



14. Redes Neuronales Convolucionales Bootcamp Visión Artificial para los ODS



M.Sc. Edwin Salcedo



8 de Junio del 2022



Zoom



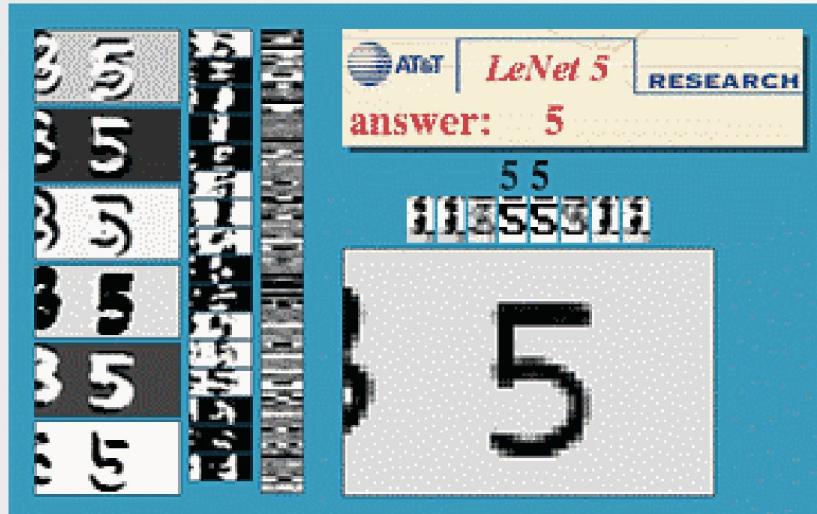
UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ



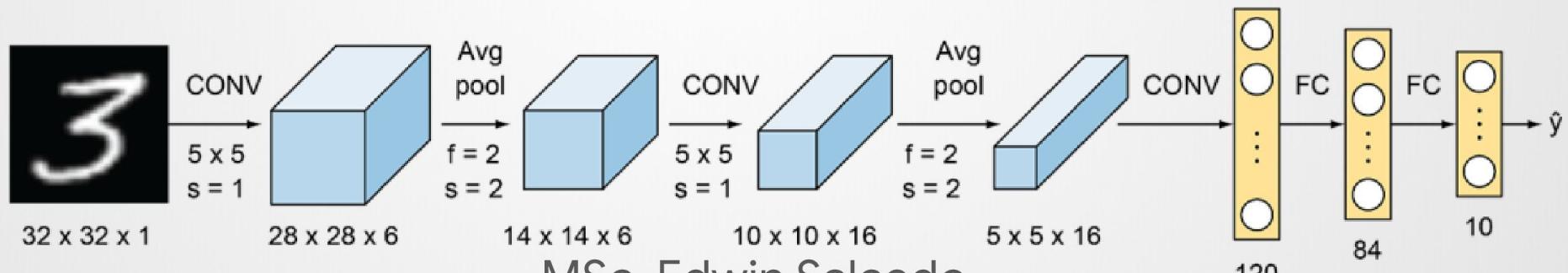
Redes Neuronales Convolucionales (CNNs)



UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ



Yann LeCun, Leon Bottou,
Yoshua Bengio and Patrick Haffner (1990)

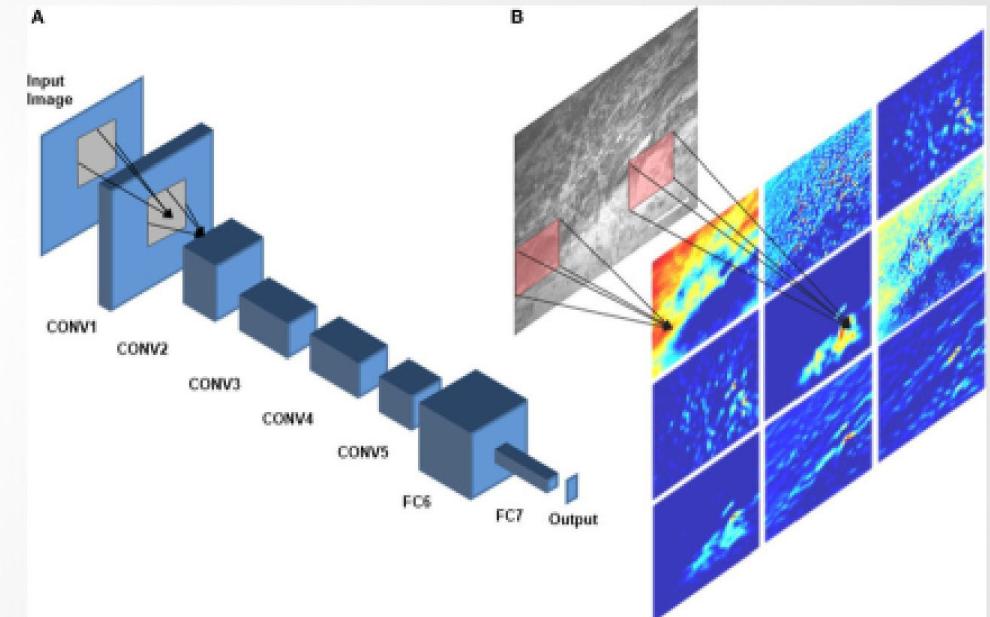
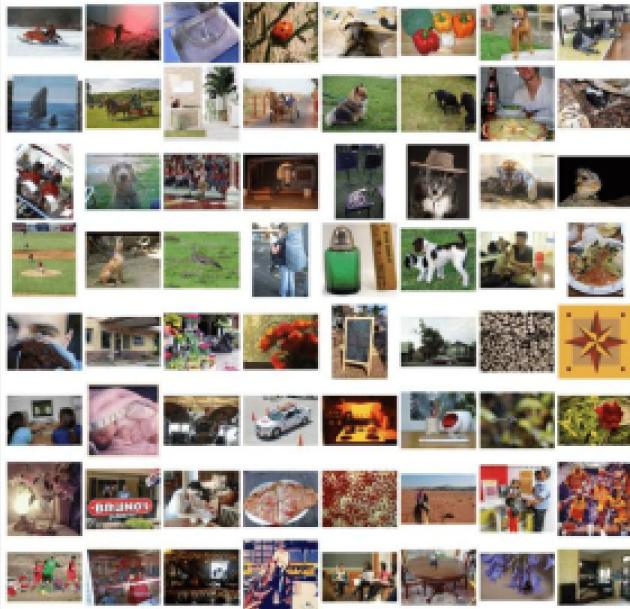


MSc. Edwin Salcedo
LeNet 5

Redes Neuronales Convolucionales (CNNs)



Conjuntos de datos anotados más grandes y computadoras más capaces dieron origen a las CNN.



Conjunto de datos de ImageNet,
iniciado en 2009, sigue creciendo

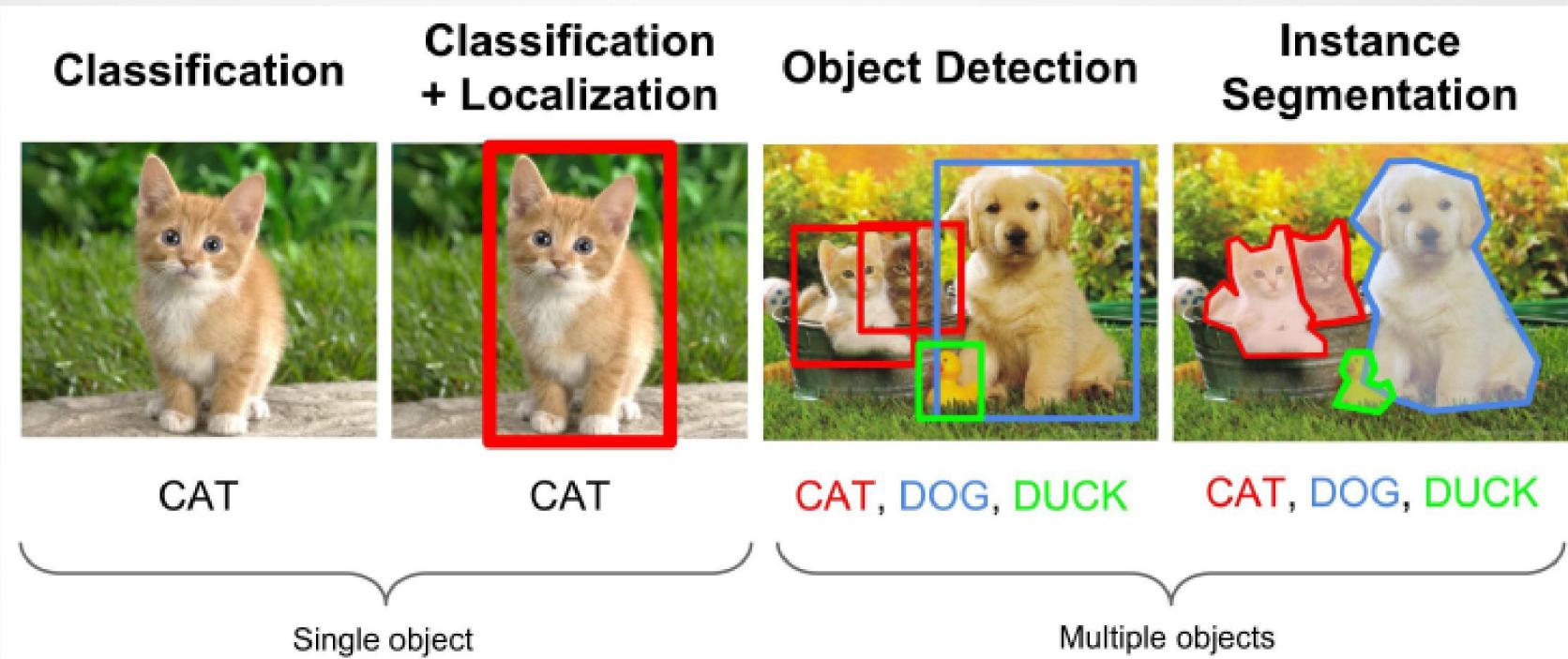
MSC. Edwin Salcedo

CNN AlexNet publicado desde la
U de Toronto, 2012

Deep Learning para Visión Artificial



UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ

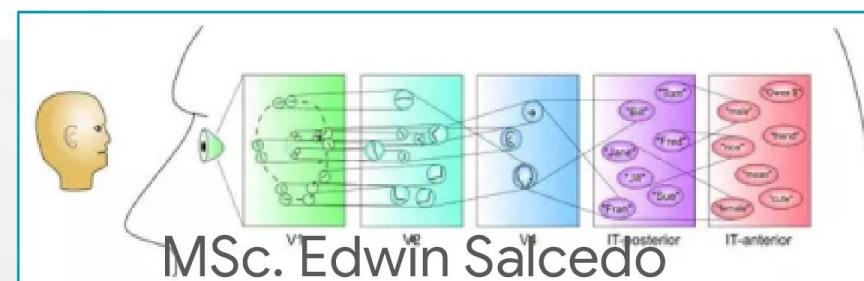
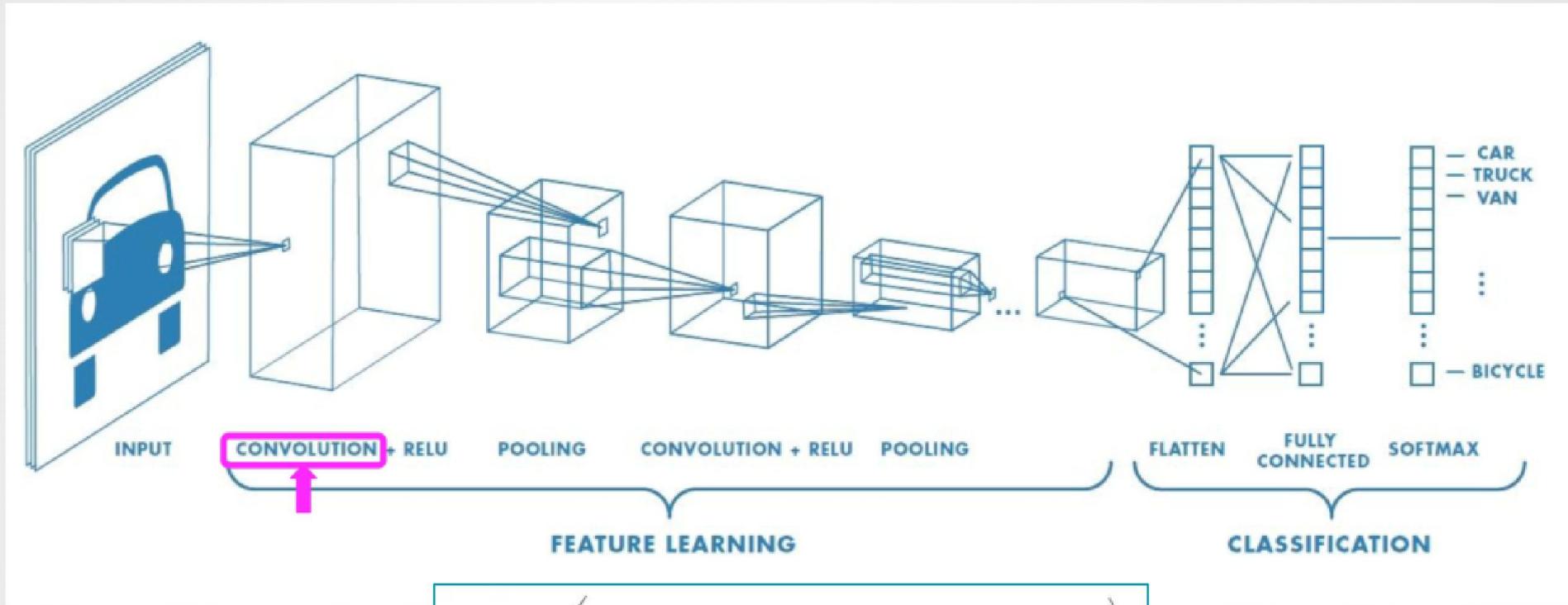


MSc. Edwin Salcedo

Redes Neuronales Convolucionales (CNN)



UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ



Redes Neuronales Convolucionales (CNN)



UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ

Las CNN consisten en cuatro operaciones principales:

1. Convolution
2. Non Linearity (ReLU)
3. Pooling or Sub Sampling
4. Classification (Fully Connected Layer)

Las CNN también consisten en estas operaciones adicionales, pero se usan cuando es necesario:

5. Dropout
6. Flatten
7. Softmax

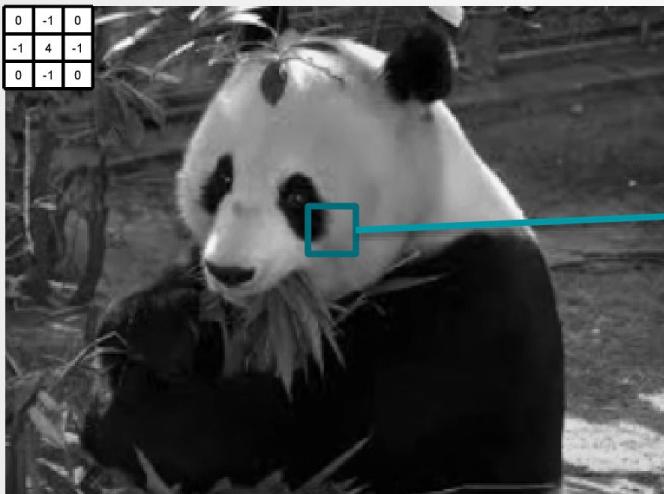
Ejemplo

Input → Conv → ReLU → Conv → ReLU → Pool → ReLU → Conv → ReLU → Pool → Fully Connected Layer

MSc. Edwin Salcedo

Aprendizaje de funciones: convolución

Imagen Original



$$f(x,y)$$

0	5	12	16
2	10	18	20
5	20	45	75
50	80	105	120
100	110	170	225
110	120	140	205
170	225	220	205
225	250	250	230
250	255	250	230



MSc. Edwin Salcedo
Output image

120	140	120
225	220	205
255	250	230

0	-1	0
-1	4	-1
0	-1	0

0	-140	0
-225	880	-205
0	-205	0

= 60
valor de píxel
en la salida

Feature Learning: Convolution



UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ

Convoluciones para CNN:

- Crea un volumen de imágenes. Aumenta la profundidad de un volumen (tensor) según el número de filtros aplicados.
- Las dimensiones (ancho y alto) del volumen debe



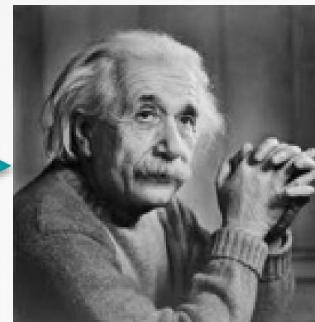
MSc. Edwin Salcedo

Convolución



Identidad

0	0	0
0	1	0
0	0	0



Vertical
Sobel

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



Vertical
Prewitt

-1	0	1
-1	0	1
-1	0	1



Laplacian

0	1	0
1	-4	1
0	1	0



MSc. Edwin Salcedo

Convolución: Stride



Stride controla cómo el filtro gira alrededor de la imagen de entrada.

Input					
4	9	2	5	8	3
		2	4	0	3
2	4	5	4	5	2
5	6	5	4	7	8
5	7	7	9	2	1
5	8	5	3	8	4

*

Filter

1	0	-1
1	0	-1
1	0	-1

Result

2	1

=

Parameters:

Size: $f = 3$

Stride: $s = 2$

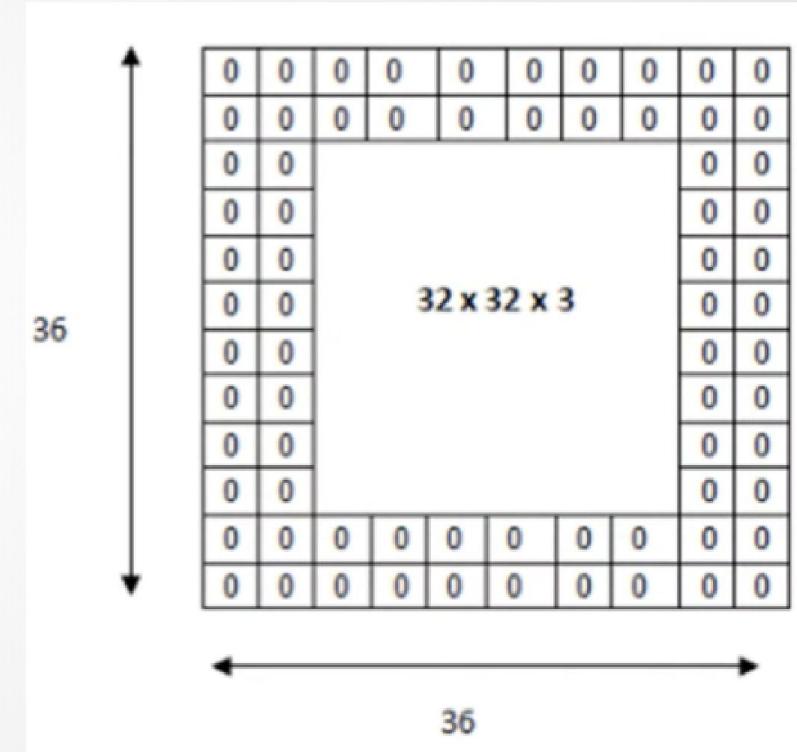
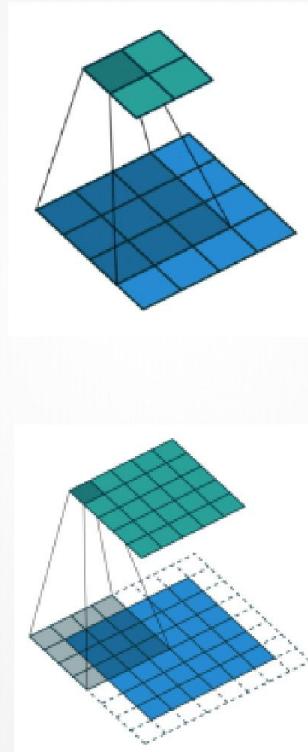
Padding: $p = 0$

$$1 = 2*1 + 5*0 + 3*(-1) + \\ 2*1 + 4*0 + 3*(-1) + \\ 5*1 + 4*0 + 2*(-1)$$

Stride

Convolution: Padding

Padding es una estrategia que incluye columnas adicionales y filas de ceros alrededor de la imagen base para que obtengamos un mapa de características con dimensiones similares después de las circunvoluciones.



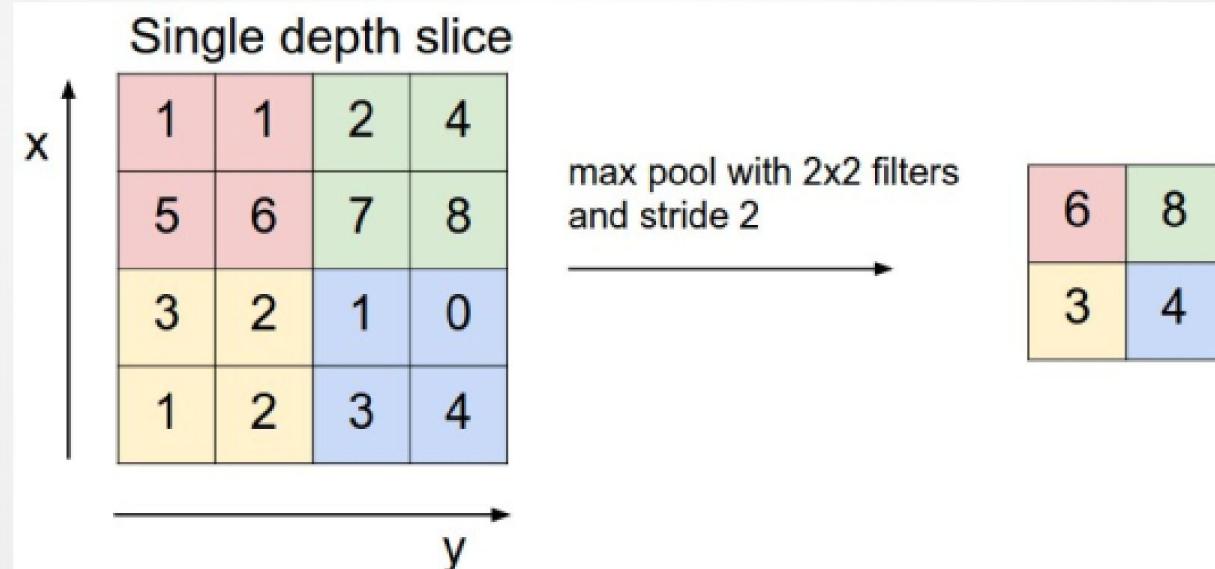
MSc. Edwin Salcedo



Pooling



- Reduce el tamaño espacial de la representación para reducir la cantidad de parámetros y cómputo en la red.
- Existen diferentes tipos de agrupación (agrupación promedio o agrupación L2), pero ninguna es tan popular como la agrupación máxima.

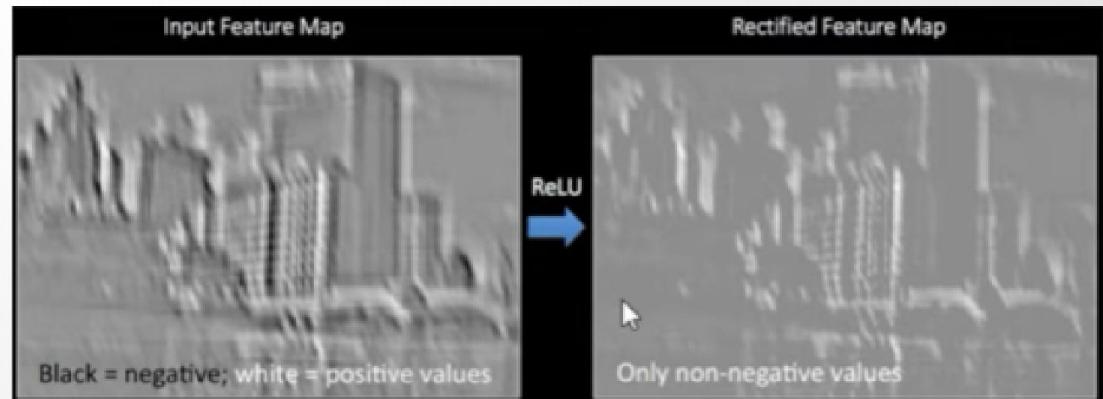
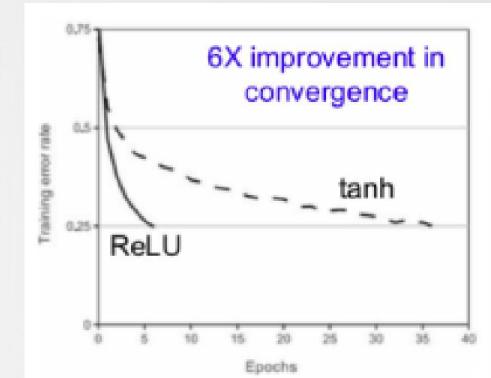
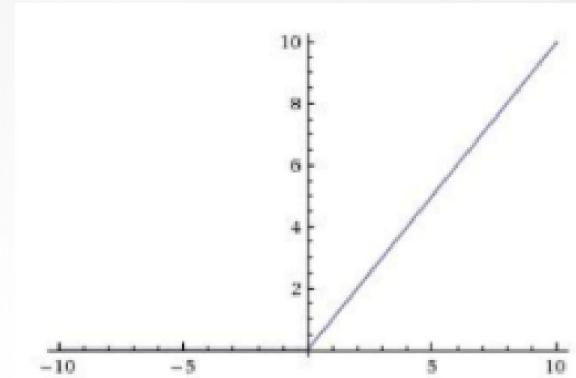


MSc. Edwin Salcedo

ReLU

ReLU, para abreviar, es una función lineal por partes que generará la entrada directamente si es positiva; de lo contrario, generará cero.

ReLU acelera enormemente la convergencia de SGD.



MSc. Edwin Salcedo

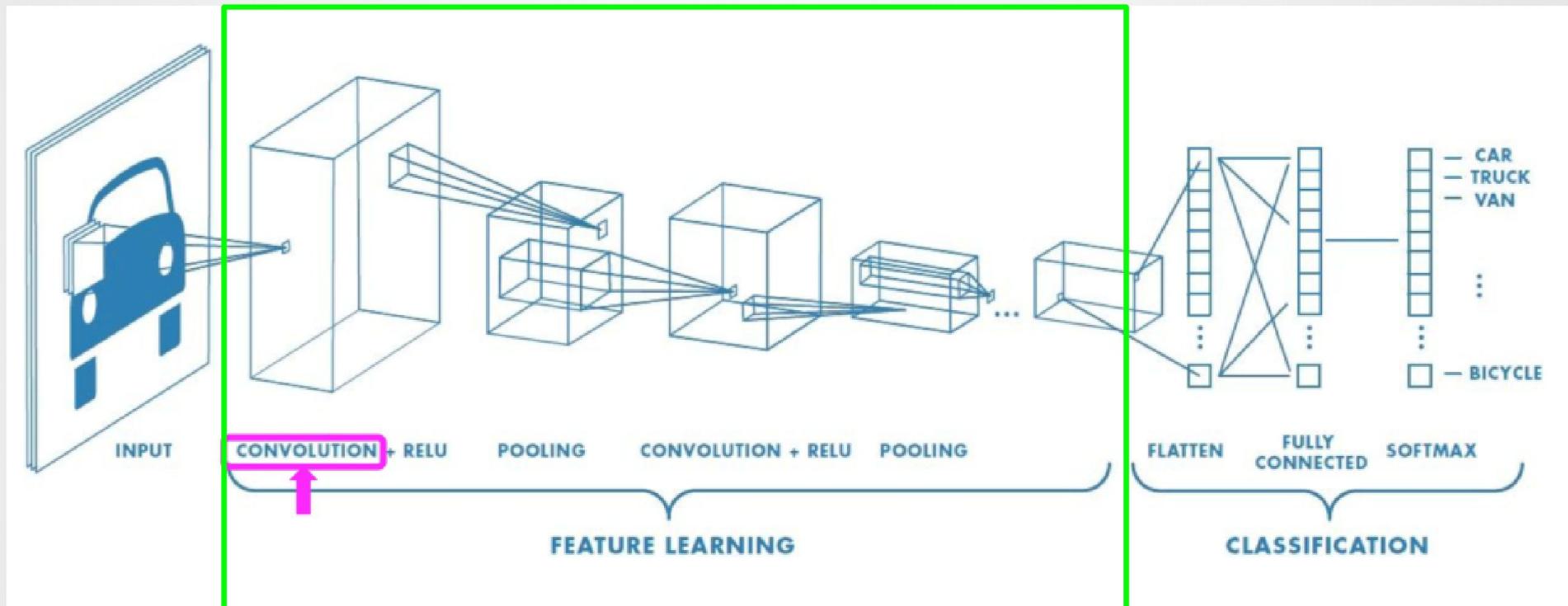


UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ

Feature Learning: Resumen



UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ



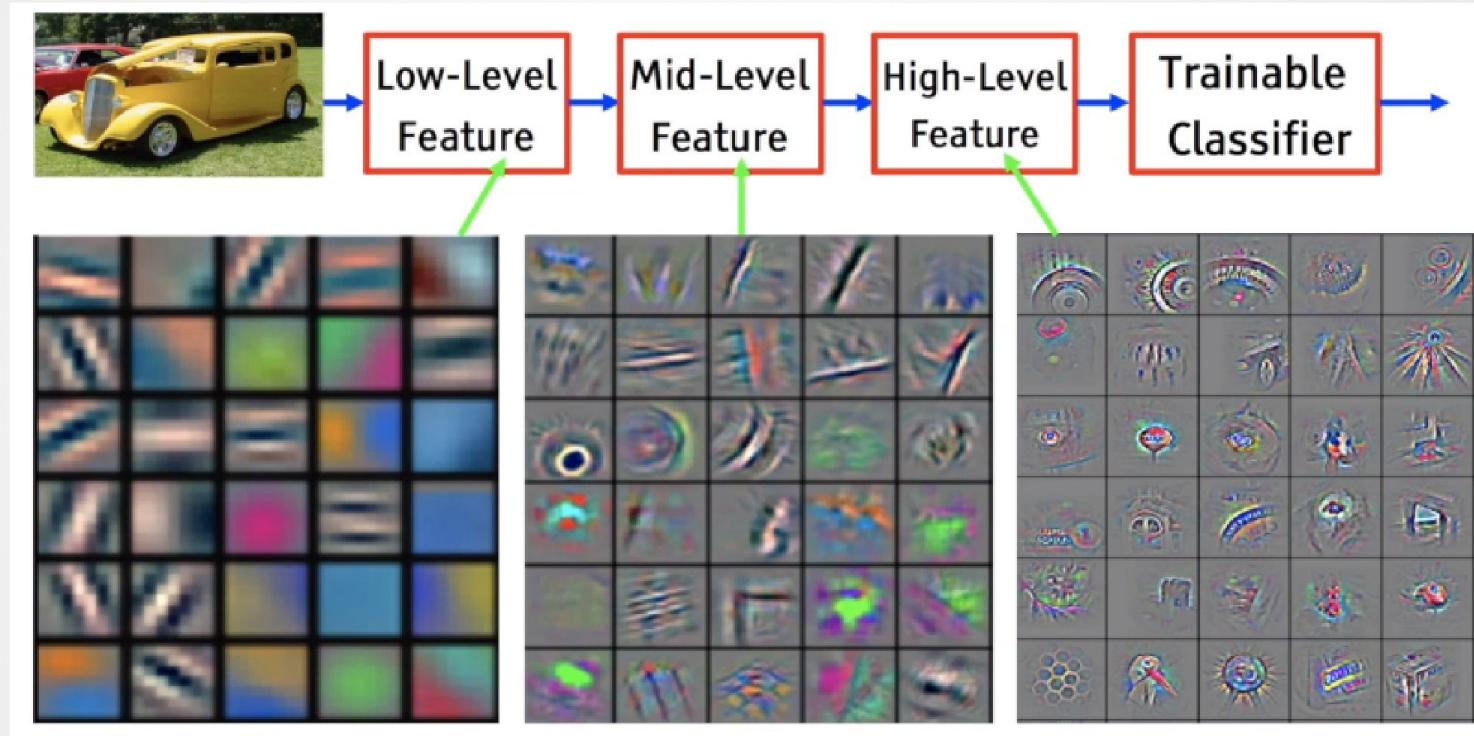
Feature learning or extraction

MSc. Edwin Salcedo

Feature Learning: Resumen



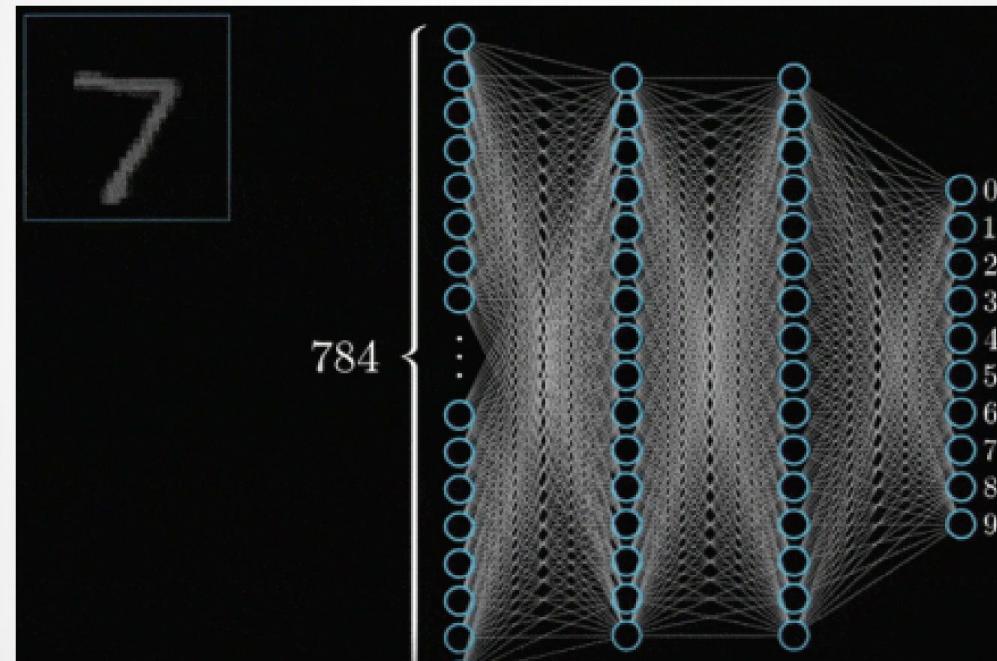
UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ



MSc. Edwin Salcedo

Clasificación

El objetivo de la capa de clasificación es obtener un vector porcentual con la ponderación correspondiente de una imagen de entrada y un conjunto de clases. Contiene una red neuronal totalmente conectada (FCN).



MSc. Edwin Salcedo

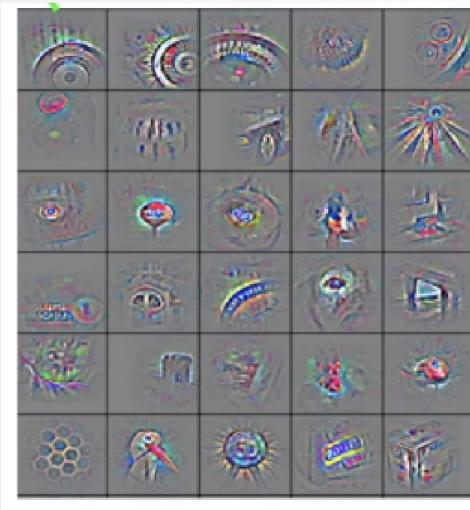
Flatten



UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ

Antes del FCN, requerimos convertir un tensor de mapas de características ($M \times N \times D$) en un vector ($MND \times 1$).

Feature maps



7	5
5	8



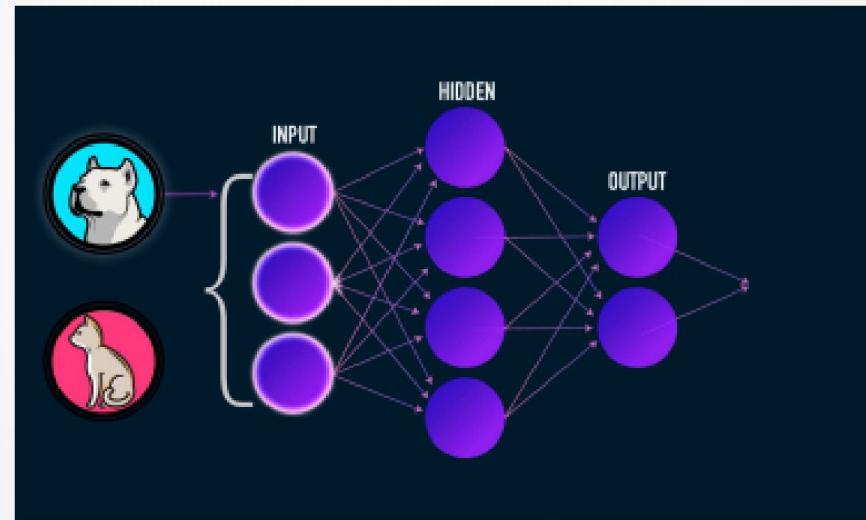
MSc. Edwin Salcedo

Fully Connected Layer



Este es un perceptrón multicapa y su arquitectura dependerá de la complejidad de la tarea de clasificación.

Input:
Flattened feature maps



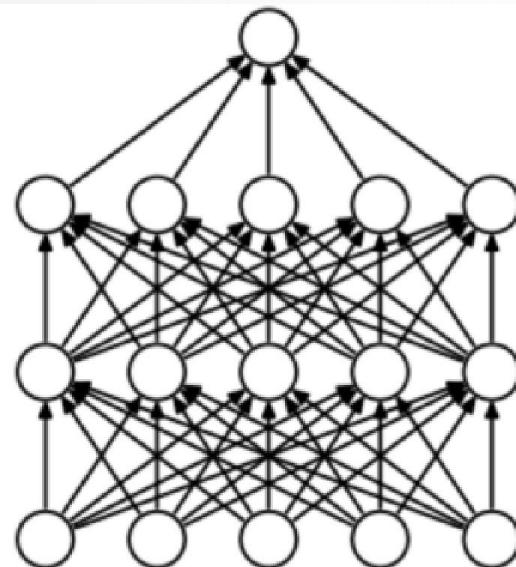
Output:
Array of percentages result
of a **Softmax activation function**
which can be processed with
OneHot encoder to obtain
a final class.

E.g. dog(0.45) and cat(0.55)

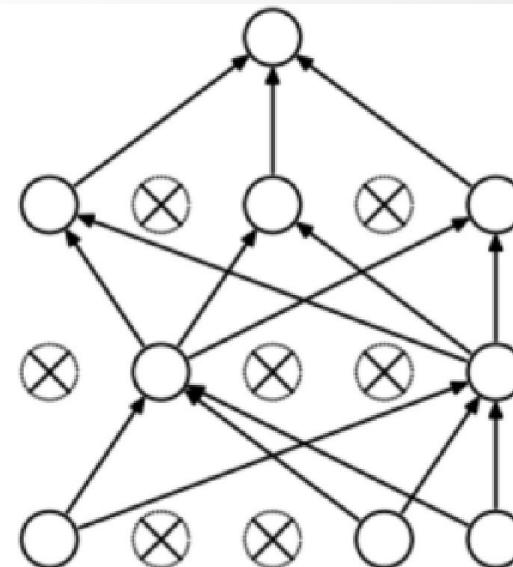
Dropout



Dropout ayuda a apagar algunos nodos al azar en una capa para que podamos evitar el sobreajuste.



(a) Standard Neural Net



(b) After applying dropout.

MSc. Edwin Salcedo

CNNs: Complete Example



UNIVERSIDAD
CATÓLICA
BOLIVIANA
LA PAZ

<https://poloclub.github.io/cnn-explainer/>

<https://www.cs.ryerson.ca/~aharley/vis/conv/>

The screenshot shows a web-based application for visualizing CNN layers. On the left, there is a drawing canvas with the placeholder text "Draw your number here". Below the canvas are three icons: a red X, a pencil, and an eraser. To the right of the drawing area are three input fields labeled "Downsampled drawing:", "First guess:", and "Second guess:", each with a small square checkbox next to it. At the bottom of the left sidebar, there is a section titled "Layer visibility" with a list of layers and "Hide" buttons:

Layer	Action
Input layer	Hide
Convolution layer 1	Hide
Downsampling layer 1	Hide
Convolution layer 2	Hide
Downsampling layer 2	Hide
Fully-connected layer 1	Hide
Fully-connected layer 2	Hide
Output layer	Hide

In the center of the screen, there is a large, dark blue rectangular area representing the input image. In the bottom right corner of this area, there is some very faint, illegible text or a watermark.

MSc. Edwin Salcedo

CNNs: LeNet 5 and VGG

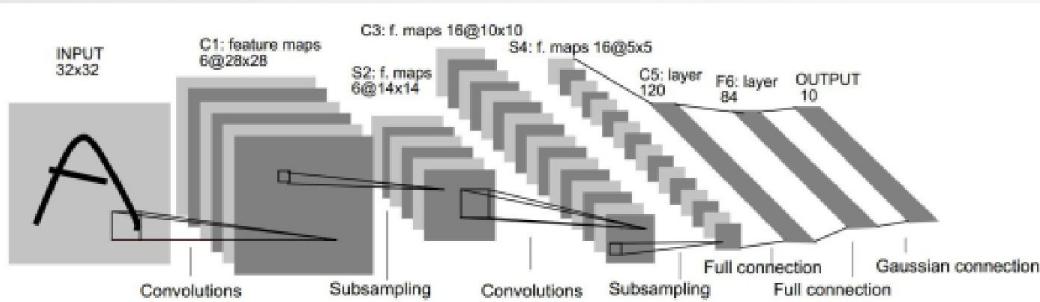
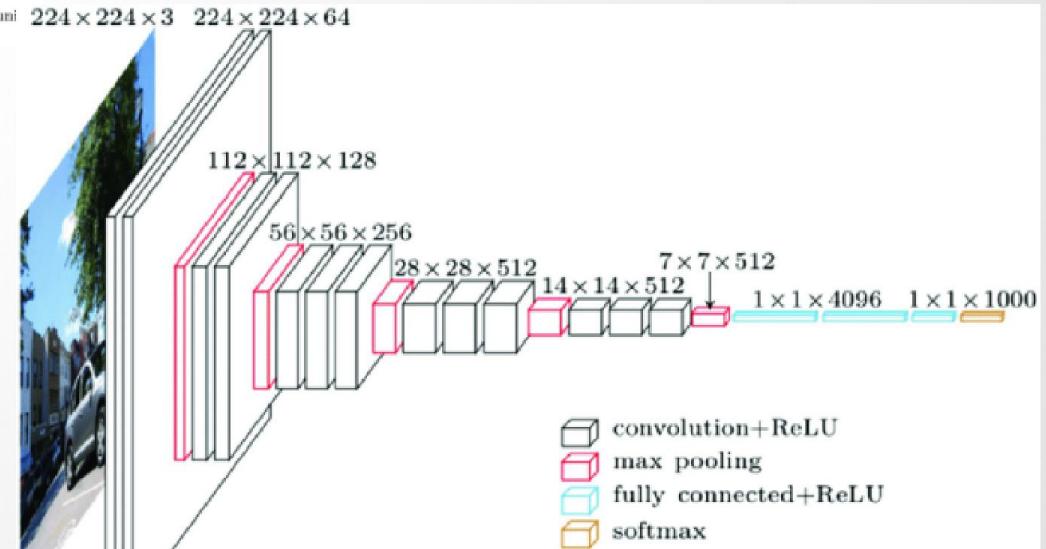


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

LeNet 5

VGG



MSc. Edwin Salcedo