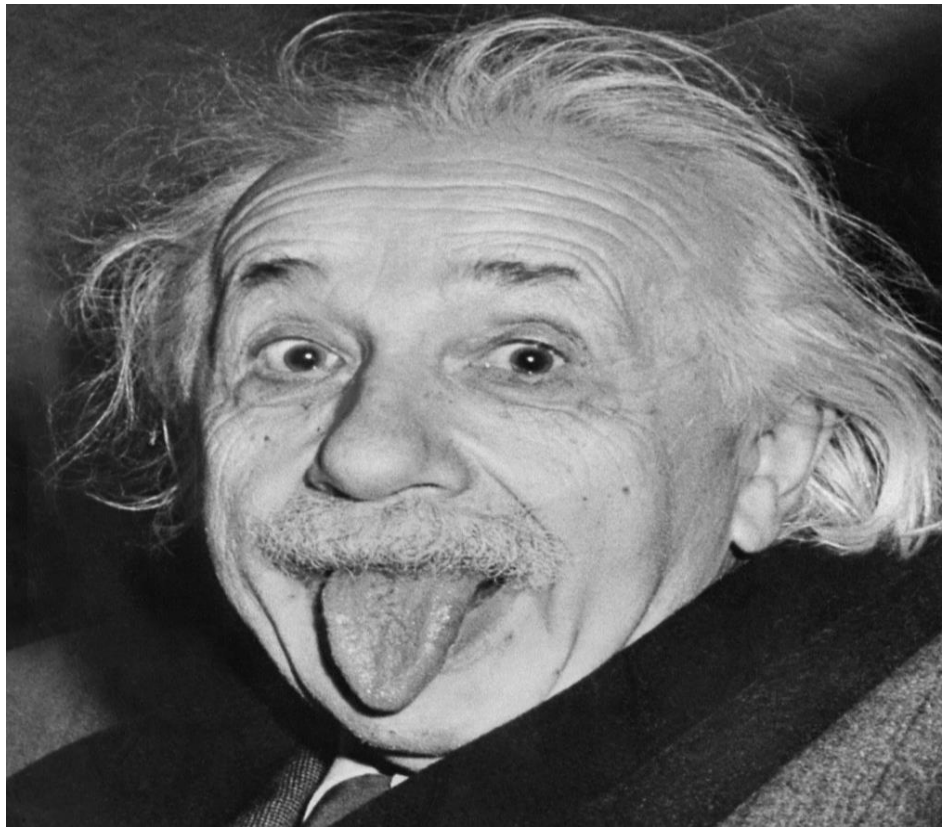
DataLab
Community

TÉCNICAS PARA UN MEJOR COMPUTER VISION

- Dr. Rubén Alvarez

- *"Everything human beings can imagine; nature has already created…"*
  - -Albert Einstein.



WHAT IS COMPUTER VISION?

DataLab Community

- Humans use our eyes and our brains to see and visually perceive the world around us. Computer Vision is the science that aims to provide a similar, if not better, capability to a machine or computer.

COMPUTER VISION

1. Optics and 3D Reconstruction
2. Image Processing
3. Machine Learning in images

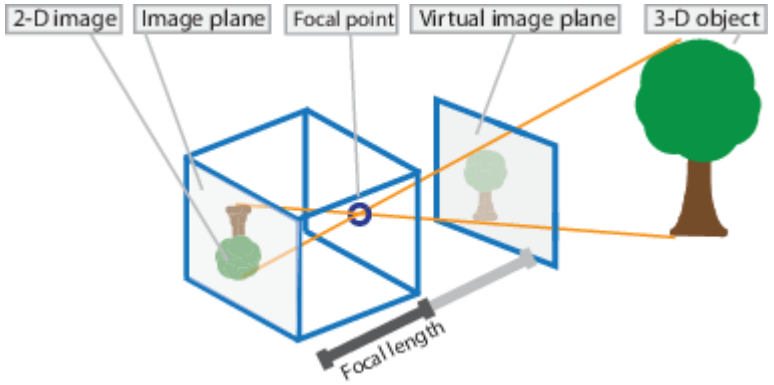
COMPUTER VISION

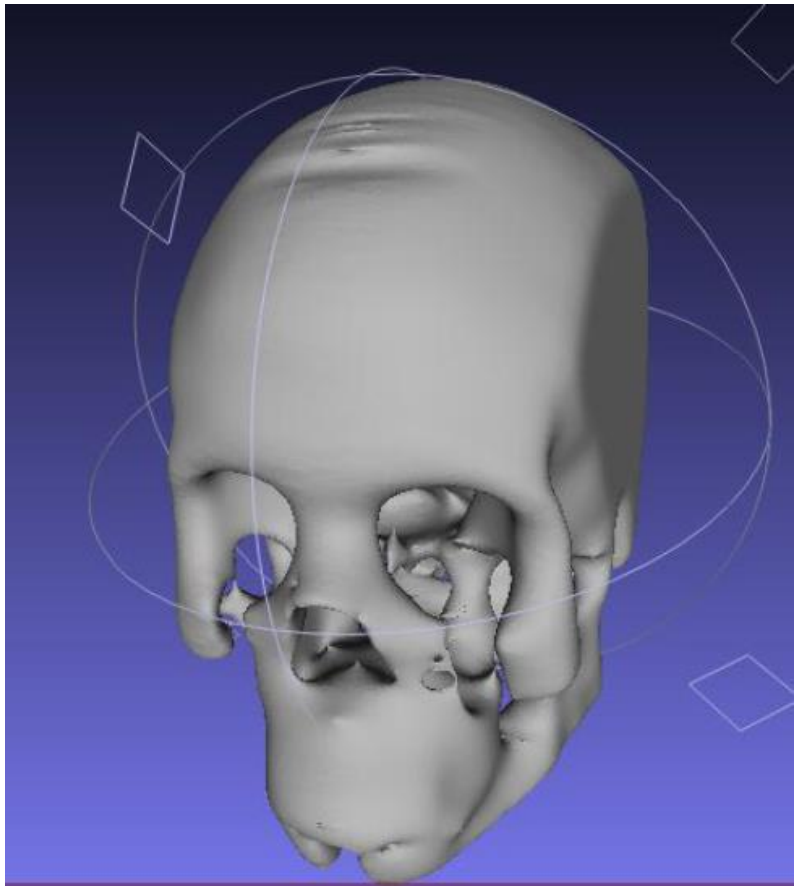
DataLab Community

1. **Optics and 3D Reconstruction**
2. Image Processing
3. Machine Learning in images

OPTICS AND 3D RECONSTRUCTION

DataLab Community





DataLab Community

COMPUTER VISION

3D CLOUD SAMPLES

COMPUTER VISION

DataLab Community

1. Optics and 3D Reconstruction
2. **Image Processing**
3. Machine Learning in images

IMAGE PROCESSING

DataLab Community

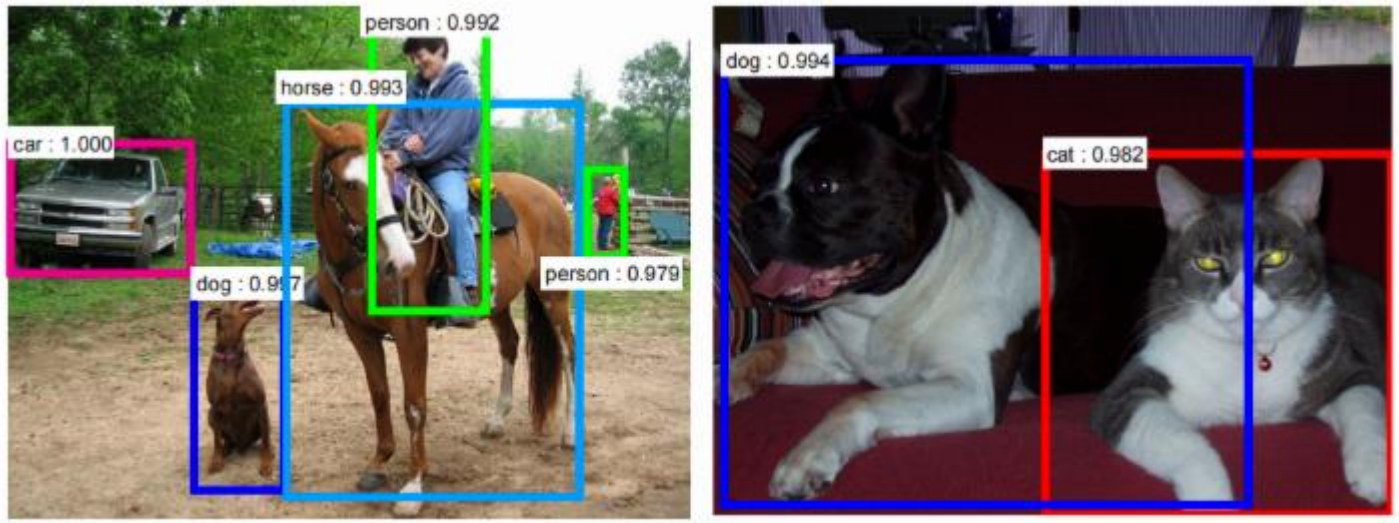
COMPUTER VISION

DataLab Community

1. Optics and 3D Reconstruction
2. Image Processing
3. **Machine Learning in images**

OBJECT RECOGNITION

DataLab Community



SEGMENTATION

DataLab Community

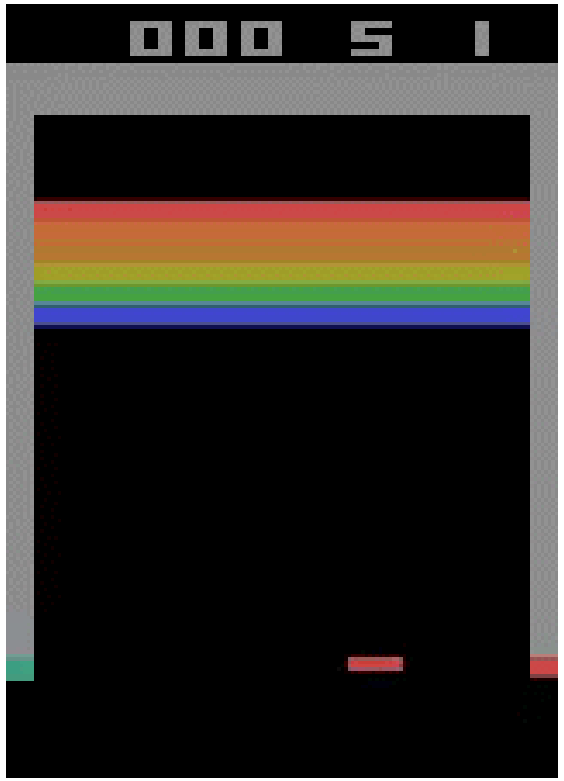
(c) Semantic segmentation

(d) Instance segmentation

STYLE TRANSFER

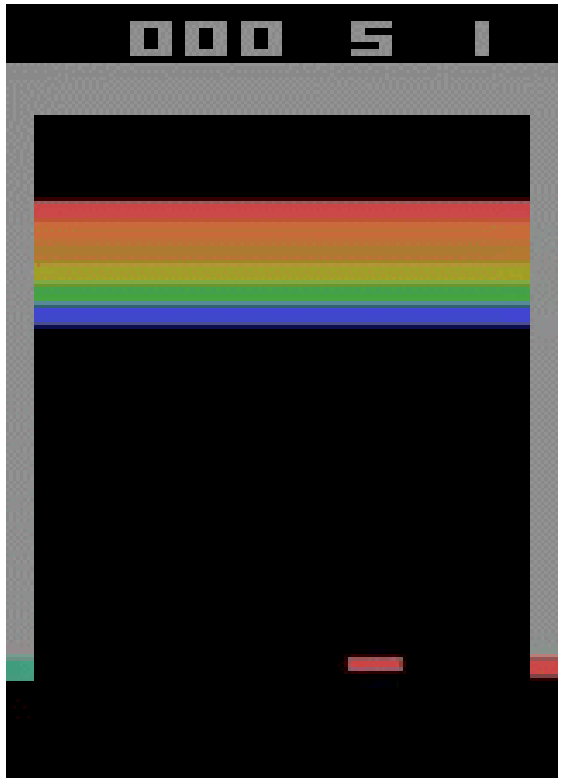
DataLab Community

VIDEO GAMES



VIDEO GAMES

DataLab Community

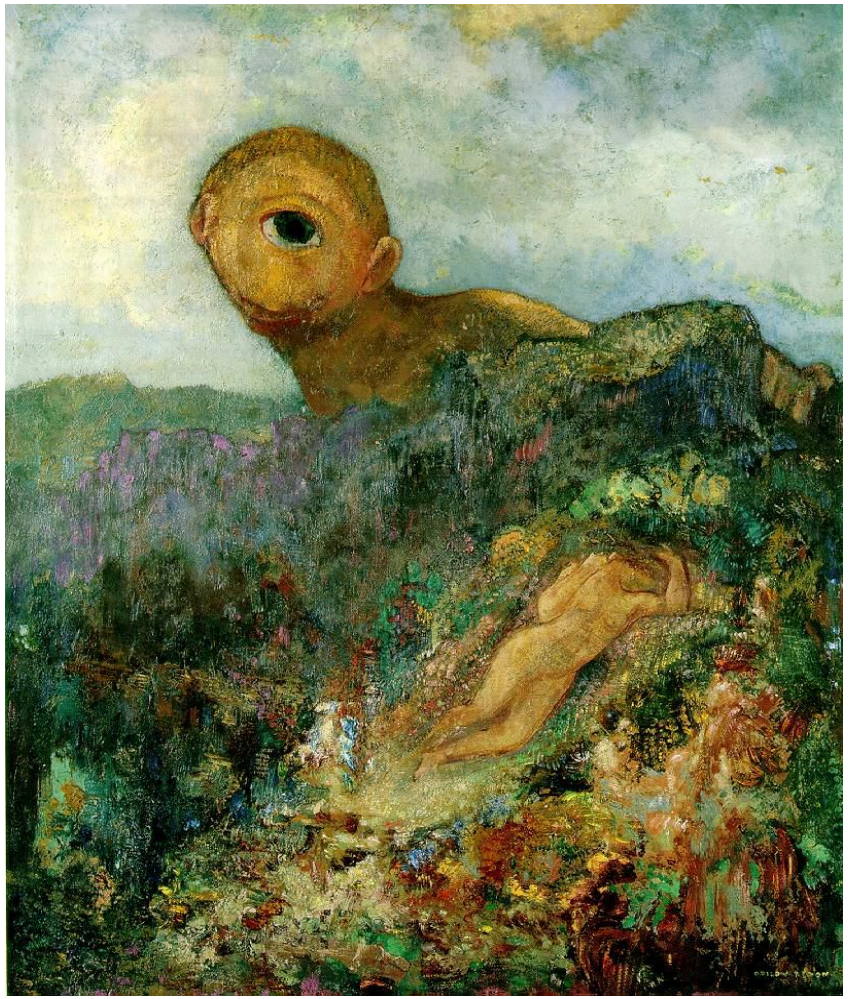


DataLab Community

COMPUTER VISION

PERCEPCIÓN 3D

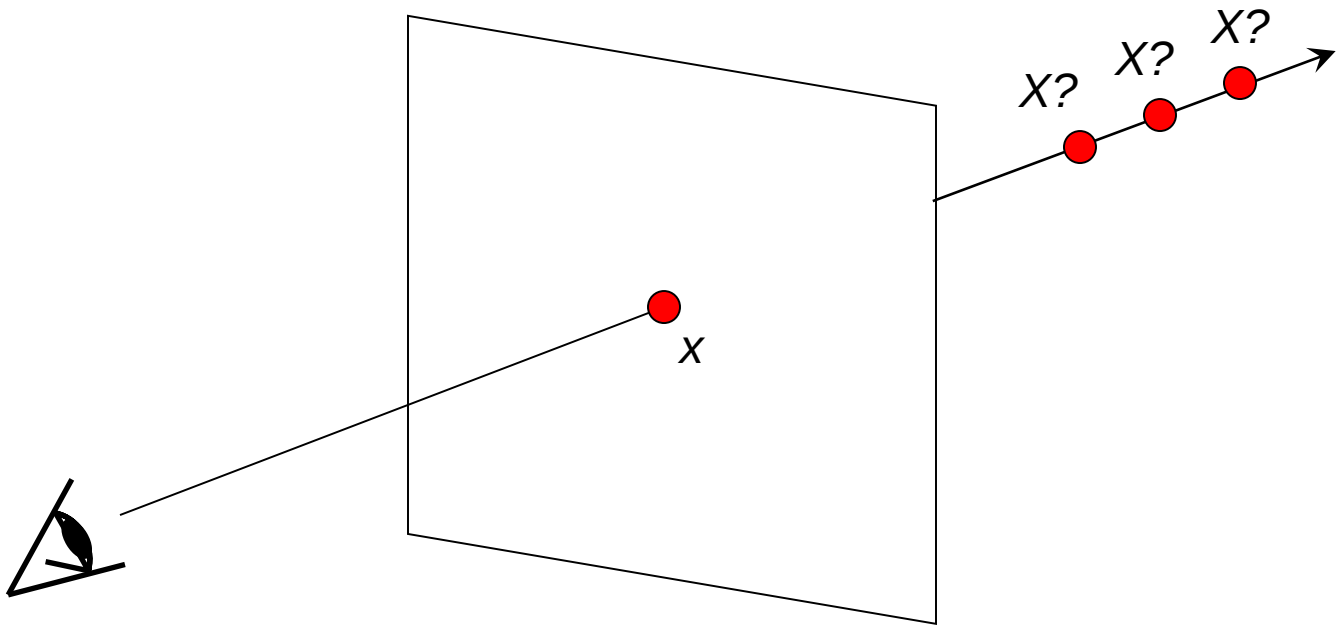
SINGLE-VIEW GEOMETRY



Odilon Redon, *Cyclops*, 1914

OUR GOAL: RECOVERY OF 3D STRUCTURE

- Is recovering the structure of an image unambiguous?

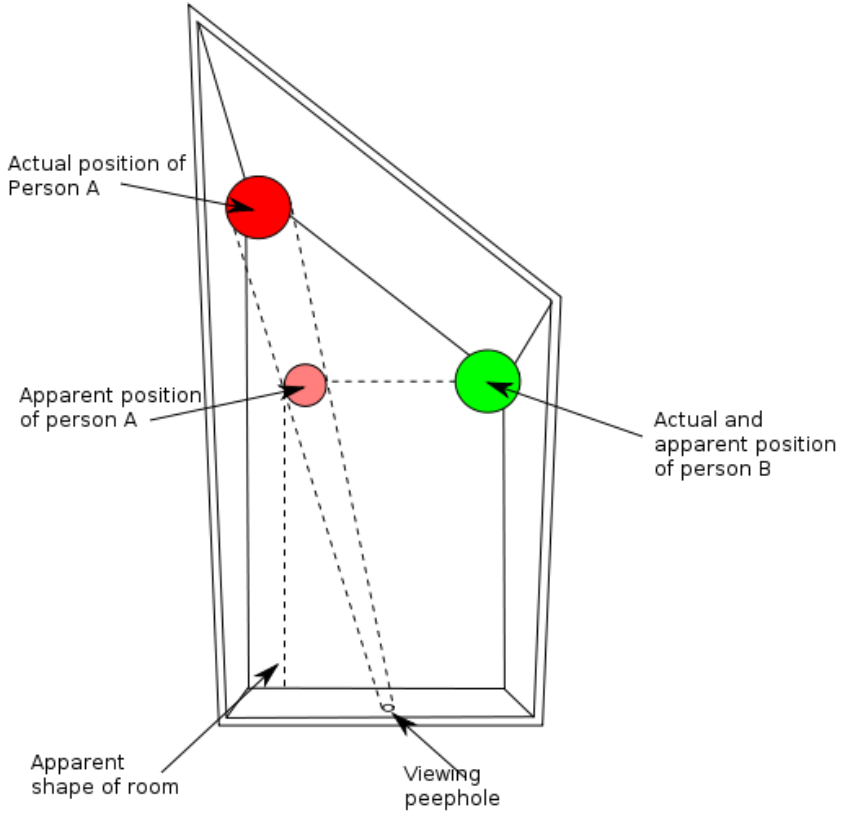


OUR GOAL: RECOVERY OF 3D STRUCTURE

DataLab Community

- What about perspective?

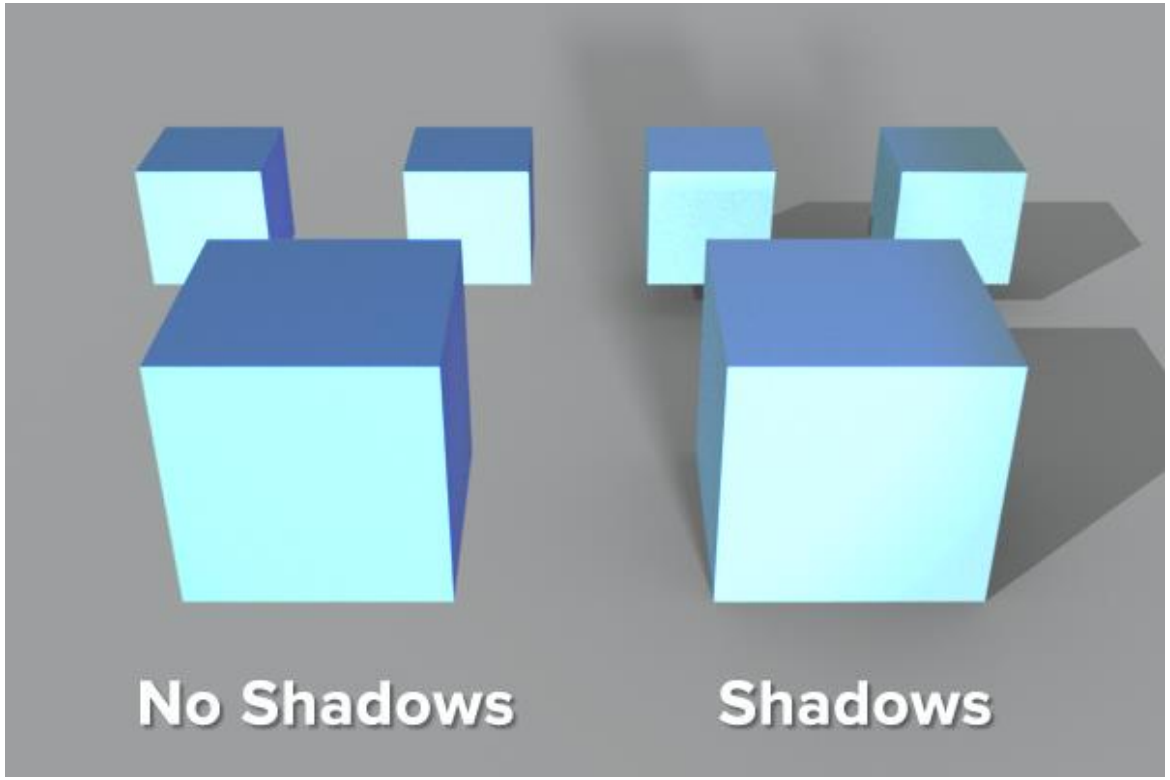
AMES ROOM


Actual position of
Person A
Apparent position
of person A
Actual and
apparent position
of person B
Apparent
shape of room
Viewing
peephole

AMES ROOM

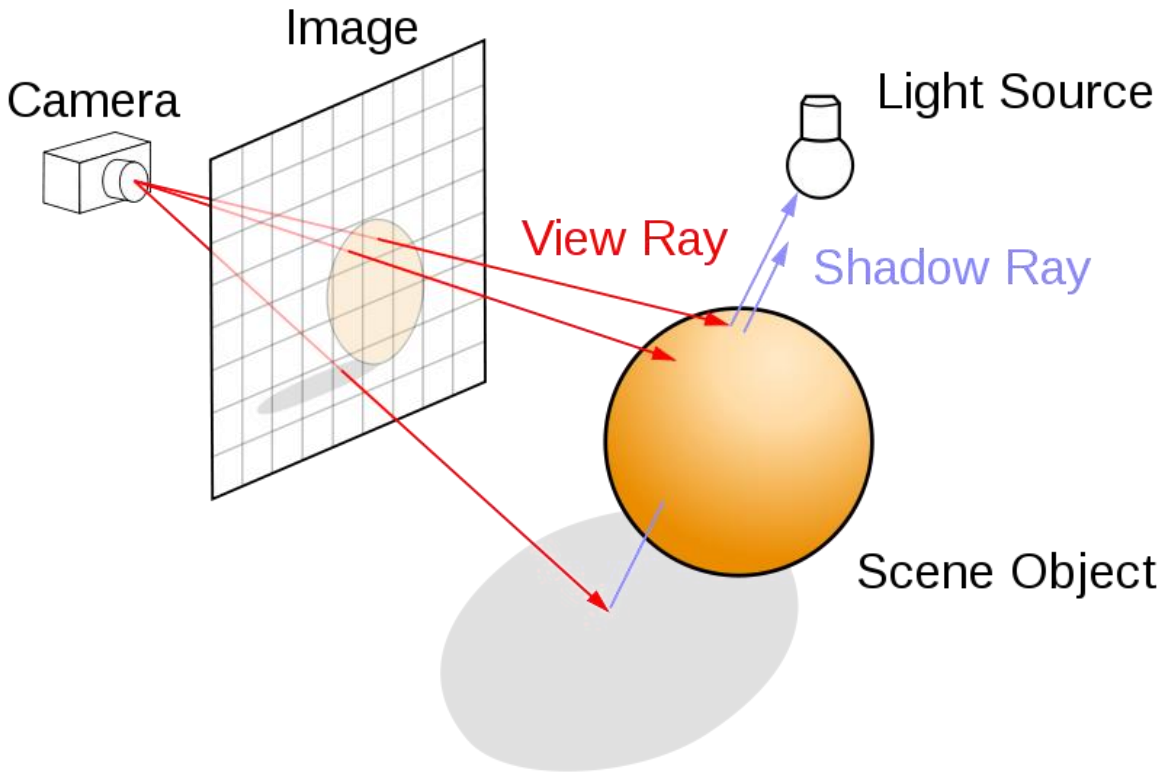
DataLab Community

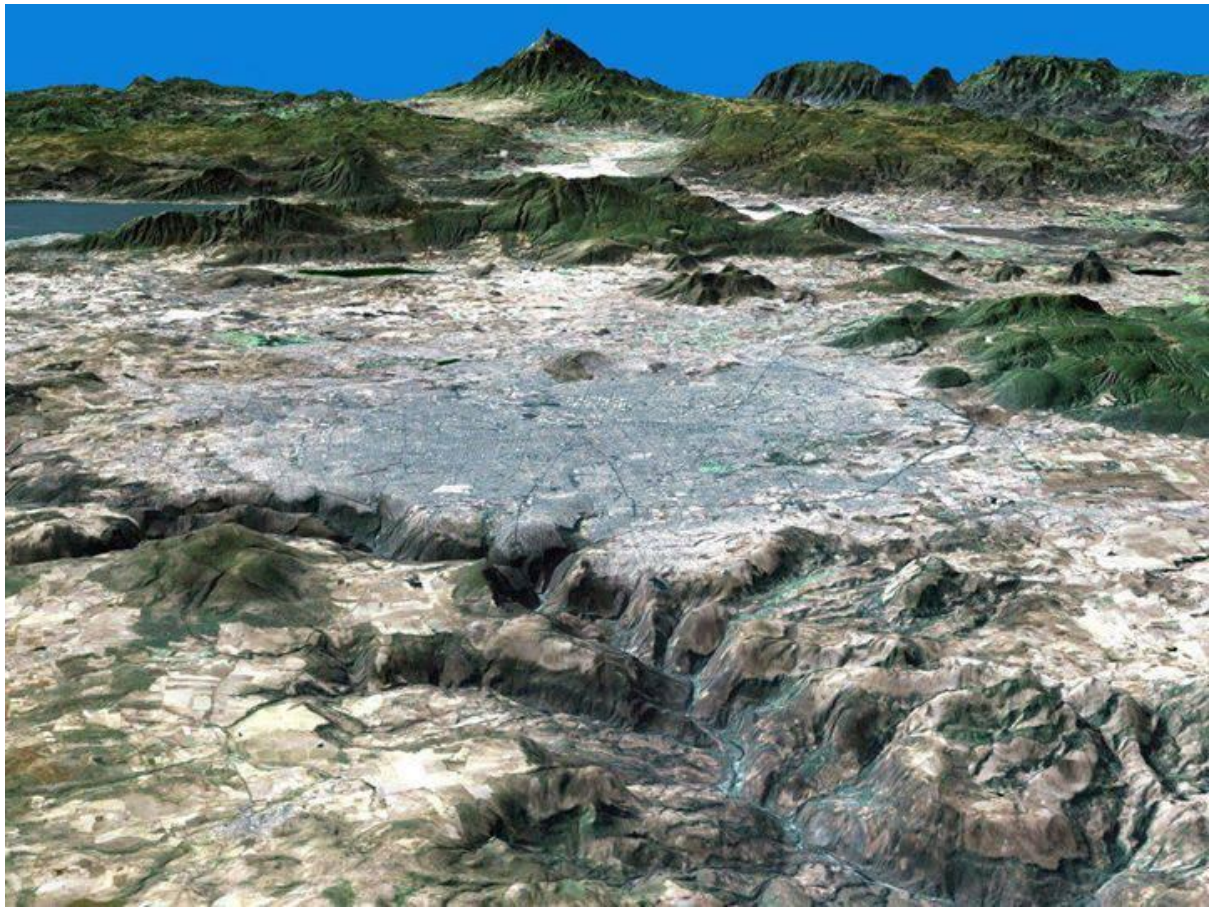
VIRTUAL REALITY LIGHTING



RAY TRACING

DataLab Community



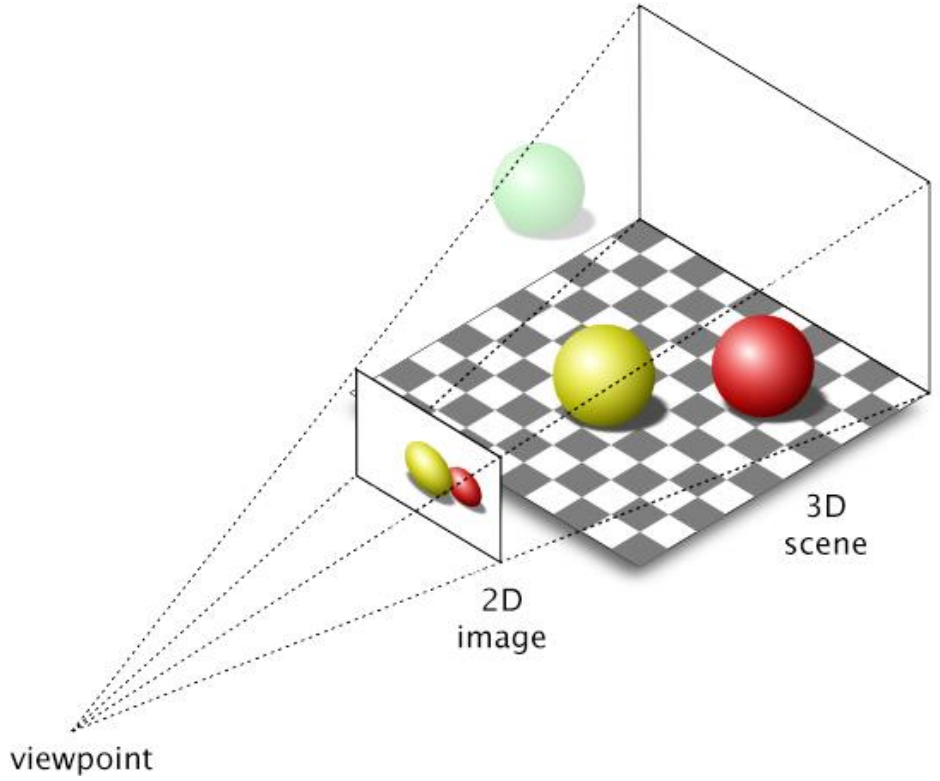


WHAT IF WE JUST WORK WITH THE CENTER OF THE LENS?

DataLab Community

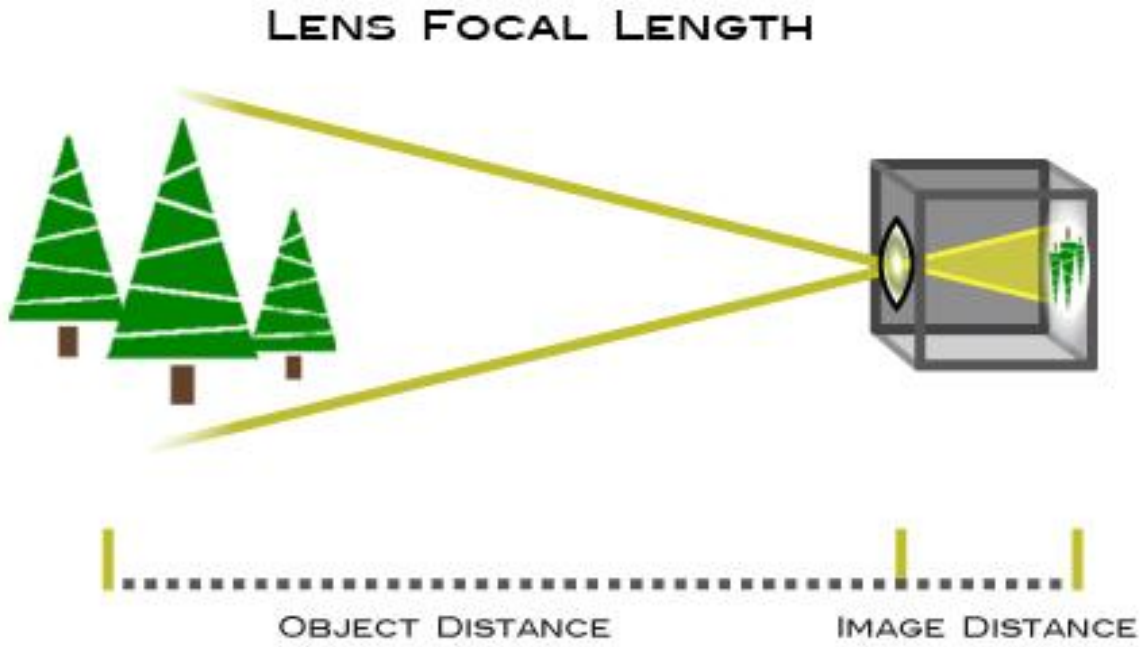
MOVE INFORMATION FROM 3D TO 2D

DataLab Community



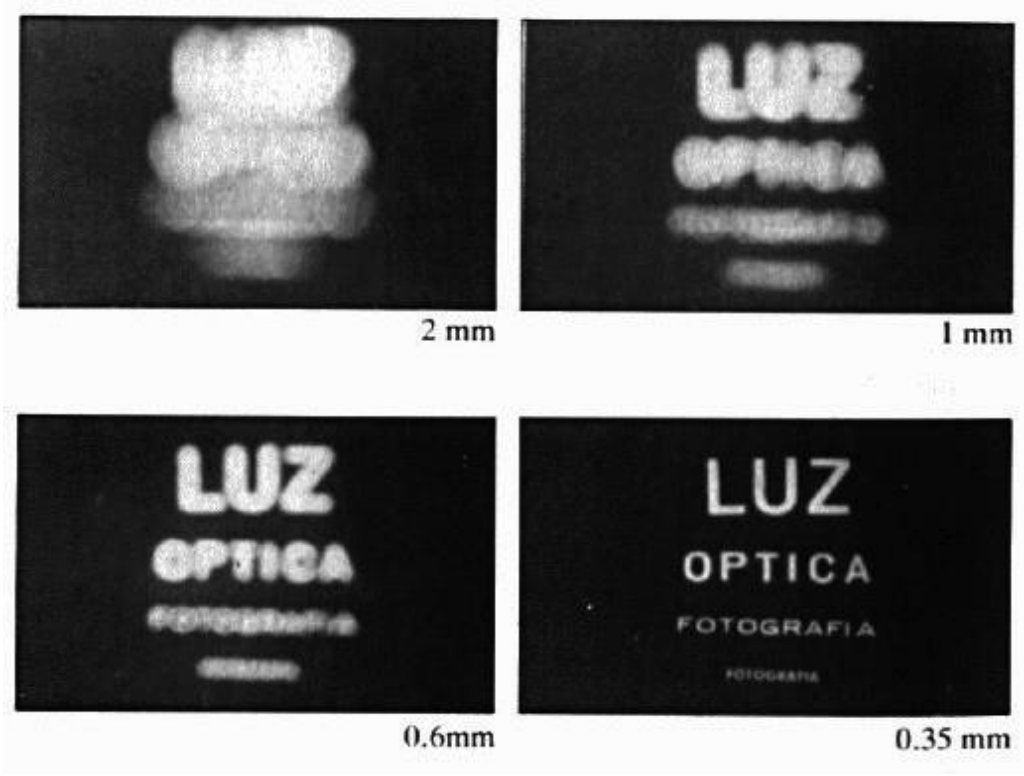
DOES OPENNESS MATTER?

- In the case of the above drawing, is it important to enter the dimensions of that opening where the light enters the box?



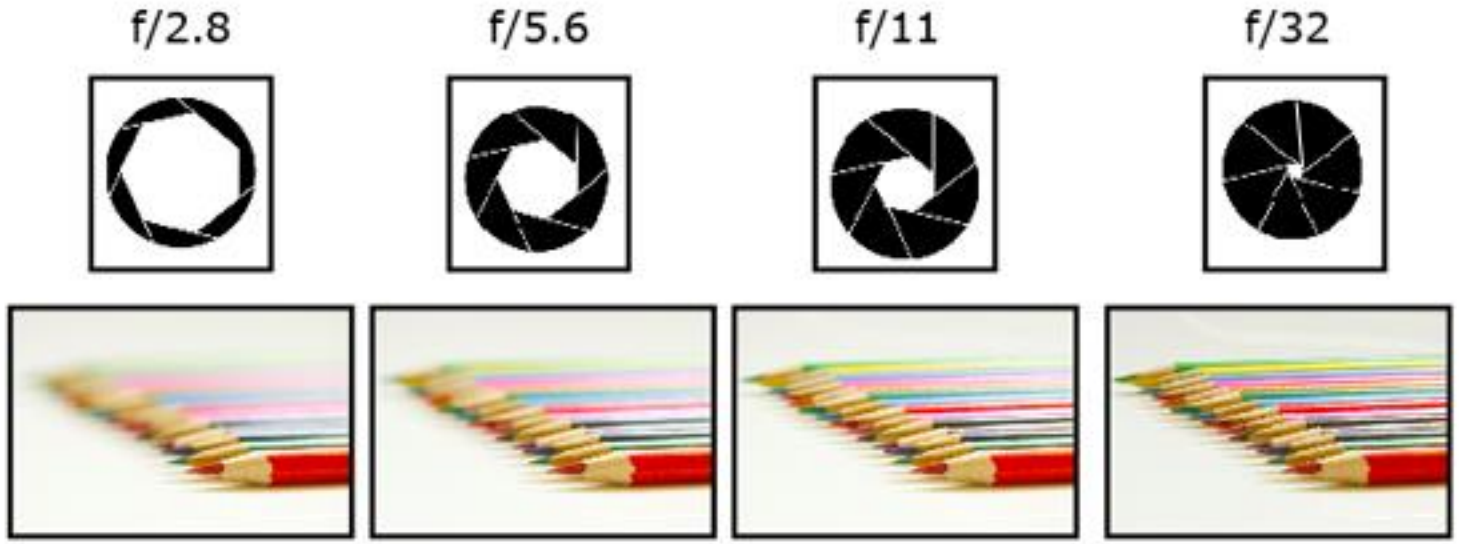
DOES OPENNESS MATTER?

DataLab Community



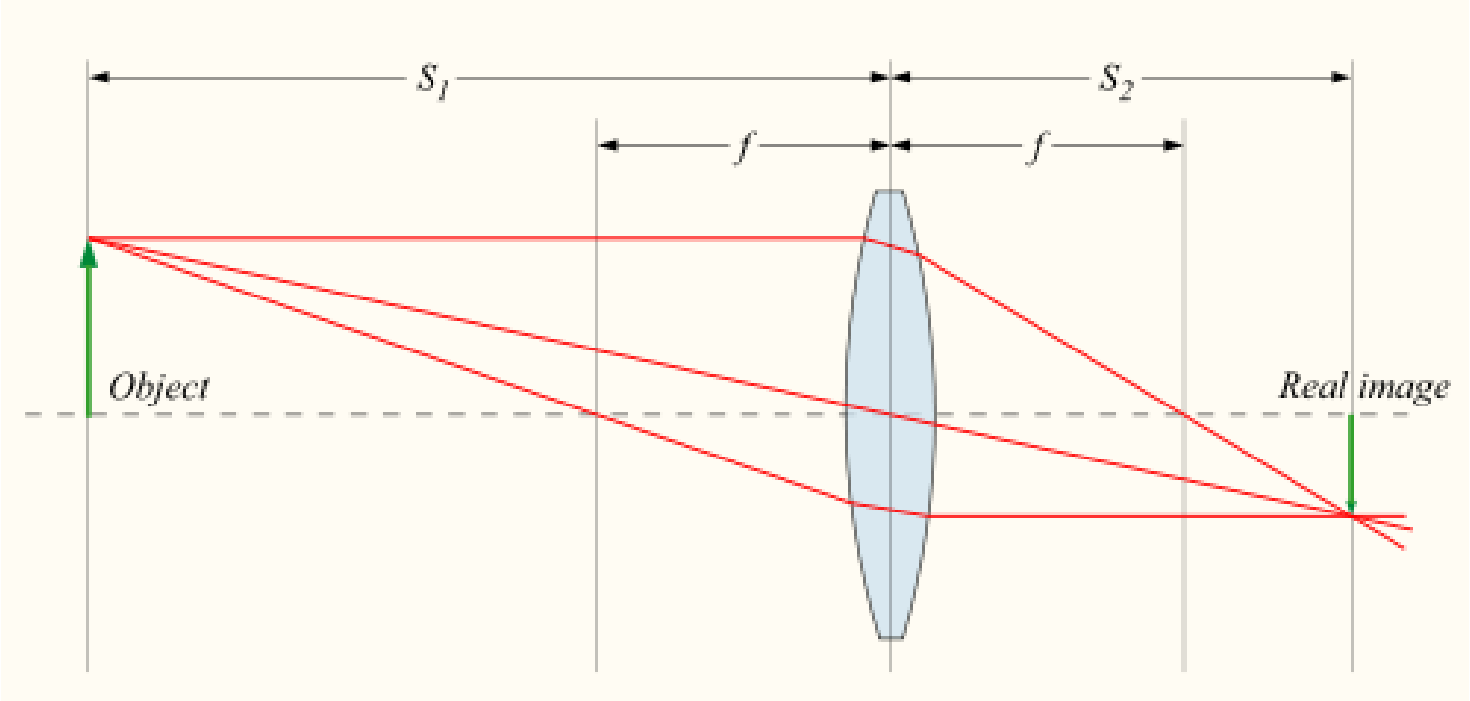
OPENNESS AND DEPTH ARE CORRELATED

DataLab Community



THIN LENSES

- Gaussian Law $\frac{1}{Z} + \frac{1}{z} = \frac{1}{f}$



CAMERA CALIBRATION

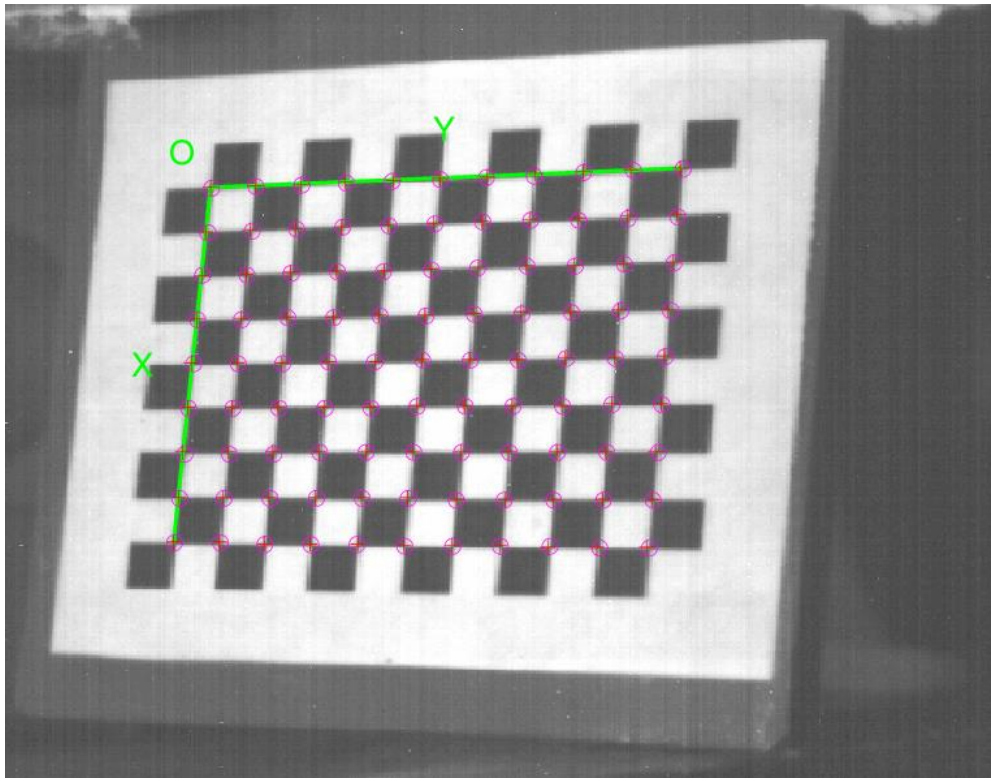
DataLab Community

- $$\begin{pmatrix} p_{xi} \\ p_{yi} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{f}{dx} & -\rho\frac{f}{dy} & c_{xi} \\ 0 & \frac{f}{d_y} & c_{yi} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_W \\ Z_W \end{pmatrix}$$

COMPUTER VISION

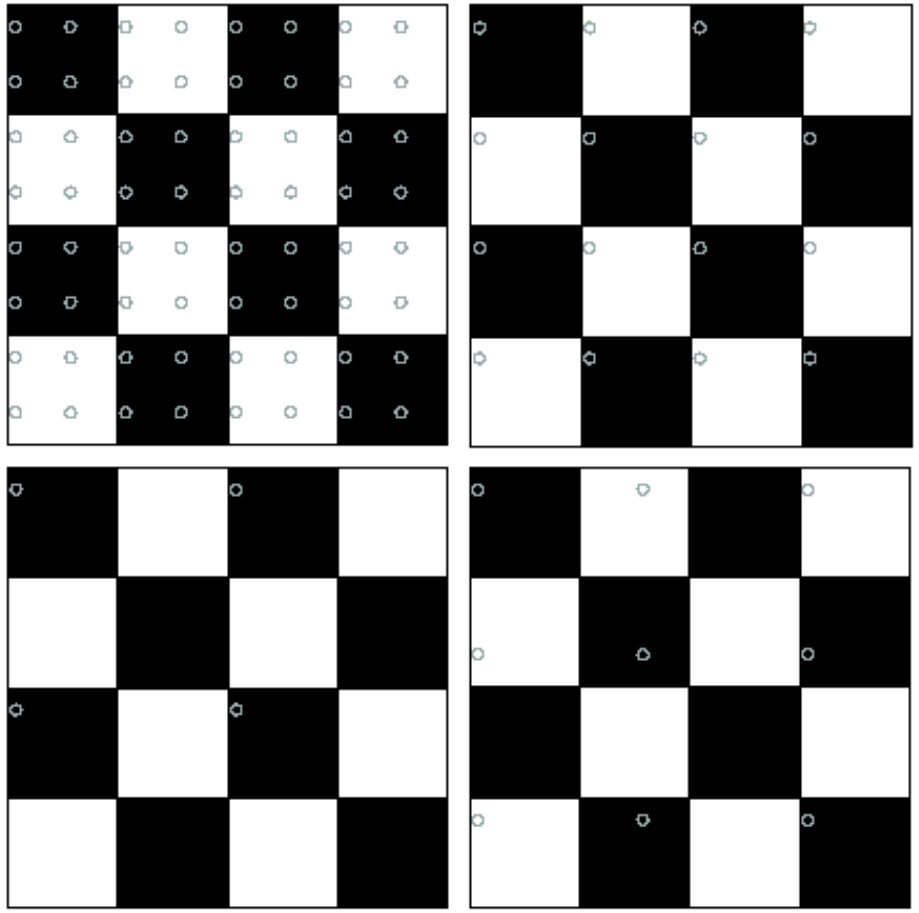
DEMO ARUCO

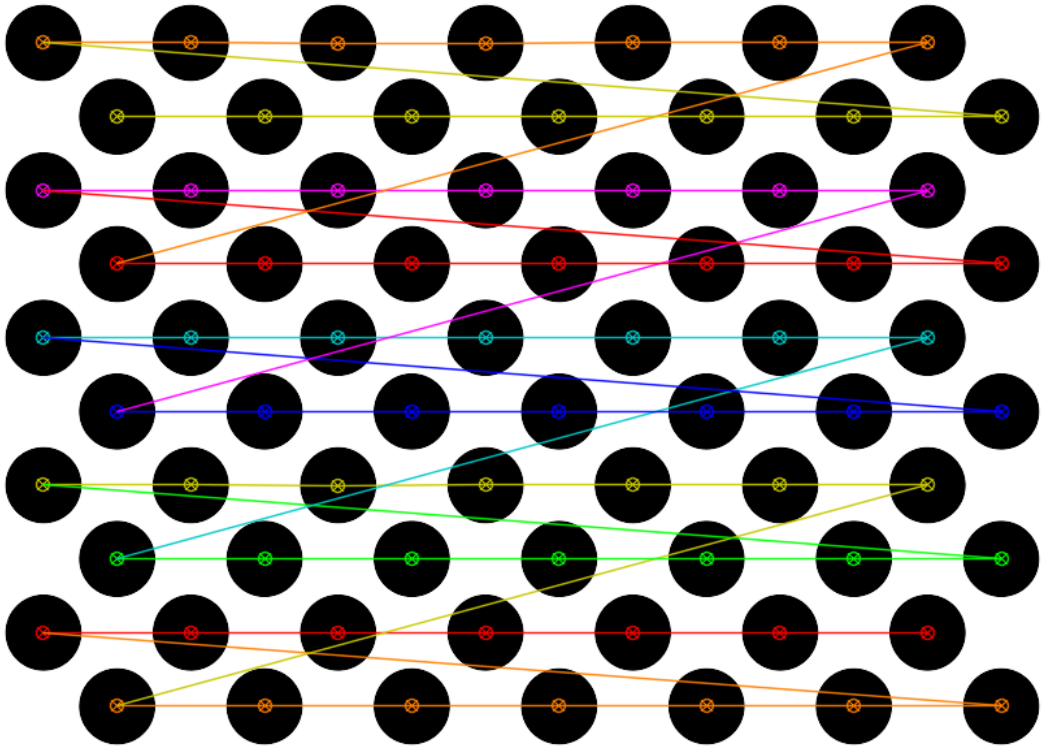
CHECKERBOARD



CHECKERBOARD

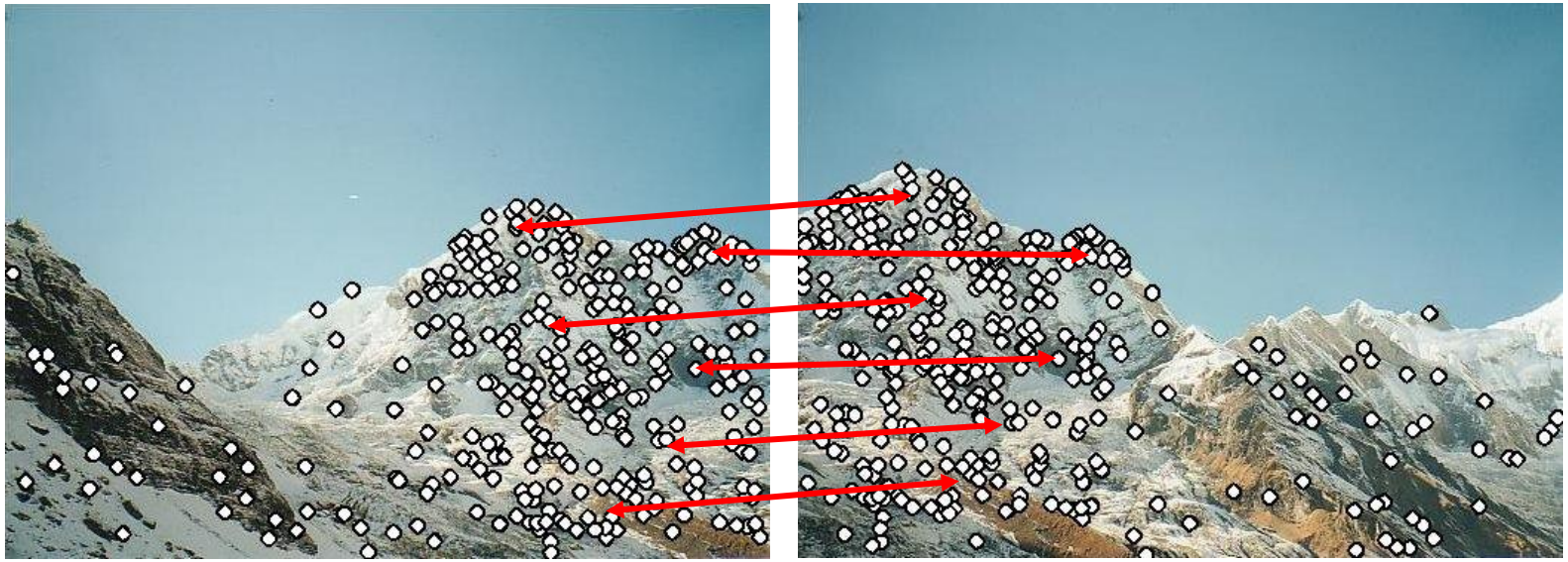
DataLab Community





FEATURE MATCHING

DataLab Community



COMPUTER VISION

DEMO MATCHING

DataLab Community

WHAT IS CONVOLUTION?

REMEMBER...

AVERAGE FILTER

DataLab Community

- Then for a neighborhood of 3...

$$\frac{1}{9}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

AND IN PICTURES...

DataLab Community

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

	0								

AND IN PICTURES...

DataLab Community

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

	0	10							

AND IN PICTURES...

DataLab Community

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

	0	10	20						

AND IN PICTURES...

DataLab Community

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

	0	10	20	30					

AND IN PICTURES...

DataLab Community

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

	0	10	20	30					

AND IN PICTURES...

DataLab Community

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

	0	10	20	30	30	30	20	10	
	0	20	40	60	60	60	40	20	
	0	30	60	90	90	90	60	30	
	0	30	50	80	80	90	60	30	
	0	30	50	80	80	90	60	30	
	0	20	30	50	50	60	40	20	
	10	20	30	30	30	30	20	10	
	10	10	10	0	0	0	0	0	

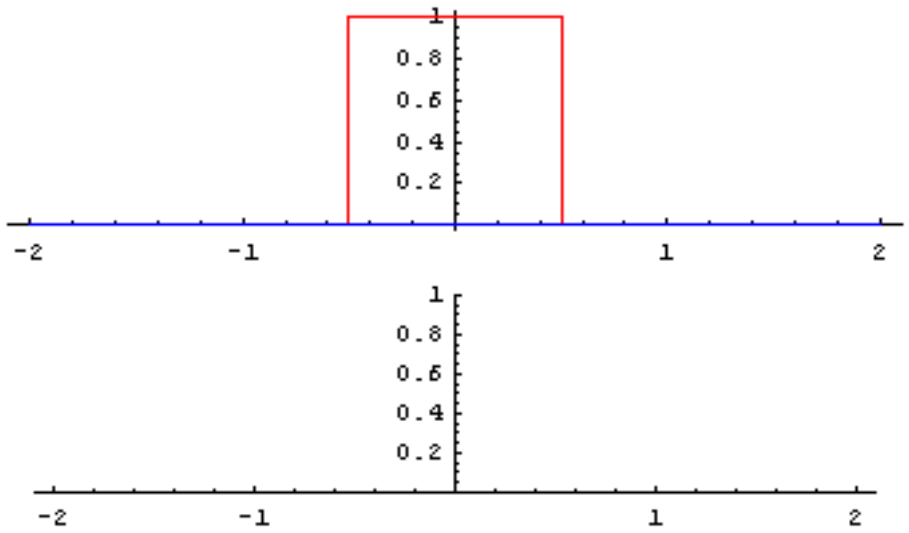
WHAT IS CONVOLUTION?

DataLab Community

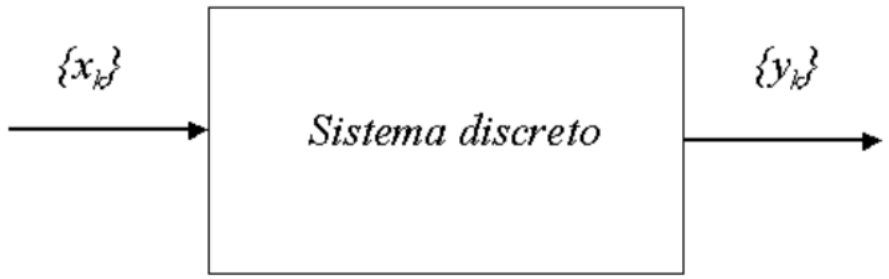
$$(f * g)(t) = \int_{-\infty}^{\infty} f(\eta) g(t - \eta)\, d\eta$$

WHAT IS CONVOLUTION?

DataLab Community



WHY THE CONVOLUTION?



WHY THE CONVOLUTION?

- Filtros FIR (Finite Impulse Response)
- Filtros IIR (Infinite Impulse Response)

CONVOLUTION THEOREM

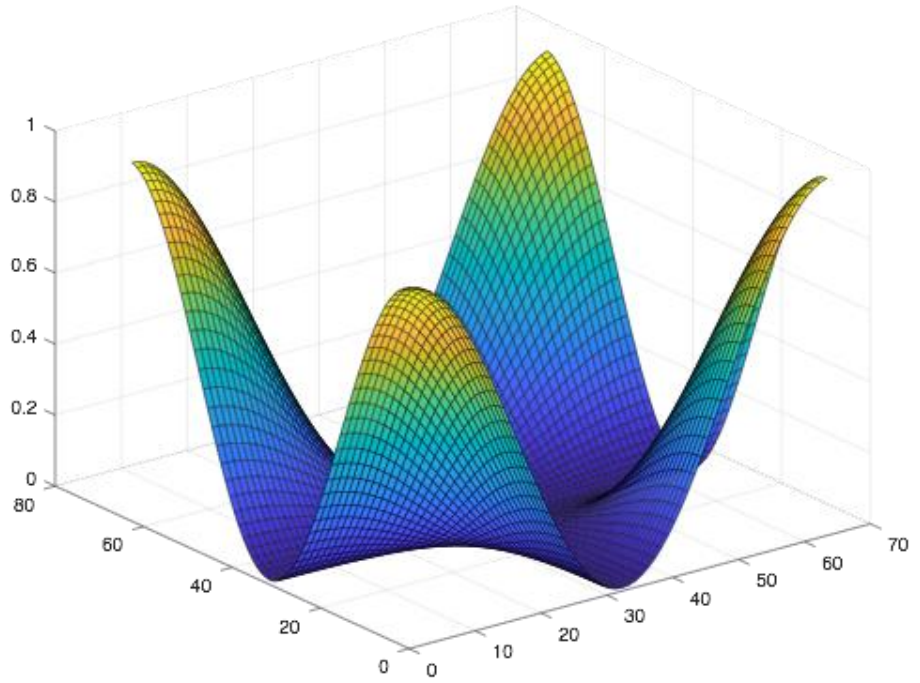
DataLab Community

$$\mathcal{F}(f * g) = F(\omega)G(\omega)$$

FILTERS IN WAVE NUMBER DOMAIN

DataLab Community

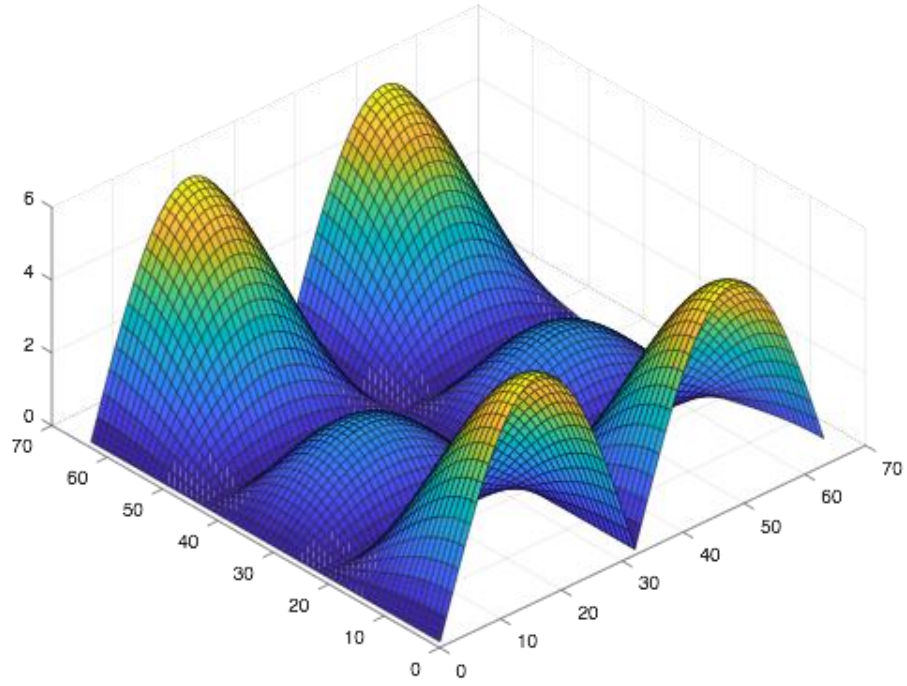
$$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$



FILTERS IN WAVE NUMBER DOMAIN

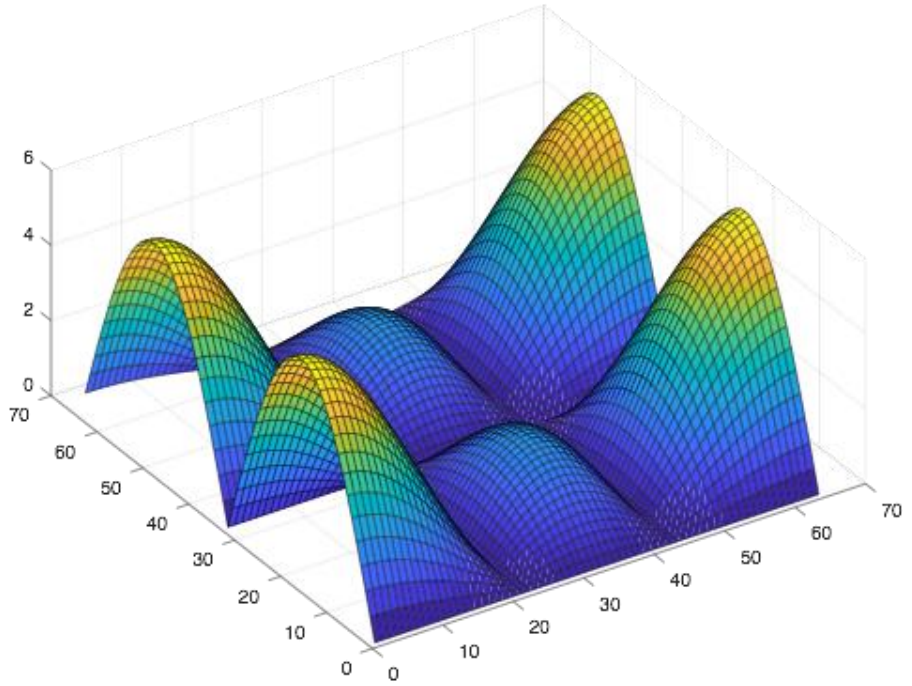
DataLab Community

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$



FILTERS IN WAVE NUMBER DOMAIN

$$\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$



COMPUTER VISION

DEMO EDGES

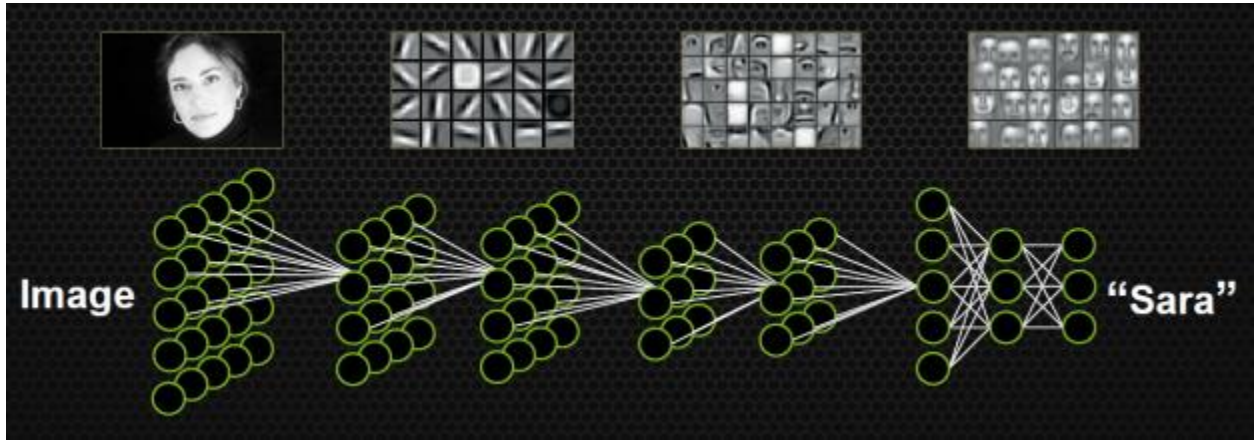
IMAGE PROCESSING VS CNN

COMPUTER VISION

DataLab Community

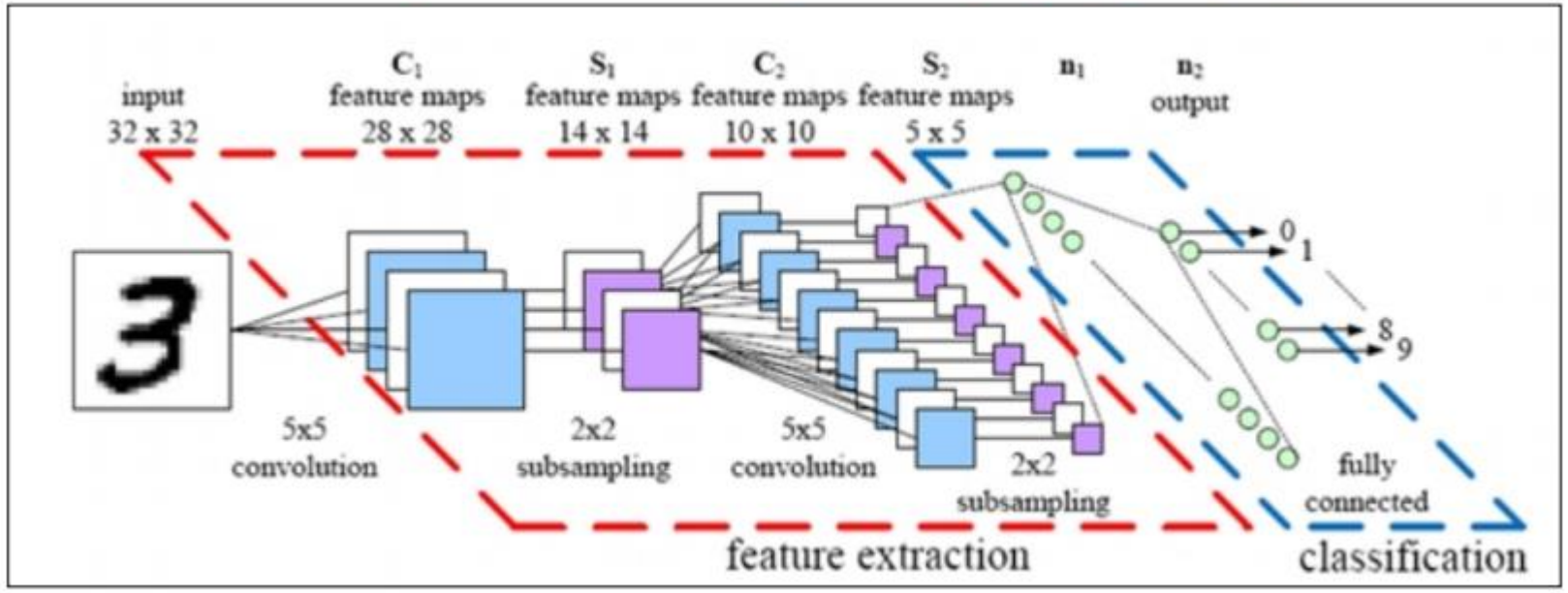
CONVOLUTIONAL NEURAL NETWORKS

DataLab Community



CNN

DataLab Community

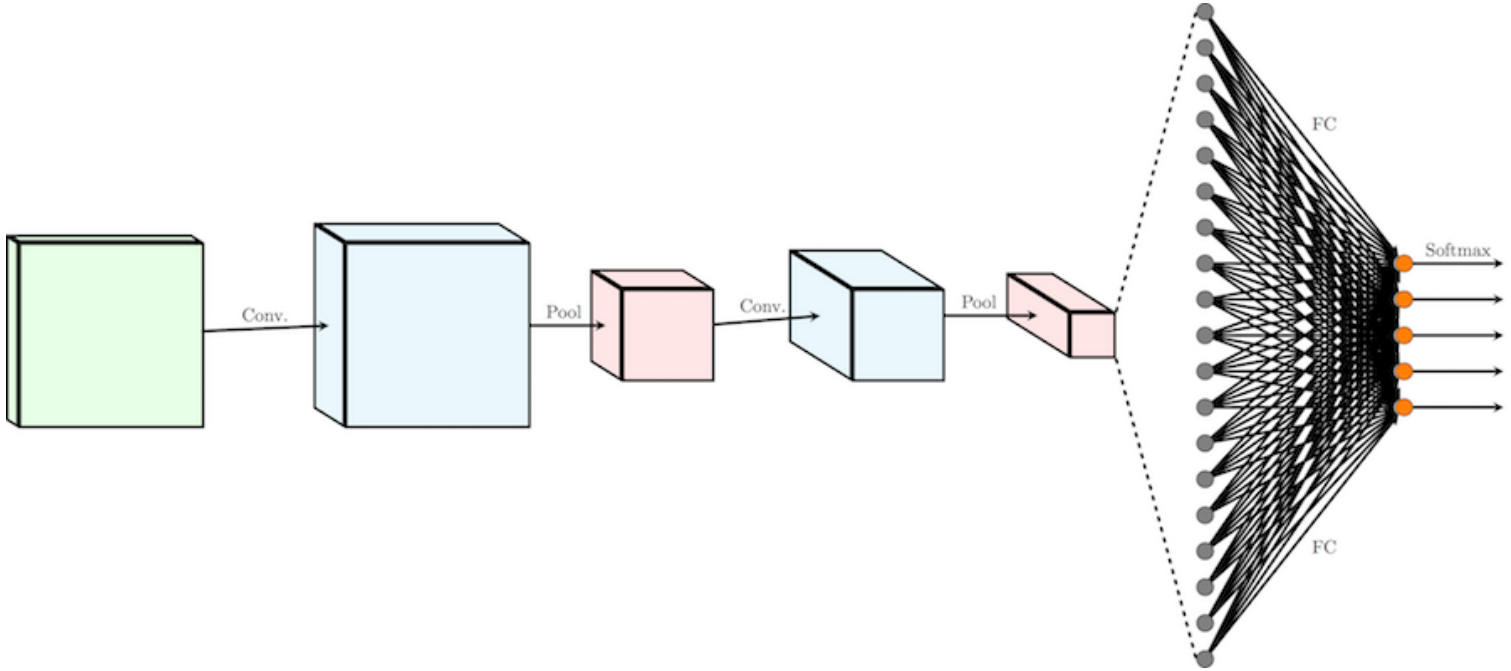


ARCHITECTURE OF A CNN

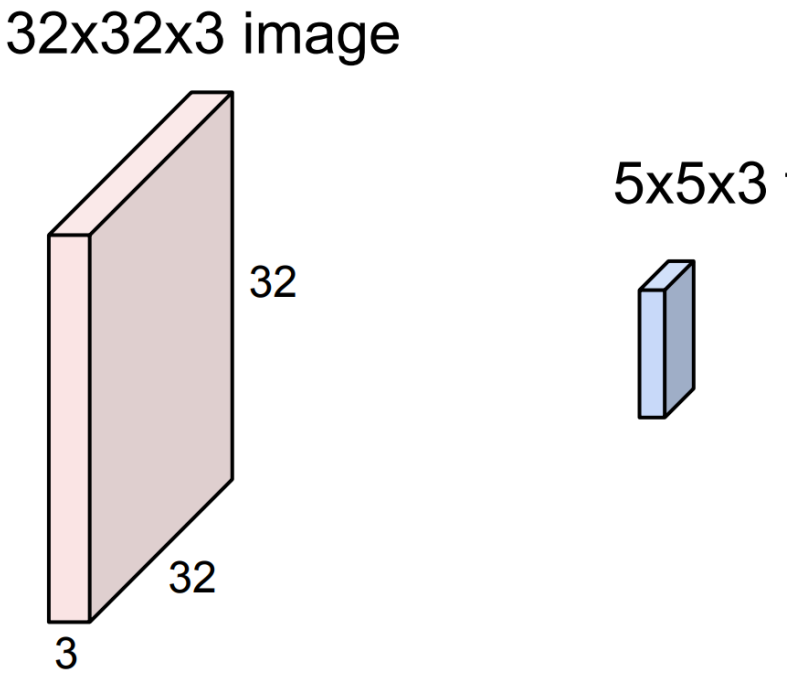
- Convolutional Layer
- Pooling Layer
- Flatten Layer
- Activation Layer
- Fully Connected Layer
- …

CONVOLUTIONAL NEURAL NETWORKS

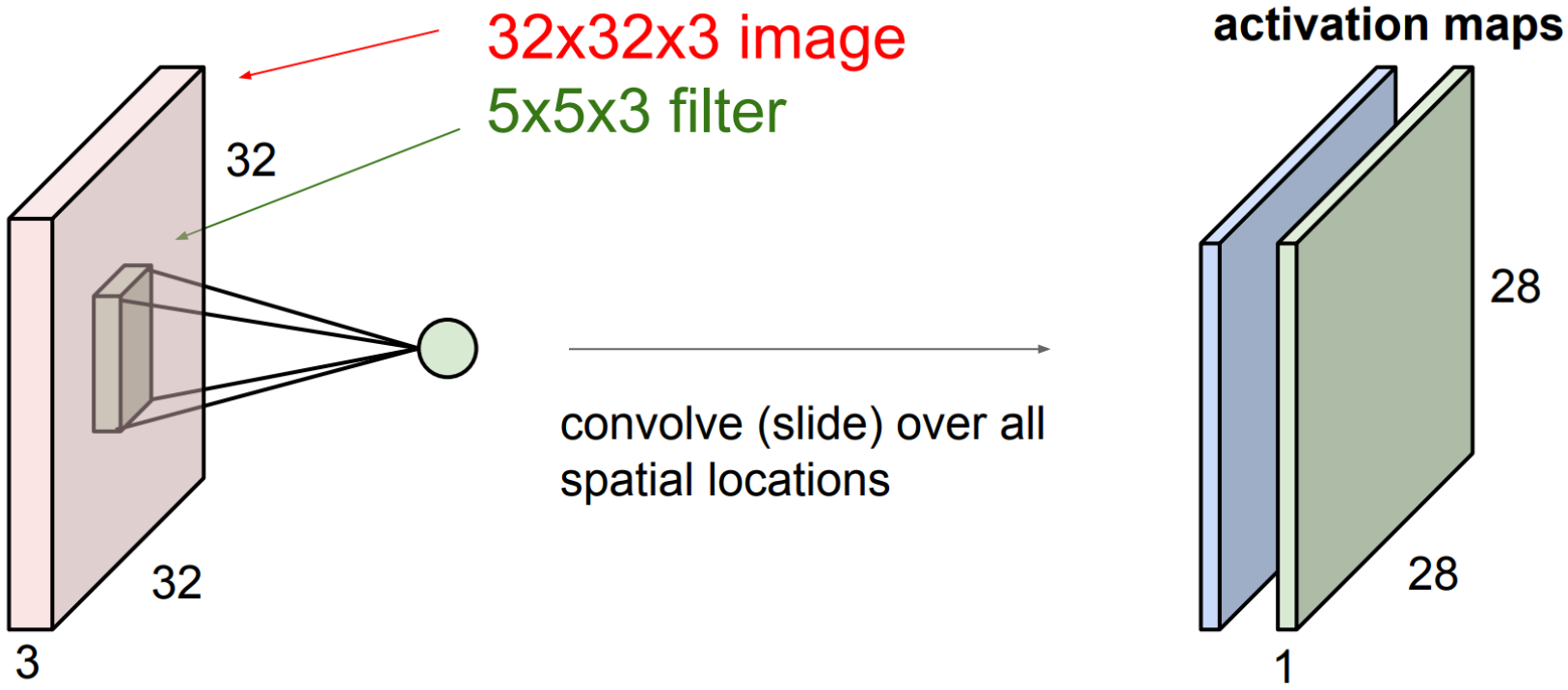
DataLab Community



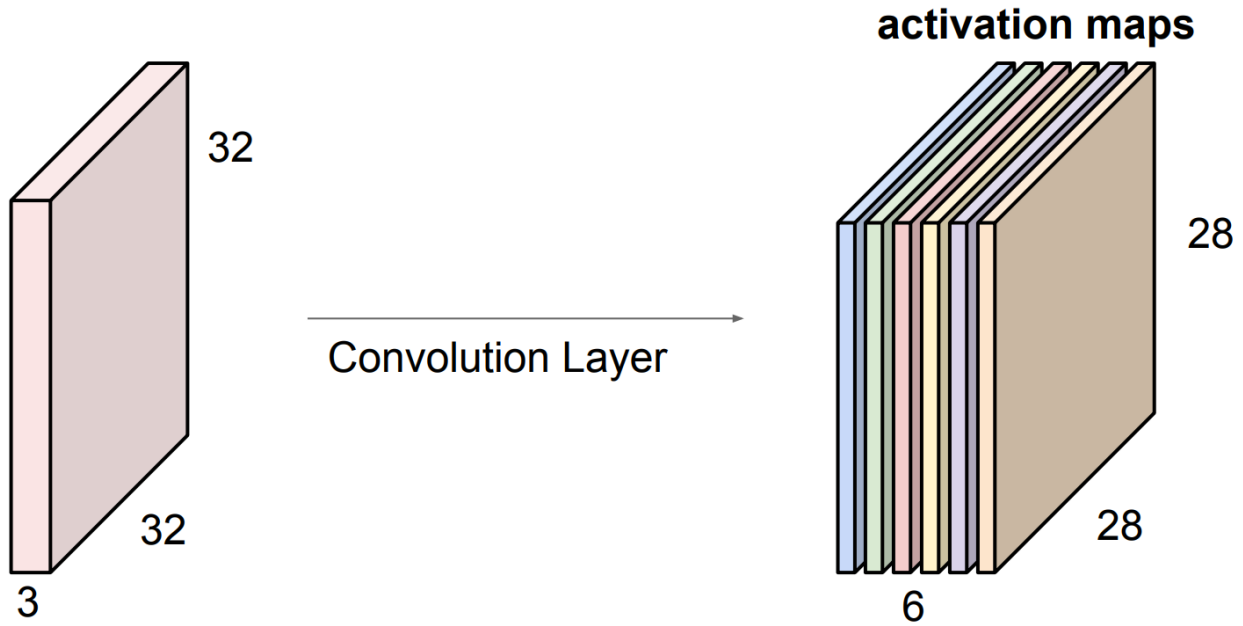
CONVOLUTIONAL LAYER



ACTIVATION MAP CNN



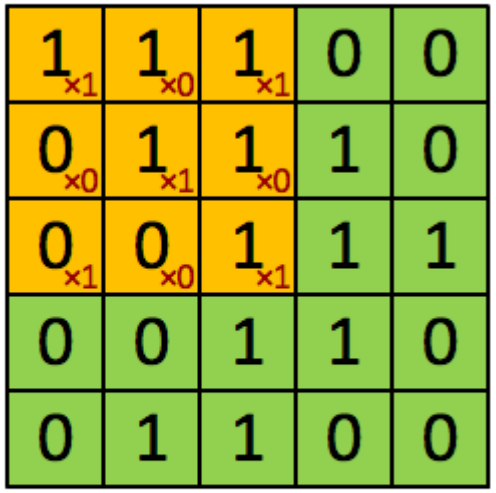
ACTIVATION MAP



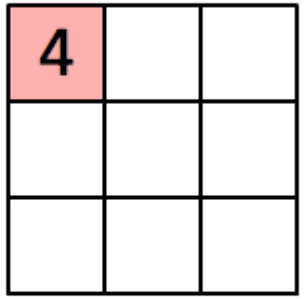
We stack these up to get a “new image” of size 28x28x6!

CONVOLUTIONAL LAYER

DataLab Community



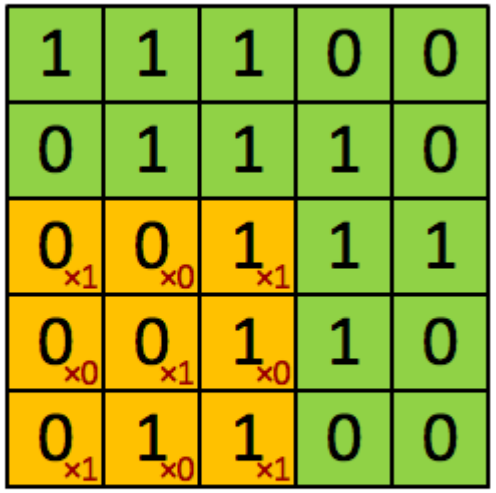
Image



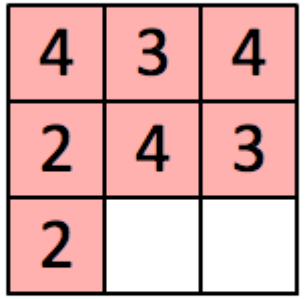
Convolved Feature

CONVOLUTIONAL LAYER

DataLab Community



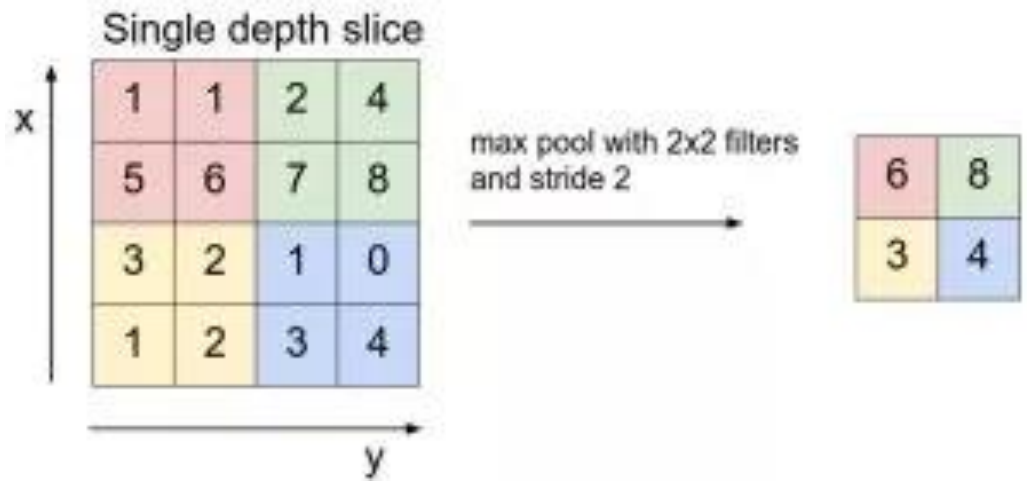
Image



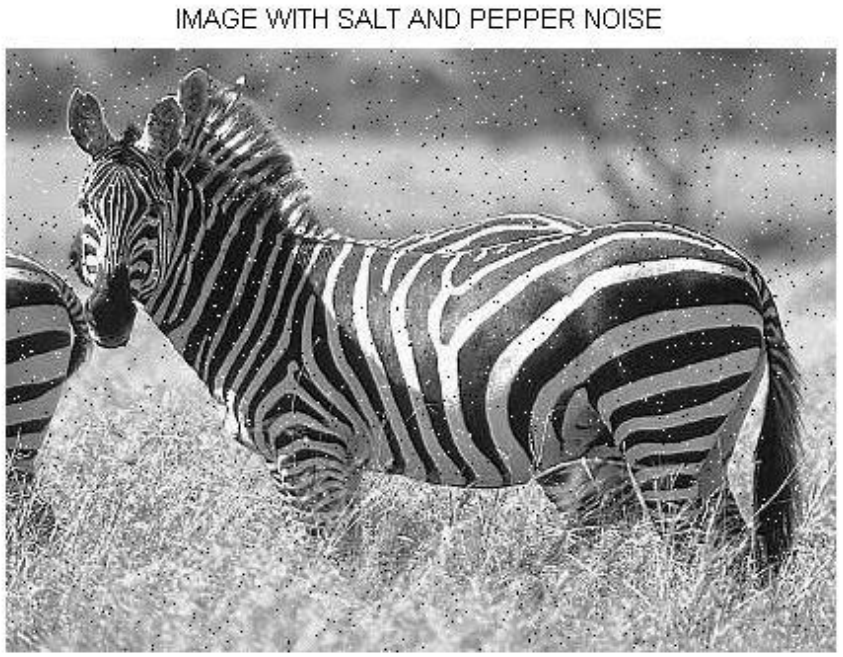
Convolved Feature

POOLING LAYER

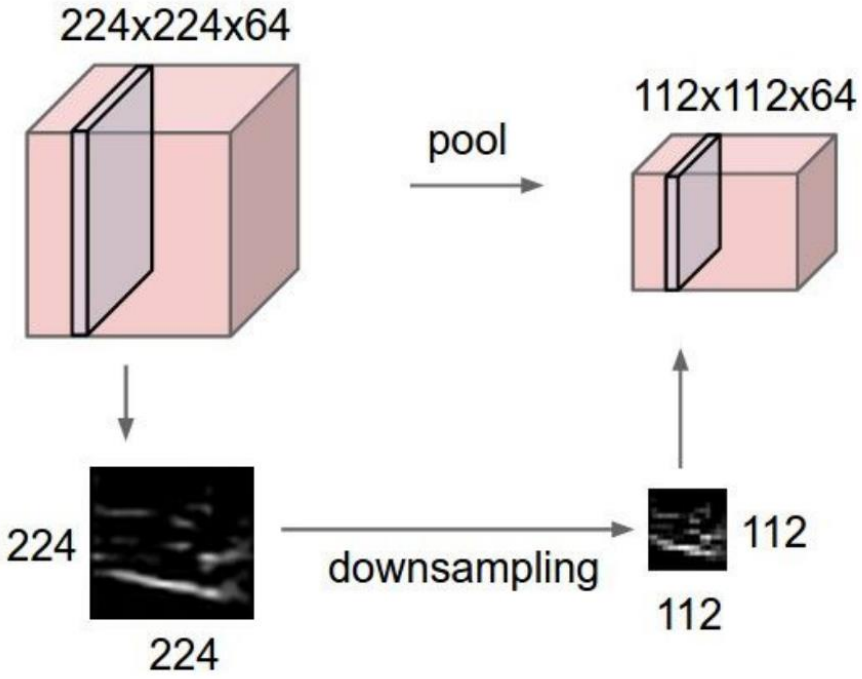
DataLab Community



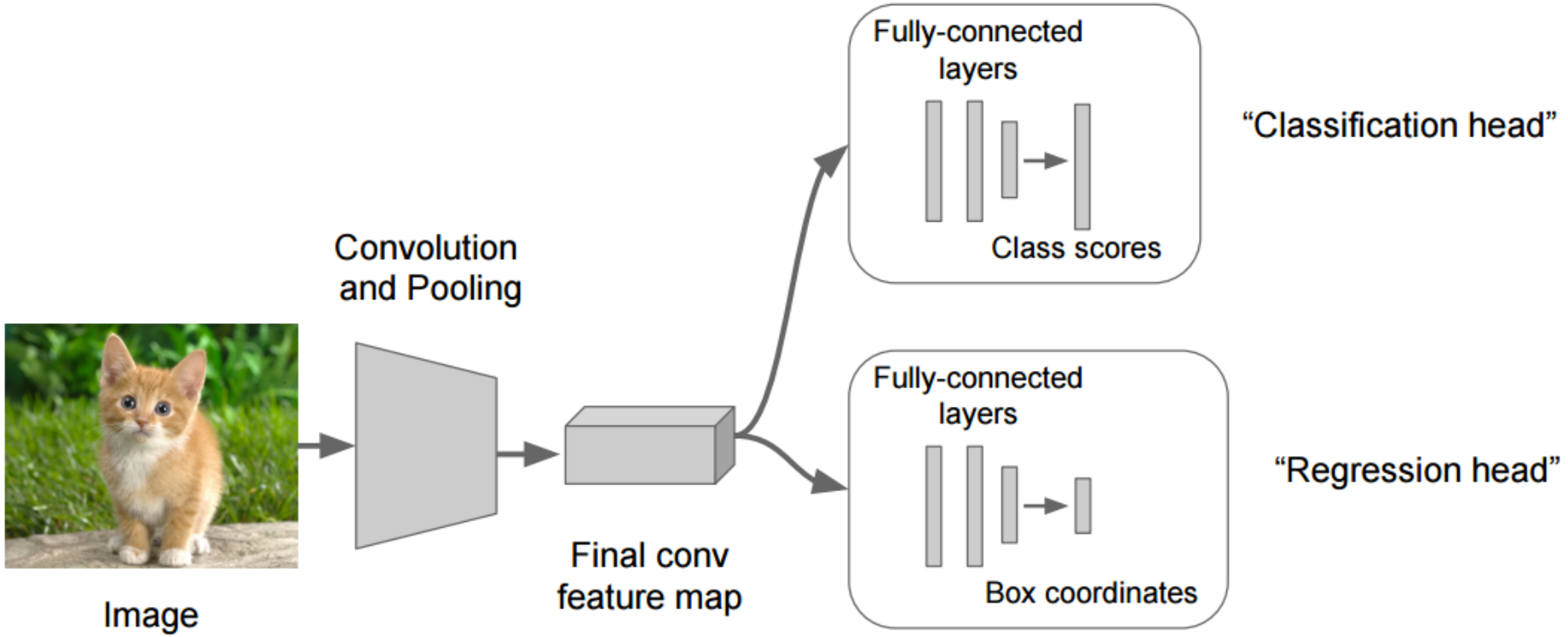
POOLING LAYER



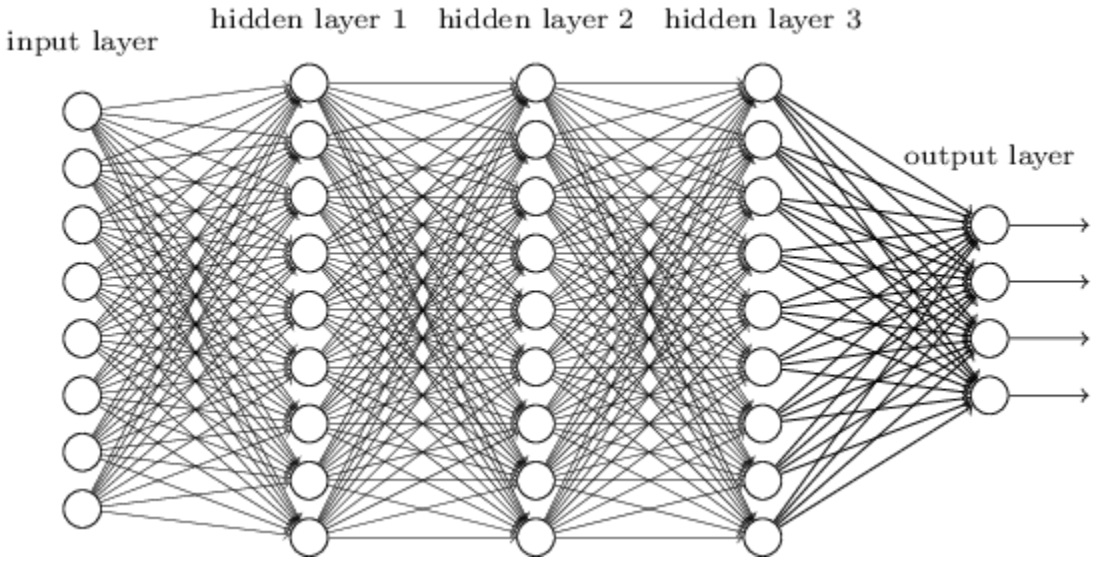
POOLING LAYER



FULLY CONNECTED LAYER

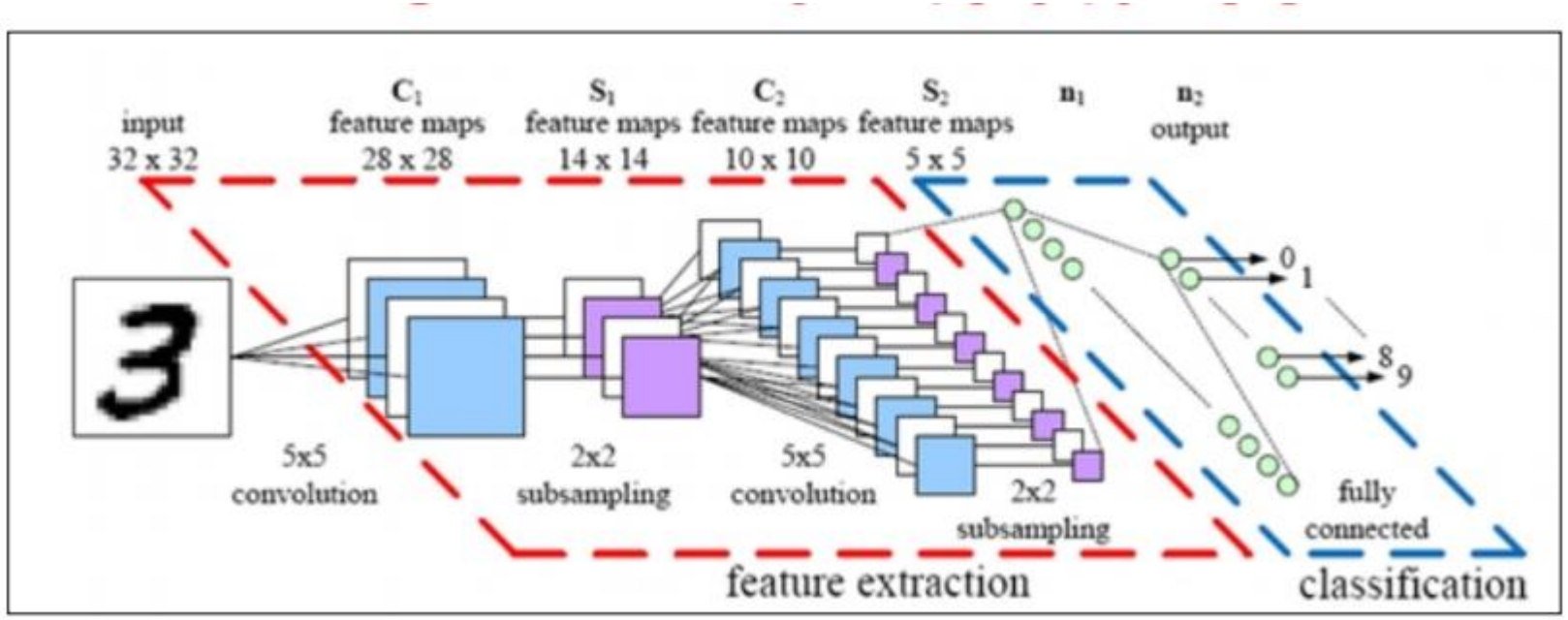


FULLY CONNECTED LAYER



DEEP LEARNING

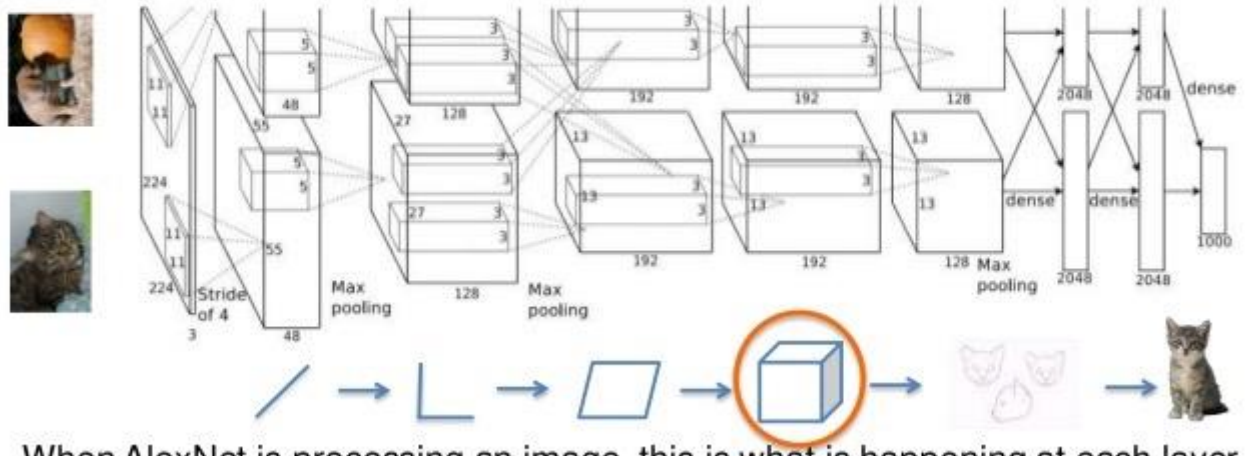
DataLab Community



DataLab Community

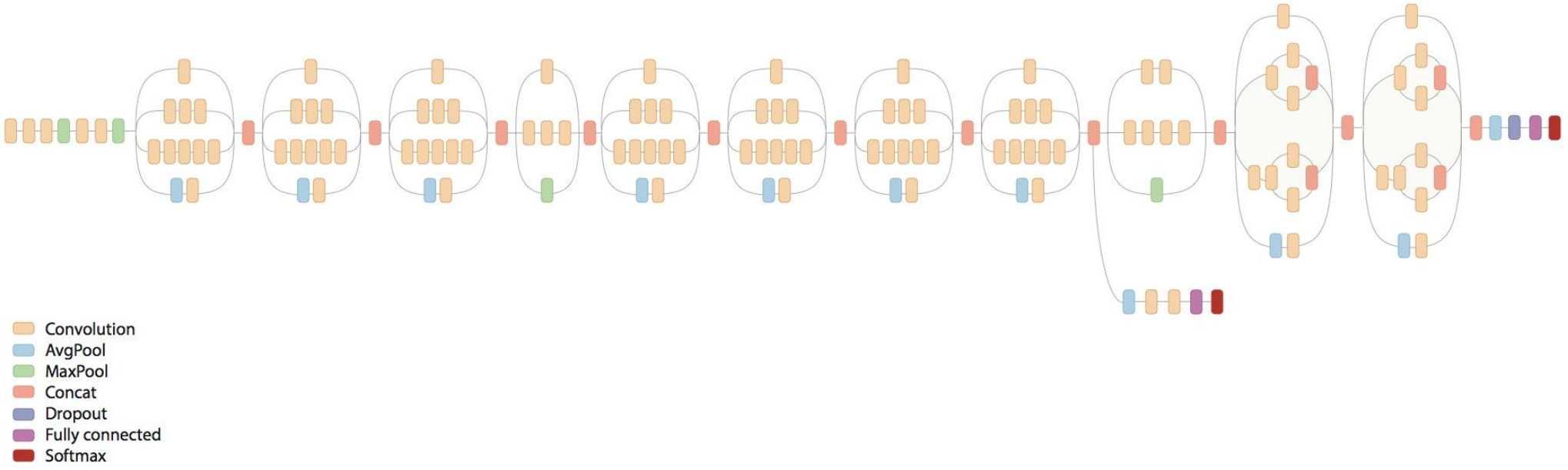
AlexNet (Krizhevsky et al. 2012)

The class with the highest likelihood is the one the DNN selects



When AlexNet is processing an image, this is what is happening at each layer.

INCEPTION V3



COMPUTER VISION

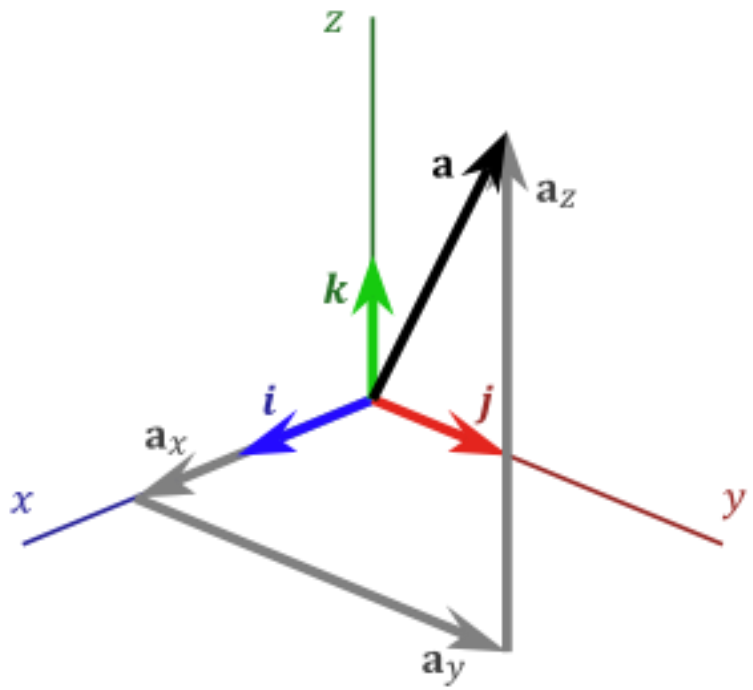
DEMO CNN

COMPUTER VISION

WHAT IS A TENSOR?

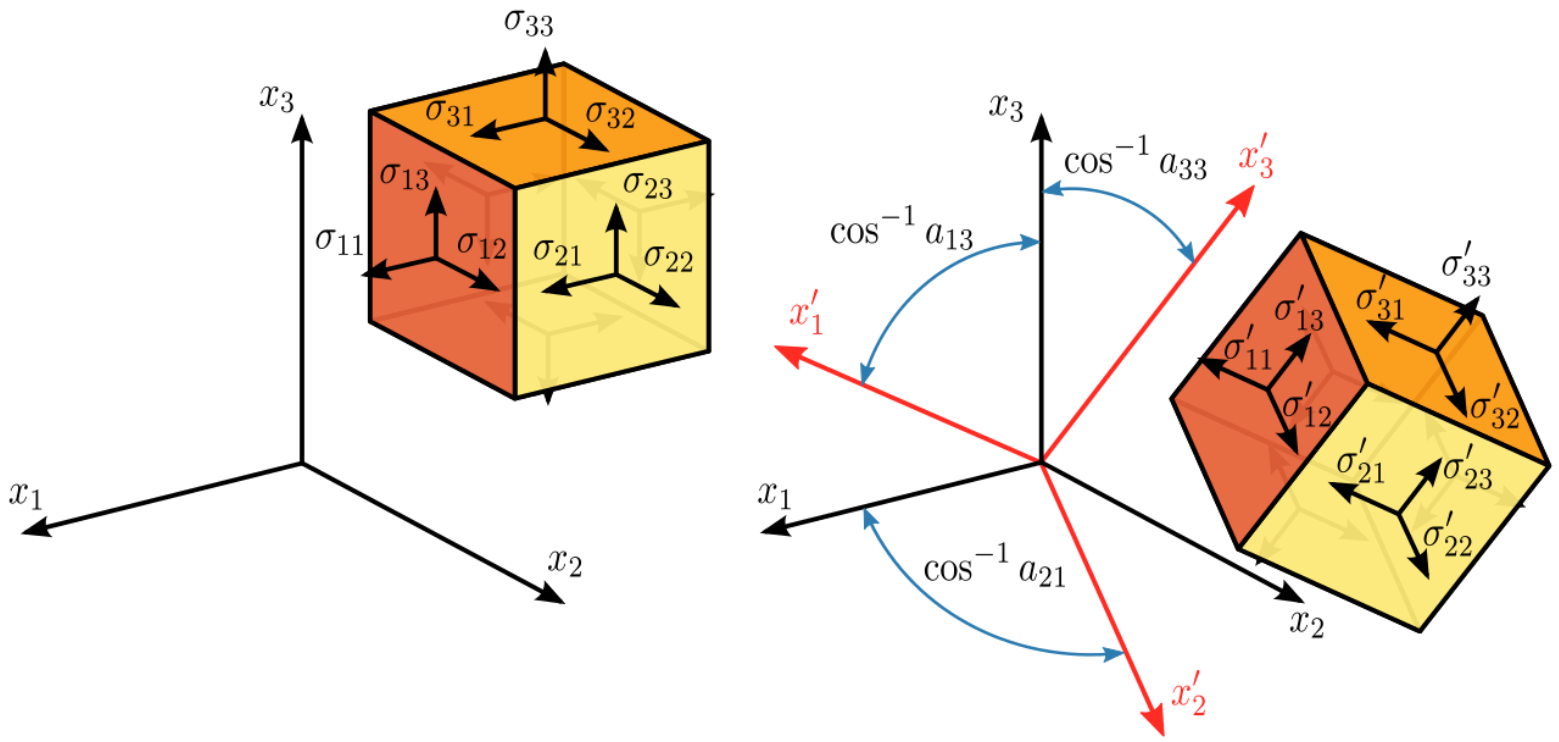
WHAT IS A TENSOR?

DataLab Community



WHAT IS A TENSOR?

DataLab Community



LILIAN LIEBER

DataLab Community

- Tensors
- “The facts of the universe”.

- *1886-1986*

LILIAN LIEBER

DataLab Community

- From "The Einstein Theory of Relativity"
- In n-dimensional space,
- a VECTOR has n components,
- a TENSOR of rank TWO has n^2 comp
- a TENSOR of rank THREE has n^3 comp
- and so on.

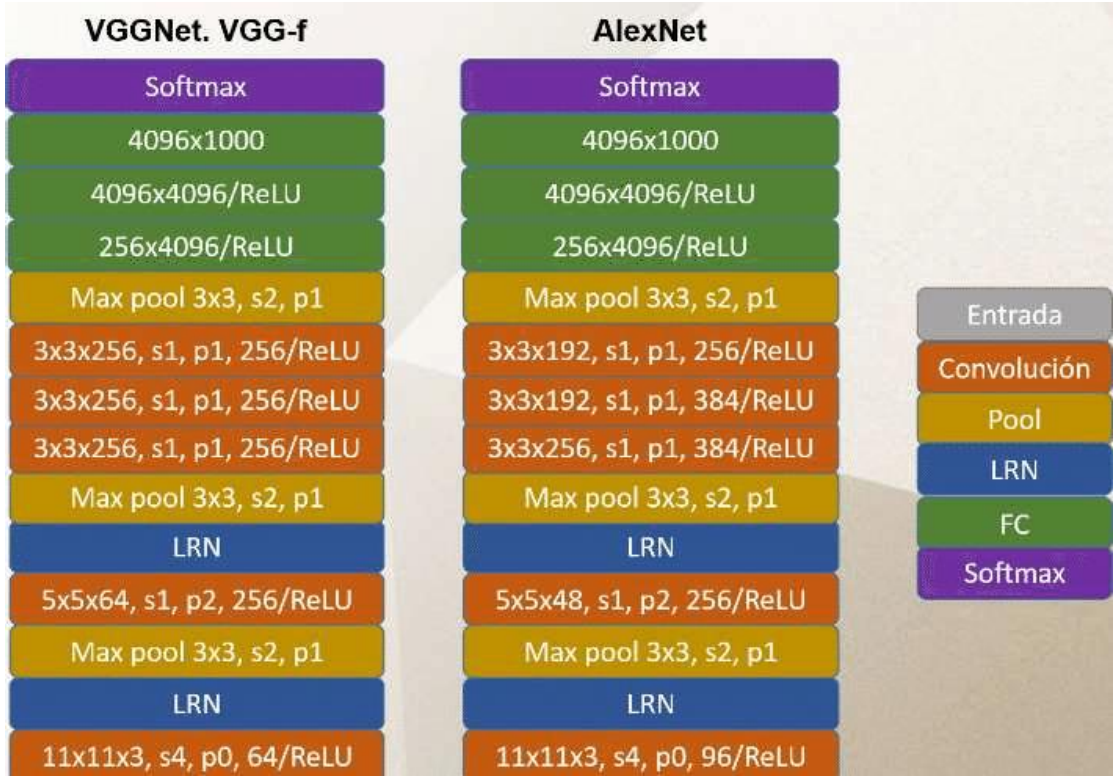
- *1886-1986*

IMAGE PROCESSING VS CNN

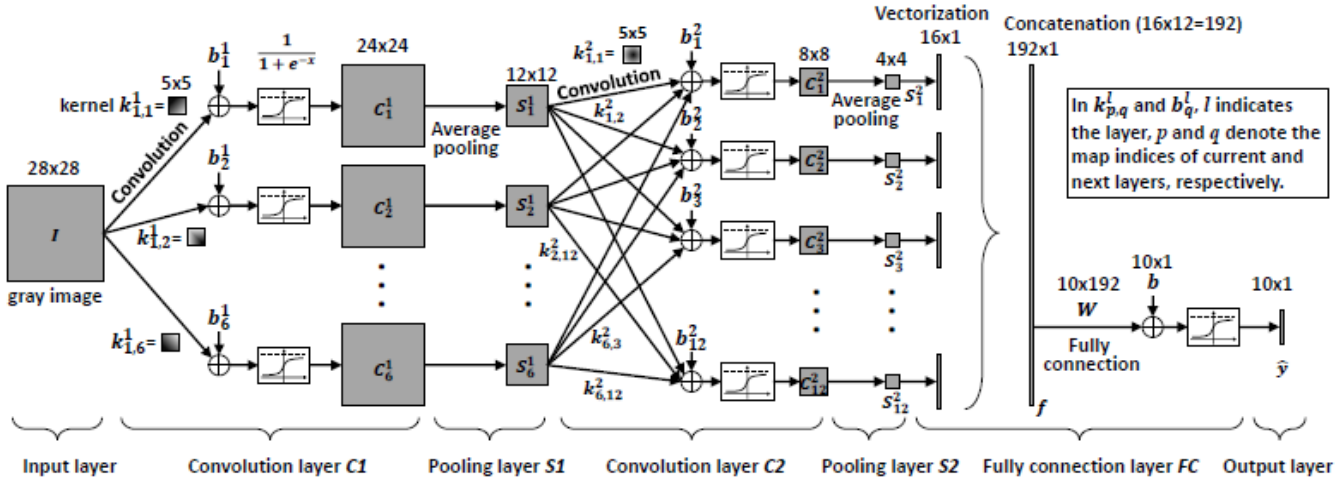
COMPUTER VISION

DataLab Community

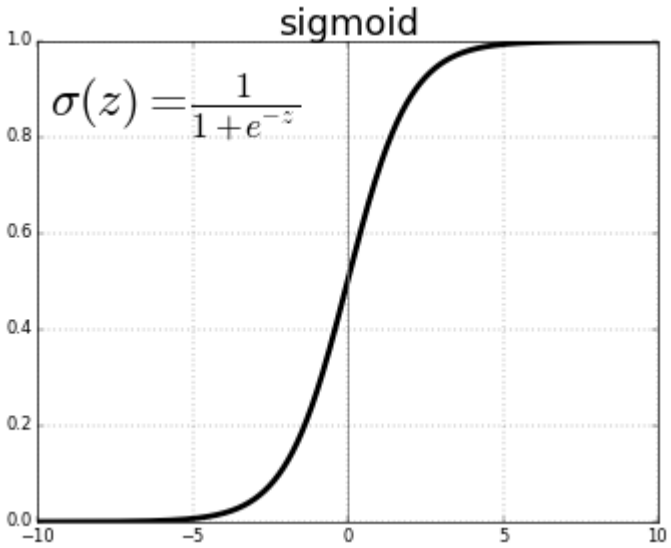
CNN ARCHITECTURE

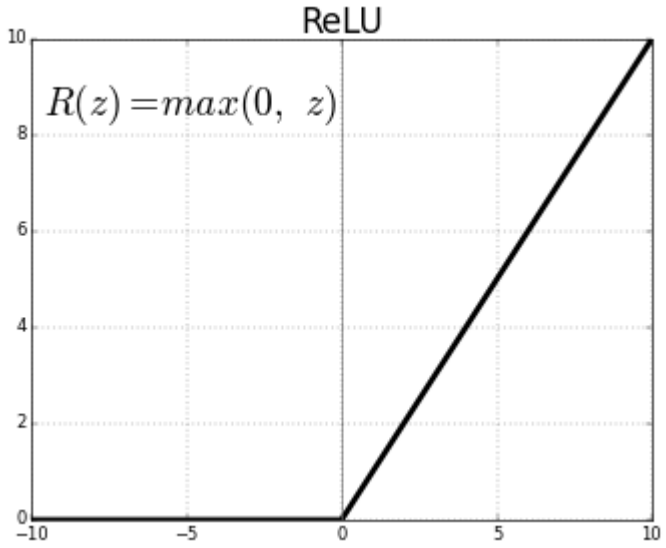


CNN ARCHITECTURE



RELU

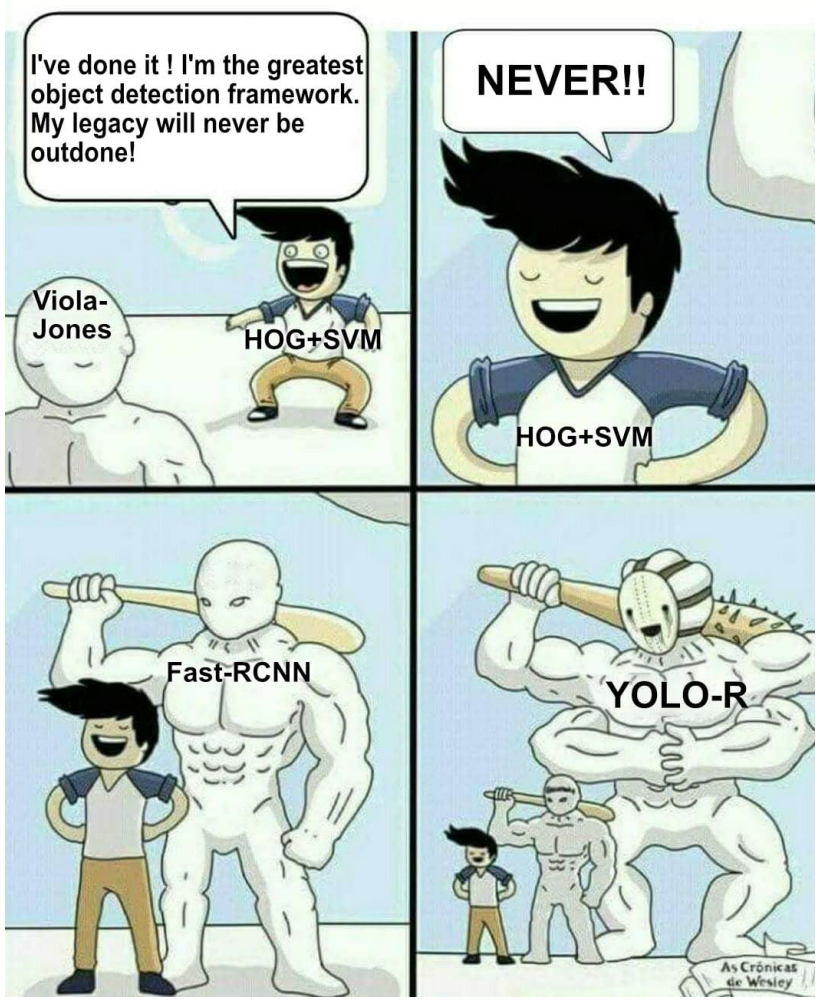

sigmoid
$\sigma(z) = \frac{1}{1+e^{-z}}$
1.0
0.8
0.6
0.4
0.2
0.0
−10
−5
0
5
10


ReLU
$R(z) = max(0,\ z)$
10
8
6
4
2
0
−10
−5
0
5
10

IF WE LOOK CLOSELY AT THE "CONVOLUTION LAYER"

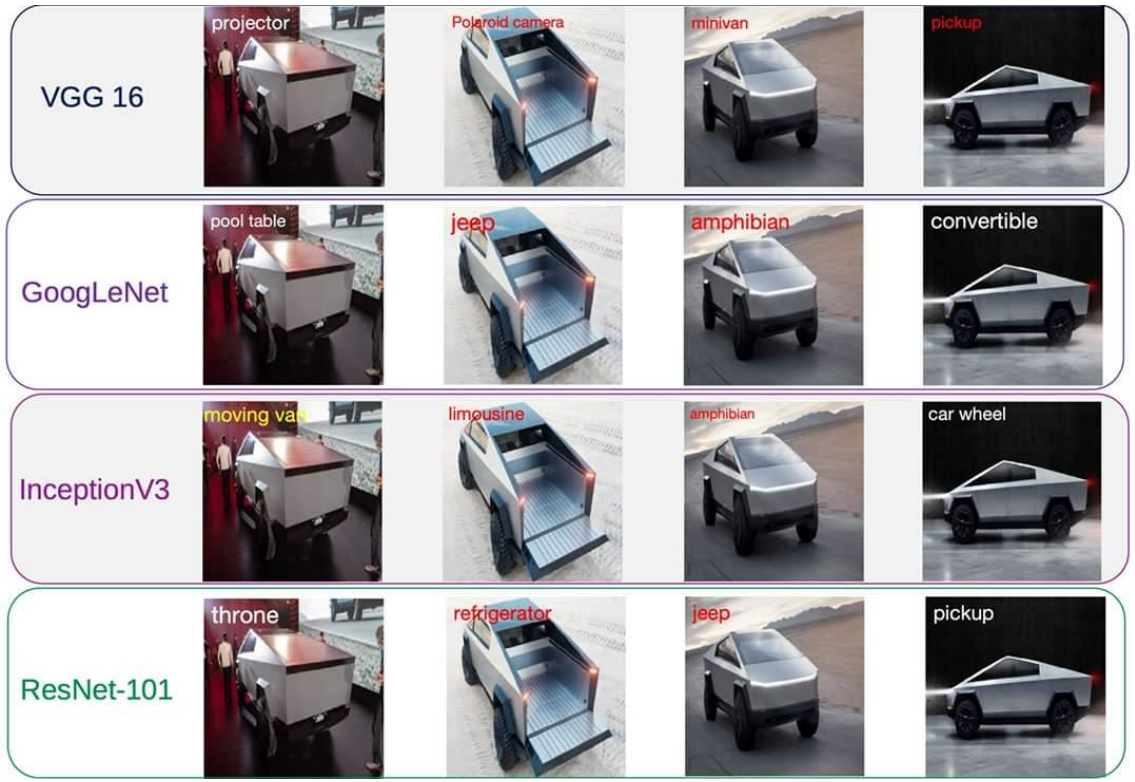
DataLab Community

$$C_p^1(i,j) = \sigma\left(\sum_{u=-2}^{2}\sum_{v=-2}^{2} I(i-u, j-v)k_{1,p}^1(u,v) + b_p^1\right)$$


I've done it ! I'm the greatest object detection framework. My legacy will never be outdone!
Viola-Jones
HOG+SVM
NEVER!!
HOG+SVM
Fast-RCNN
YOLO-R
As Crônicas de Wesley

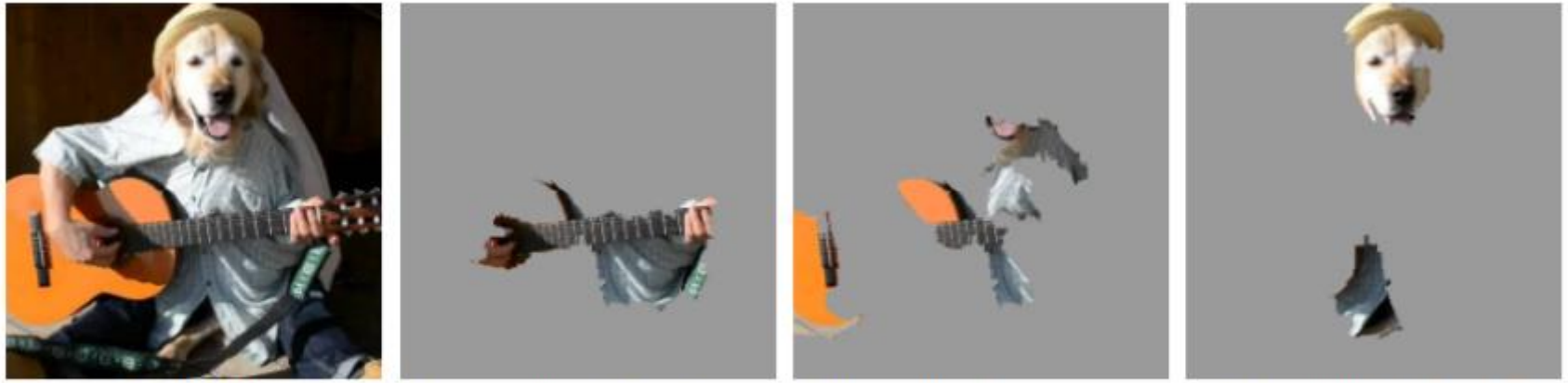
COMPUTER VISION

WHY DO YOU NEED SO MANY IMAGES FOR TRAINING?


VGG 16
projector
Polaroid camera
minivan
pickup
GoogLeNet
pool table
jeep
amphibian
convertible
InceptionV3
moving van
limousine
amphibian
car wheel
ResNet-101
throne
refrigerator
jeep
pickup

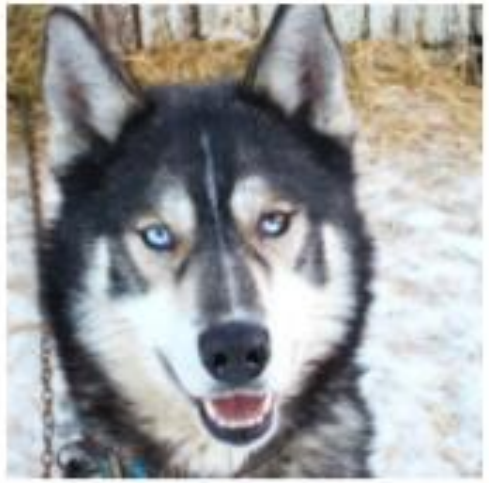
DataLab Community

How to confuse machine learning



(a) Original Image (b) Explaining *Electric guitar* (c) Explaining *Acoustic guitar* (d) Explaining *Labrador*

Figure 4: Explaining an image classification prediction made by Google's Inception neural network. The top 3 classes predicted are "Electric Guitar" ($p = 0.32$), "Acoustic guitar" ($p = 0.24$) and "Labrador" ($p = 0.21$)



(a) Husky classified as wolf

(b) Explanation

Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.

> “Those who can imagine anything, can create the impossible.”
>
> *-Alan Turing*

Questions?

Rubén Alvarez - @bio_ruben

@datalabmx