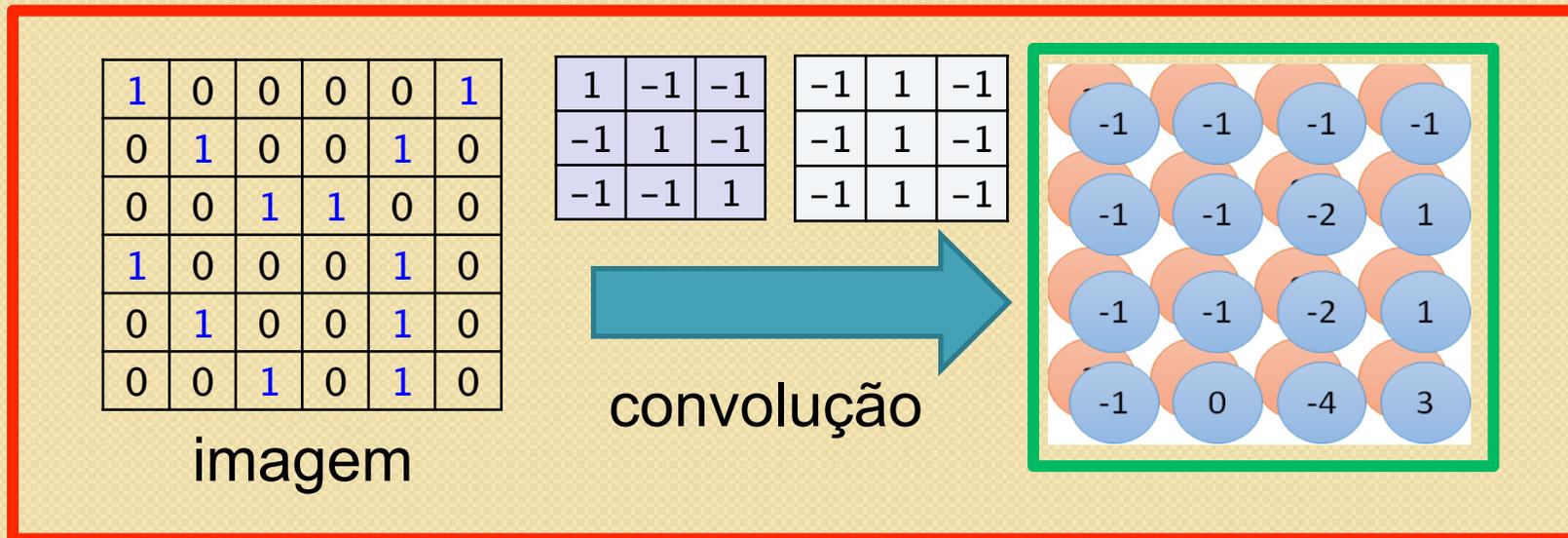
Filtro 1



Filtro 2

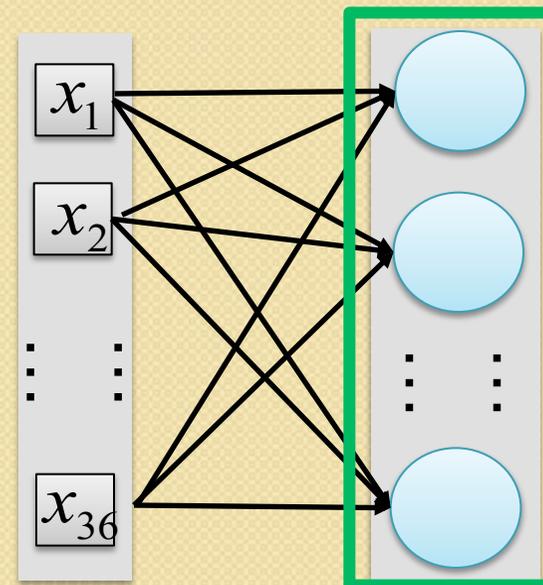


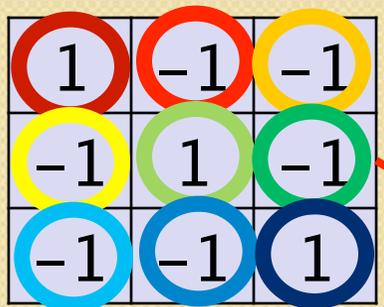
Convolução vs Totalmente Conectada



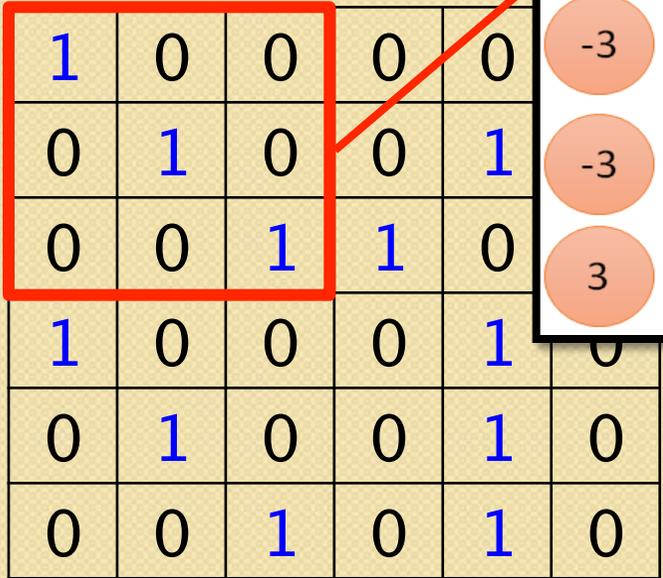
Fully-connected

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0



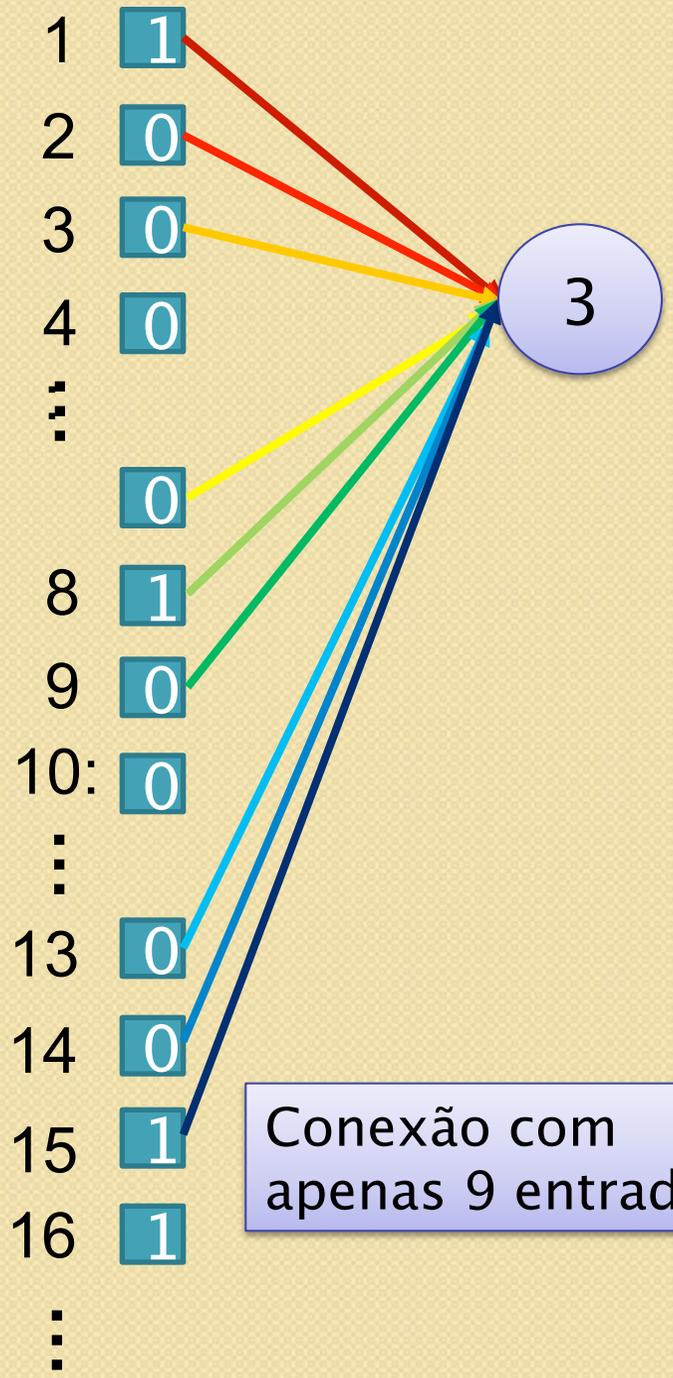
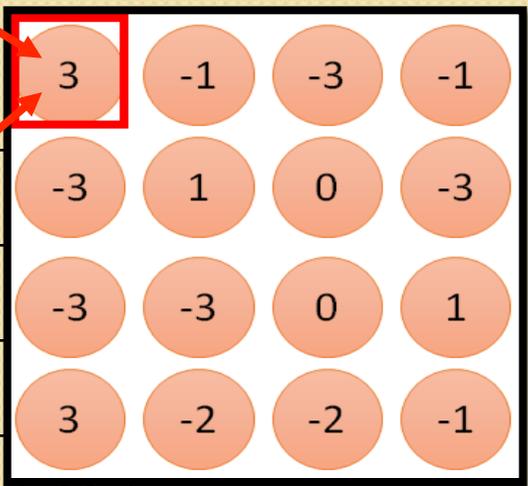


Filter 1

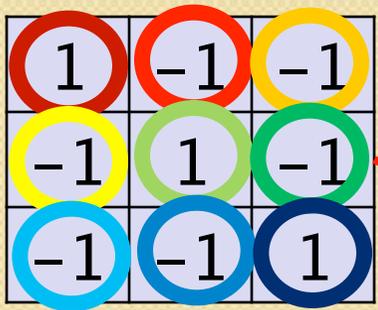


6 x 6 image

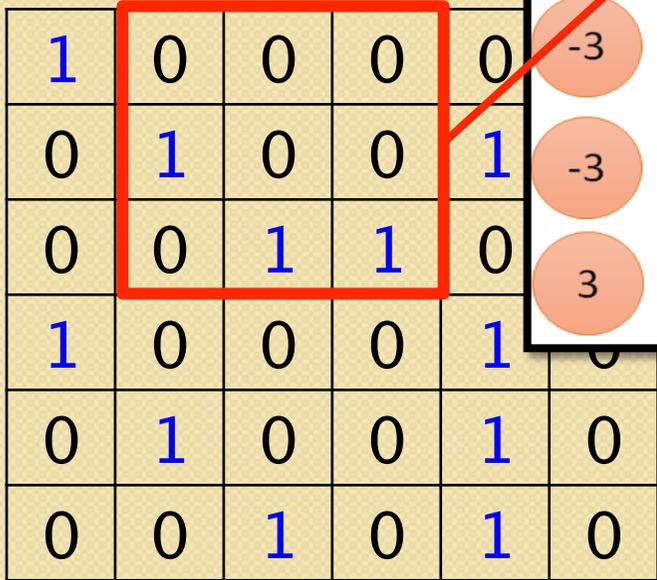
Menos parâmetros



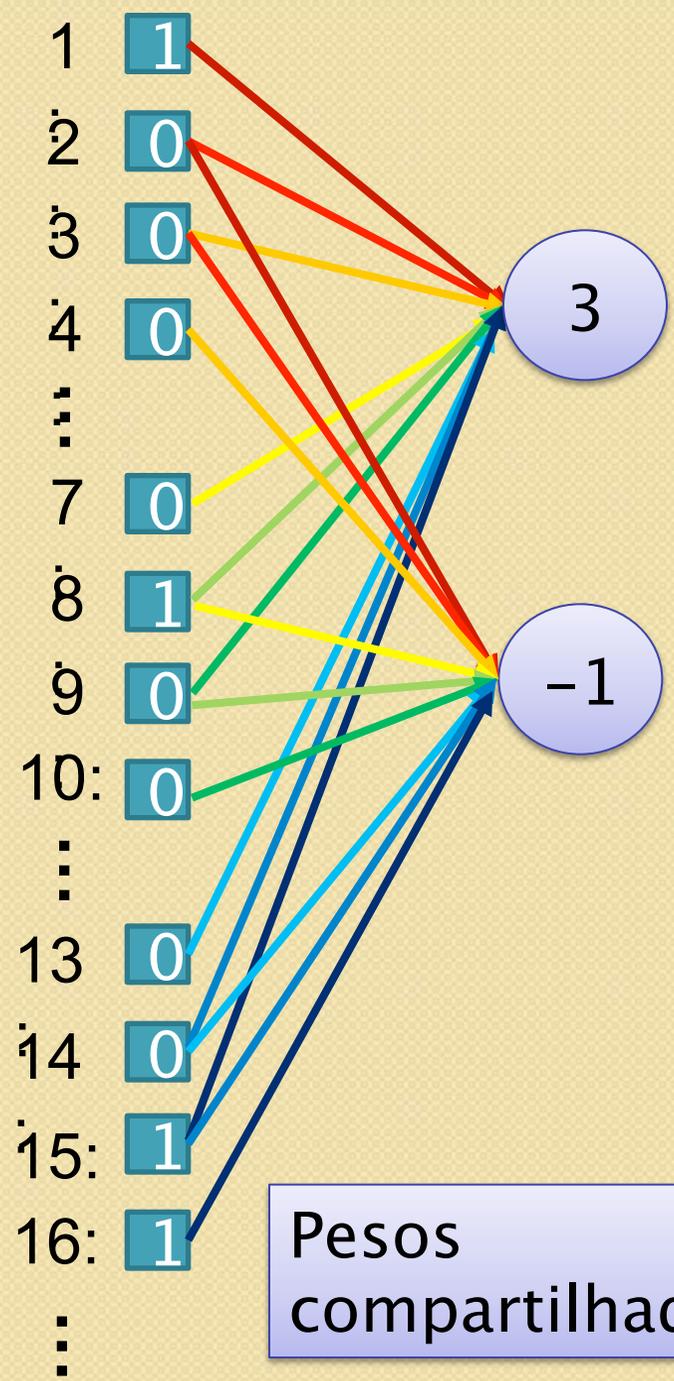
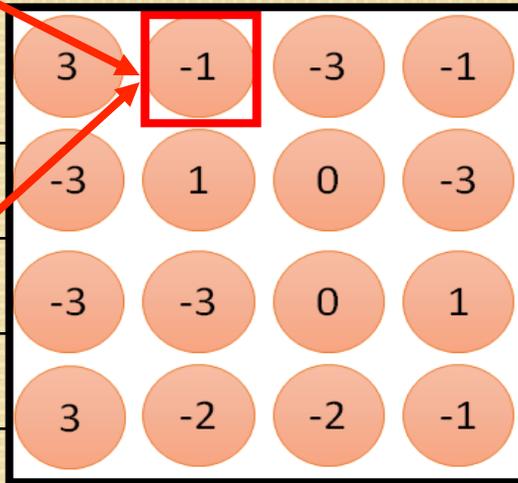
Conexão com apenas 9 entradas



Filter 1



6 x 6 image



Pesos compartilhados

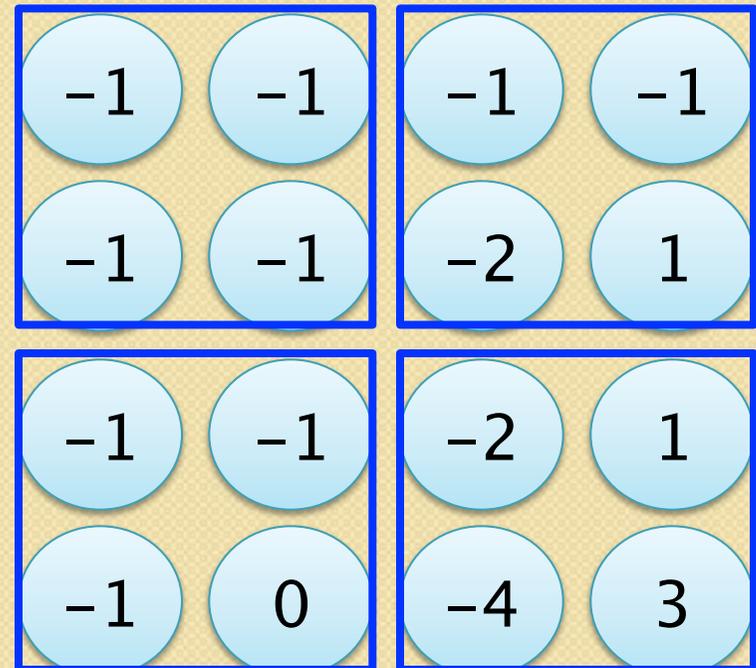
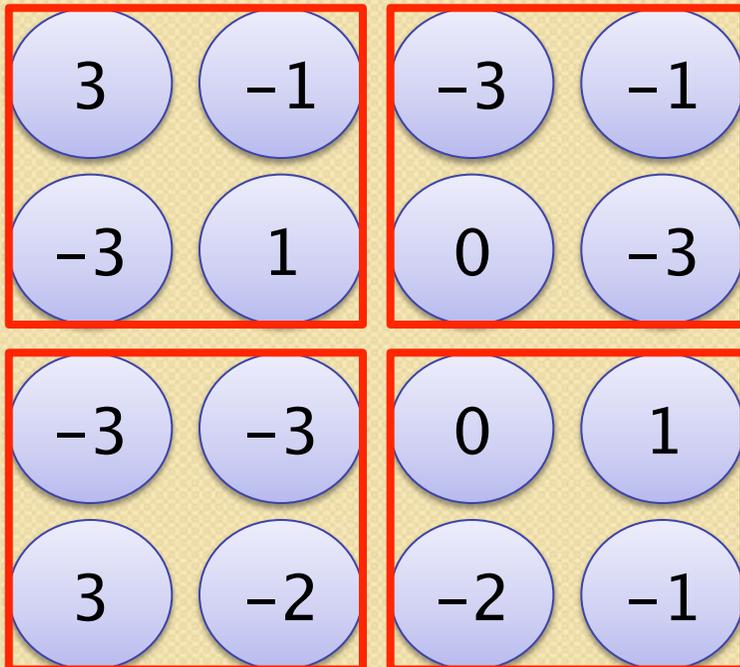
Max Pooling

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

Filtro 1

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

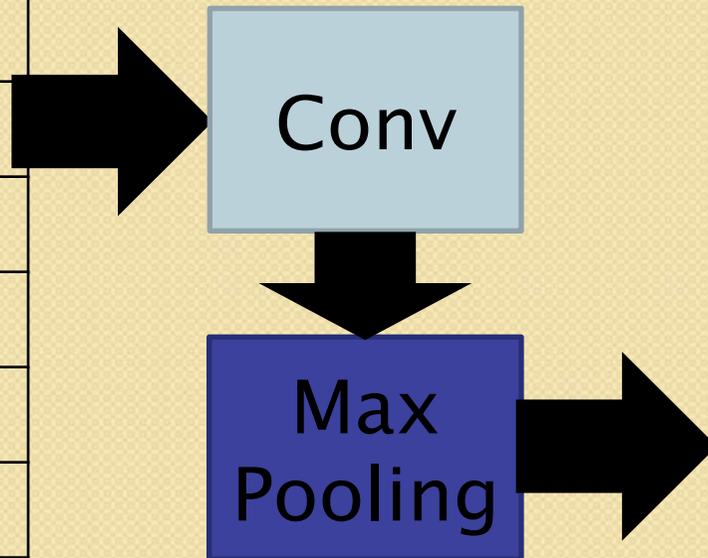
Filtro 2



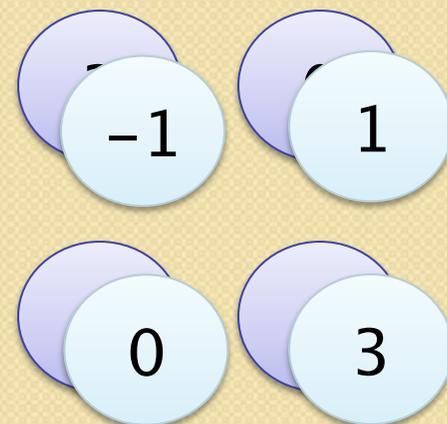
Max Pooling

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image



Nova imagem menor



2 x 2 image

Cada filtro é um canal

Convolução...

Main CNN idea for text:

Compute vectors for n-grams and group them afterwards

Example: "this takes too long" compute vectors for:

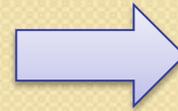
This takes, takes too, too long, this takes too, takes too long, this takes too long

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Input matrix

1	0	1
0	1	0
1	0	1

Convolutional
3x3 filter



1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Image

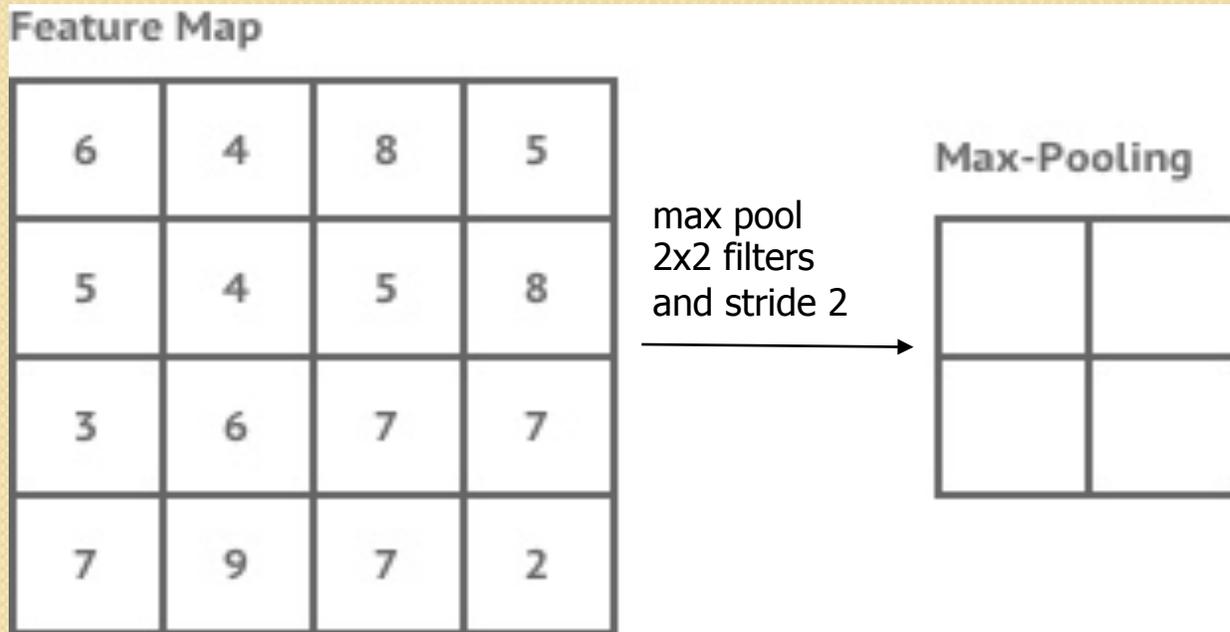
4		

Convolved
Feature

http://deeplearning.stanford.edu/wiki/index.php/Feature_extraction_using_convolution

Max Pooling

Main CNN idea for text:
Compute vectors for n-grams and **group them afterwards**



Por Que Fazer *Pooling*?

- Subsampling não muda o objeto

pássaro



Subsampling

pássaro

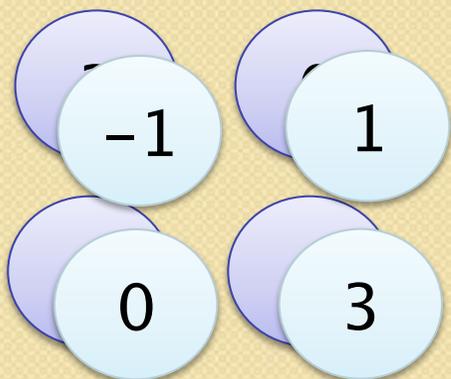


Subamostrar pixels para tornar imagem menor



Menos parâmetros para caracterizar a imagem

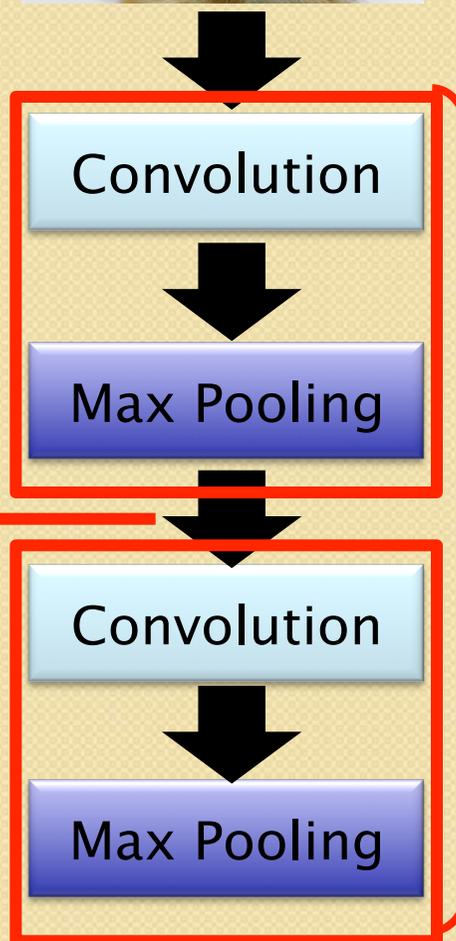
CNN Completa



Uma nova imagem

Menor que a original

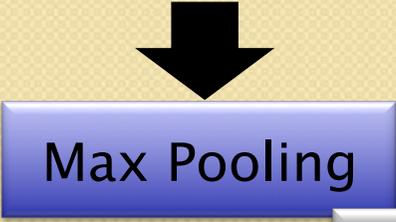
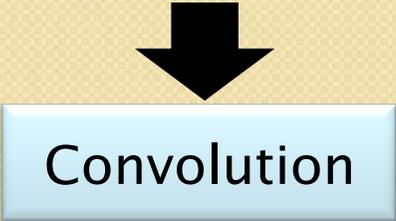
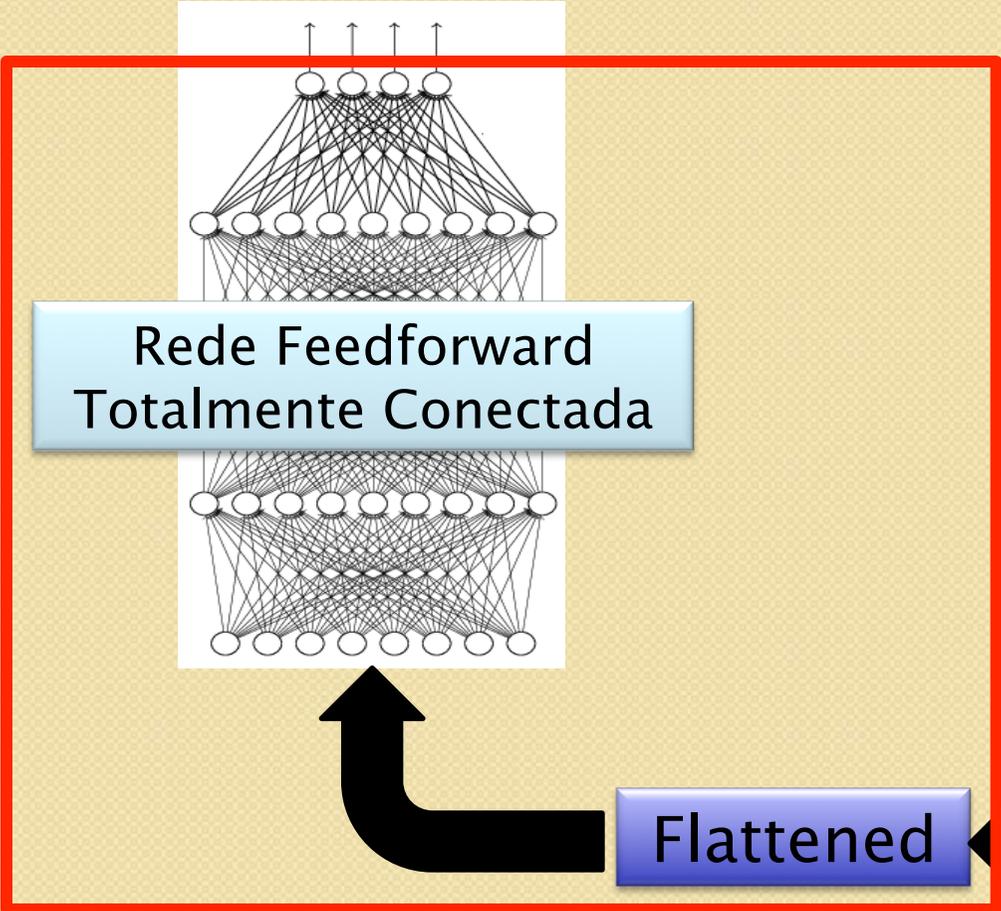
Número de canais=Número de Filtros



Pode se repetir várias vezes

CNN Completa

cat dog



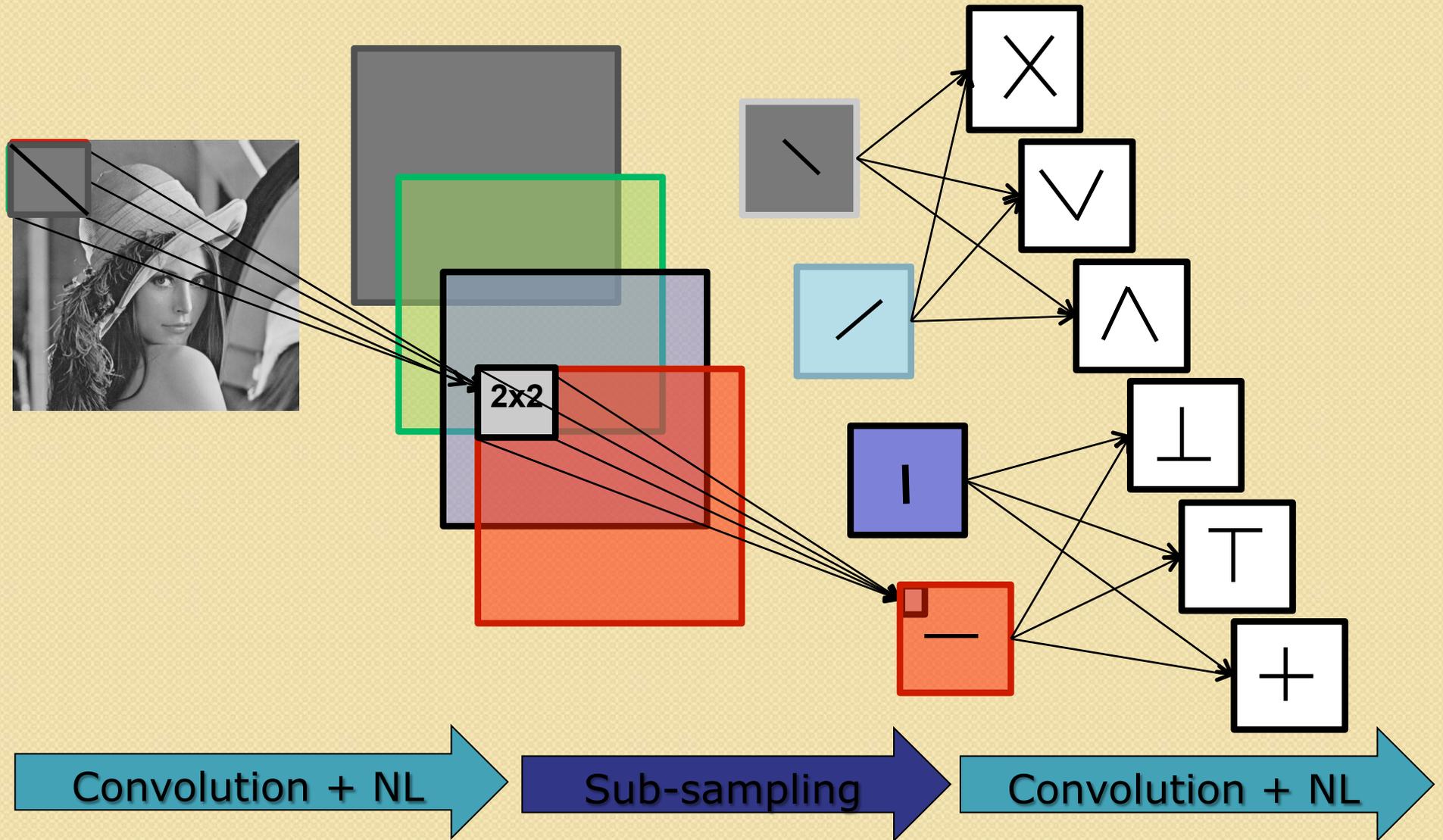
Nova imagem



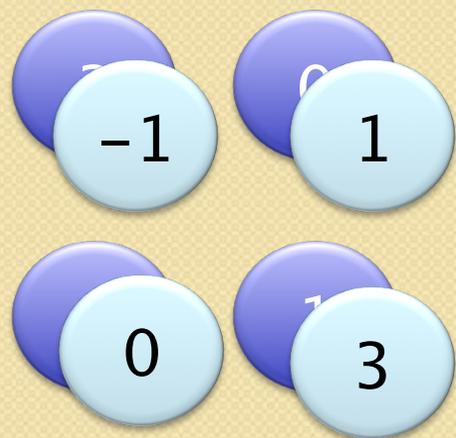
Nova imagem

Flattened

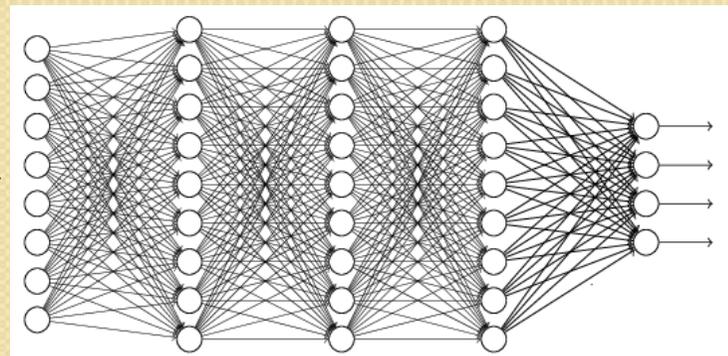
Rede Neural CNN



Achatamento (Flattening)

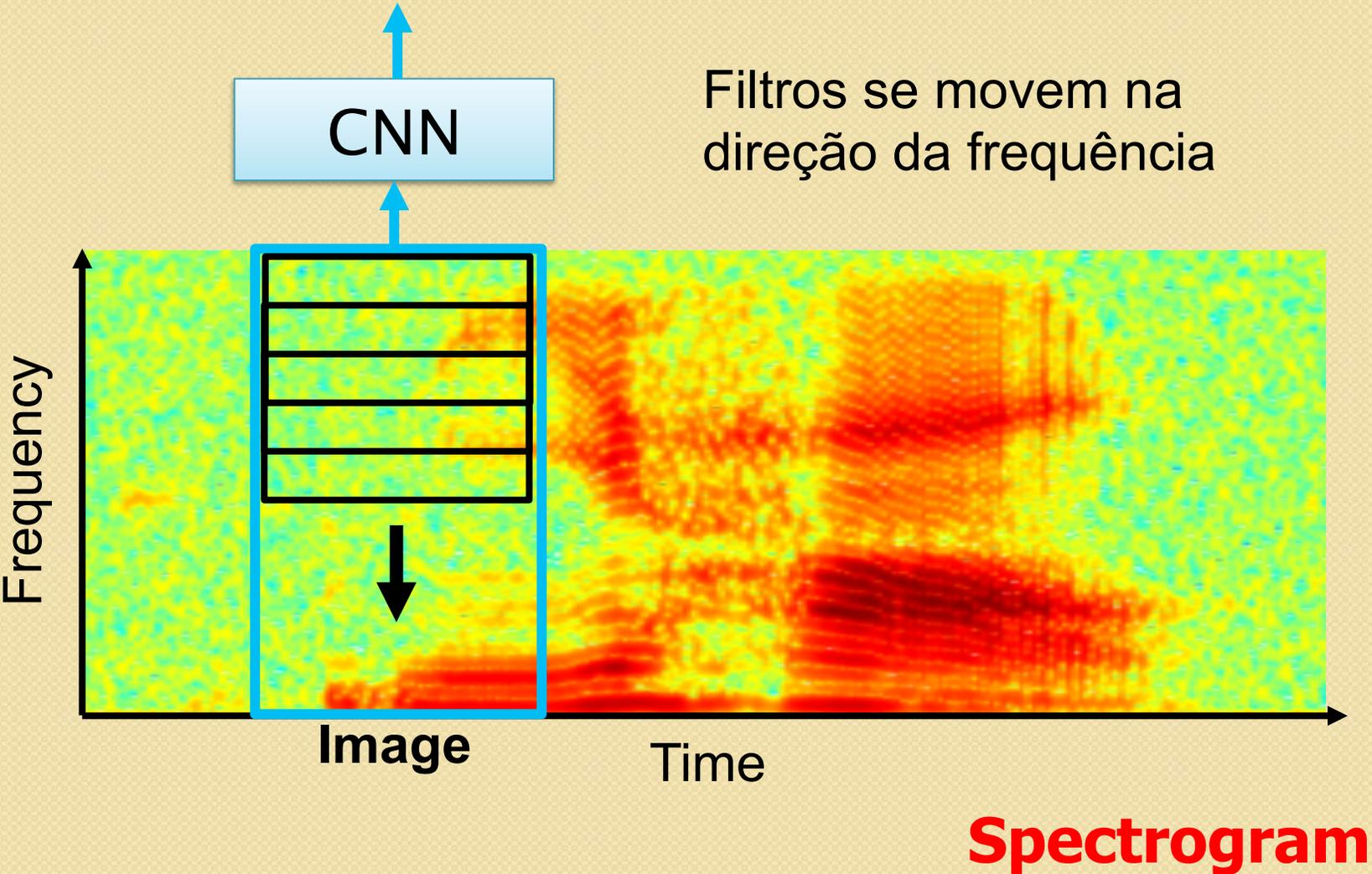


Achatada



Rede Feedfoward
Totalmente Conectada

CNN em Reconhecimento da Fala



Filtros se movem na direção da frequência

CNN

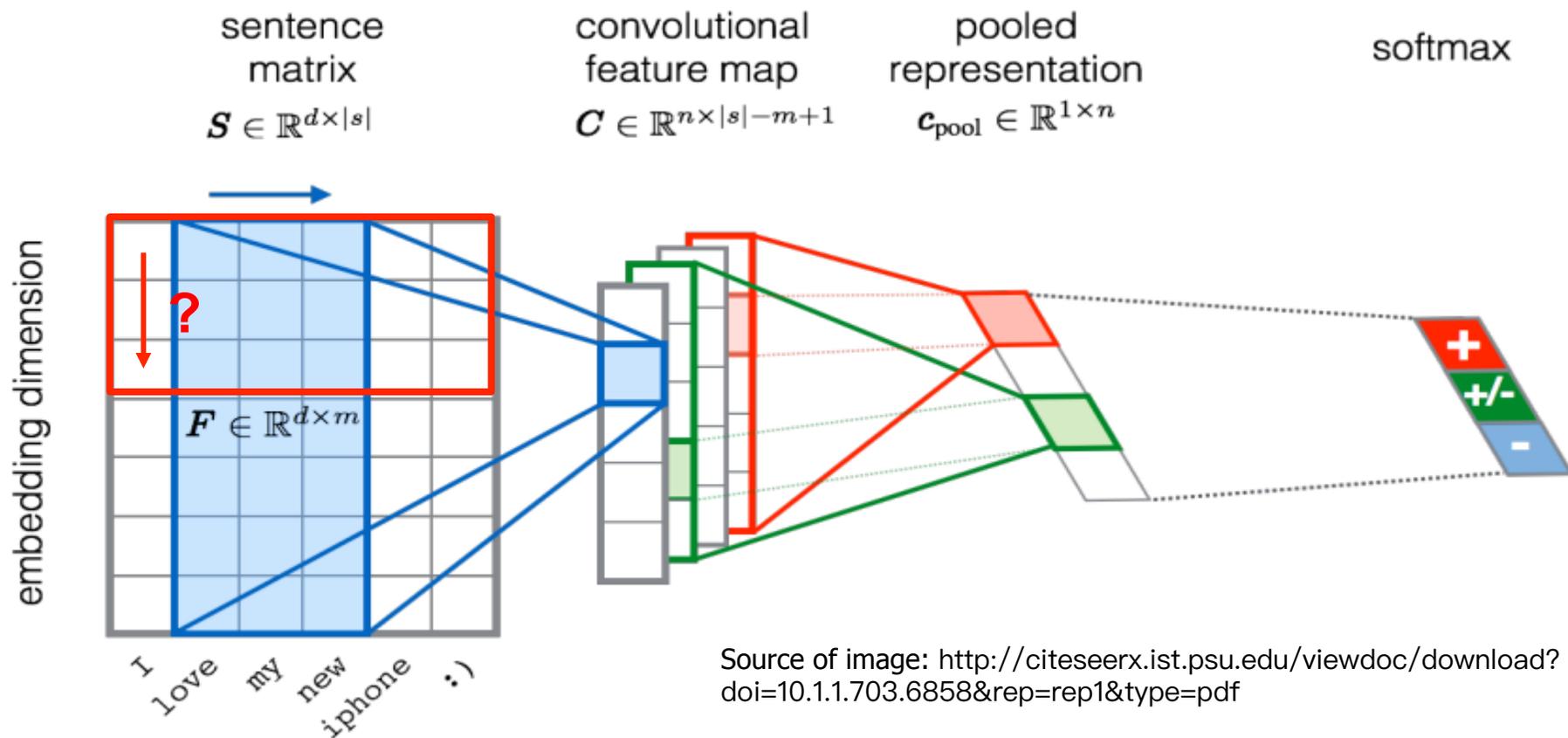
Frequency

Image

Time

Spectrogram

CNN em Classificação de Textos



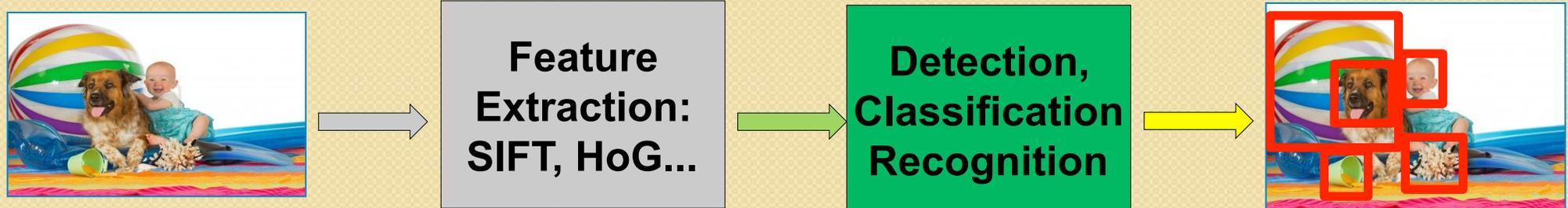
Solução Típica em Visão Computacional



Para reconhecer coisas?

Solução Típica em Visão Computacional

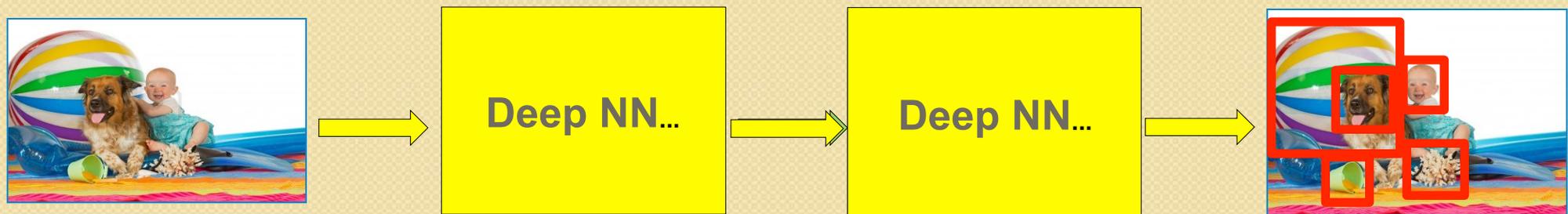
1. Definir e selecionar características:
2. SURF, HoG, SIFT, RIFT.
3. Acrescentar classificadores



Definição de características é dependente do domínio e consome tempo

Solução em Deep Learning: Pipeline Baseado em Visão

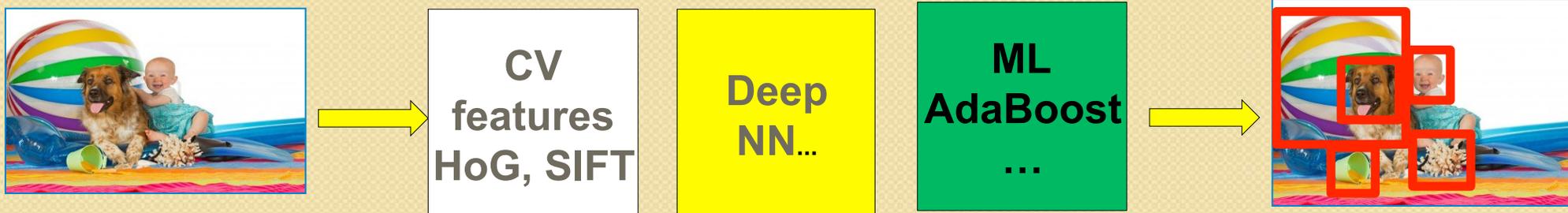
- Extrair características automaticamente baseado nos dados
- Combina a extração com classificação
- Papel do especialista: definir topologia e treinamento



Promessa do Deep Learning?
Treinar boas características automaticamente
Mesmo método para diferentes domínios

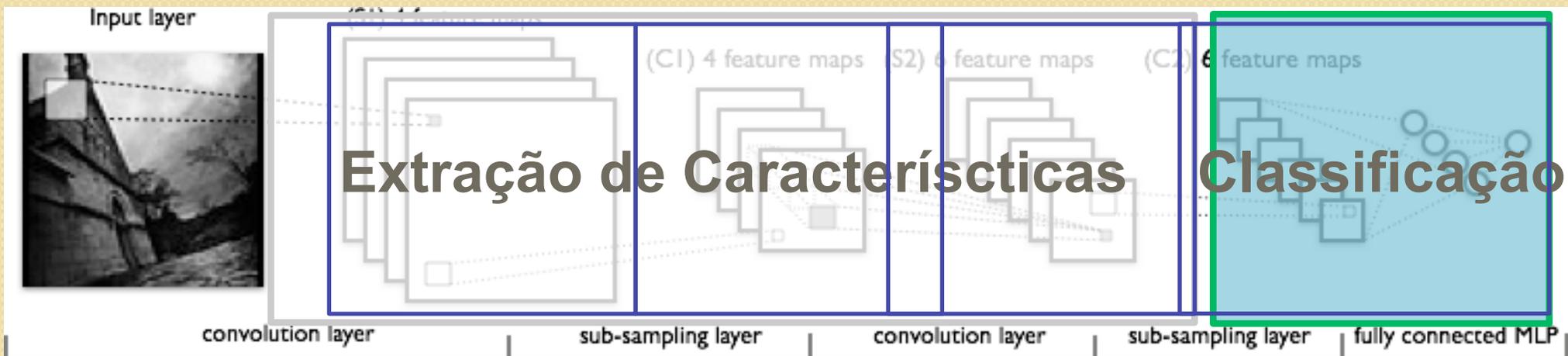
Visão Computacional + Deep Learning + Machine Learning

- Combinar características pré-definidas com características aprendidas
- Melhores métodos para classificação de múltiplas classes
- Especialistas em CV+DL+ML necessários para criar melhores soluções

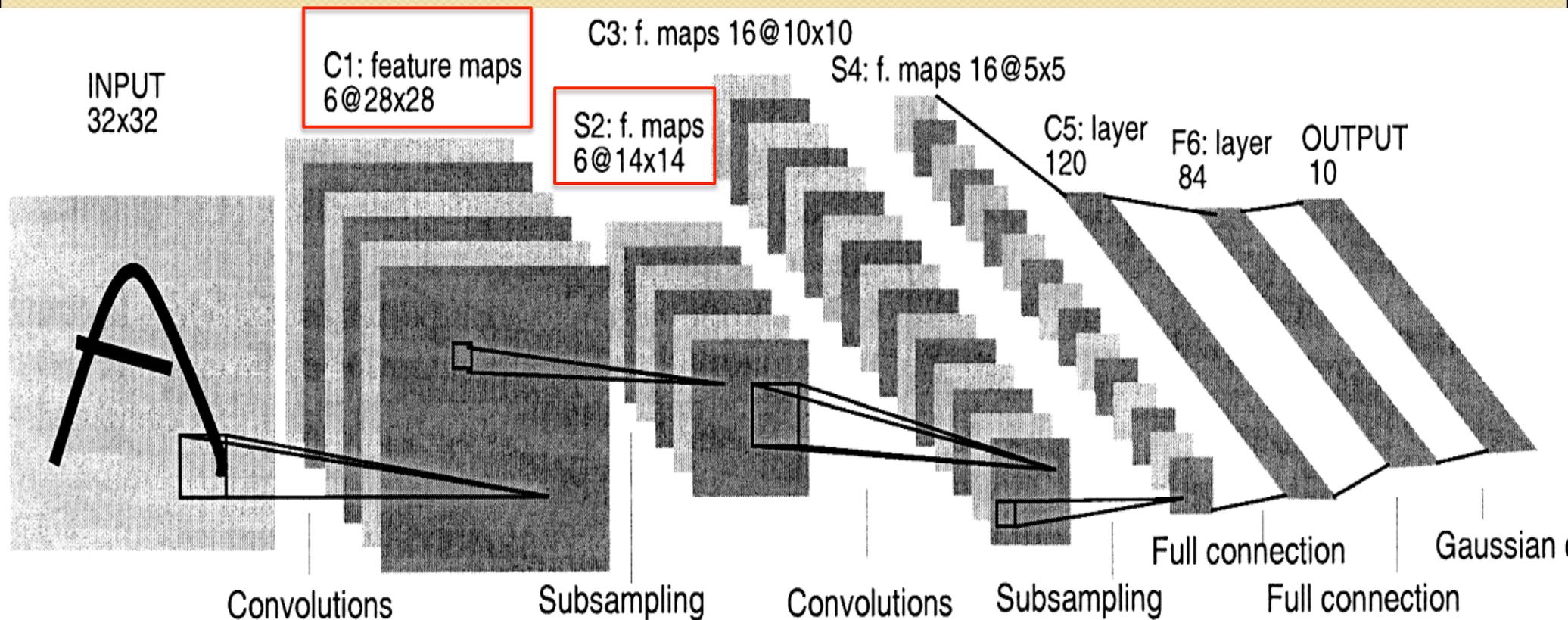


CNN Multi-camadas

- Convolutional + Non-Linear Layer
- Sub-sampling Layer
- Convolutional + Non-Linear Layer
- Fully connected layers
- Supervised



Como Tudo Começou? LeNet5



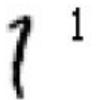
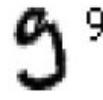
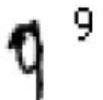
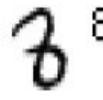
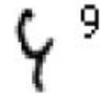
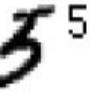
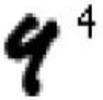
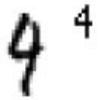
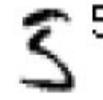
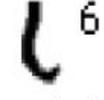
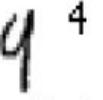
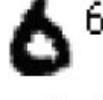
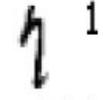
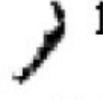
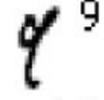
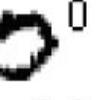
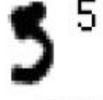
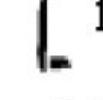
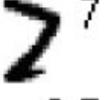
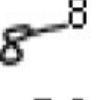
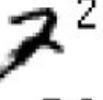
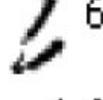
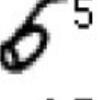
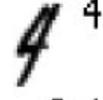
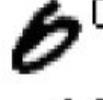
4	3	2	1	3	4	2	3	6	1
4->6	3->5	8->2	2->1	5->3	4->8	2->8	3->5	6->5	7->3
4	8	7	5	7	6	7	2	3	4
9->4	8->0	7->8	5->3	8->7	0->6	3->7	2->7	8->3	9->4
8	3	4	3	6	2	9	1	4	1
8->2	5->3	4->8	3->9	6->0	9->8	4->9	6->1	9->4	9->1
9	0	1	3	3	9	6	6	6	8
9->4	2->0	6->1	3->5	3->2	9->5	6->0	6->0	6->0	6->8
4	7	9	4	2	9	4	4	9	9
4->6	7->3	9->4	4->6	2->7	9->7	4->3	9->4	9->4	9->4
7	4	8	3	4	6	8	3	3	9
8->7	4->2	8->4	3->5	8->4	6->5	8->5	3->8	3->8	9->8
1	9	6	0	6	7	0	1	4	2
1->5	9->8	6->3	0->2	6->5	9->5	0->7	1->6	4->9	2->1
2	8	9	7	7	6	9	1	6	5
2->8	8->5	4->9	7->2	7->2	6->5	9->7	6->1	5->6	5->0
4	2								
4->9	2->8								

Os 82 erros da LeNet5

Muitos casos facilmente reconhecidos por humanos

Taxa de erro humana estimada em 20 a 30 erros

Erros na Rede de Ciresan *et.al.*

 2 17	 1 71	 9 98	 9 59	 9 79	 5 35	 8 23
 4 49	 3 35	 9 97	 4 49	 4 94	 0 02	 3 35
 6 16	 4 94	 0 60	 0 06	 8 86	 1 79	 1 71
 9 49	 0 50	 3 35	 8 98	 9 79	 7 17	 6 61
 2 27	 8 58	 2 78	 6 16	 6 65	 4 94	 0 60

Ígito em cima: resposta certa. Dígitos de baixo: 2 melhores apostas da rede.

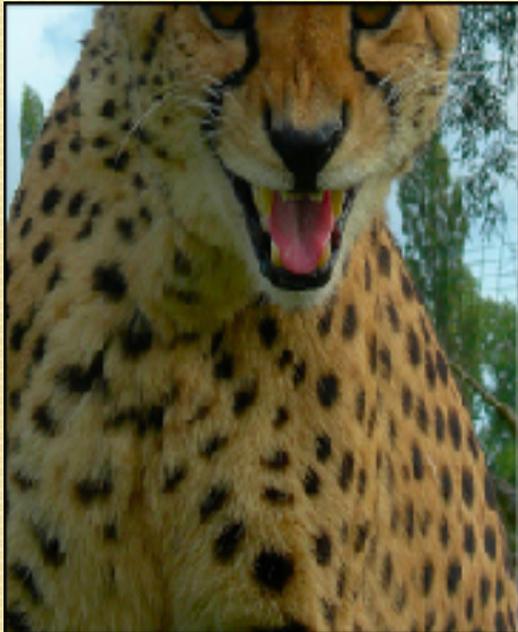
Resposta certa quase sempre nas 2 melhores apostas.

Tirando a média, performance aumentou para 25 erros.

Competição ILSVRC-2012 no Dataset ImageNet

- 1.2 milhões de imagens em alta resolução.
- **Tarefa de classificação:**
 - Encontre a classe correta nas suas 5 melhores apostas. 1000 classes.
- **Tarefa de localização:**
 - Para cada aposta, desenhe uma caixa entorno do objeto. Sua caixa tem de ter pelo menos 50% de sobreposição com a caixa correta.
- Alguns dos melhores métodos de visão computacional (VC) foram usados: Oxford, INRIA, XRCE, Tokio
 - Sistemas de VC tarefas complicadas de múltiplos estágios
 - Estágios iniciais envolvem refinamento a mão de parâmetros

Exemplos de Apostas



cheetah

cheetah

leopard

snow leopard

Egyptian cat



bullet train is like a plane, with in-train magazine and a job that you can plug your headphones into and listen to

bullet train

bullet train

passenger car

subway train

electric locomotive



hand glass

scissors

hand glass

frying pan

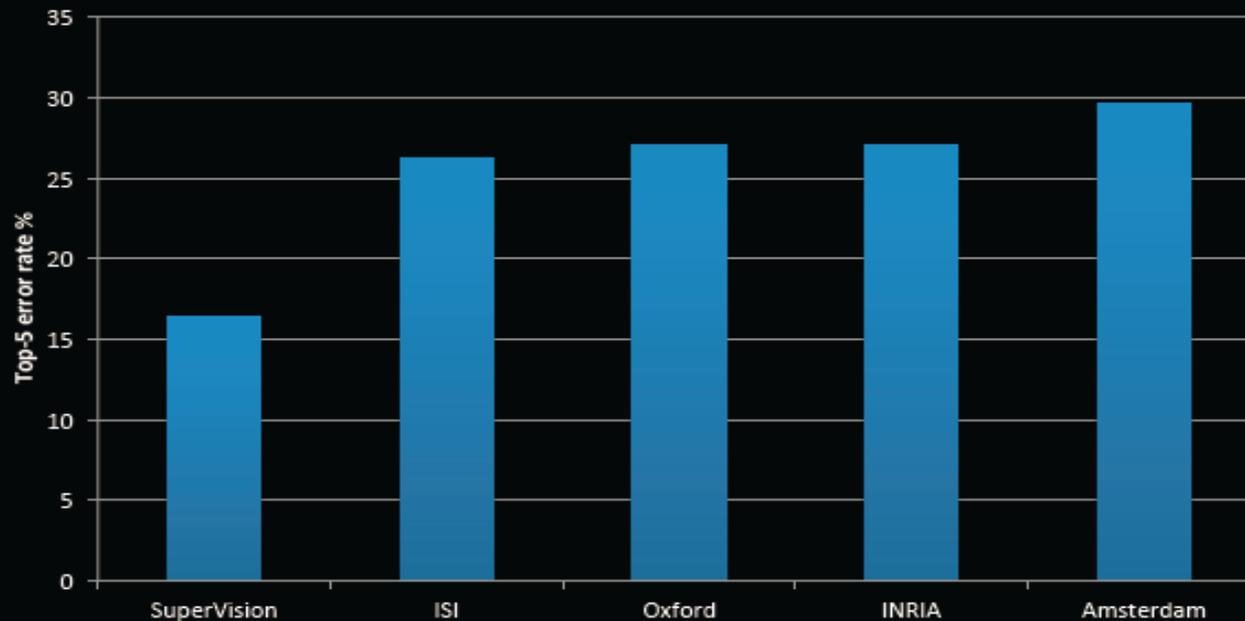
stethoscope

Taxas de Erros: ILSVRC-2012

	classification	classification & localization
• University of Tokyo	• 26.1%	53.6%
• Oxford University Computer Vision Group	• 26.9%	50.0%
• INRIA (French national research institute in CS) + XRCE (Xerox Research Center Europe)	• 27.0%	
• University of Amsterdam	• 29.5%	
• University of Toronto (Alex Krizhevsky)	• 16.4%	34.1%

Classificações na Imagenet 2012

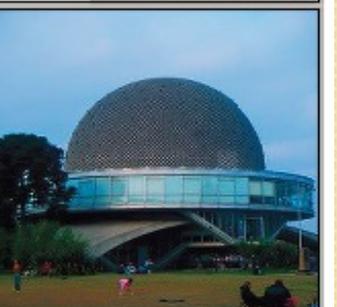
- Krizhevsky et al. -- 16.4% error (top-5)
- Next best (non-convnet) – 26.2% error



Rede Neural CNN para ImageNet

- Alex Krizhevsky (NIPS 2012): rede muito profunda baseada no trabalho de Yann Le Cun
 - 7 camadas intermediárias sem contar camadas de max pooling
 - Primeiras camadas convolucionais
 - Últimas 2 totalmente conectadas

Alguns Exemplos de Classificação: ILSVRC

			
<p>lens cap</p>	<p>abacus</p>	<p>slug</p>	<p>hen</p>
<p>reflex camera Polaroid camera pencil sharpener switch combination lock</p>	<p>abacus typewriter keyboard space bar computer keyboard accordion</p>	<p>slug zucchini ground beetle common newt water snake</p>	<p>hen cock cocker spaniel partridge English setter</p>
			
<p>tiger</p>	<p>chambered nautilus</p>	<p>tape player</p>	<p>planetarium</p>
<p>tiger tiger cat tabby boxer Saint Bernard</p>	<p>lampshade throne goblet table lamp hamper</p>	<p>cellular telephone slot reflex camera dial telephone iPod</p>	<p>planetarium dome mosque radio telescope steel arch bridge</p>

Outros



mite

container ship

motor scooter

leopard

	mite
	black widow
	cockroach
	tick
	starfish

	container ship
	lifeboat
	amphibian
	fireboat
	drilling platform

	motor scooter
	go-kart
	moped
	bumper car
	golfcart

	leopard
	jaguar
	cheetah
	snow leopard
	Egyptian cat



grille

mushroom

cherry

Madagascar cat

	convertible
	grille
	pickup
	beach wagon
	fire engine

	agaric
	mushroom
	jelly fungus
	gill fungus
	dead-man's-fingers

	dalmatian
	grape
	elderberry
	ffordshire bullterrier
	currant

	squirrel monkey
	spider monkey
	titi
	indri
	howler monkey

Topologia da CNN de Krizhevsky

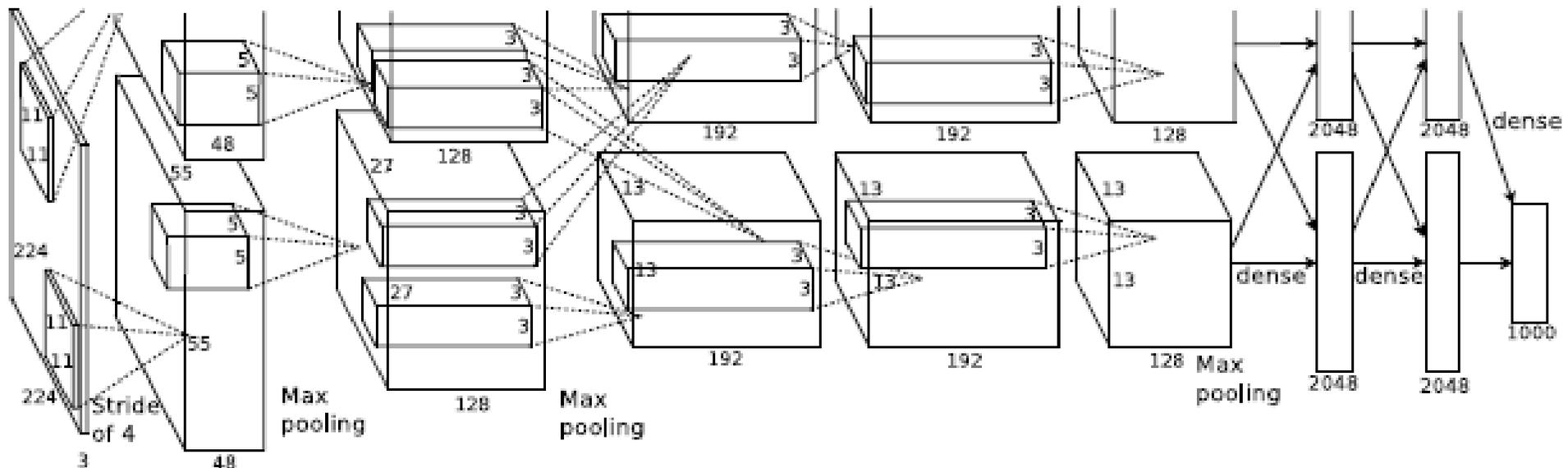
- 5 convolutional layers
- 3 fully connected layers + soft-max
- 650K neurons , 60 Mln weights

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca



Explosão

The image shows a screenshot of the MIT Technology Review website's '10 Breakthrough Technologies 2013' page. The page features a dark navigation bar at the top with 'HOME', 'MENU', and 'CONNECT' on the left, and 'THE LATEST', 'POPULAR', and 'MOST SHARED' on the right. Below the navigation bar, the main heading reads '10 BREAKTHROUGH TECHNOLOGIES 2013' with the MIT Technology Review logo to its left. Navigation links for 'Introduction', 'The 10 Technologies', and 'Past Years' are positioned to the right of the heading. The main content area is a grid of ten light blue cards, each representing a technology. The first card, 'Deep Learning', is highlighted with a red rounded rectangle. The text on this card reads: 'With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.' The other cards include: 'Temporary Social Media', 'Prenatal DNA Sequencing', 'Additive Manufacturing', 'Baxter: The Blue-Collar Robot', 'Memory Implants', 'Smart Watches', 'Ultra-Efficient Solar Power', 'Big Data from Cheap Phones', and 'Supergrids'. Each card has a right-pointing arrow at the bottom right corner.

MIT Technology Review

10 BREAKTHROUGH TECHNOLOGIES 2013

Introduction The 10 Technologies Past Years

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart. →

Temporary Social Media

Messages that quickly self-destruct could enhance the privacy of online communications and make people freer to be spontaneous. →

Prenatal DNA Sequencing

Reading the DNA of fetuses will be the next frontier of the genomic revolution. But do you really want to know about the genetic problems or musical aptitude of your unborn child? →

Additive Manufacturing

Skeptical about 3-D printing? GE, the world's largest manufacturer, is on the verge of using the technology to make jet parts. →

Baxter: The Blue-Collar Robot

Rodney Brooks's newest creation is easy to interact with, but the complex innovations behind the robot show just how hard it is to get along with people. →

Memory Implants

Smart Watches

Ultra-Efficient Solar Power

Big Data from Cheap Phones

Supergrids

MIT Technology Review, April 23rd, 2013

ILSVRC 2012: Top Rankers

<http://www.image-net.org/challenges/LSVRC/2012/results.html>

N	Error-5	Algorithm	Team	Authors
1	0.153	Deep Conv. Neural Network	Univ. of Toronto	Krizhevsky et al
2	0.262	Features + Fisher Vectors + Linear classifier	ISI	Gunji et al
3	0.270	Features + FV + SVM	OXFORD_VGG	Simonyan et al
4	0.271	SIFT + FV + PQ + SVM	XRCE/INRIA	Perronin et al
5	0.300	Color desc. + SVM	Univ. of Amsterdam	van de Sande et al

Imagenet 2013: Top Rankers

<http://www.image-net.org/challenges/LSVRC/2013/results.php>

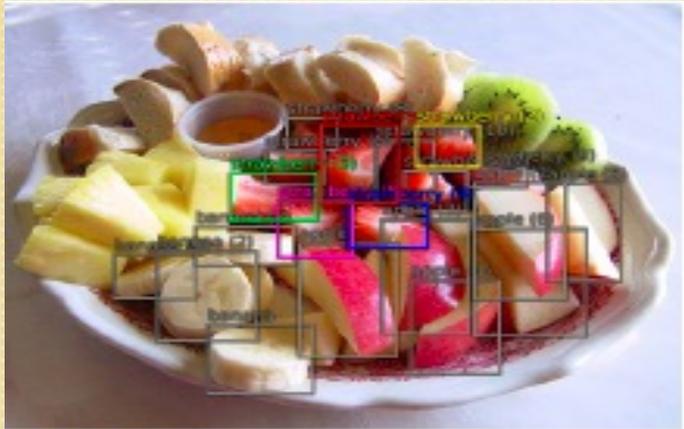
N	Error-5	Algorithm	Team	Authors
1	0.117	Deep Convolutional Neural Network	Clarifi	Zeiler
2	0.129	Deep Convolutional Neural Networks	Nat.Univ Singapore	Min LIN
3	0.135	Deep Convolutional Neural Networks	NYU	Zeiler Fergus
4	0.135	Deep Convolutional Neural Networks		Andrew Howard
5	0.137	Deep Convolutional Neural Networks	Overfeat NYU	Pierre Sermanet et al

CNN: Detecção de Ações



Taylor, ECCV 2010

CNN: Detecção



Groundtruth:

strawberry
strawberry (2)
strawberry (3)
strawberry (4)
strawberry (5)
strawberry (6)
strawberry (7)
strawberry (8)
strawberry (9)
strawberry (10)
apple
apple (2)
apple (3)

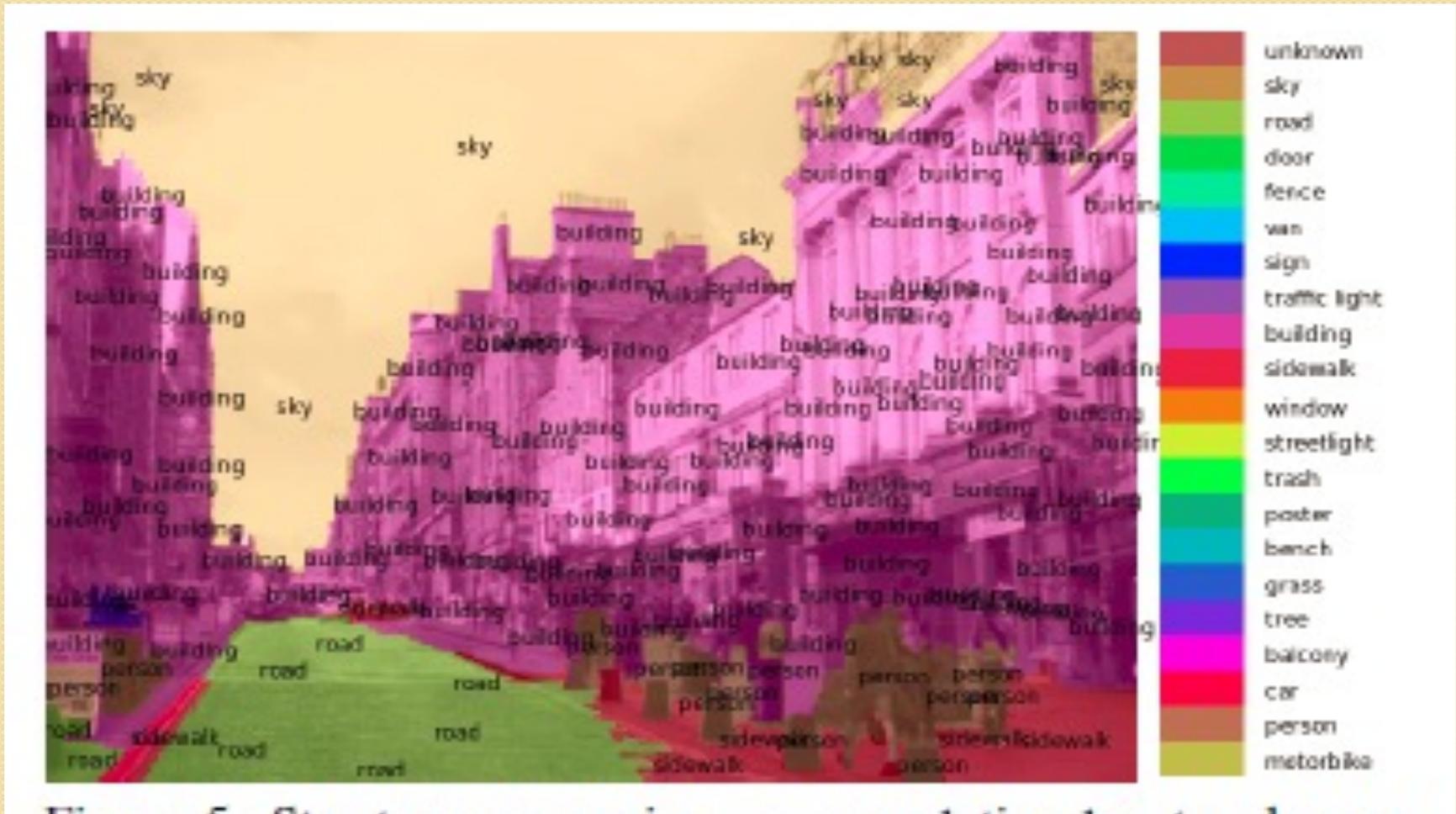


Groundtruth:

tv or monitor
tv or monitor (2)
tv or monitor (3)
person
remote control
remote control (2)

Sermanet, CVPR 2014

CNN: Descrição de Cenários



Farabet, PAMI 2013

CNN: Rotulamento Semântico Indoor em RGBD

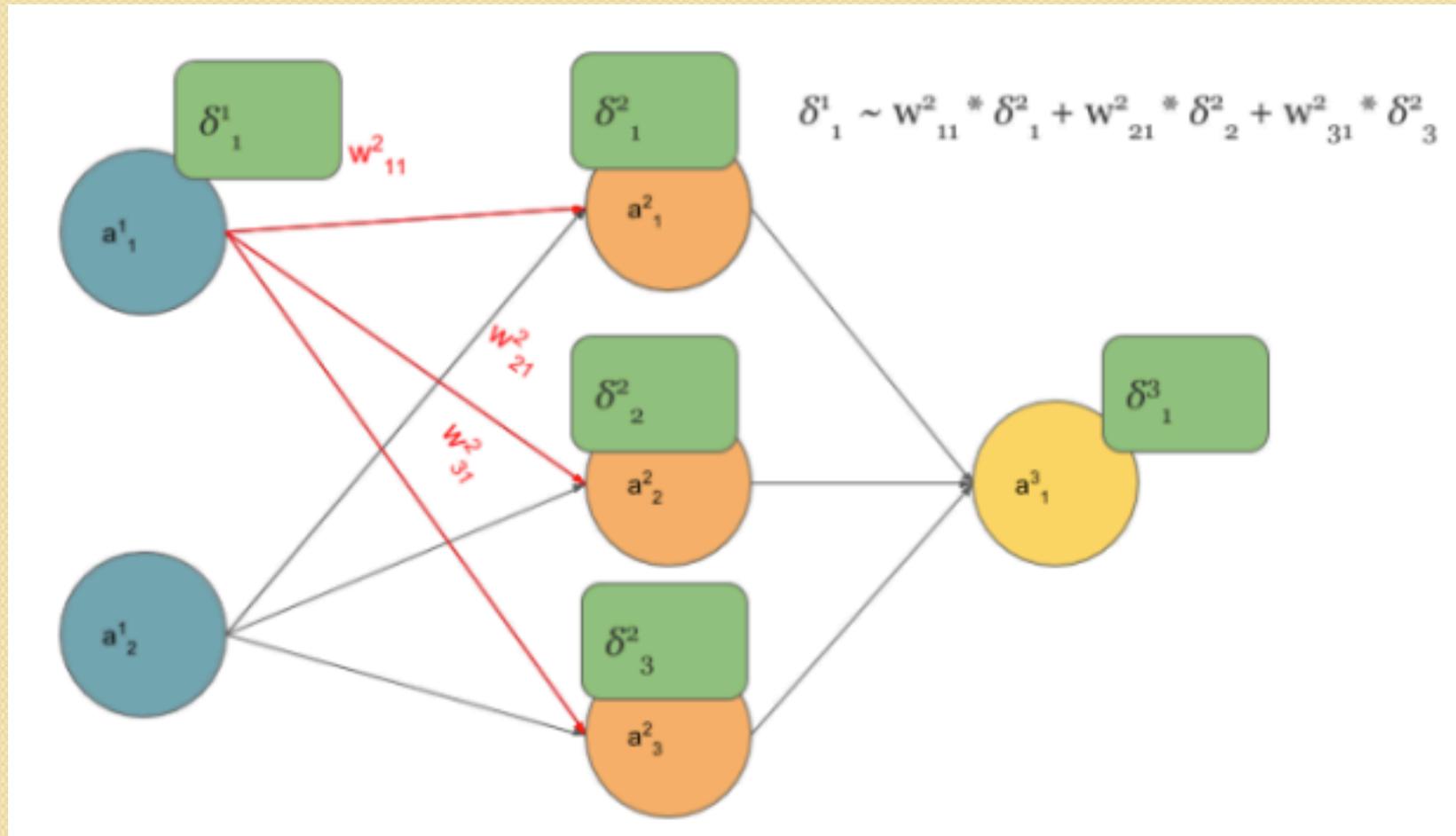


Figure 2: Some scene labelings using our Multiscale Convolutional Network trained on RGBD images.

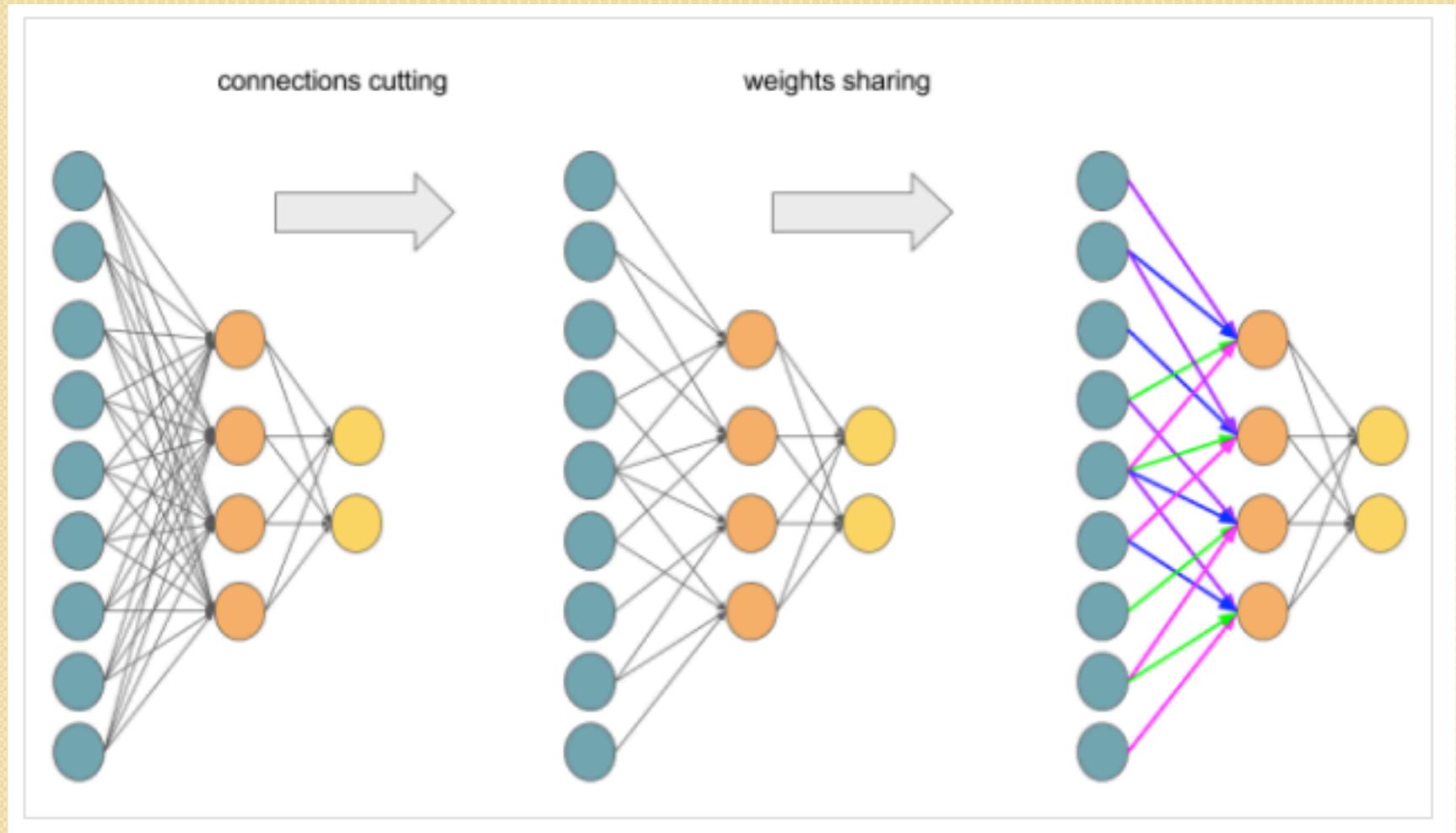
Farabet, 2013

Como se Treina uma CNN?

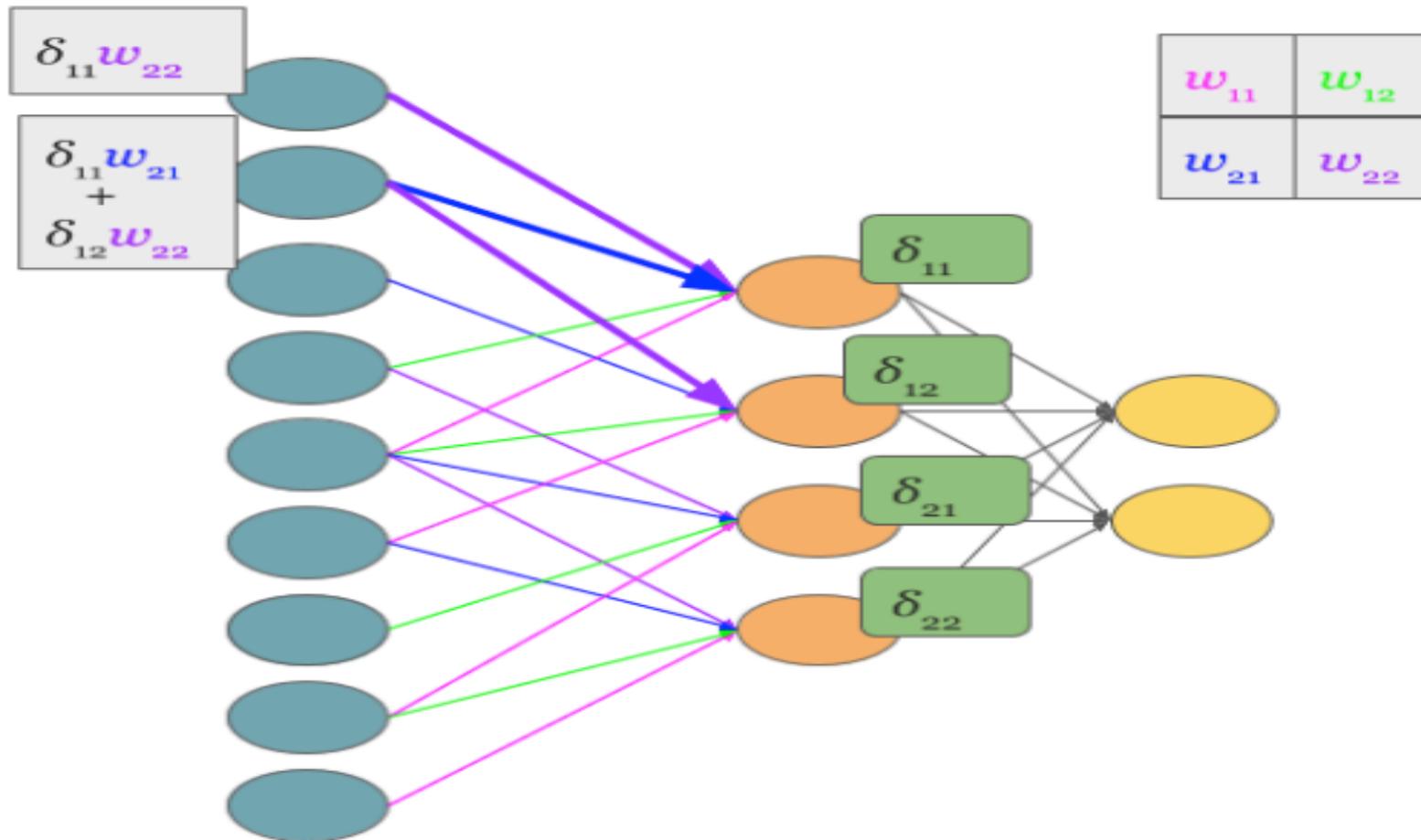
Lembrando Gradiente Descendente (Backpropagation)



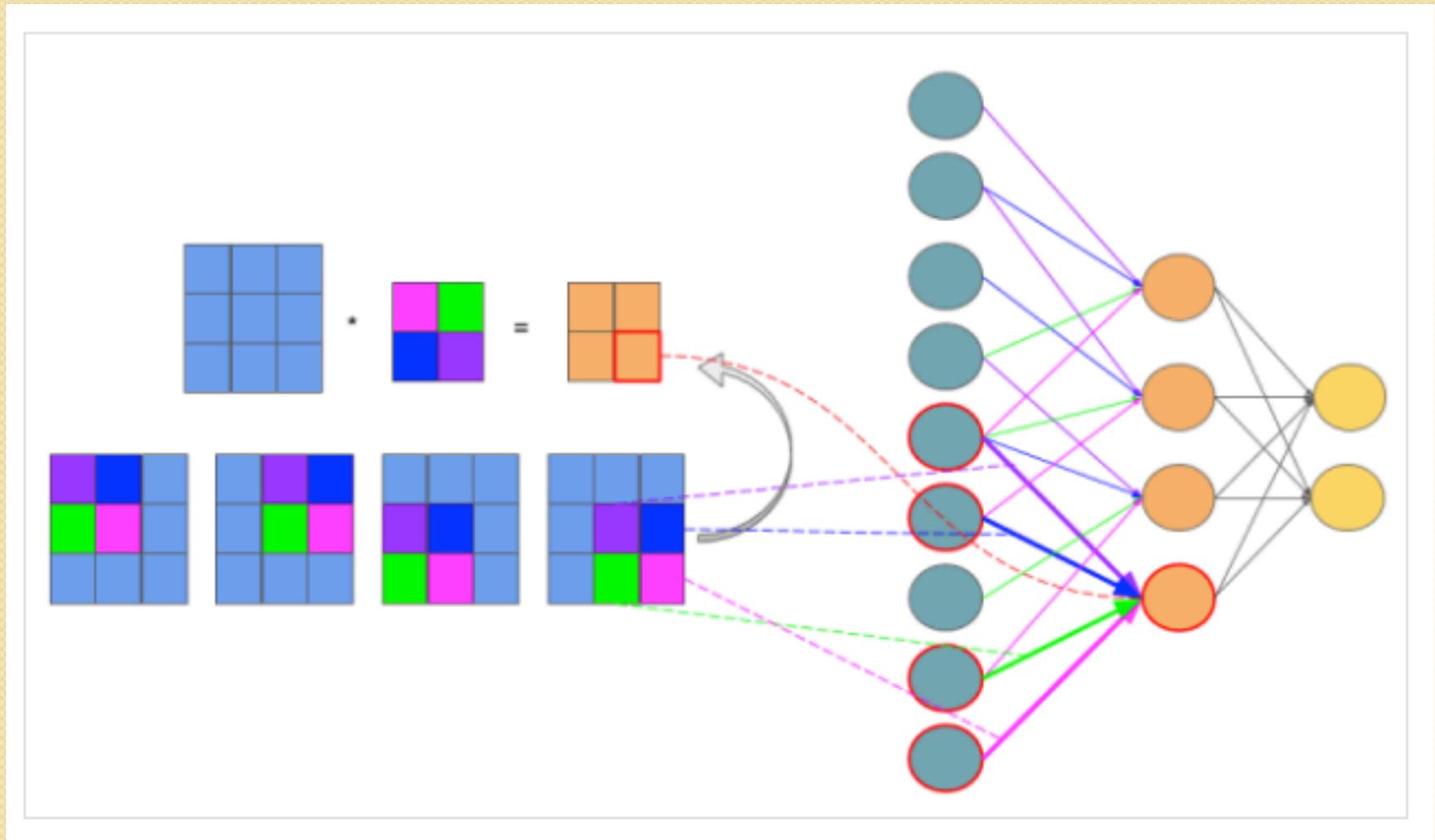
Transformando Rede Totalmente Conectada em CNN



Transformando Rede Totalmente Conectada em CNN



Operação de Convolução



Definição do Erro para Ajuste de Cada Conexão

w_{22}	w_{21}
w_{12}	w_{11}

rot_180(w)

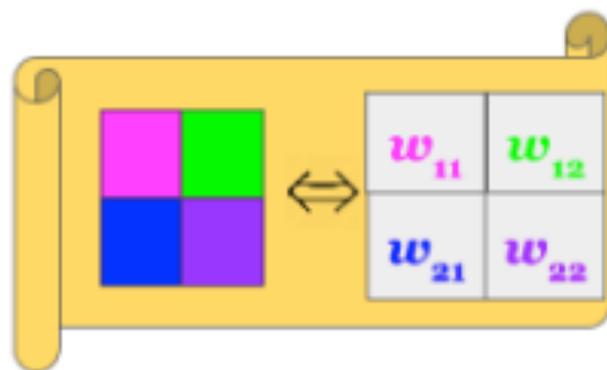
*

δ_{11}	δ_{12}
δ_{21}	δ_{22}

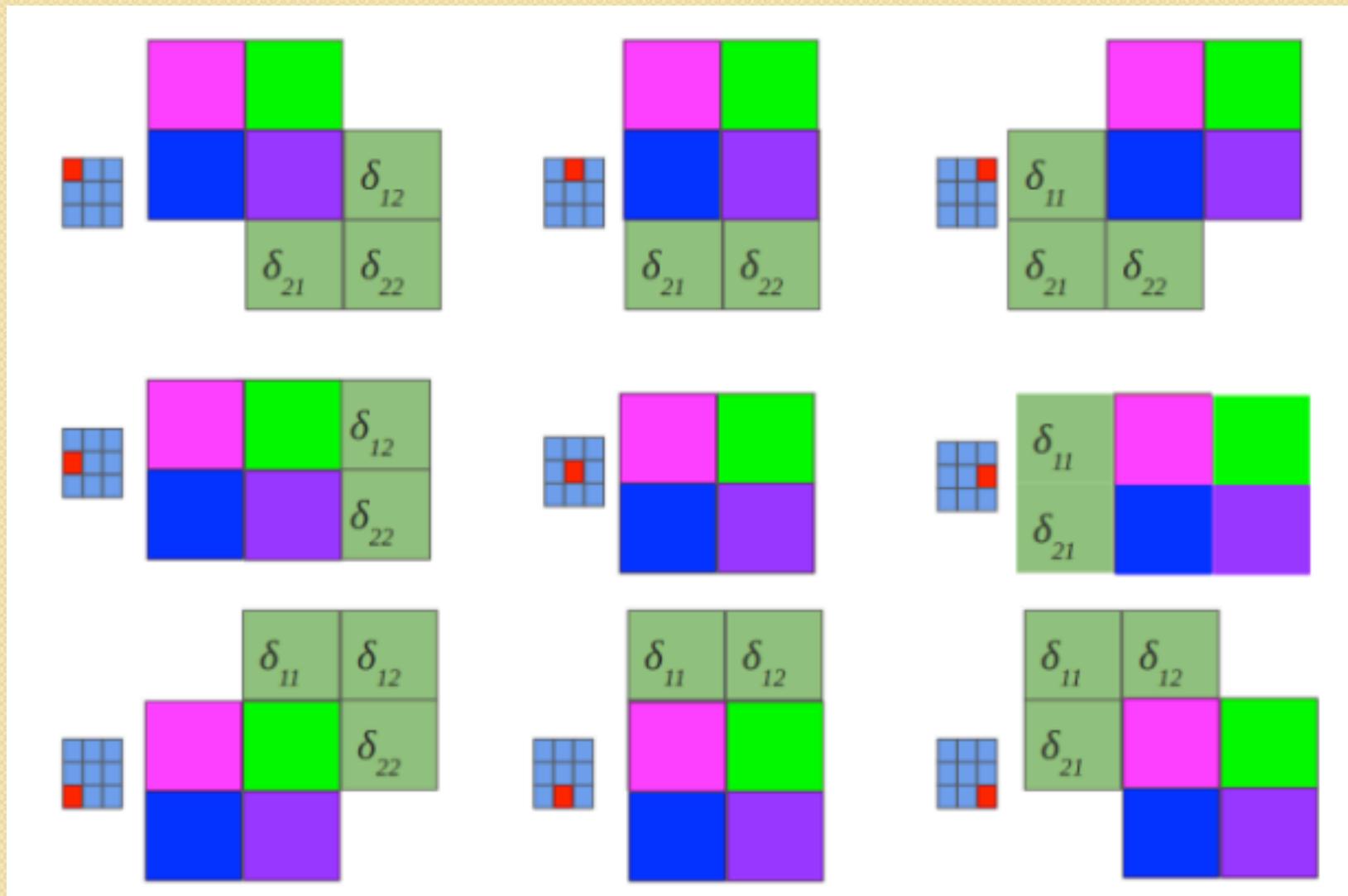
grads from orange layer

=

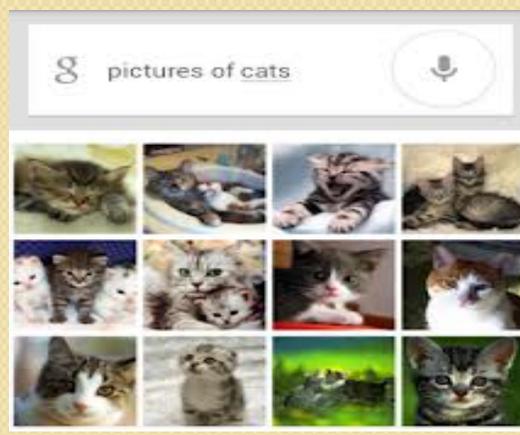
$\delta_{11} w_{22}$	$\delta_{11} w_{21} + \delta_{12} w_{22}$	$\delta_{12} w_{21}$
$\delta_{11} w_{12} + \delta_{21} w_{22}$	$\delta_{11} w_{11} + \delta_{12} w_{12} + \delta_{21} w_{21} + \delta_{22} w_{22}$	$\delta_{12} w_{11} + \delta_{22} w_{21}$
$\delta_{21} w_{12}$	$\delta_{21} w_{11} + \delta_{22} w_{12}$	$\delta_{22} w_{11}$



Definição do Erro para Ajuste de Cada Conexão



Da Pesquisa para Tecnologia



Deep Learning – breakthrough reconhecimento visual e da fala

Buzz

- July 2012 – Started DL lab
- Nov 2012 – Big improvement in OCR:
 - Speech – reduce Error rate
 - OCR – reduce Error rate by
- 2013 launch of Baidu Visual Search
 - Voice search
 - Photo Wonder
 - Visual search

Baidu map voice search

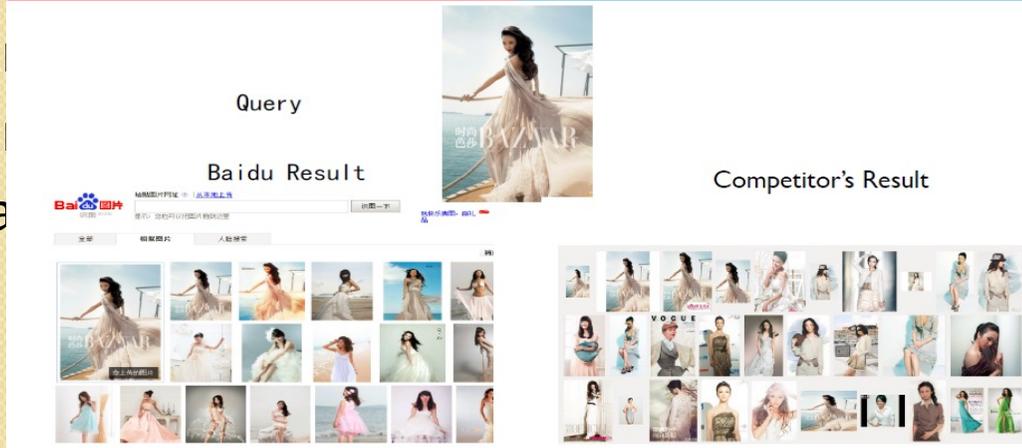


PhotoWonder

A magic photo beautifier, cool and easy to use!

Baidu 百度

Baidu Visual Search



Baidu 百度

Buzz

Scientists See Promise in Deep-Learning Programs



Han Zhong/The New York Times

A voice recognition program translated a speech
Chinese.

By JOHN MARKOFF

Published: November 23, 2012

The New York Times

[Microsoft On Deep Learning for Speech](#) goto 3:00-5:10

Buzz



Google Scoops Up Neural Networks Startup DNNresearch To Boost Its Voice And Image Search Tech

 RIP EMPSON ✓

Tuesday, March 12th, 2013

5 Comments



[Why Google invest in Deep Learning](#)

Buzz

Courant's LeCun to Lead Facebook's New Artificial Intelligence Group

December 9, 2013

143

Facebook has named New York University Professor Yann LeCun the director of a new laboratory devoted to research in artificial intelligence and deep learning.

"As one of the most respected thinkers in this field, Yann has done groundbreaking research in deep learning and computer vision," said Mike Schroepfer, Facebook's chief technology officer. "We're thrilled to welcome him to Facebook."

Facebook is building the team across three locations: Facebook's headquarters in Menlo Park, Calif., New York City, and London.

Machine learning is a branch of artificial intelligence that involves computers "learning" to extract knowledge from massive data sets and rendering informed analyses and judgments, often predicting outcomes.

LeCun, a professor at NYU's Courant Institute of Mathematical Sciences, is a pioneer in the growing field. In the 1980s, LeCun proposed one of the early versions of the back-propagation algorithm, the most popular method for training artificial neural networks. In the late 1980s and early 1990s at AT&T Bell Laboratories, he developed the convolutional network model, a pattern-recognition model whose architecture mimics, in part, the visual cortex of animals.

ENTERPRISE

research

uncategorized

Facebook Taps 'Deep Learning' Giant for New AI Lab

BY CADE METZ 12.09.13 3:14 PM

Follow @cademetz

Share 239

Tweet 214

+1 43

Share 48

Pin it



Facebook is building a research lab dedicated to the new breed of artificial intelligence, after hiring one of the preeminent researchers in the field, New York University professor Yann LeCun.

NYU "Deep Learning" Professor LeCun Will Head Facebook's New Artificial Intelligence Lab, Dec 10, 2013

Por Hoje é Só...