

Introducción a las Redes Neuronales Convolucionales

Aprendizaje de Máquina Aplicado

Yomin Jaramillo Munera
yejaramilm@eafit.edu.co

2025

Juan David Martínez Vargas
jdmartinev@eafit.edu.co

Agenda

- Convolución de imágenes
- ANN densas y visión por computadora
- Redes neuronales convolucionales (CNN)
- Aumento de datos
- Ejemplo práctico



Convolución de Imágenes

Imágenes como Datos



Blue channel

Green channel

Red channel

		171	200	19	6	...	26
		24	56	230	1	...	8
1	120	67	89	107	...	13	89
2	12	216	145	26	...	181	18
3	0	16	4	45	...	44	81
4	0	78	90	167	...	25	71
...	56
...
...	7
...	12
64	12	67	82	141	...	12	
	1	2	3	4	...	64	

Image array: $[64 \times 64 \times 3]$

Imágenes como Datos

Una imagen en
escala de grises
tiene un único
canal



Una imagen...

		0.6	0.6			
	0.6				0.6	
	0.6	0.6	0.6	0.6		
	0.6				0.6	

...es una matriz de pixeles.

El valor de los pixeles va de 0 a 255 pero se normaliza para la red neuronal de 0 a 1

Convolución de Imágenes

7	2	3	3	8
4	5	3	8	4
3	3	2	8	4
2	8	7	2	7
5	4	4	5	4

Imagen

*

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

$$\begin{aligned} &7 \times 1 + 4 \times 1 + 3 \times 1 + \\ &2 \times 0 + 5 \times 0 + 3 \times 0 + \\ &3 \times -1 + 3 \times -1 + 2 \times -1 \\ &= 6 \end{aligned}$$

Filtro o
kernel

=

6		

Resultado de
la convolución

Convolución de Imágenes

Las convoluciones son utilizadas en procesamiento de imágenes clásico para detectar ciertas características de las imágenes como líneas en cierta dirección o bordes en general.

Convolución: Detección de Líneas Verticales



$$* \begin{pmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{pmatrix} =$$



Convolución: Detección de Líneas Horizontales



$$* \begin{pmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{pmatrix} =$$



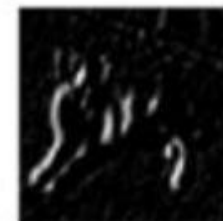
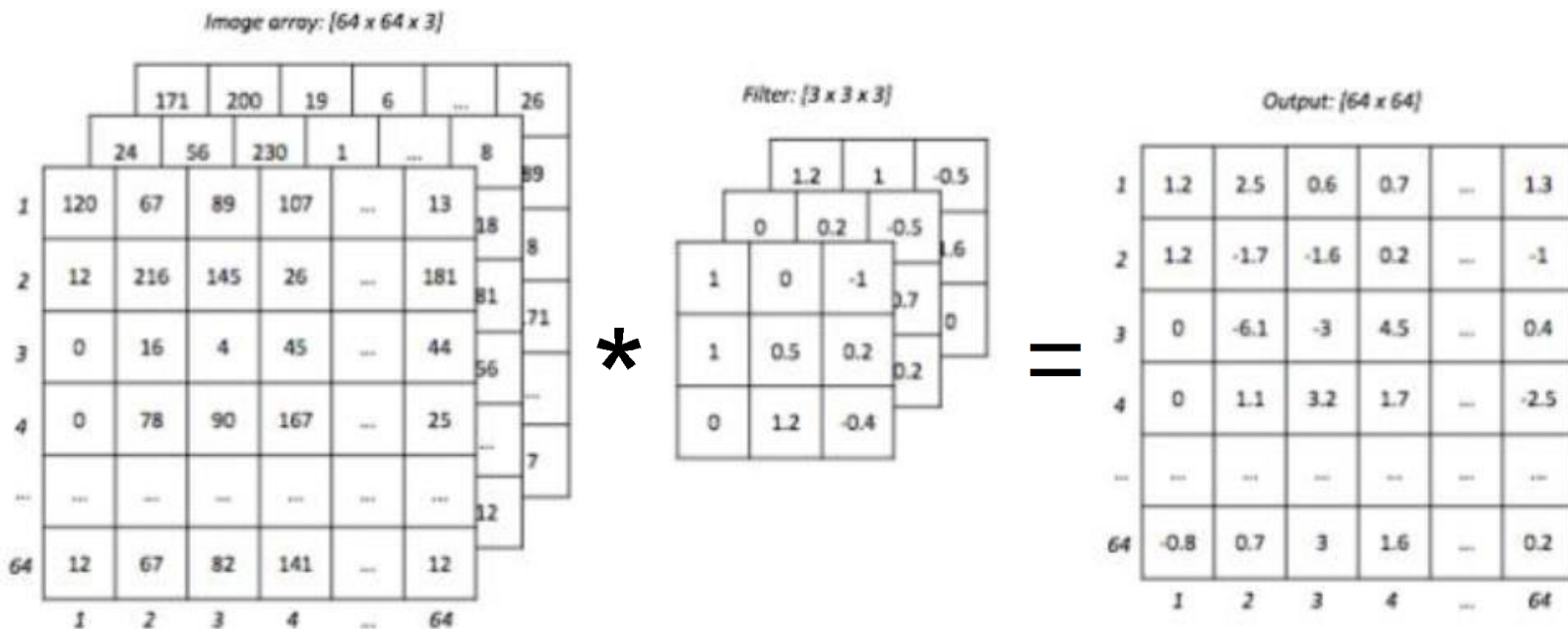
Convolución: Detección de Bordes

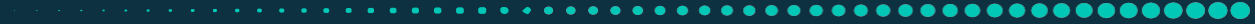


$$* \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix} =$$



Convolución de Imágenes con Múltiples Canales





Redes Neuronales Densas y Visión por Computadora

ANN Densas y Visión por Computadora

En particular queremos responder la siguiente pregunta:

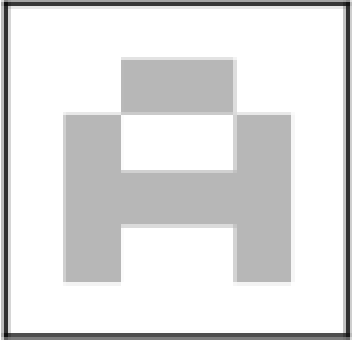
¿Por qué es mala idea utilizar redes neuronales densas (también llamadas completamente conectadas) para procesar imágenes?

ANN Densas y Visión por Computadora

En general, utilizar redes neuronales densas para procesar imágenes es mala idea por dos razones principales:

- Pérdida del **contexto** de los pixeles.
- Exceso de parámetros para obtener un número decente de características intermedias.

Contexto de un Pixel



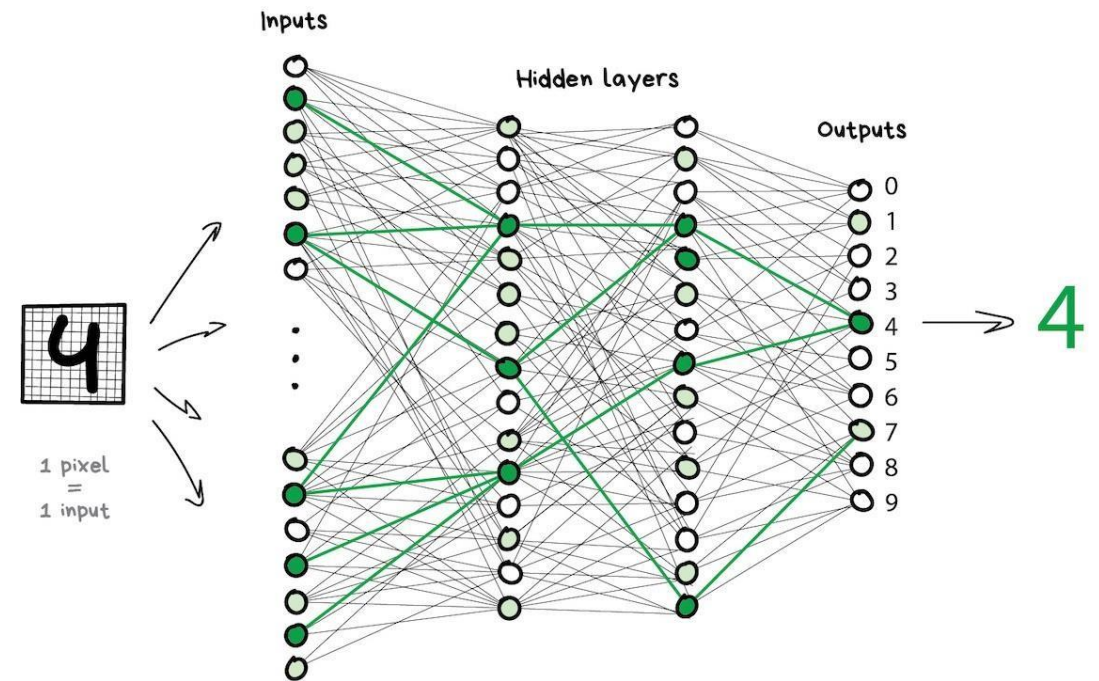
		0.6	0.6		
	0.6			0.6	
	0.6	0.6	0.6	0.6	
	0.6			0.6	

Una red neuronal densa no recibe información sobre si dos pixeles están cerca o lejos, lo que tiene consecuencias indeseables:

- Es difícil identificar las figuras formadas por grupos de pixeles cercanos entre sí.
- Se considera la interacción entre dos pixeles así se encuentren muy lejos.
- La forma en la que se procesan las características de la imagen es dependiente su posición en la misma.

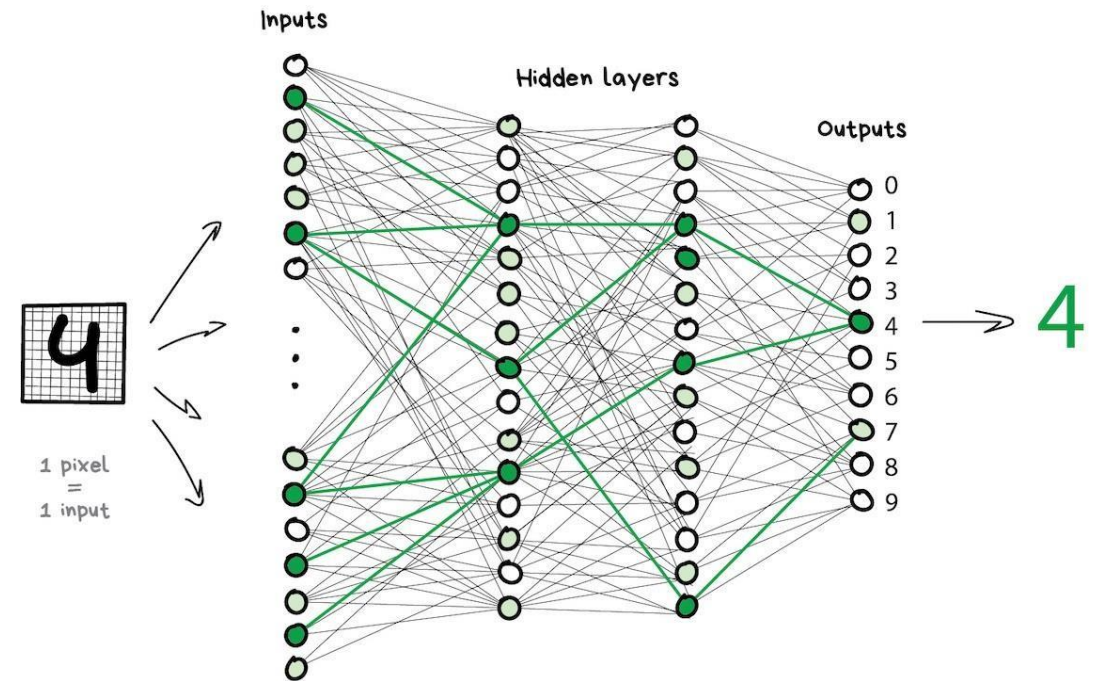
ANN Densas y Visión por Computadora: Exceso de Parámetros

Supongamos que le queremos aplicar una red neuronal densa a una imagen de 256 x 256 píxeles.



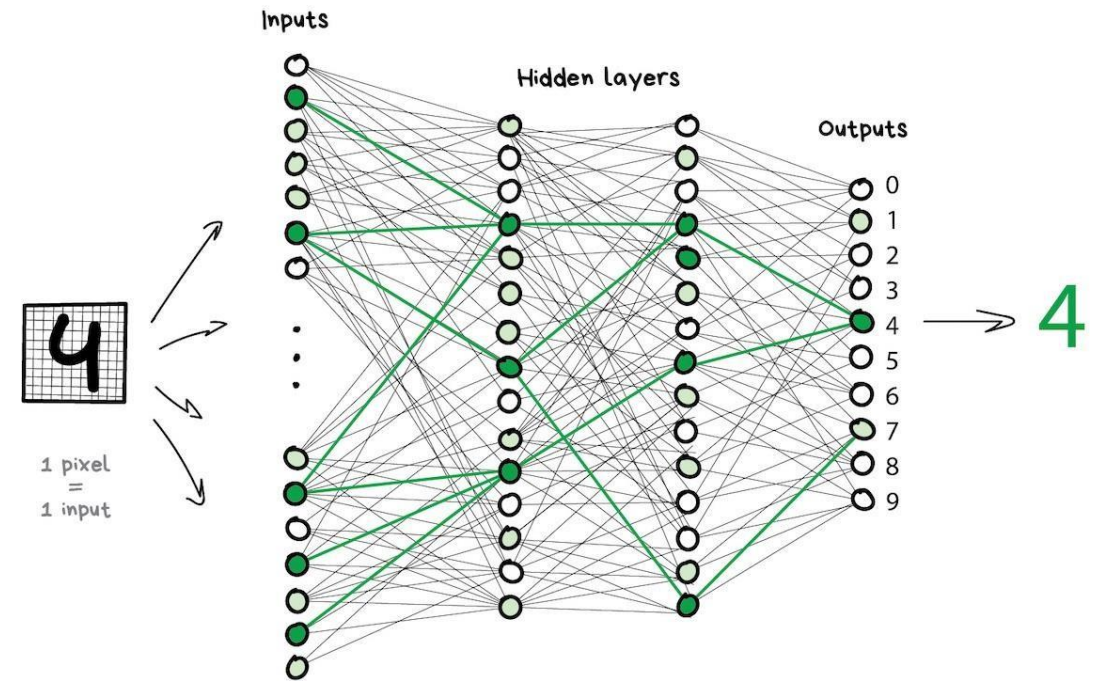
ANN Densas y Visión por Computadora: Exceso de Parámetros

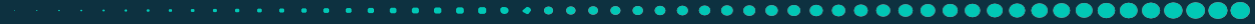
Para no reducir demasiado la dimensionalidad de las características en la primera capa oculta, supongamos que usamos 2048 neuronas en dicha capa.



ANN Densas y Visión por Computadora: Exceso de Parámetros

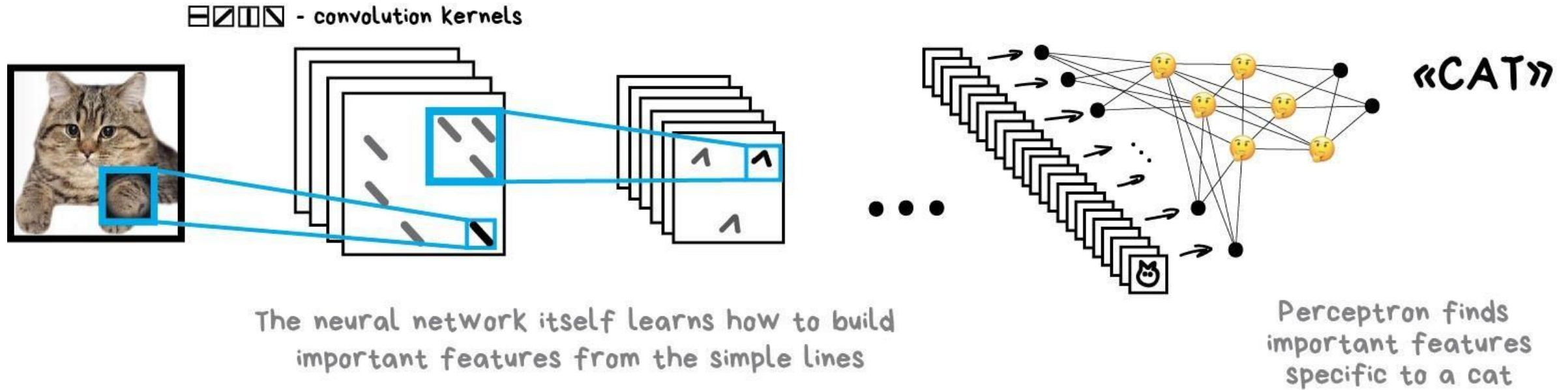
Entonces, el número de pesos de la primera capa sería $256 \times 256 \times 2048$, que es mayor que 134 millones.





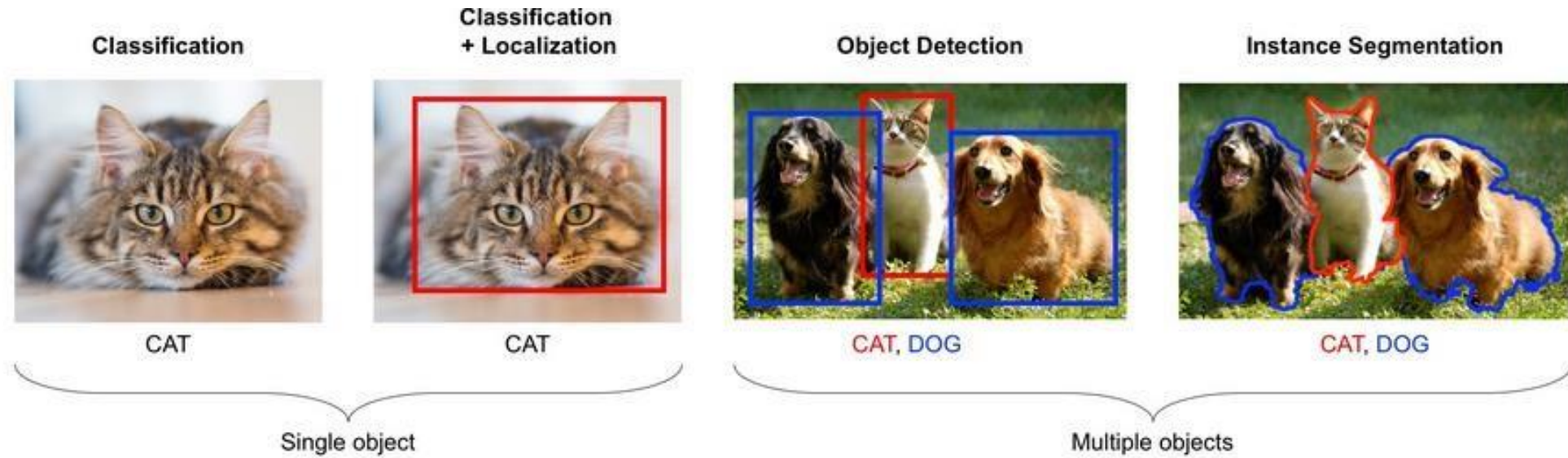
Redes Neuronales Convolucionales (CNN)

Redes Neuronales Convolucionales (CNN)



CONVOLUTIONAL NEURAL NETWORK (CNN)

Tareas en visión por computador



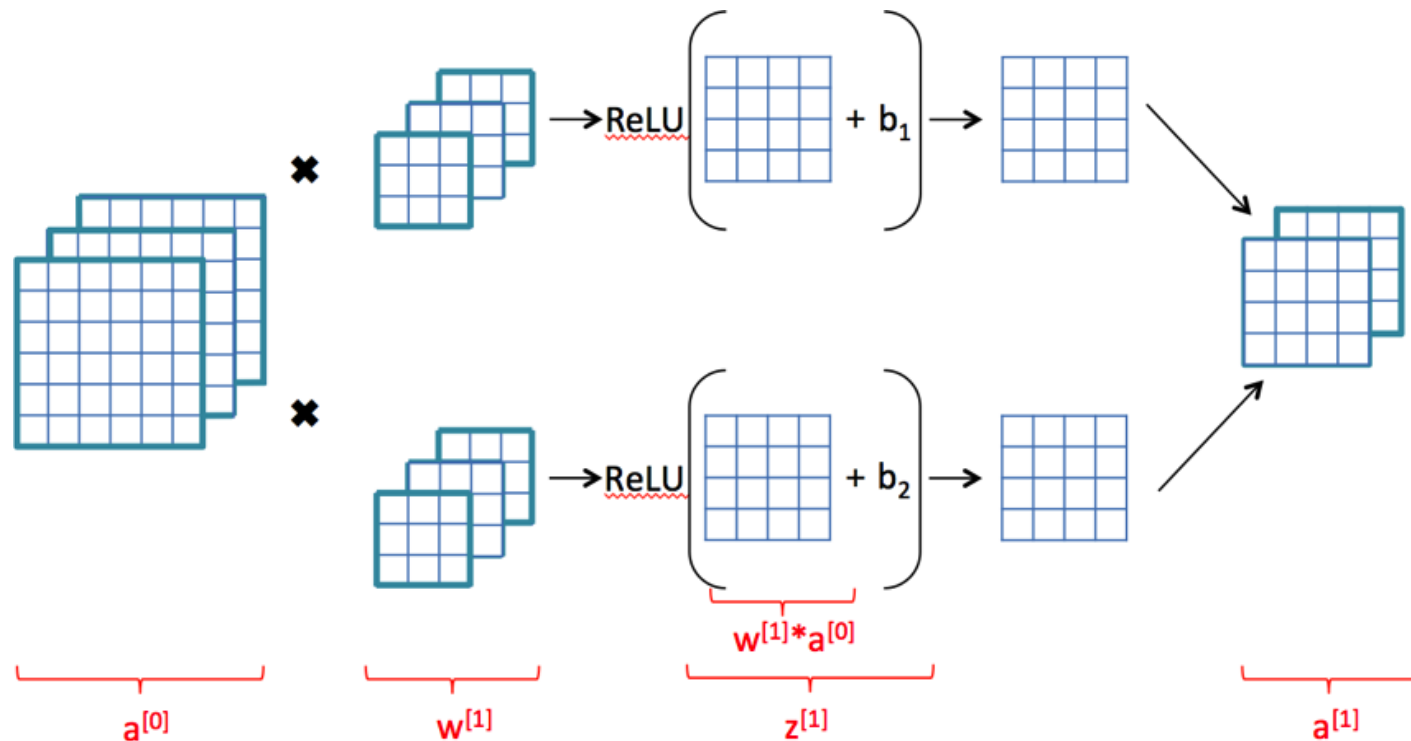
Redes Neuronales Convolucionales (CNN)

Una red neuronal convolucional utiliza principalmente tres tipos de capas:

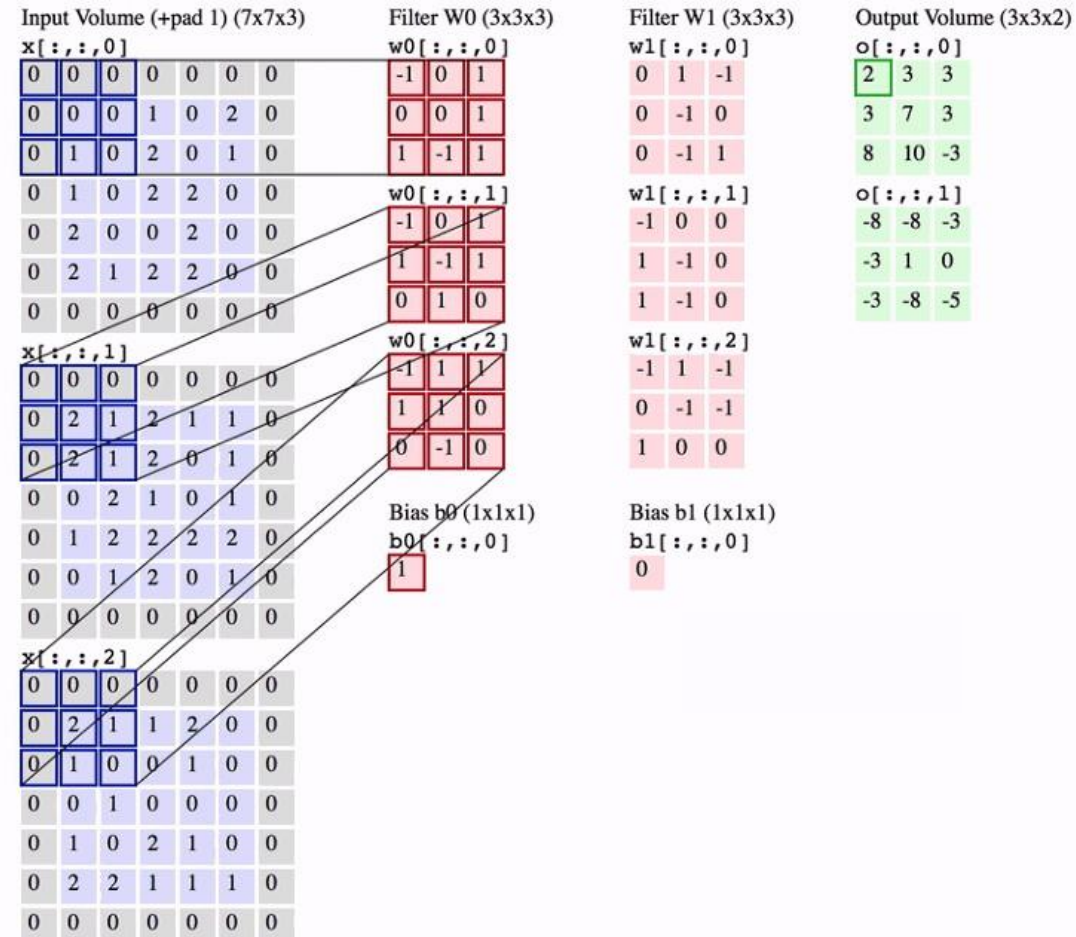
- Capas convolucionales
- Capas de pooling
- Capas densas (sólo al final)

Capa Convolucional

Una capa convolucional aplica múltiples filtros convolucionales a una imagen (o a la salida de la capa anterior), suma un término de “bias” y luego aplica una función de activación no lineal (usualmente ReLU). Los valores que se usan para los filtros y para los términos de “bias” son parámetros entrenables de la red.

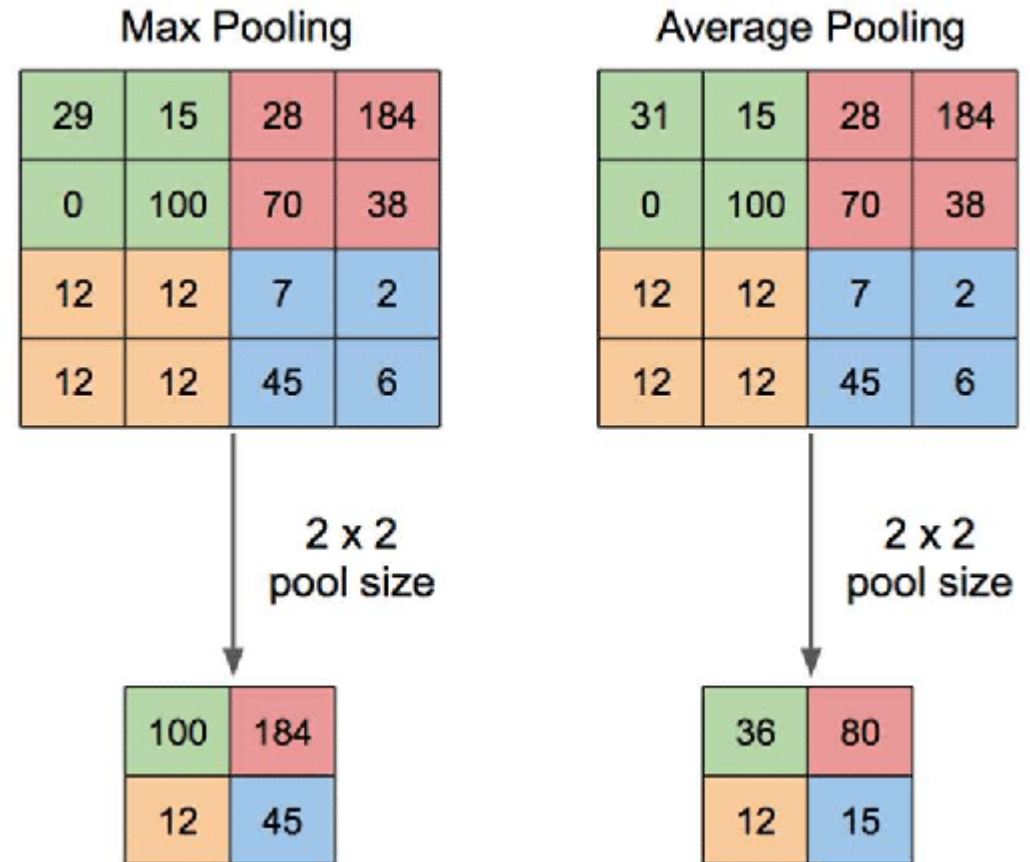


Capa Convolucional

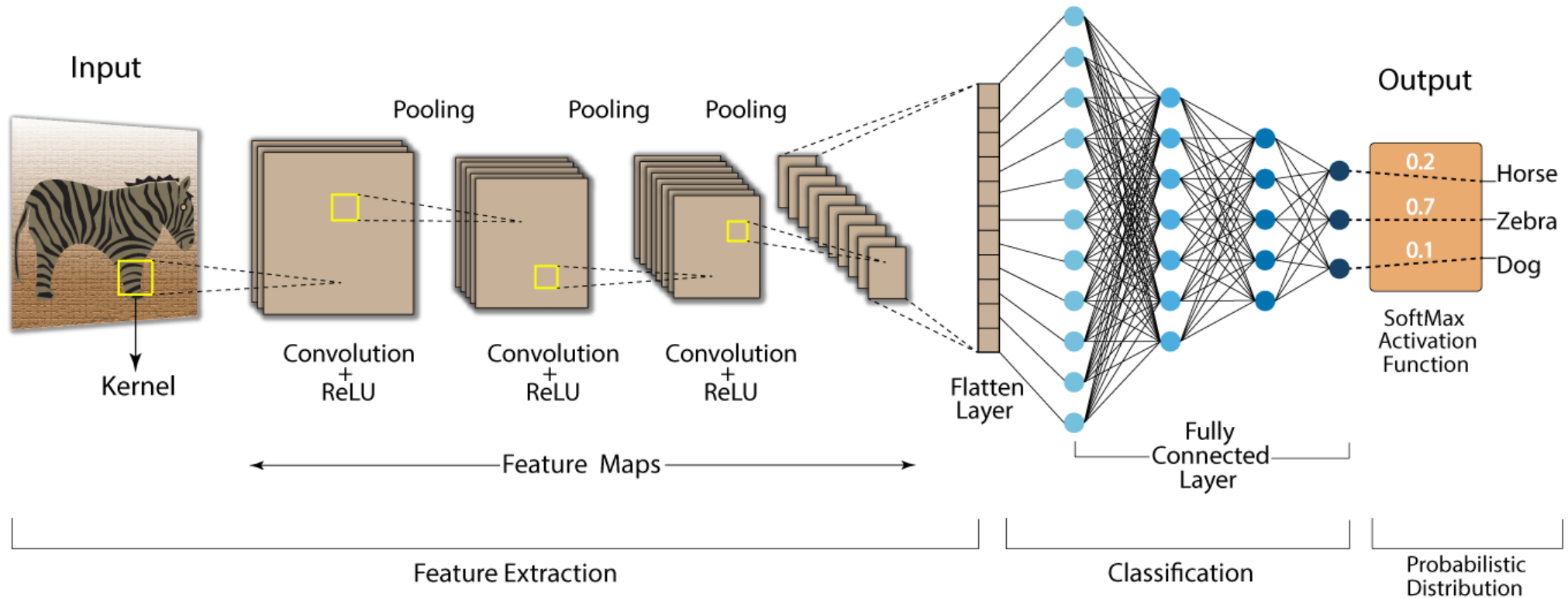


Capa de Pooling

Una capa de pooling hace un reducción de resolución (downsampling) de una imagen o matriz de datos de acuerdo a una ventana de tamaño fijo. Comúnmente la reducción se hace aplicando un promedio (average pooling) o hallando el máximo (max pooling).



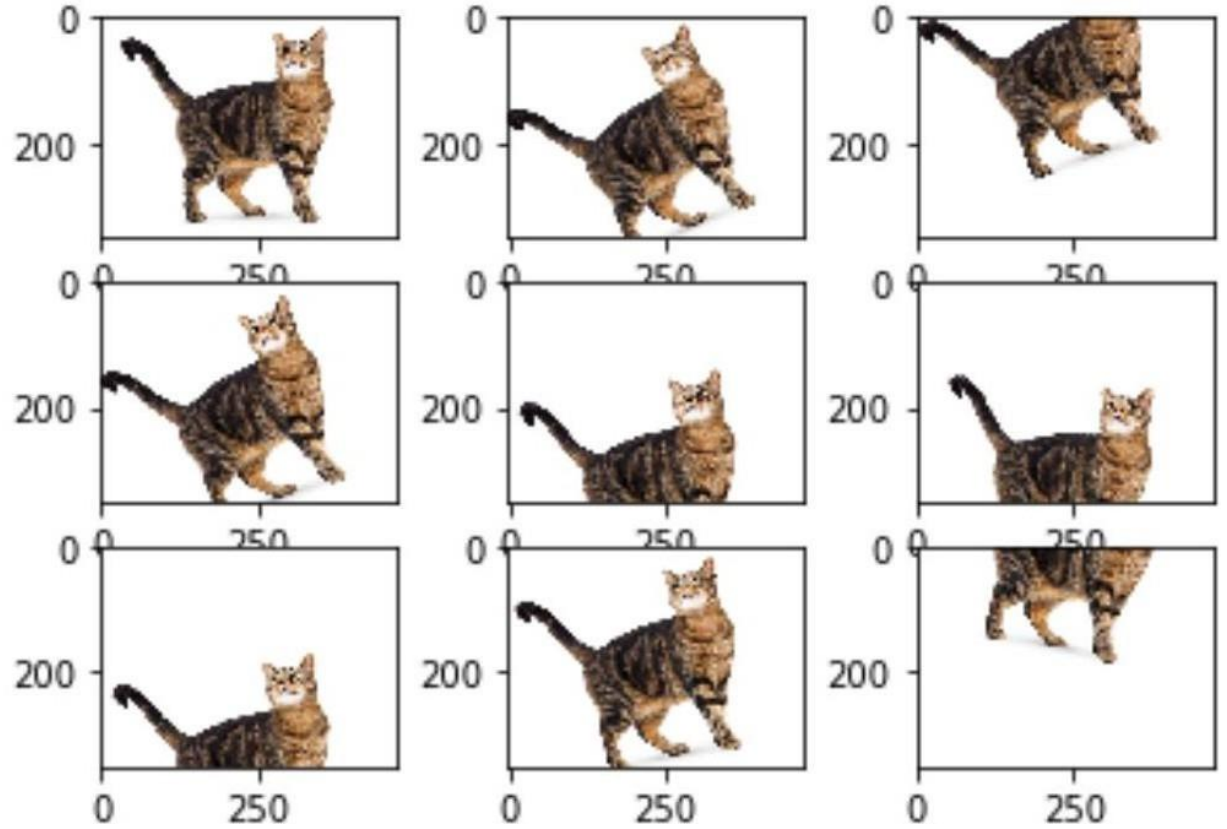
Redes Neuronales Convolucionales (CNN)





Aumento de Datos

Aumento de Datos



Tipos Comunes de Aumento de Datos en Imágenes

Geometry based



rotate



shear



vertical-flip



horizontal-flip



crop



crop-and-pad



Perspective-transform



Elastic-transformation

Color based



sharpen



brighten



Gamma-contrast



invert

Noise / occlusion



gaussian-blur



additive-gaussian-noise



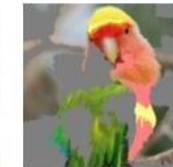
translate-x



translate-y



coarse-salt



super-pixel



emboss

Weather



clouds



fog



snow-flakes



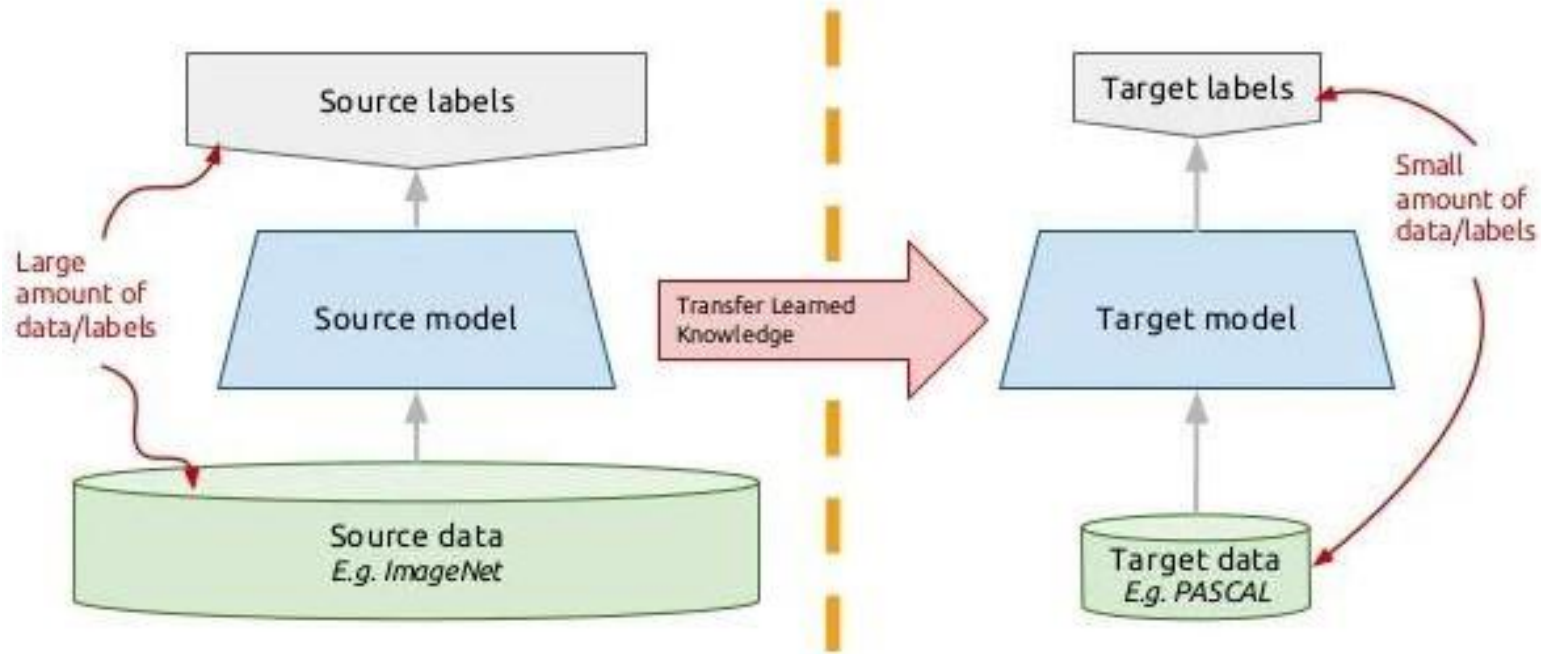
Fast-snowy-landscape



Transferencia de aprendizaje

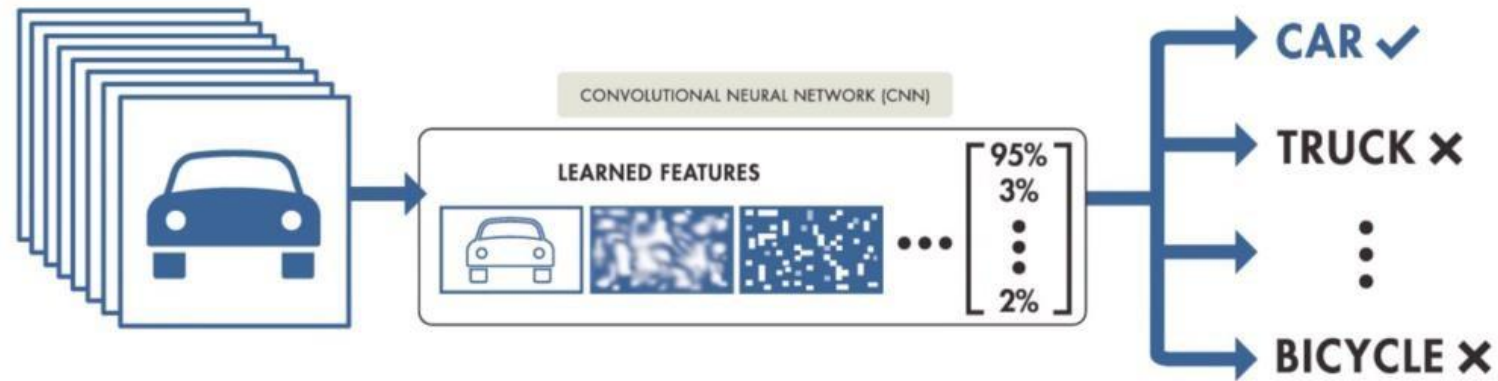
Transferencia de aprendizaje

Transfer learning: idea

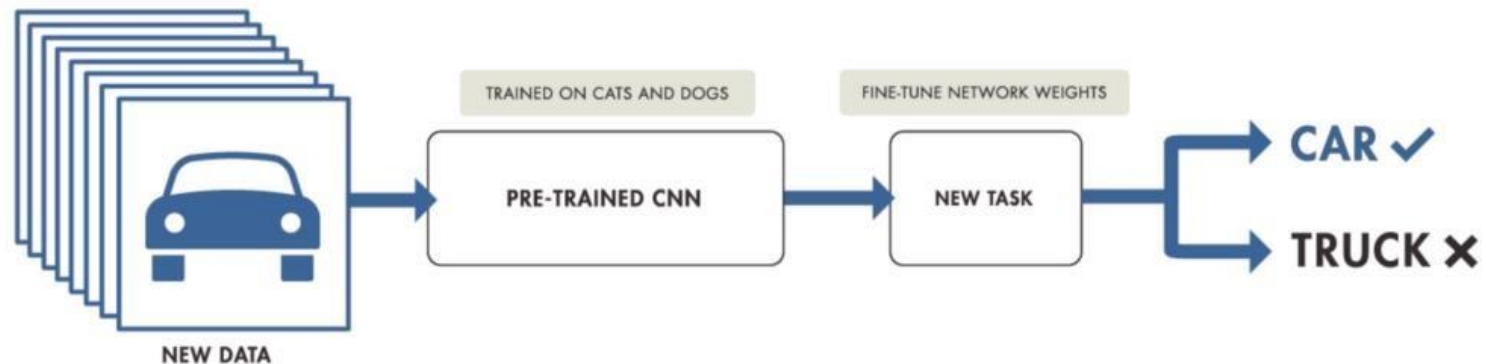


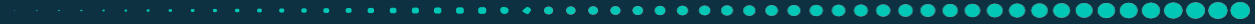
Transferencia de aprendizaje

TRAINING FROM SCRATCH



TRANSFER LEARNING





Ejemplo de Redes Neuronales Convolucionales



¡Muchas Gracias!