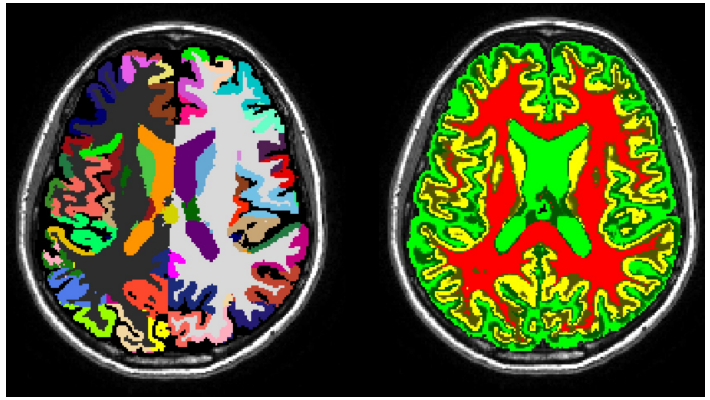
Clase 6

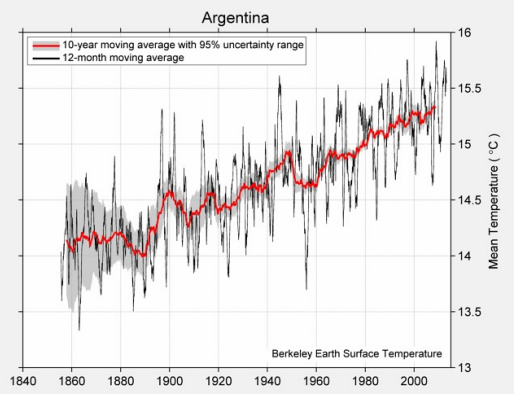
Redes neuronales más allá de la clasificación de imágenes

Enzo Ferrante

eferrante@sinc.unl.edu.ar

@enzoferrante



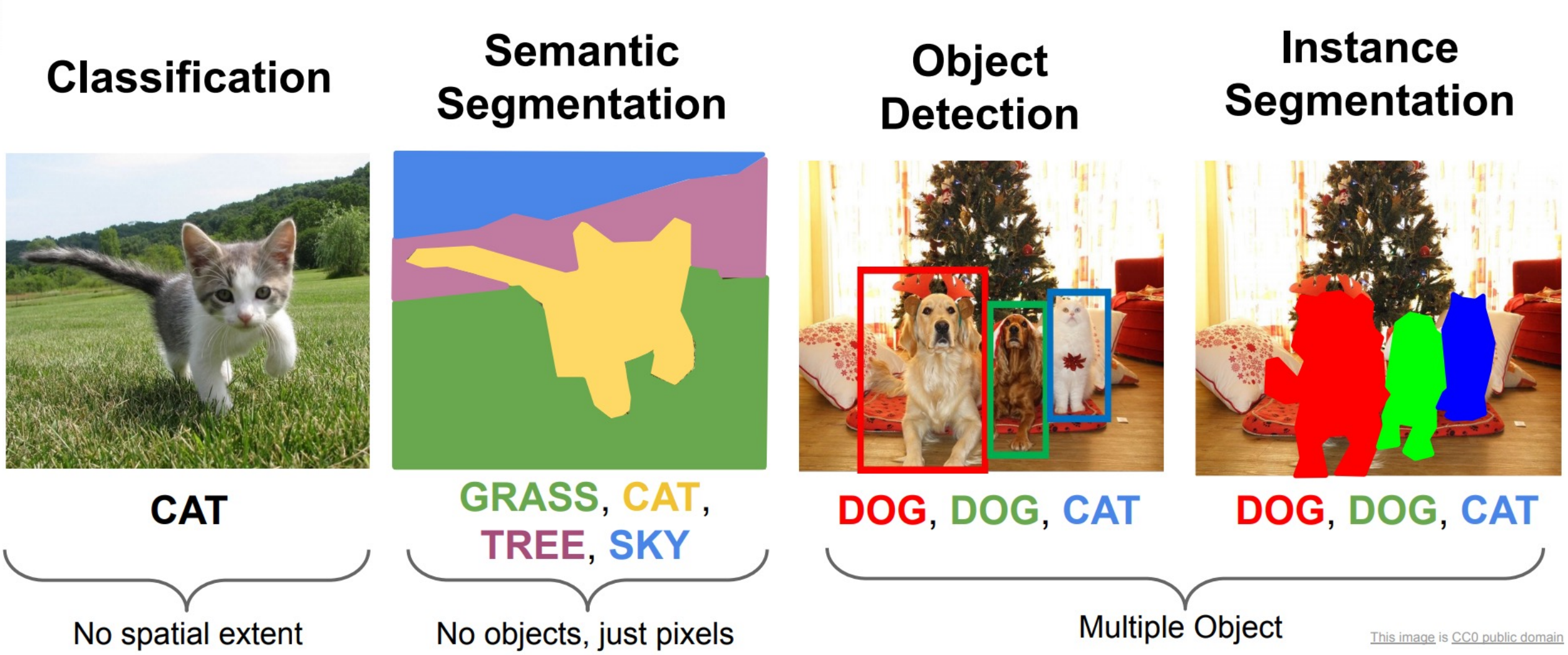


Universidad de SanAndrés

UNL CONICET

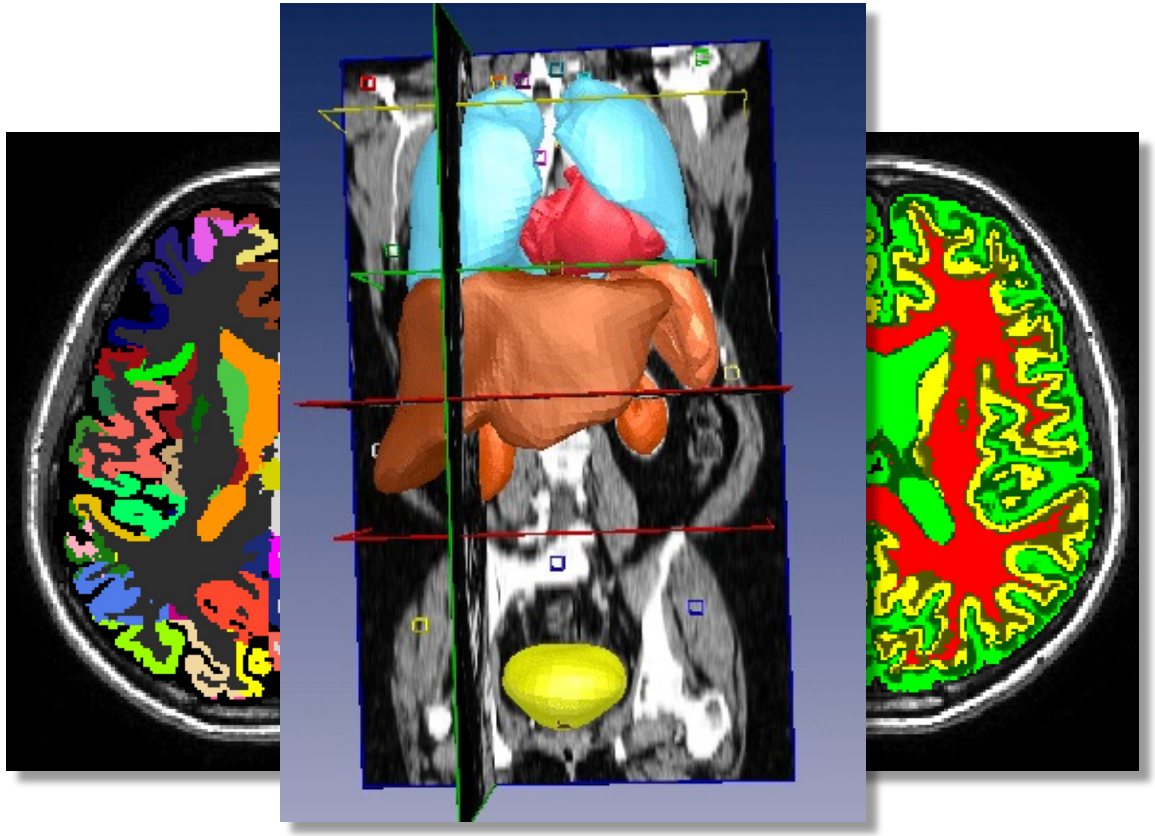
s i n c (i)

Problemas en computer vision



Source: http://cs231n.stanford.edu/slides/2020/lecture_12.pdf

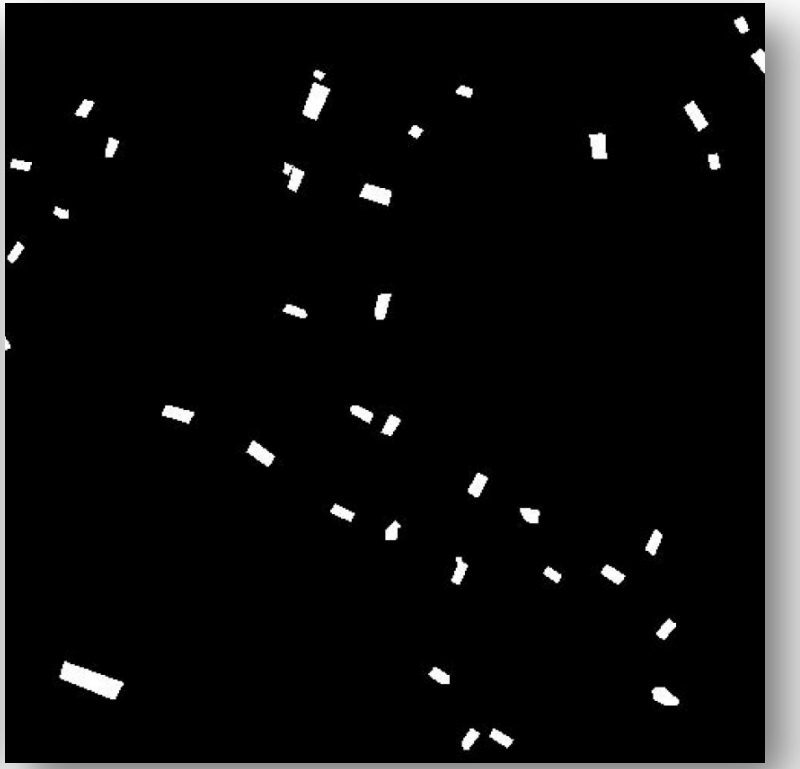
Segmentación de imágenes médicas



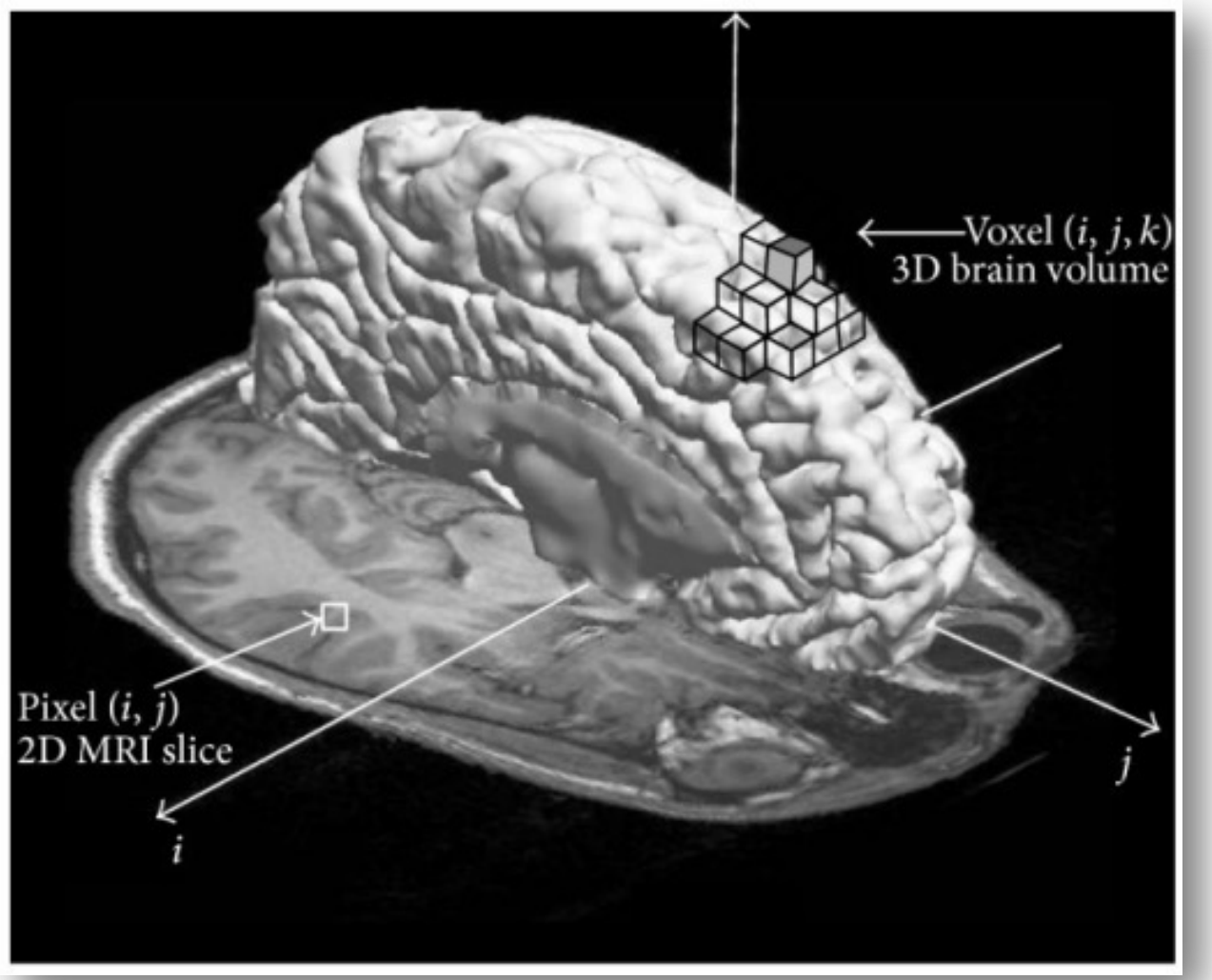
Estructuras anatómicas

Estructuras patológicas

Segmentación de imágenes satelitales



Imágenes 2D vs Imágenes 3D

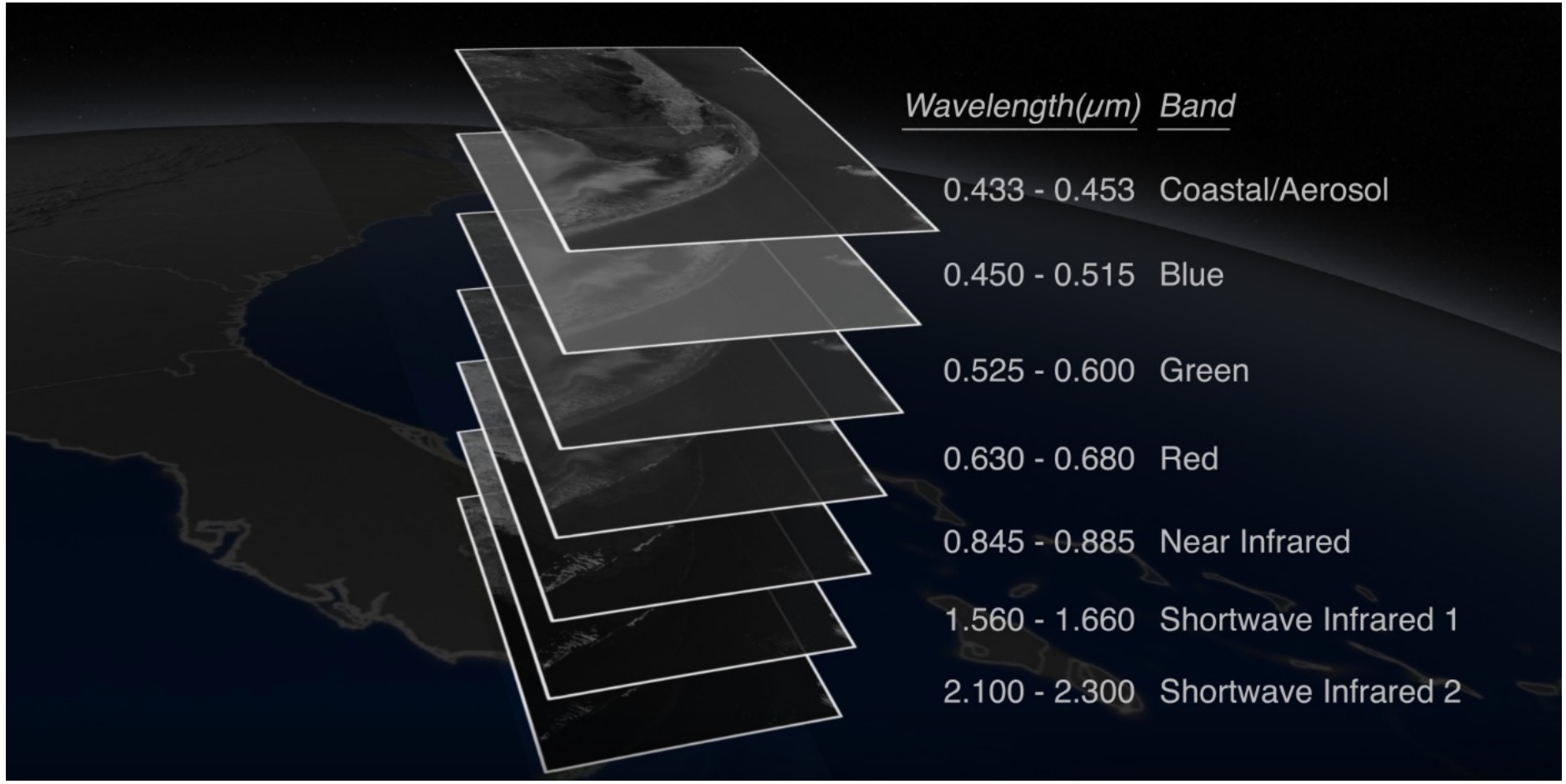


- Las imágenes pueden ser procesadas en 2D o en 3D.

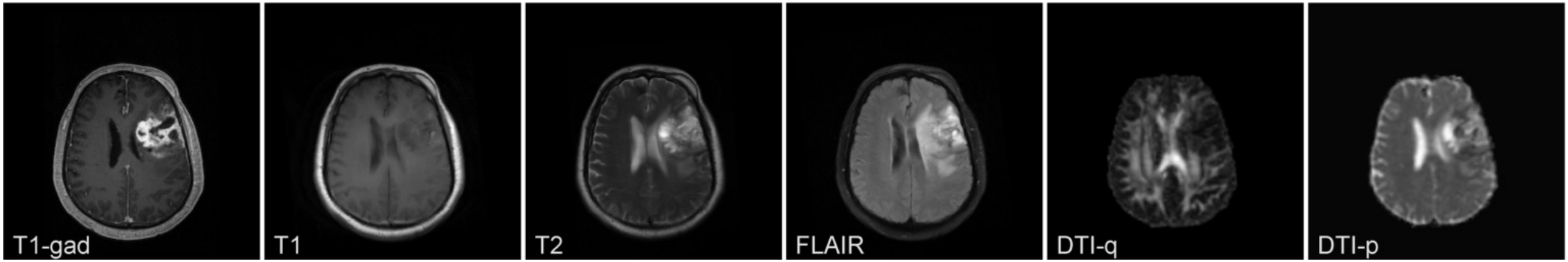
- Diferencia espacio pixel y físico.

Múltiples canales de entrada

Múltiples canales de entrada

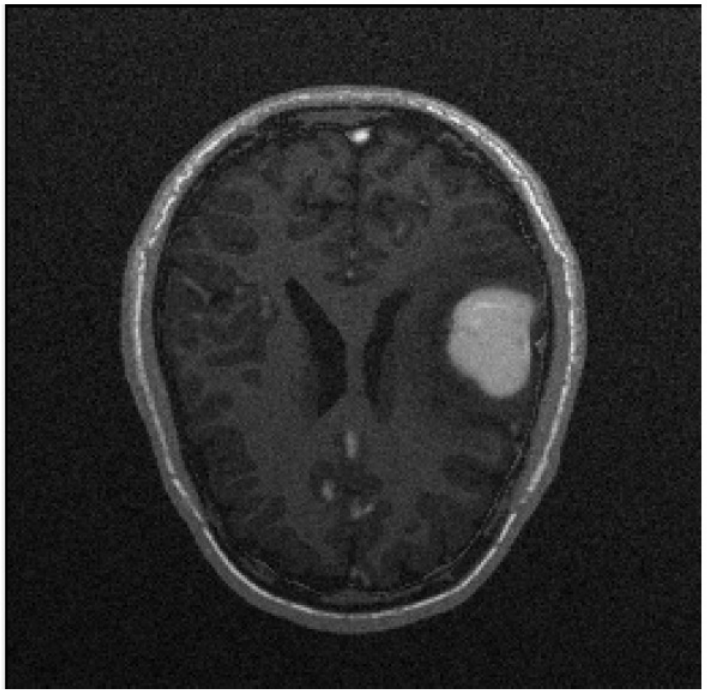


Múltiples canales de entrada



Técnicas clásicas de segmentación de imágenes

Umbralado



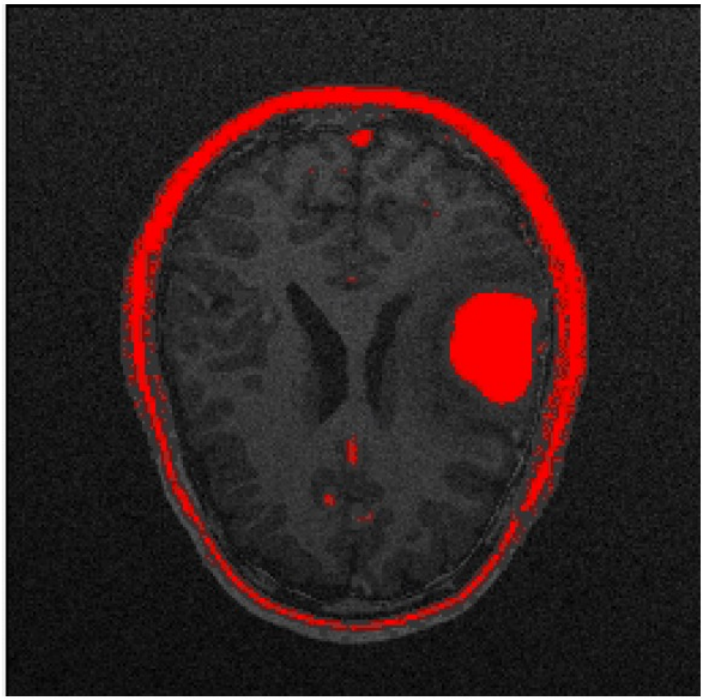


Imagen extraida de Ferrante, 2012

Crecimiento de regiones

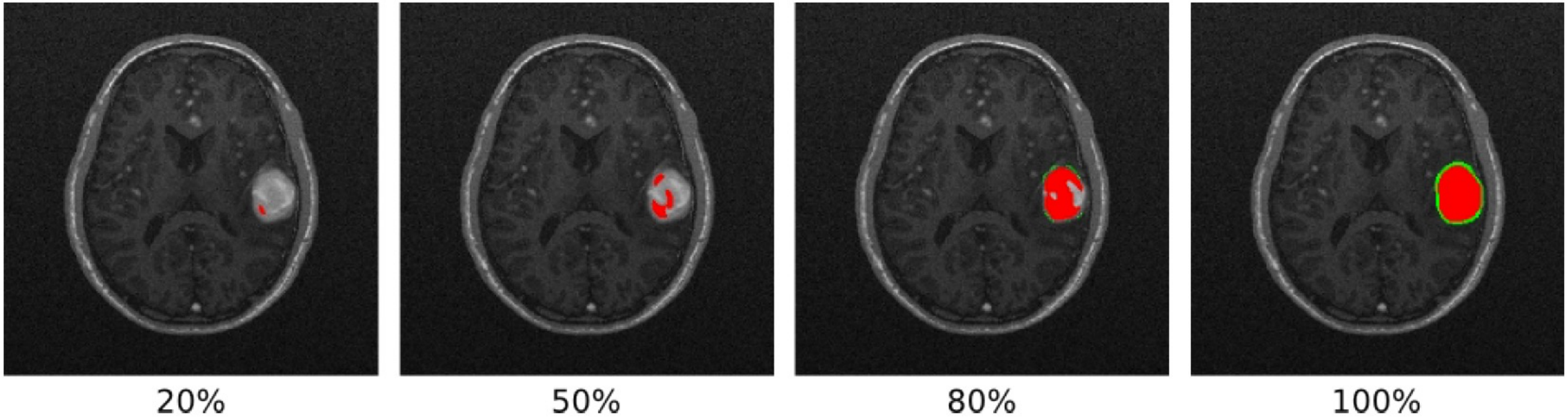


Imagen extraida de Ferrante, 2012

Watershed

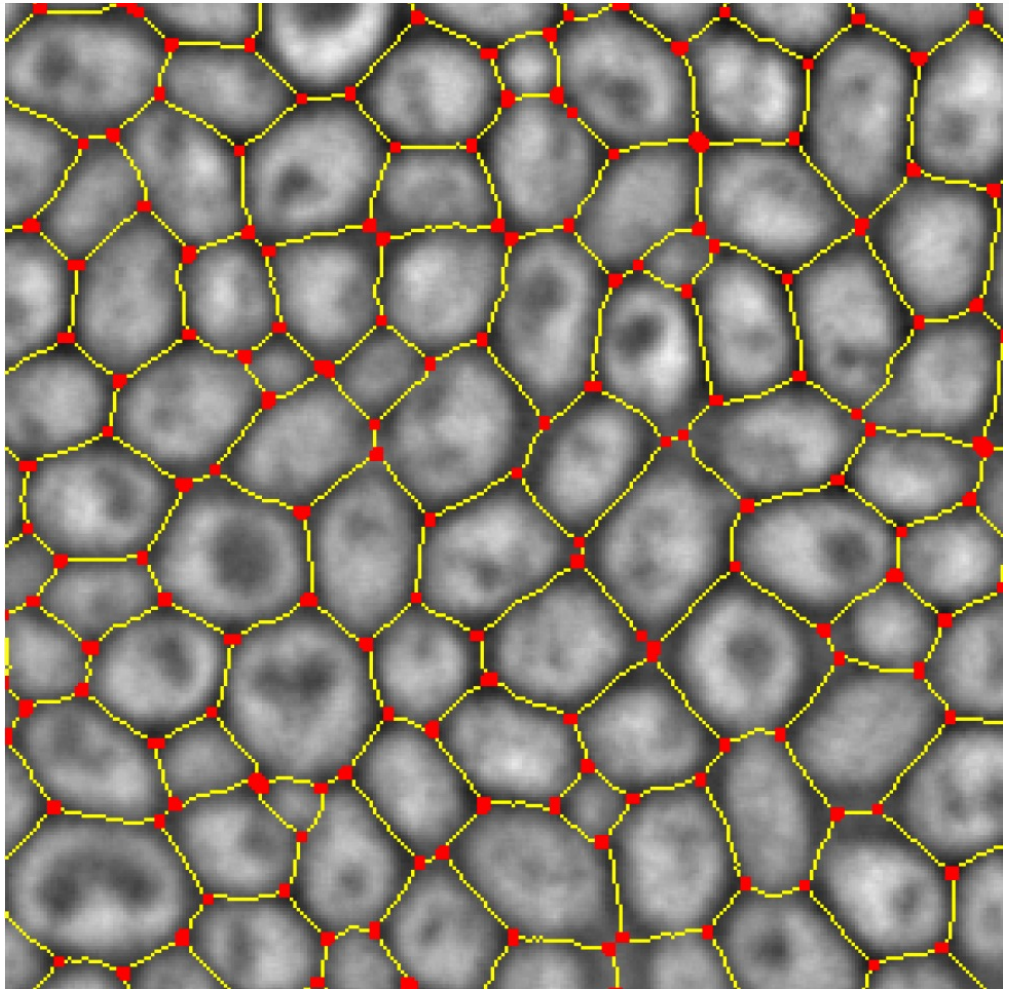
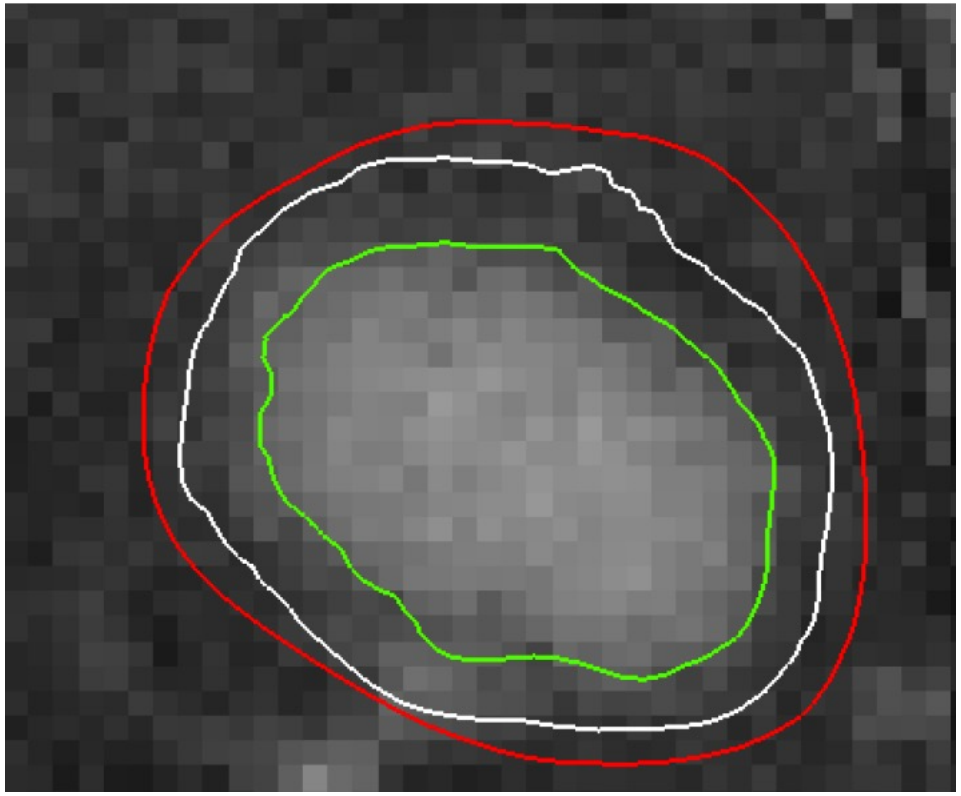
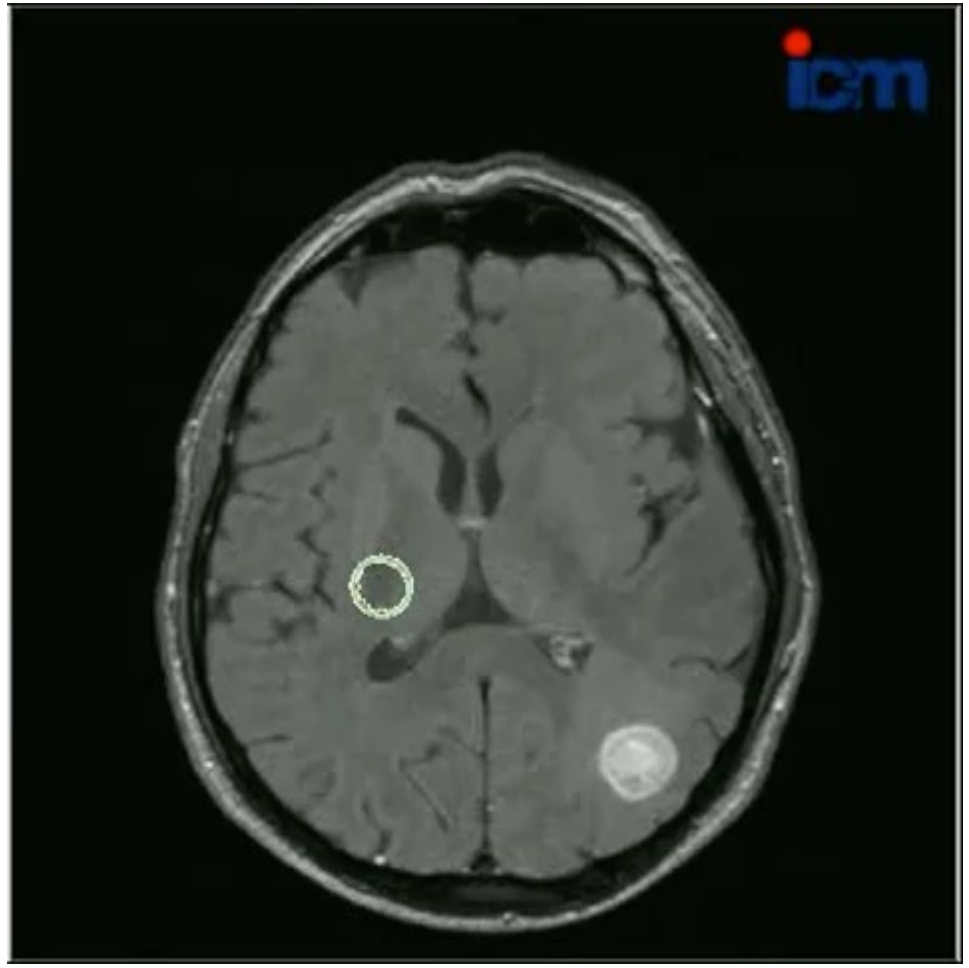


Imagen extraida de http://biodynamics.ucsd.edu/ir

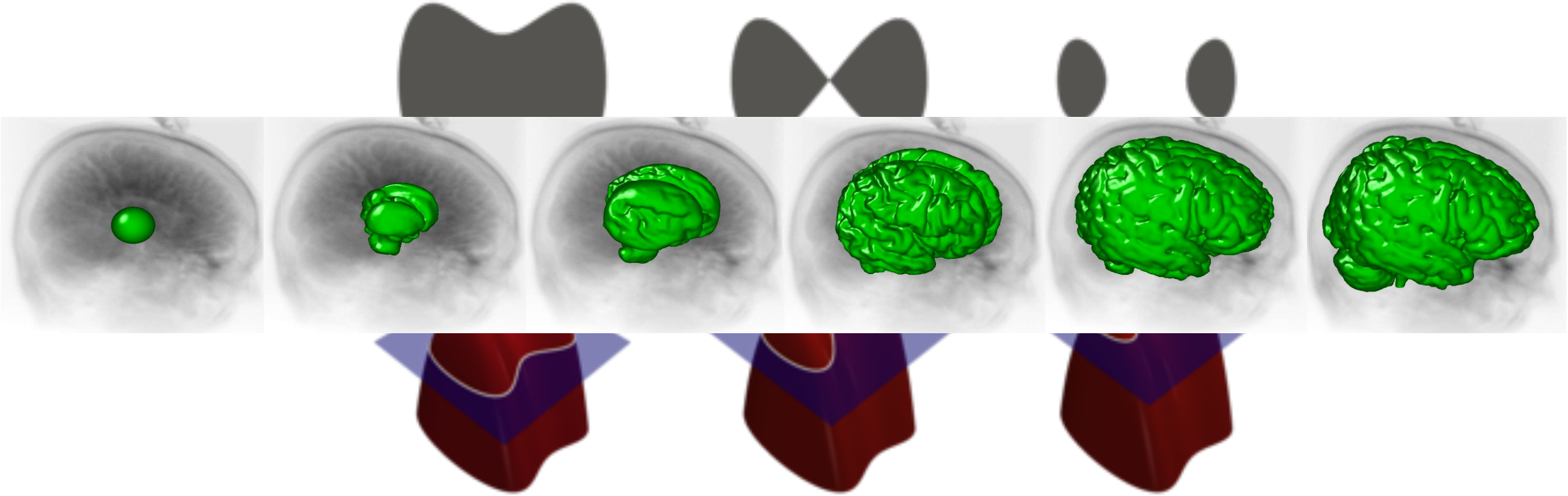
Modelos deformables explícitos

Ej. Contornos activos o Snakes

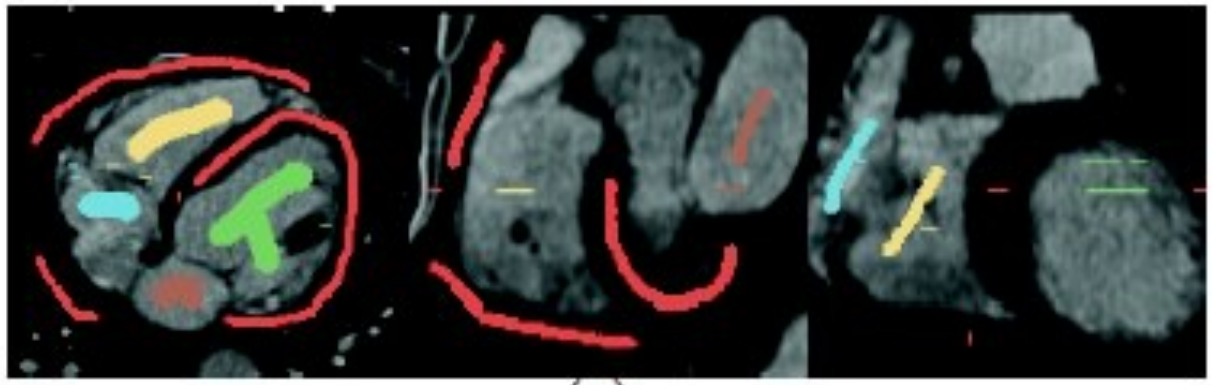




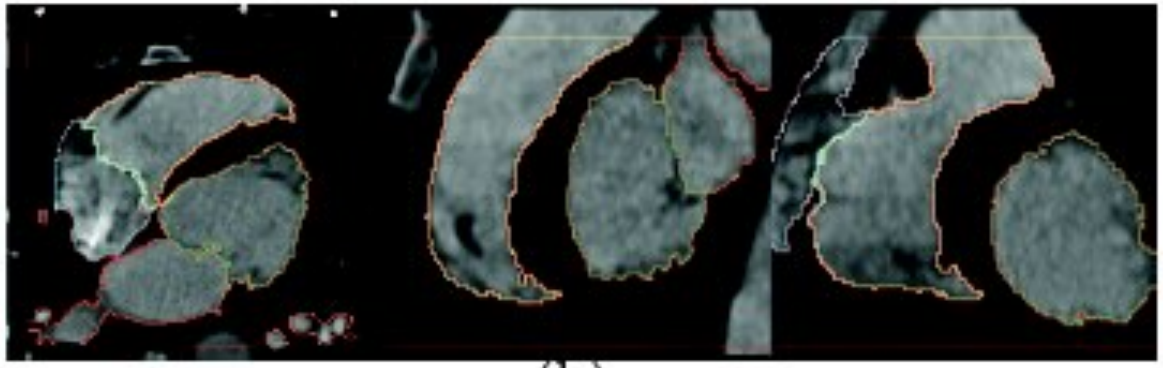
Level-sets



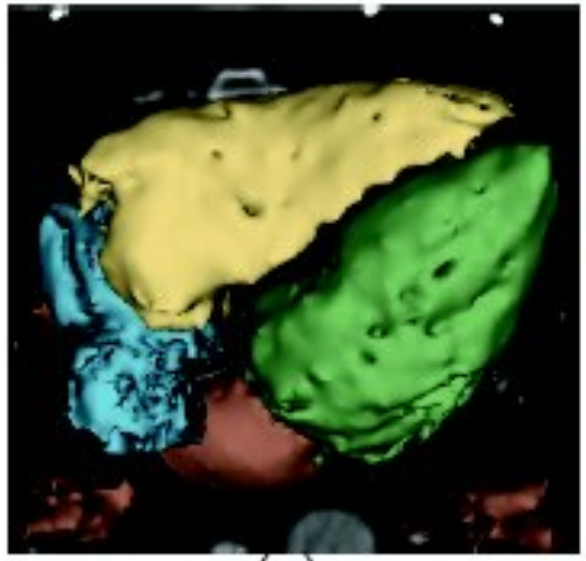
Segmentación interactiva



(a)



(b)



(c)

Donde encontrar todos estos métodos implementados?

OpenCV

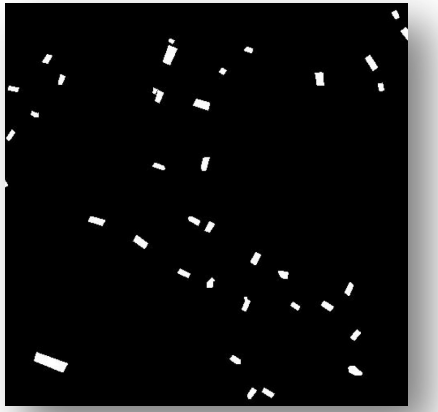
itk

Simple itk

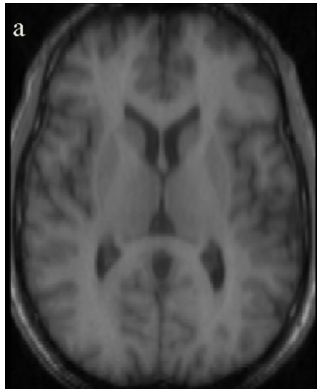
Redes neuronales convolucionales para la segmentación de imágenes

Segmentación de imágenes binaria

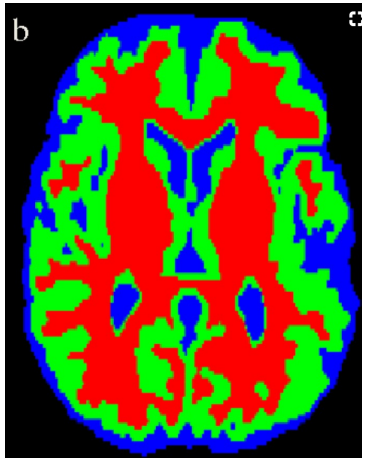
CNN



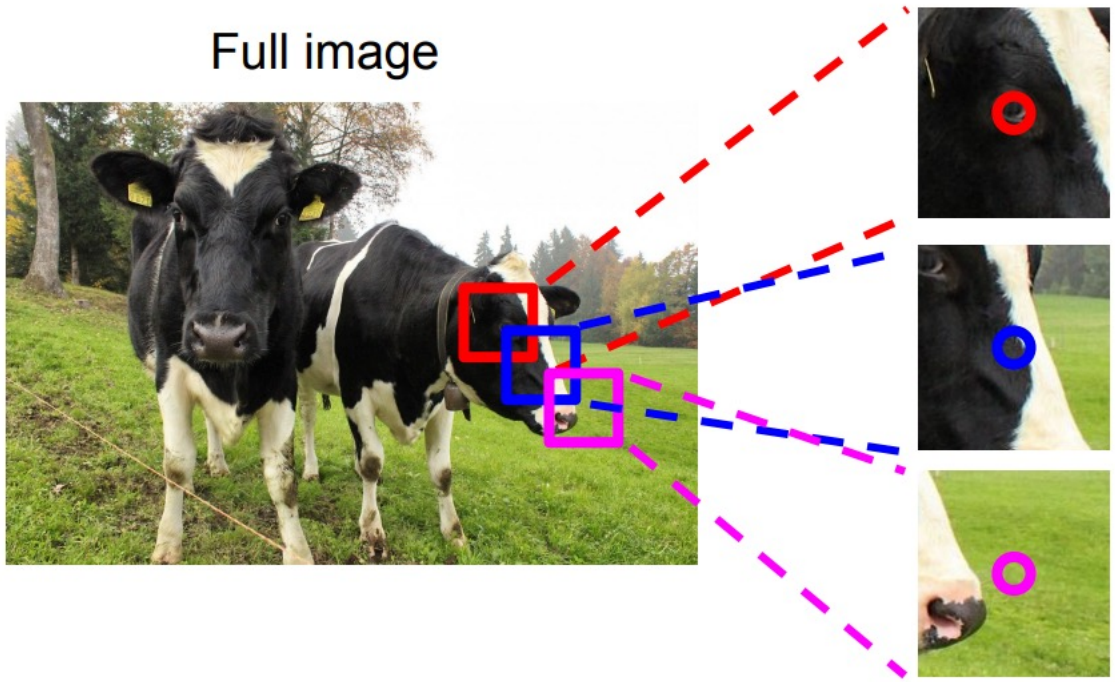
Segmentación de imágenes multiclase



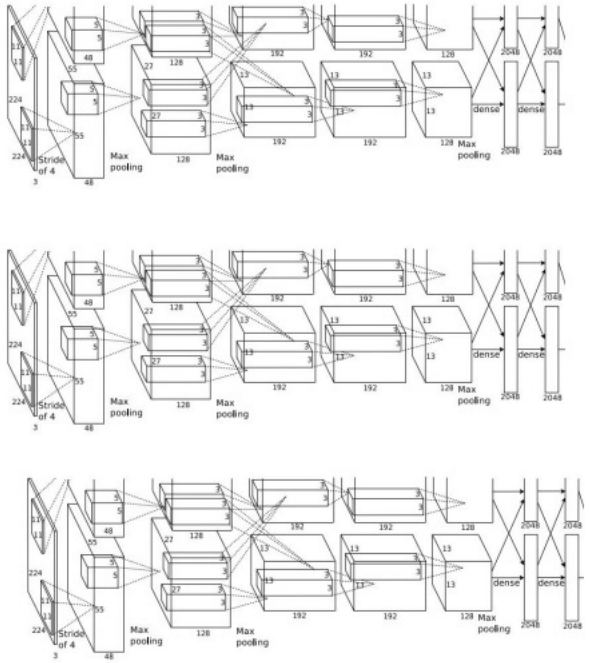
CNN



Primera solución: sliding window



AlexNet



Vaca

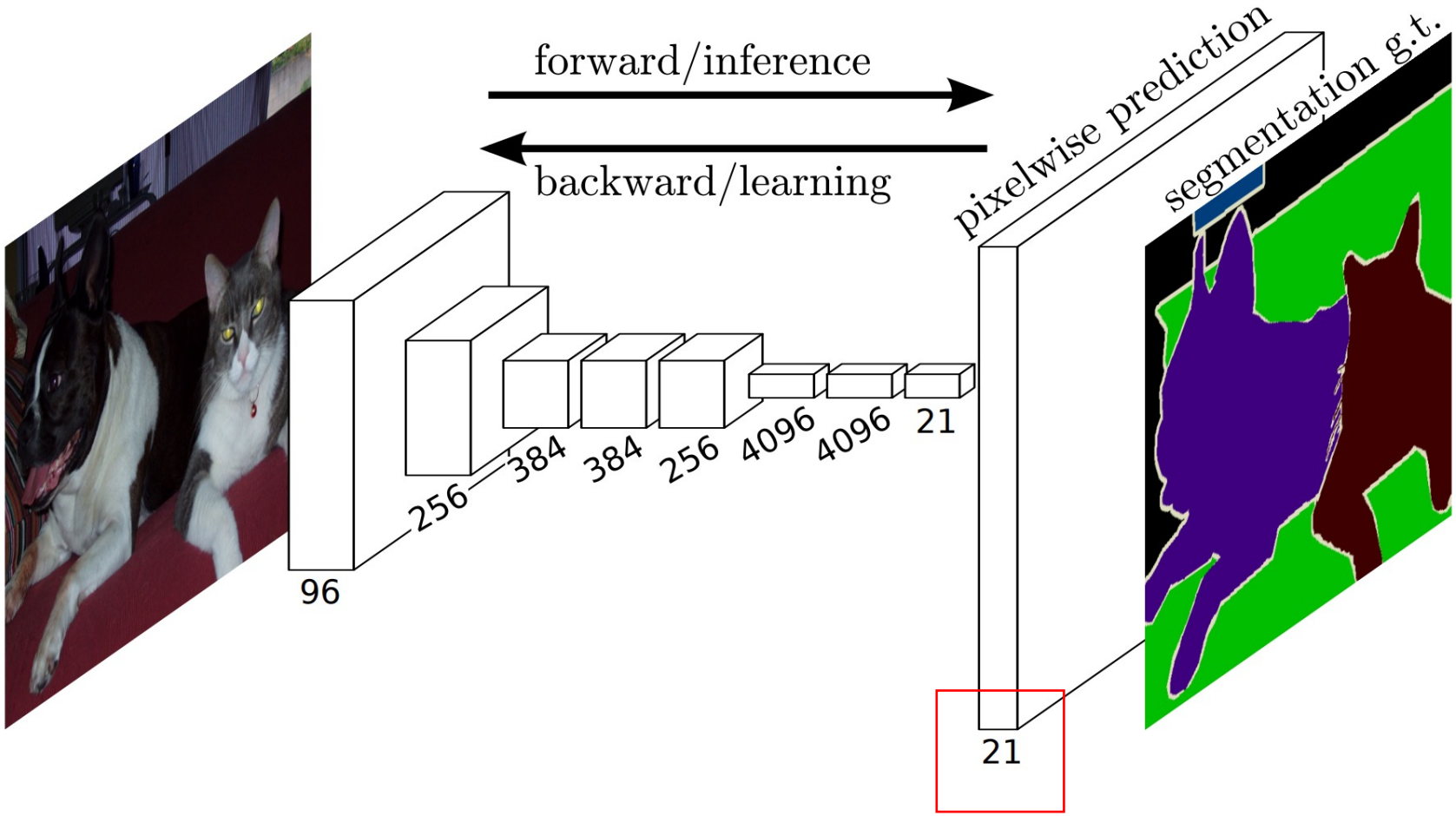
Vaca

Pasto

Muy ineficiente porque no hace uso de las features compartidas entre parches

Source: http://cs231n.stanford.edu/slides/2020/lecture_12.pdf

Fully Convolutional Networks



- El último feature map tiene tantos canales como posibles clases

- Computamos la función de pérdida para cada pixel y promediamos.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

Fully Convolutional Networks

Fully Convolutional Networks for Semantic Segmentation

Jonathan Long* Evan Shelhamer* Trevor Darrell
UC Berkeley
{jonlong,shelhamer,trevor}@cs.berkeley.edu

Abstract

Convolutional networks are powerful visual models that yield hierarchies of features. We show that convolutional networks by themselves, trained end-to-end, pixels-to-pixels, exceed the state-of-the-art in semantic segmentation. Our key insight is to build "fully convolutional" networks that take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. We define and detail the space of fully convolutional networks, explain their application to spatially dense prediction tasks, and draw connections to prior models. We
contemporary classification networks (AlexNet [20],
[32]) into fully convolu-

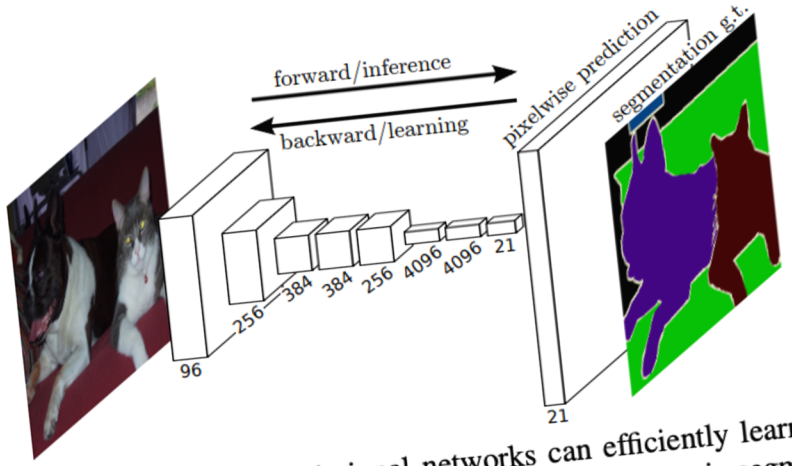
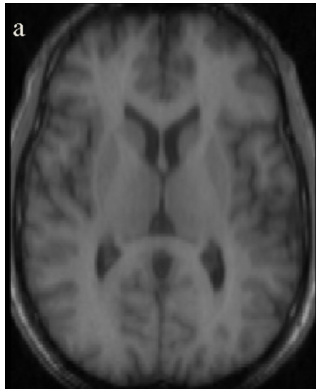


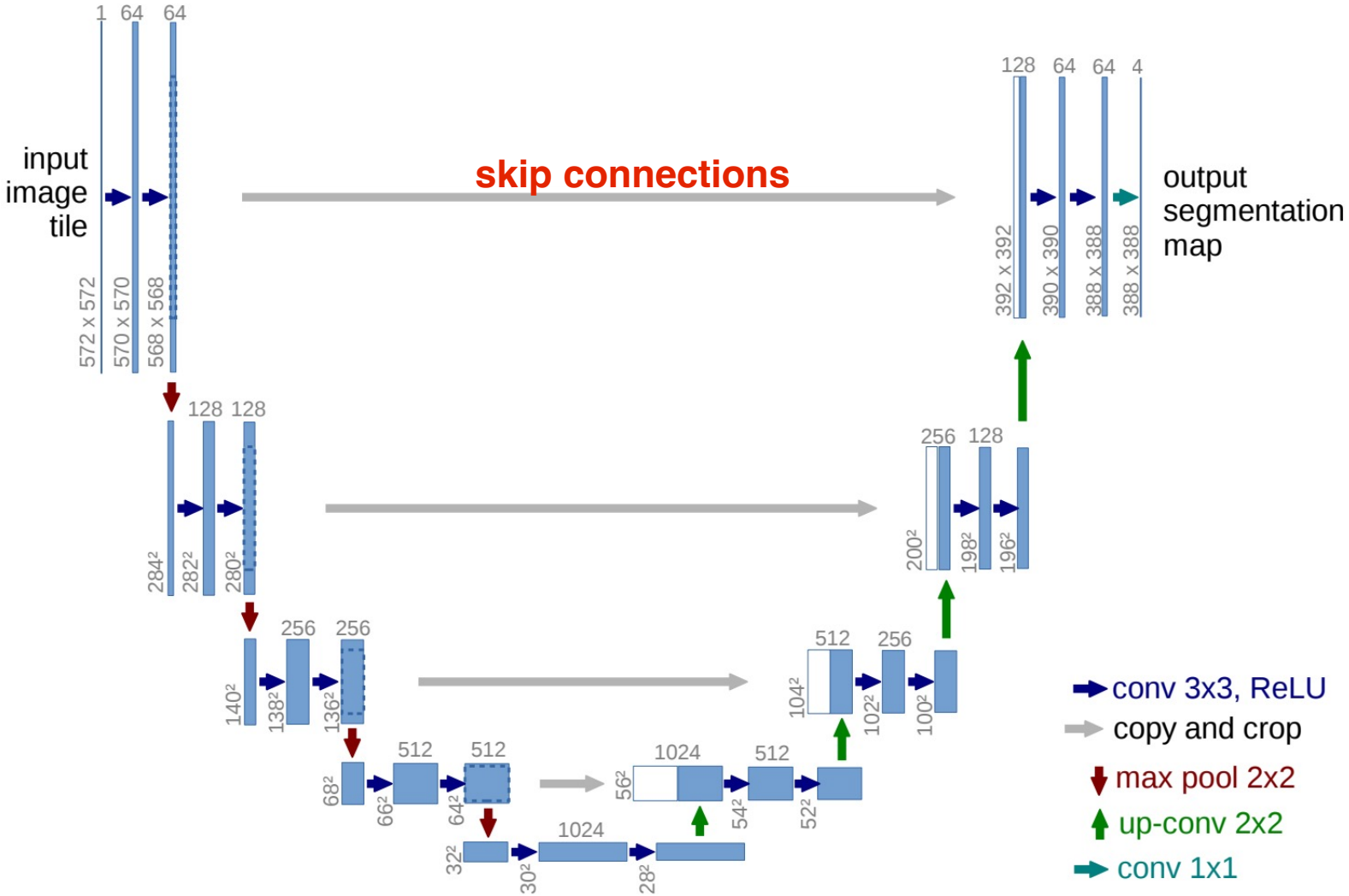
Figure 1. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation.

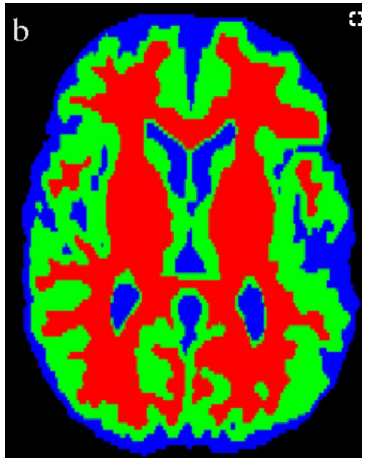
We show that a fully convolutional network (FCN) trained end-to-end, pixels-to-pixels on semantic segmen-
state-of-the-art without further machin-
train FCNs

Long, Jonathan, Ev ... tion. 2015.

UNet: Arquitectura encoder-decoder para segmentación de imágenes







UNet: Arquitectura encoder-decoder para segmentación de imágenes

U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, and Thomas Brox

Computer Science Department and BIOSS Centre for Biological Signalling Studies,
University of Freiburg, Germany
ronneber@informatik.uni-freiburg.de,
WWW home page: http://lmb.informatik.uni-freiburg.de/

Abstract. There is large consent that successful training of deep networks requires many thousand annotated training samples. In this paper, we present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. We show that such a network can be trained end-to-end from very few images and outperforms the prior best method (a sliding-window convolutional network) on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks. Using the same network trained on transmitted light microscopy images (phase contrast and DIC) we won the ISBI cell tracking challenge 2015 in these categories by a large margin. Mo
of a 512×512

Operaciones de upsampling: Unpooling

Nearest Neighbor

1	2
3	4

Input: 2 x 2

1	1	2	2
1	1	2	2
3	3	4	4
3	3	4	4

Output: 4 x 4

"Bed of Nails"

1	2
3	4

Input: 2 x 2

1	0	2	0
0	0	0	0
3	0	4	0
0	0	0	0

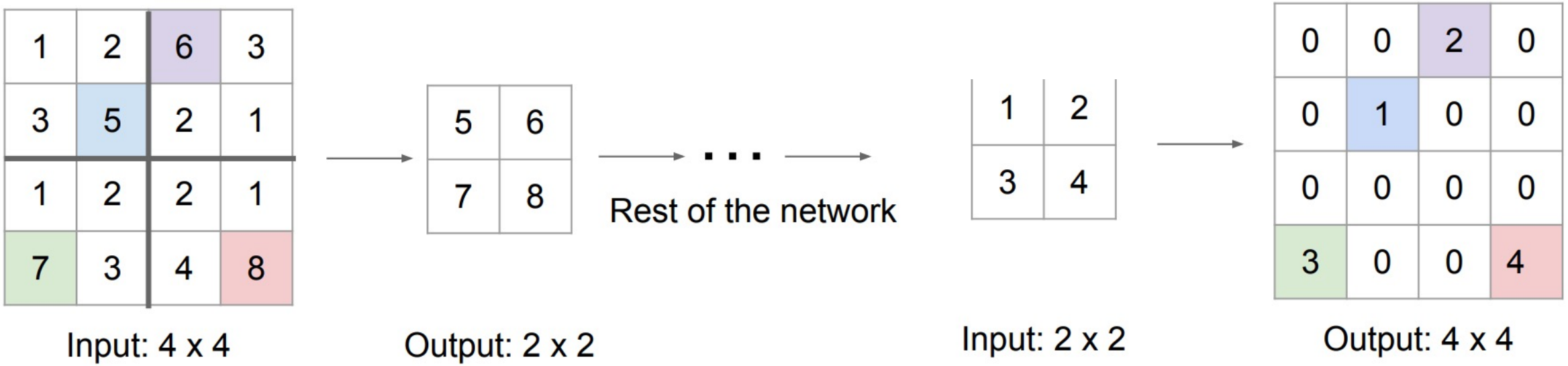
Output: 4 x 4

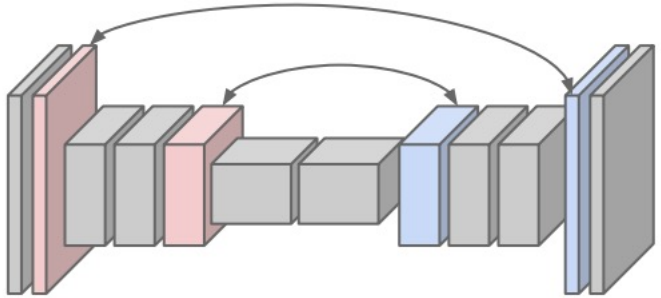
Source: http://cs231n.stanford.edu/slides/2020/lecture_12.pdf

Operaciones de upsampling: Unpooling

Durante **MaxPooling**,
nos acordamos la posición original del valor

Durante **MaxUnPooling**,
usamos esa posición

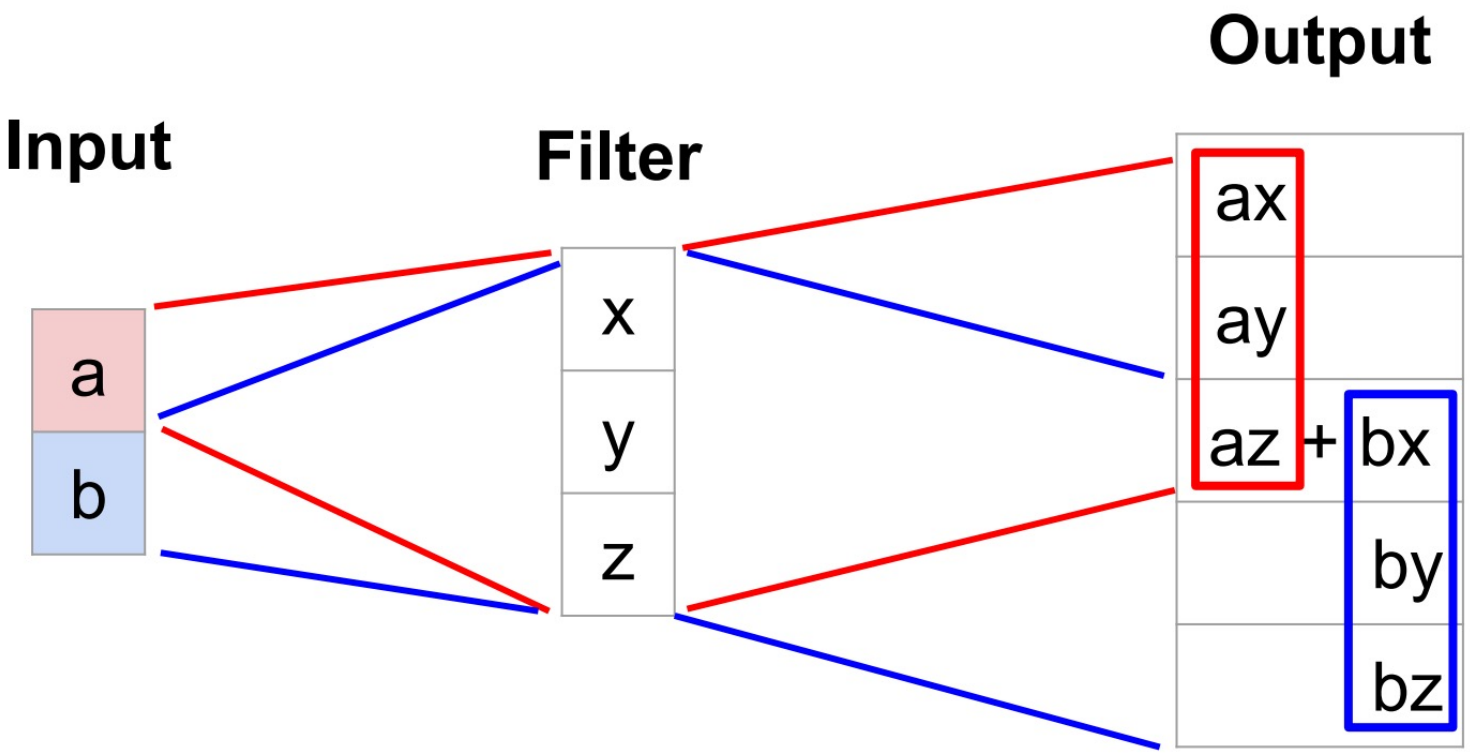




Source: http://cs231n.stanford.edu/slides/2020/lecture_12.pdf

Transpose convolution

Ejemplo para una convolución 1D



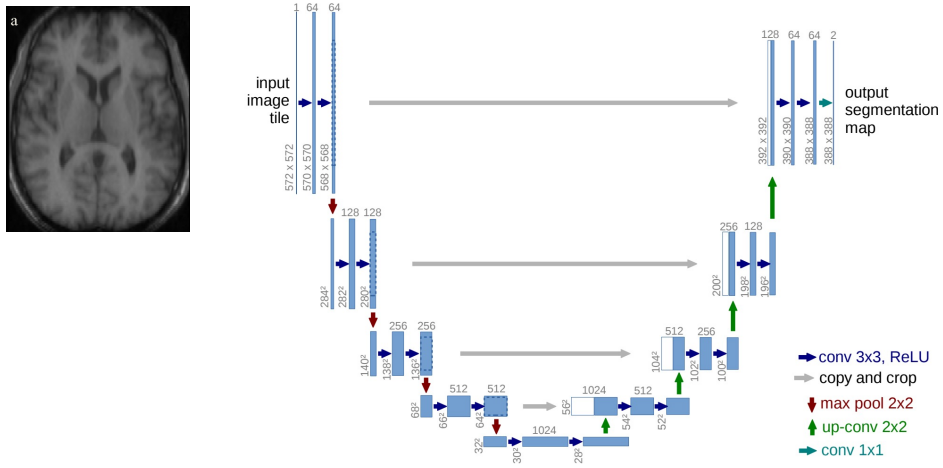
La salida se construye pesando los valores de la entrada con los distintos pesos de un kernel

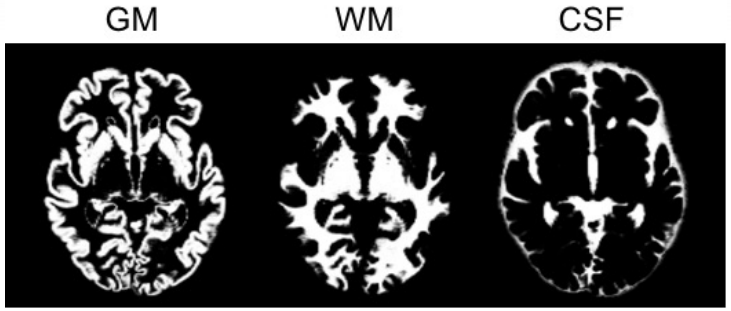
En las zonas de overlap, se suma

Por cada 1 pixel de la entrada, nos desplazamos 2 en la salida (stride = 2)

Source: http://cs231n.stanford.edu/slides/2020/lecture_12.pdf

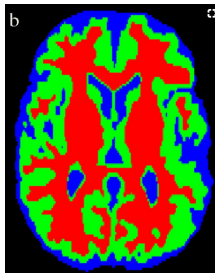
Cálculo de la función de pérdida para segmentación





Mapa de probabilidad por clase
(cada voxel suma 1 entre los 3 mapas)

ArgMax



- Se calcula la **función de pérdida por pixel** (ej. Entropía Cruzada) de los mapas de probabilidad de salida comprándolos con la versión One-hot del ground-truth, ' **y luego se promedia por todos los pixeles de la imagen**

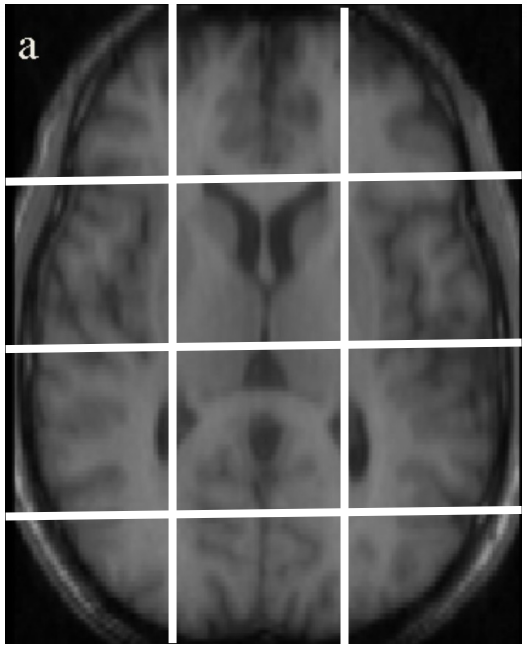
Redes totalmente convolucionales

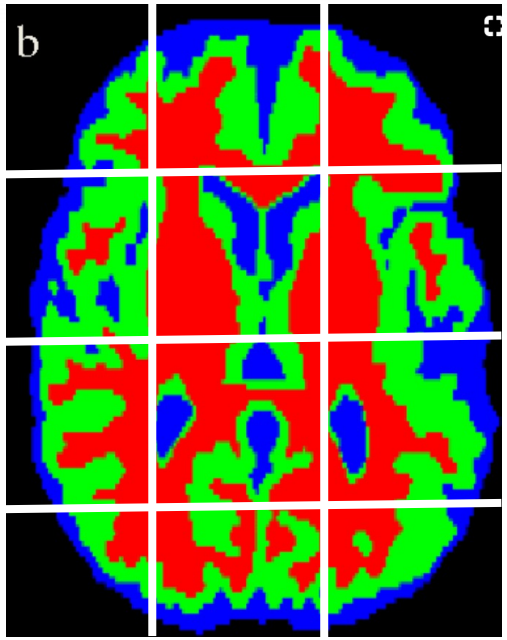
- Si la red no cuenta con ninguna capa totalmente conectada (densa), entonces se dice totalmente convolucional

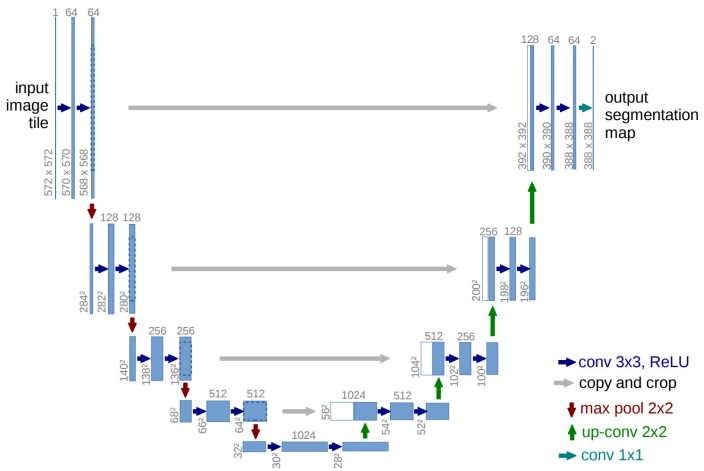
- Las capas densas pueden ser implementadas como capas convolucionales → Cómo?

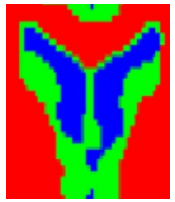
- La gran ventaja de una red totalmente convolucional es que puede procesar imágenes de cualquier tamaño, y puede ser entrenada y evaluada en imágenes de diferentes tamaños.

Entrenamiento por parches





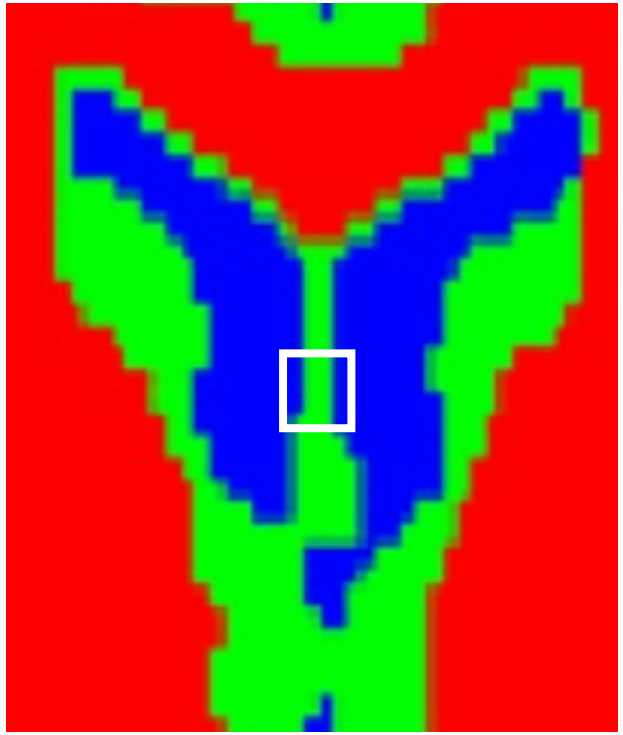




- Los mini-batches se componen de parches, no de imágenes completas
- En test time, si la red es totalmente convolucional, basta con insertar la nueva imagen y la predicción será del tamaño acorde *(cuidado con lo múltiplos de).*
- Si la red no es del tamaño acorde, se puede hacer el tiling fuera de la red, y luego reconstruir la segmentación

Estrategias de muestreo de parches

Para garantizar un entrenamiento equilibrado, es importante que cada mini-batch esté compuesto por parches centrados en diferentes etiquetas de forma equilibrada



Ayuda a solucionar el problema del desbalanceo de etiquetas

Procesando imágenes 3D

- Las arquitecturas son las mismas, pero utilizando convoluciones 3D en lugar de 2D
- Si la entrada era antes un tensor de dimensiones

(BatchSize, Channels, Width, Height,)

- Pasará a ser

(BatchSize, Channels , Depth, Width, Height)

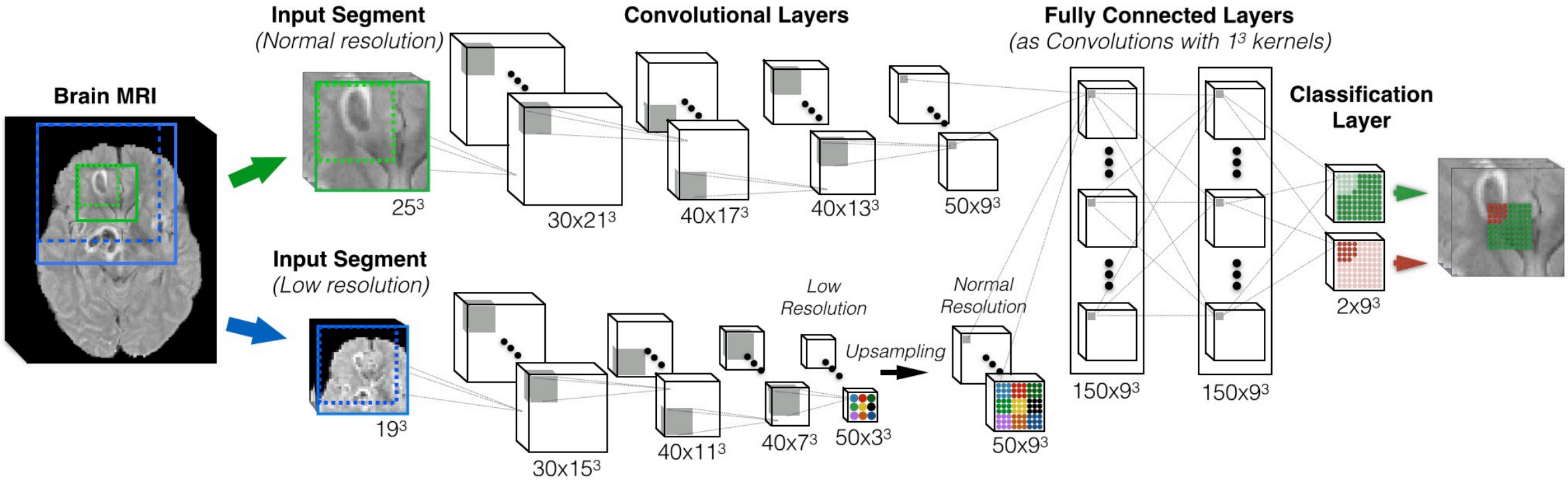
CONV3D

CLASS `torch.nn.Conv3d(in_channels, out_channels, kernel_size, stride=1, padding=0, dilation=1, groups=1, bias=True, padding_mode='zeros')` [SOURCE]

Applies a 3D convolution over an input signal composed of several input planes.

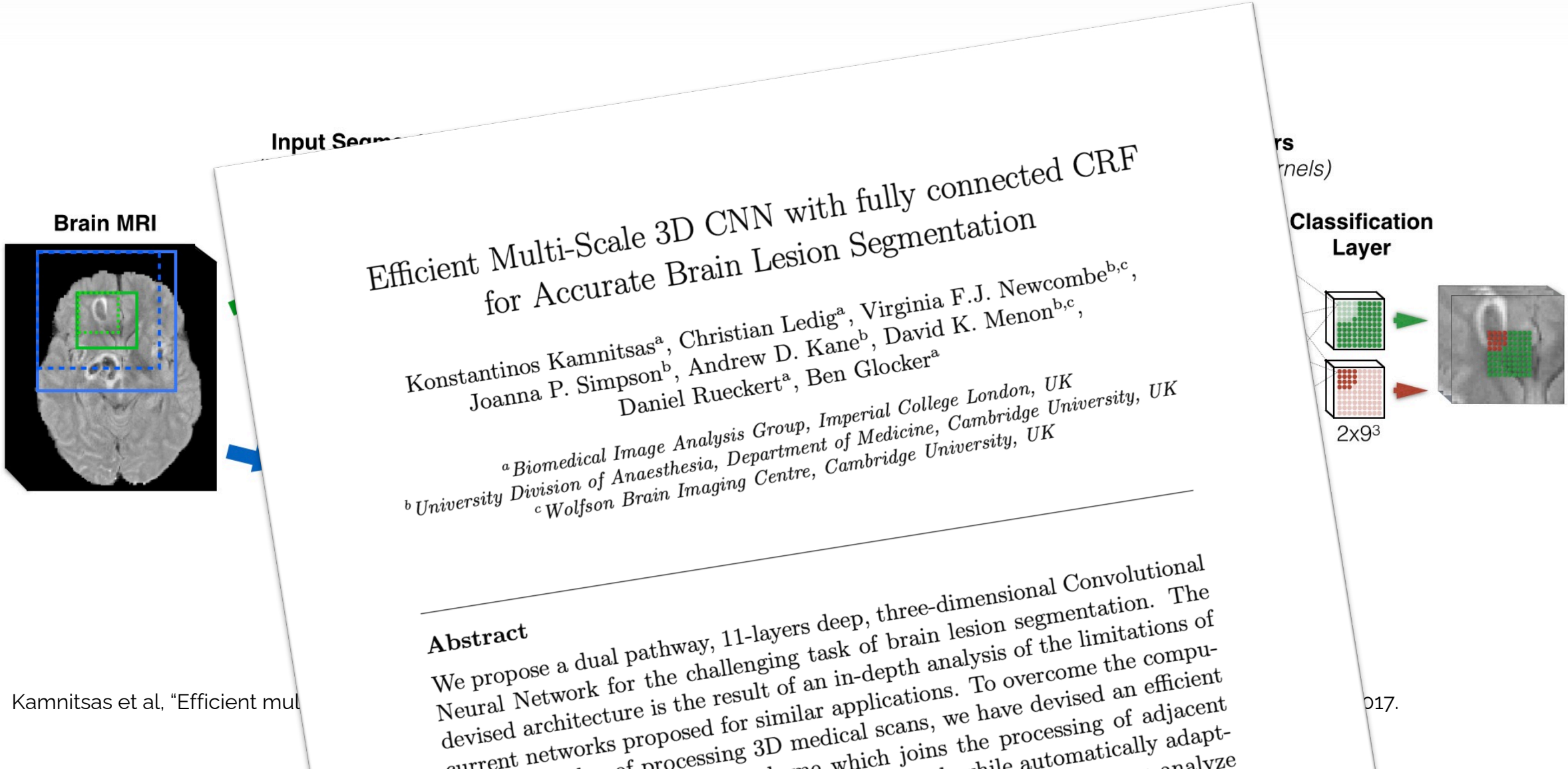
In the simplest case, the output value of the layer with input size (N, C_{in}, D, H, W) and output $(N, C_{out}, D_{out}, H_{out}, W_{out})$ can be precisely described as:

DeepMedic: ConvNets con branches multi-resolución



Kamnitsas et al, "Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation", Medical Image Analysis, 2017.

DeepMedic: ConvNets con branches multi-resolución



Efficient Multi-Scale 3D CNN with fully connected CRF for Accurate Brain Lesion Segmentation

Konstantinos Kamnitsas[a], Christian Ledig[a], Virginia F.J. Newcombe[b,c], Joanna P. Simpson[b], Andrew D. Kane[b], David K. Menon[b,c], Daniel Rueckert[a], Ben Glocker[a]

[a] Biomedical Image Analysis Group, Imperial College London, UK
[b] University Division of Anaesthesia, Department of Medicine, Cambridge University, UK
[c] Wolfson Brain Imaging Centre, Cambridge University, UK

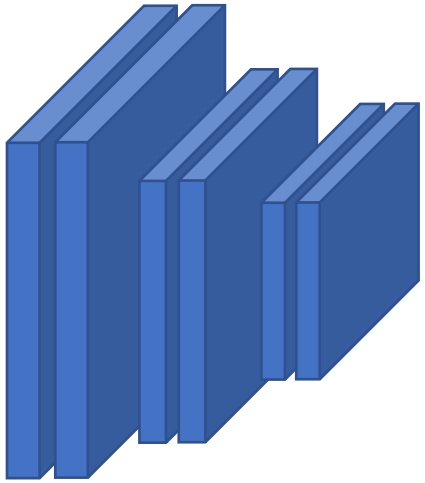
Abstract

We propose a dual pathway, 11-layers deep, three-dimensional Convolutional Neural Network for the challenging task of brain lesion segmentation. The devised architecture is the result of an in-depth analysis of the limitations of current networks proposed for similar applications. To overcome the compu-

Kamnitsas et al, "Efficient mul … 017.

Arquitectura básica para localización

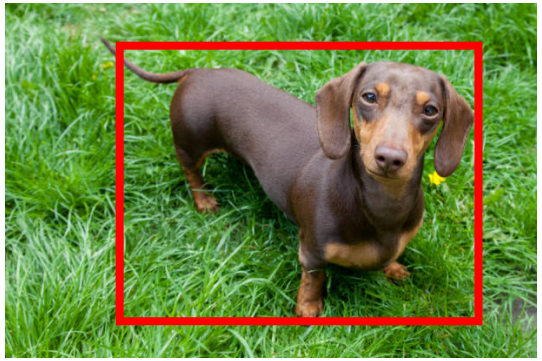
Red con salidas múltiples



Clasificación
(Probabilidad por clase)

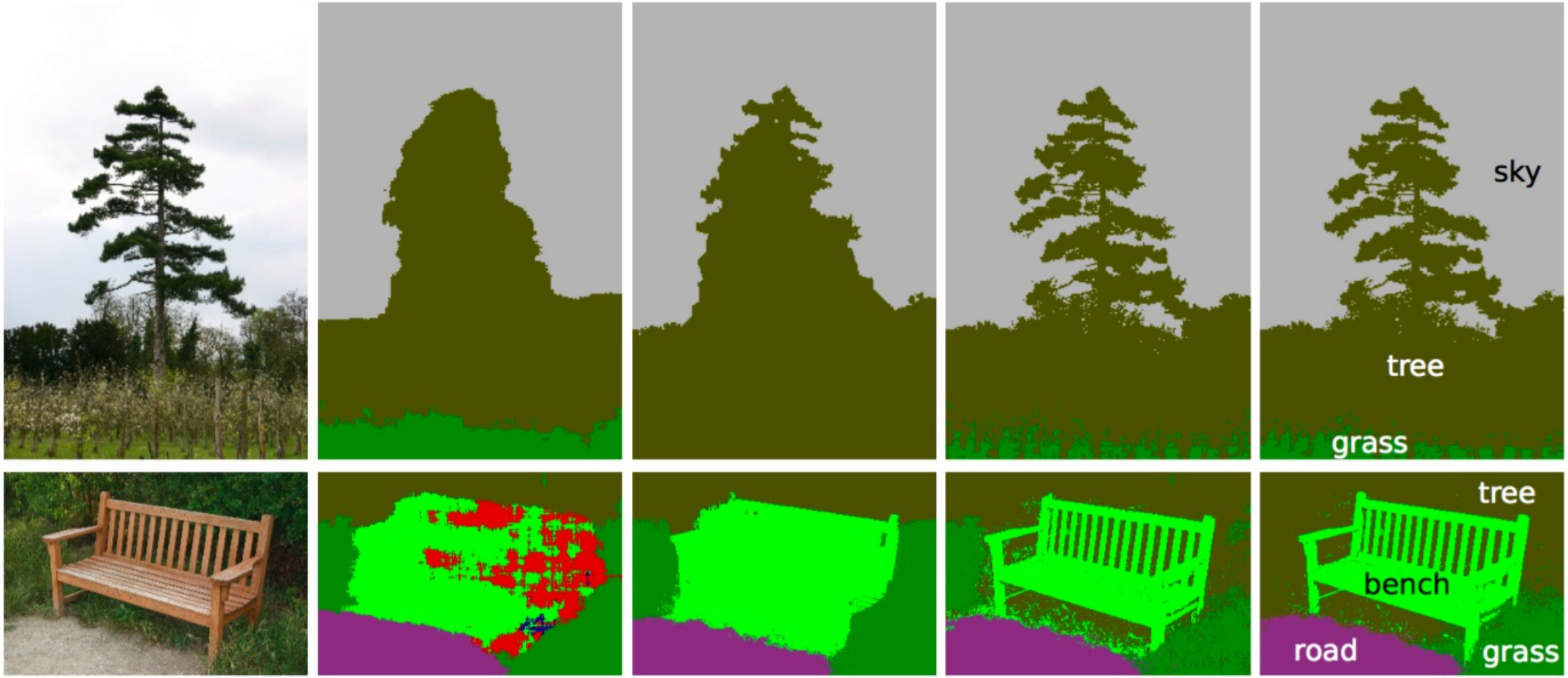
Regresión (x, y, w, h)

Perro: 0.8



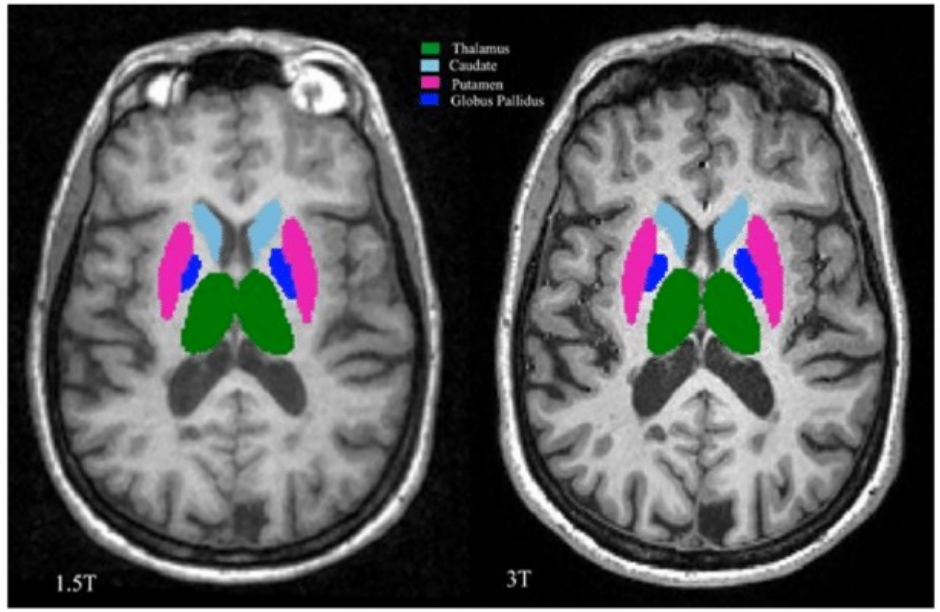
Medidas para evaluar la calidad de la segmentación

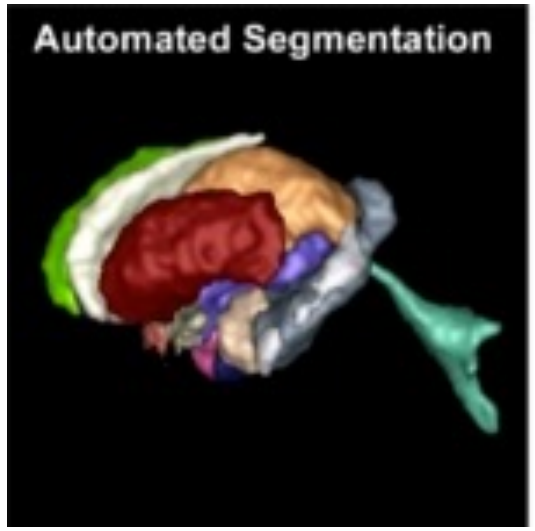
Medidas para evaluar la calidad de la segmentación

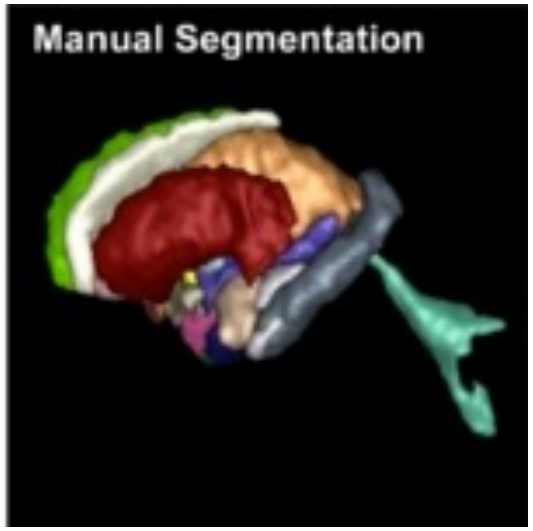


(a) Image (b) Unary classifiers (c) Robust P^n CRF (d) Fully connected CRF, MCMC inference, 36 hrs (e) Fully connected CRF, our approach, 0.2 seconds

Medidas para evaluar la calidad de la segmentación

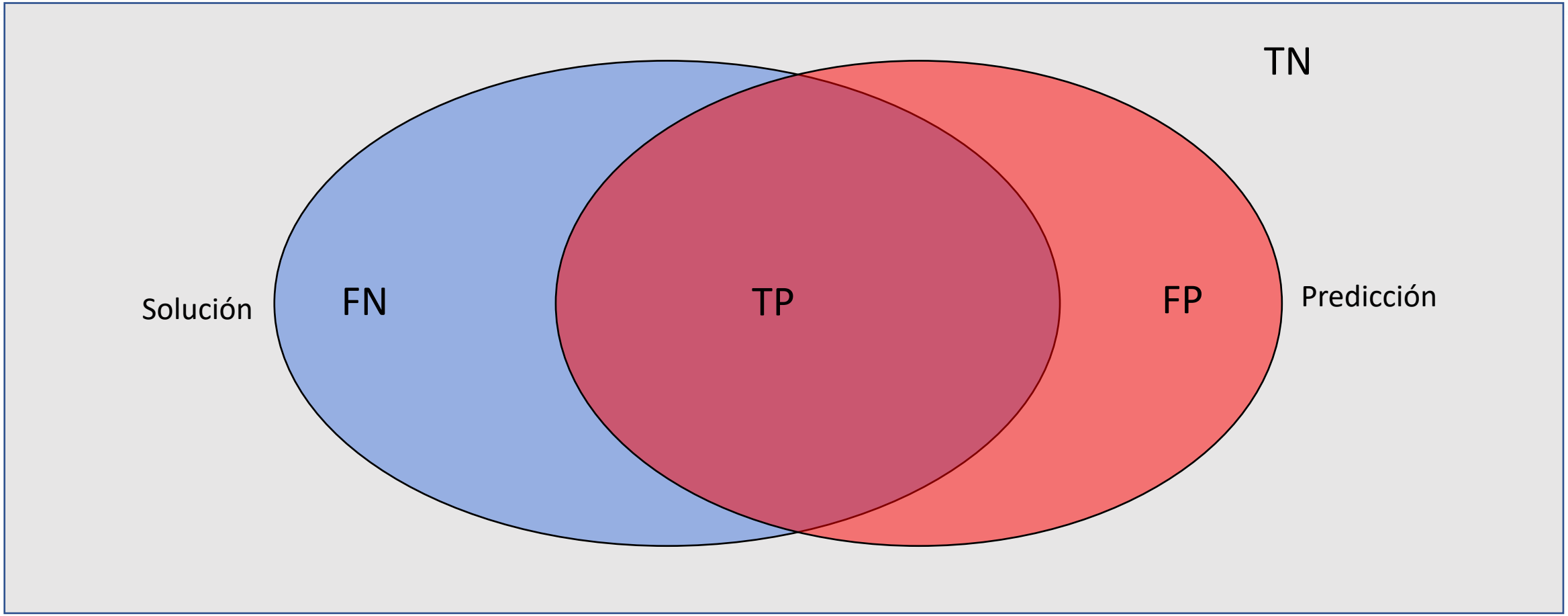






¿Intuición?

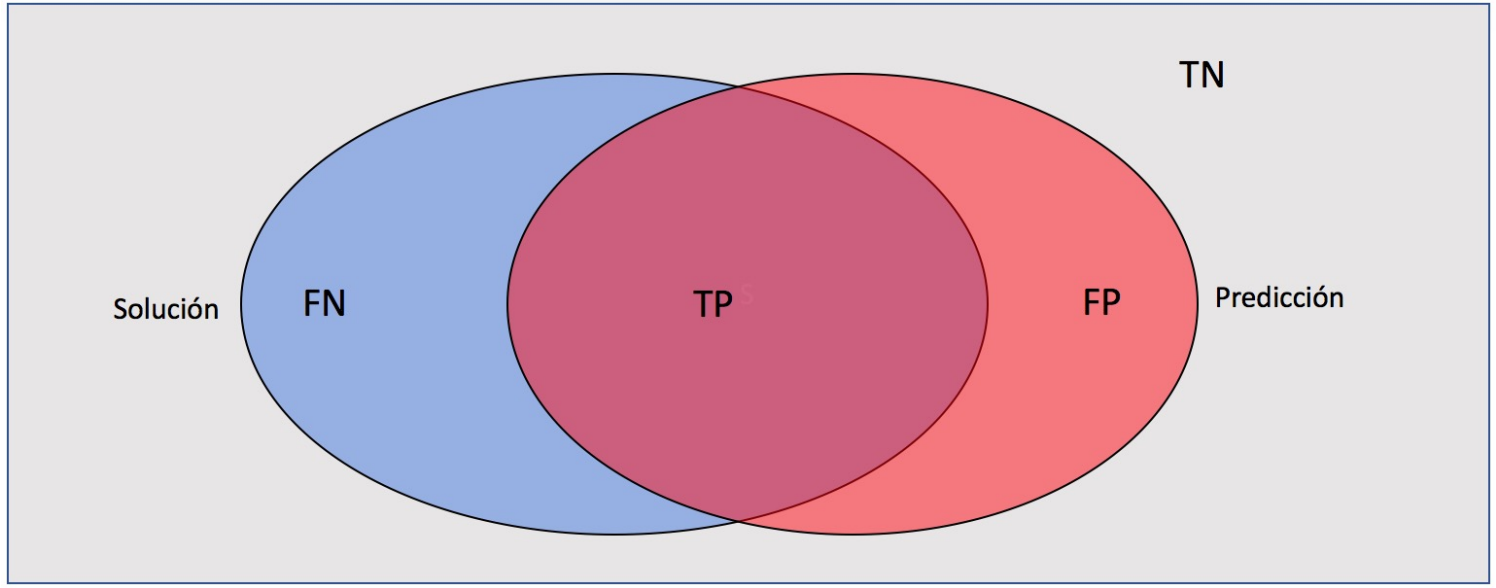
Intuición



Tasa de acierto (accuracy)

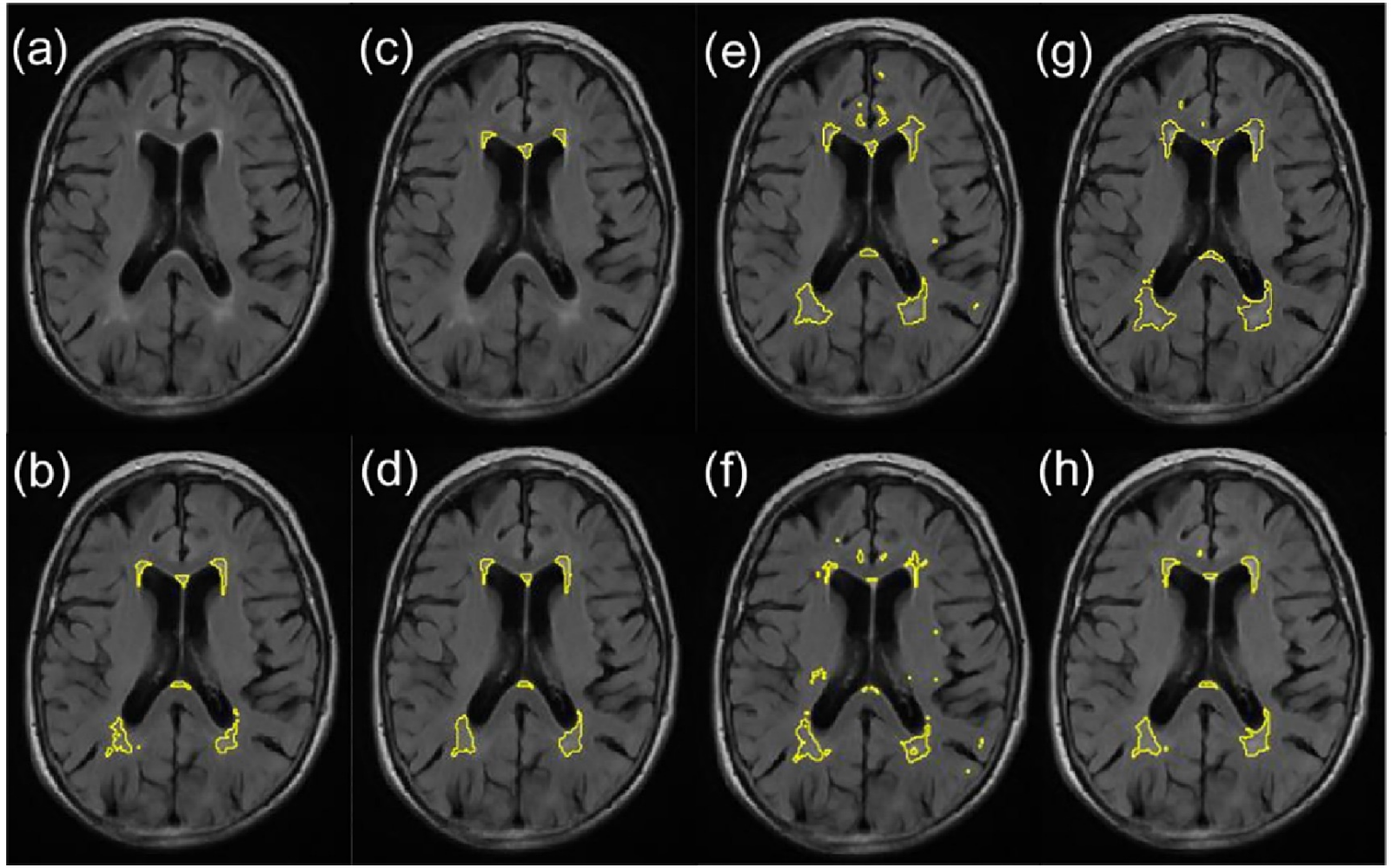
Tasa de acierto sobre el total de los píxeles

$$Accuracy = \frac{TP + TN}{FN + TP + FP + TN}$$



Tasa de acierto (accuracy)

Problema: Requiere datos balanceados



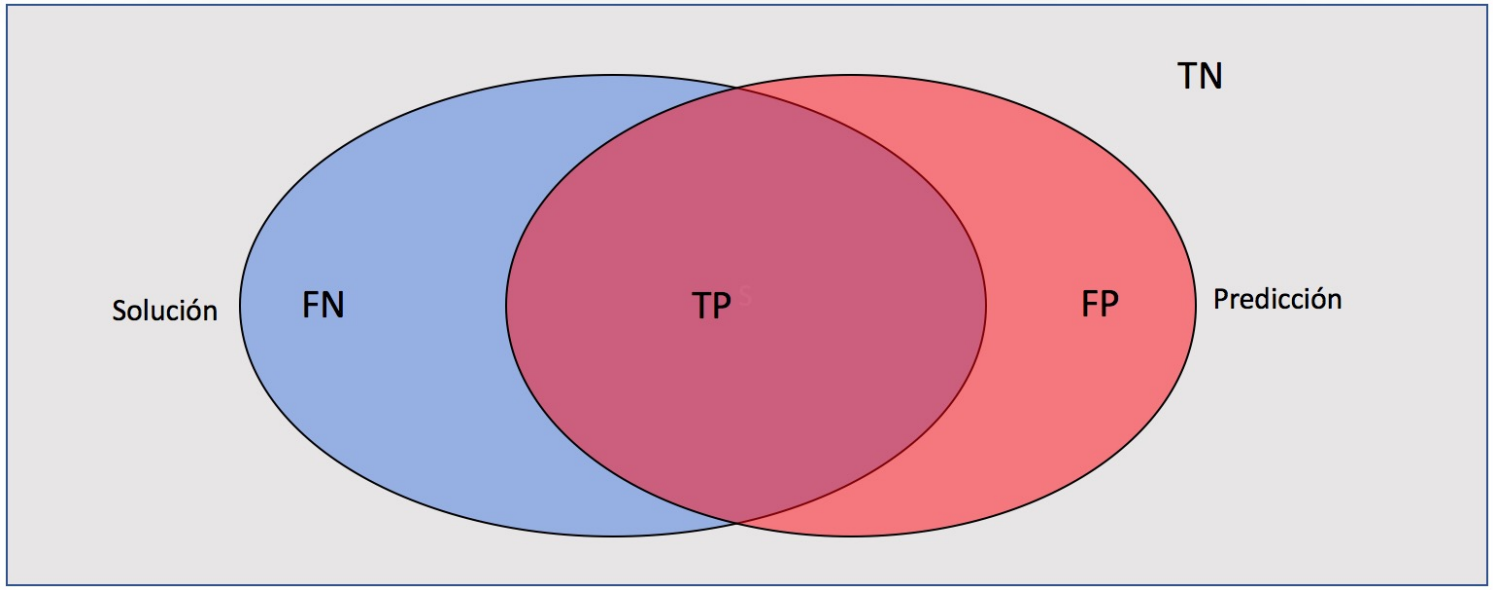
Leucoaraiosis (hiperintesidad en la materia blanca)

Medidas de solapamiento

Coeficiente Dice/ F1-Score

Se aplica en datos binarios

$$Dice = \frac{2\,|\,Solución\ \cap Predicción\,|}{|Solución| + |Predicción|} = \frac{2\,TP}{FN + 2\,TP + FP}$$

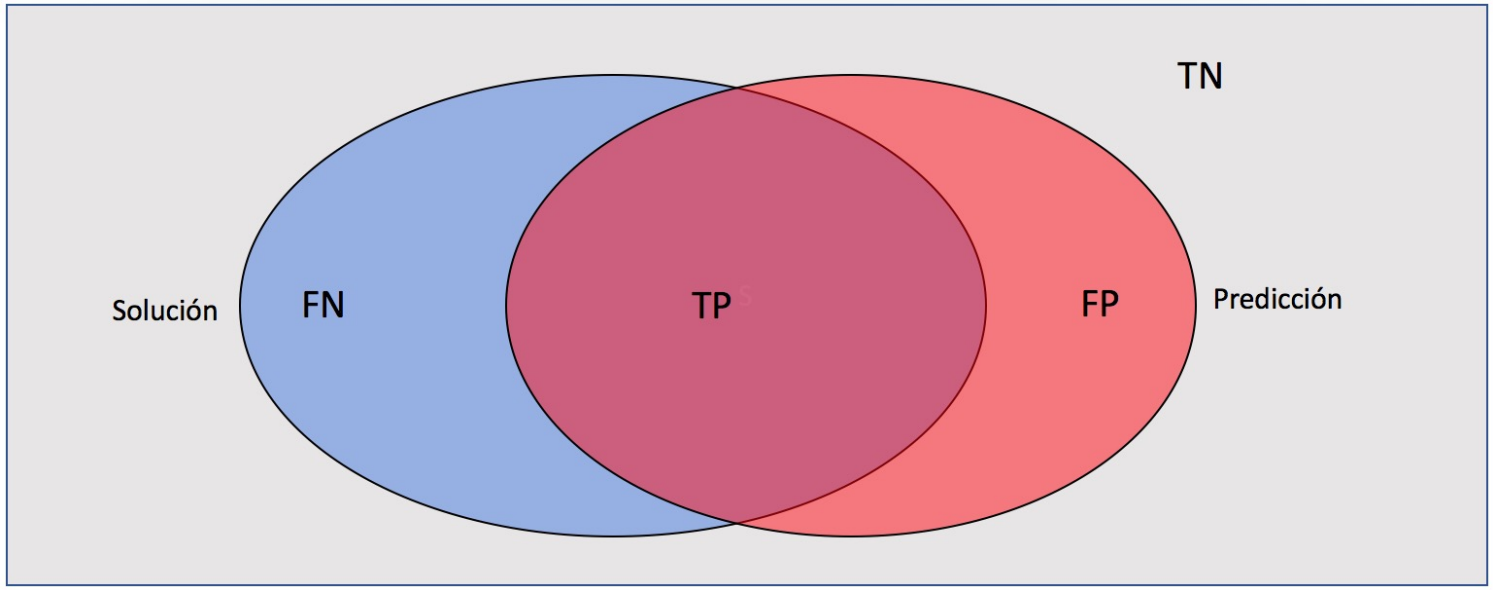


También suele usarse como función de pérdida!

Coeficiente Jaccard / IoU (Intersection over union)

Se aplica en datos binarios

$$Jaccard = \frac{|Solución \ \cap Predicción|}{|Solución \ \cup \ Predicción|} = \frac{TP}{FN + TP + FP}$$



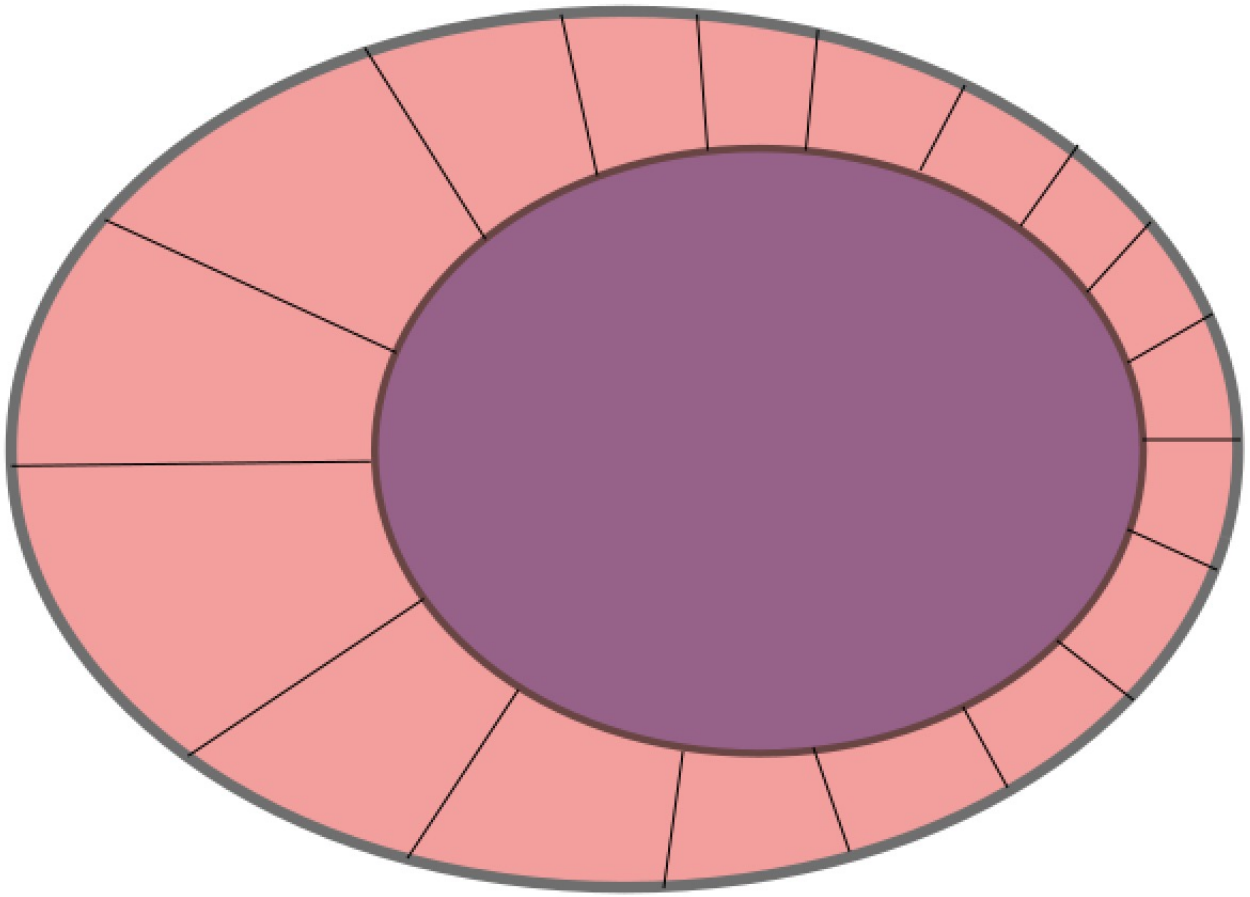
Relación entre Dice/F1 y Jaccard/IoU

$$Jaccard = \frac{Dice}{2 - Dice}$$

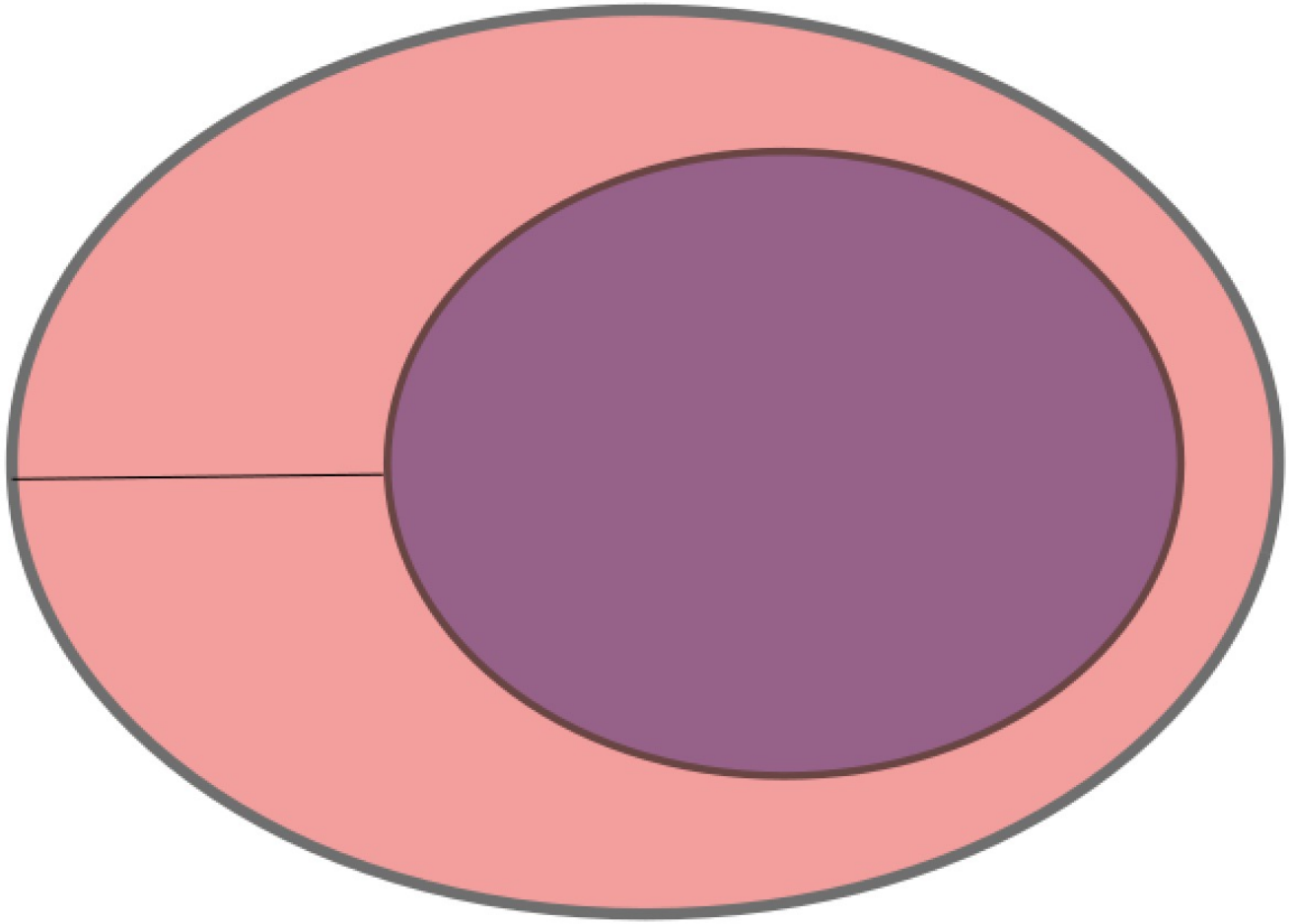
- En general, Jaccard tiende a penalizar las malas clasificaciones más que Dice en términos cuantitativos.
- Dice y Jaccard están correlacionados positivamente

Medidas distancia entre contornos

Distancia media entre contornos



Distancia de Haussdorf
(máxima distancia entre contornos)



Distancia de Haussdorf

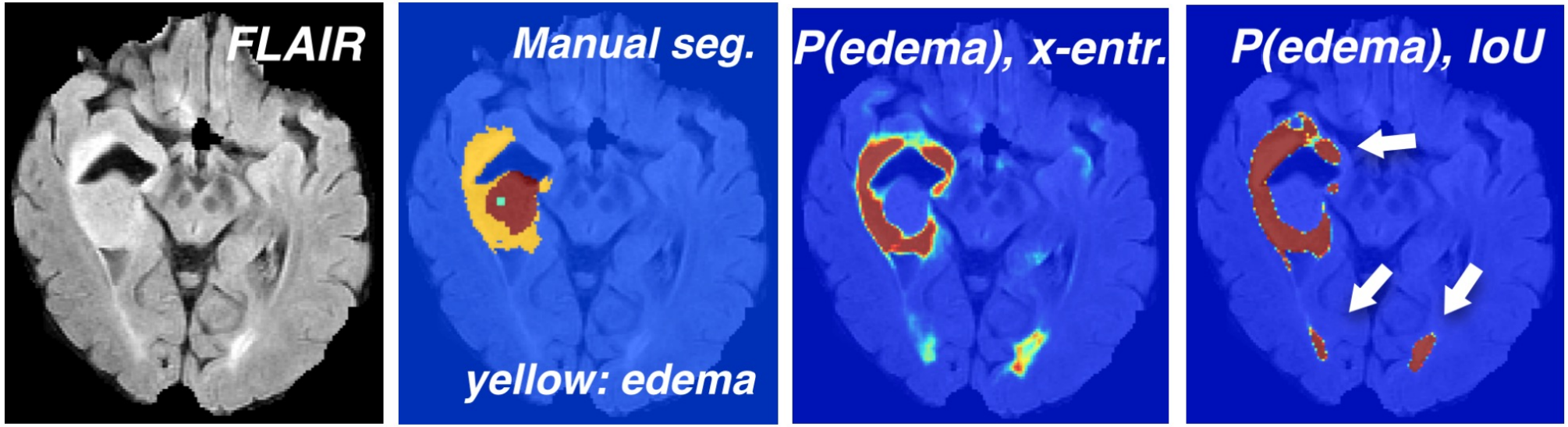
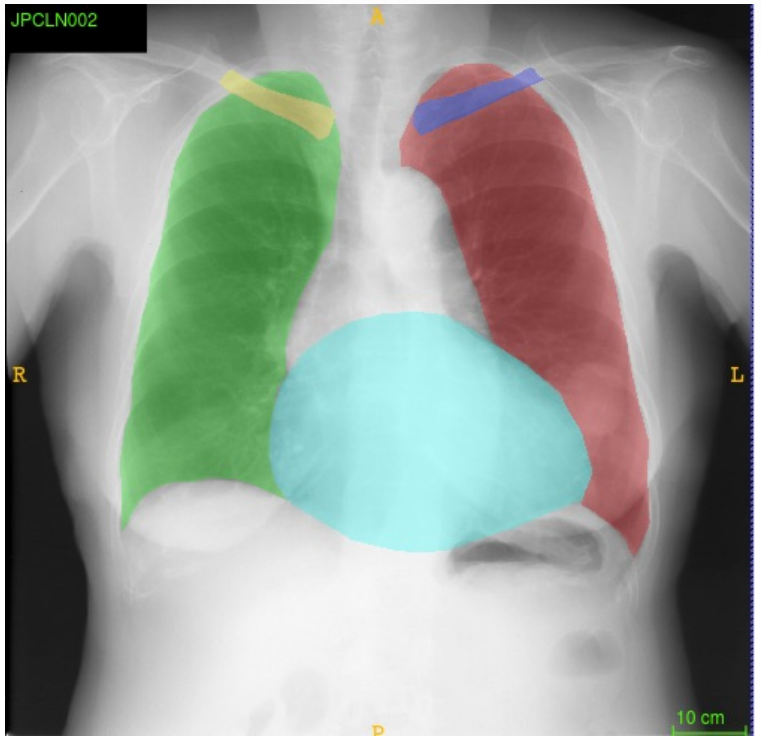
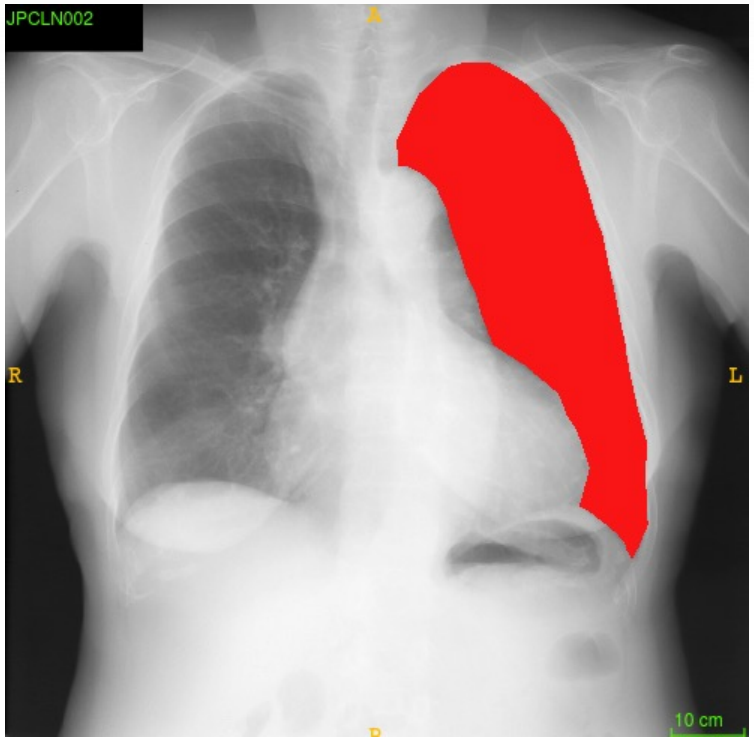


Fig. 1: Left to right: FLAIR; manual annotation of a BRATS'17 subject, where yellow depicts oedema surrounding tumour core; confidence of a CNN predicting oedema, trained with cross-entropy or IoU loss. Although overall performance is similar, training with IoU (or Dice, not shown) loss alters the CNN's behaviour, which tends to output only highly confident predictions, even when false.

Qué hacer en los casos multiclase?

Medidas de calidad en segmentación multiclase

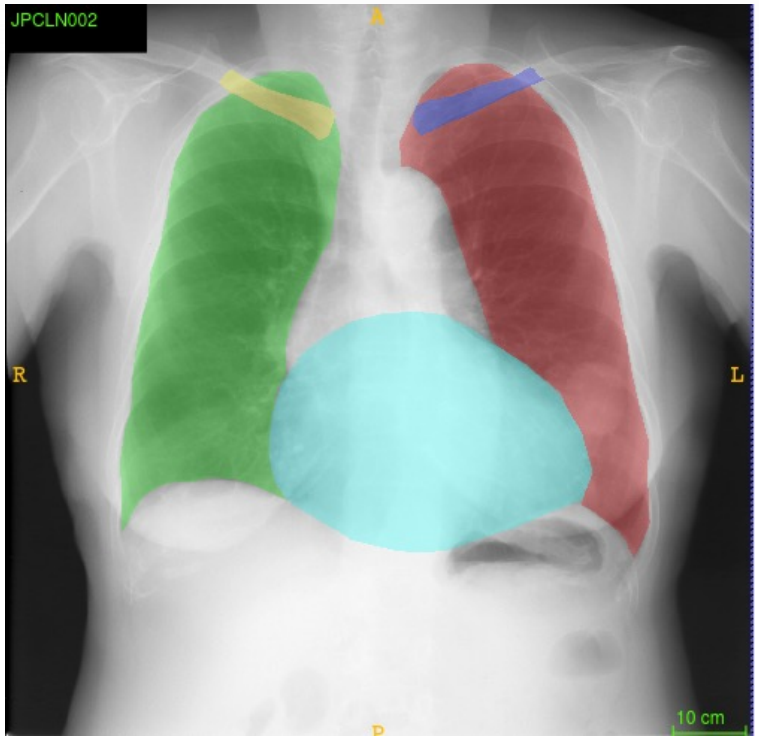


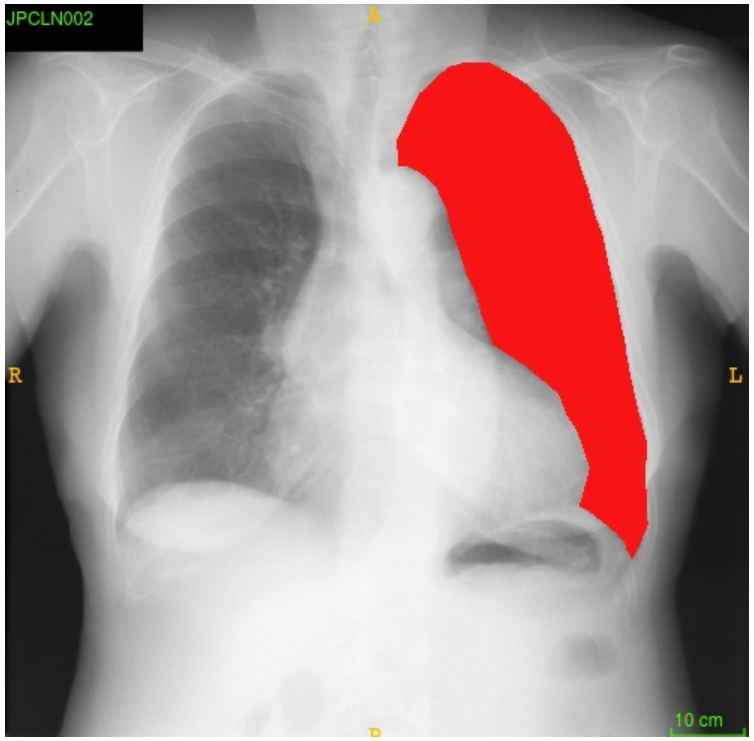


Para cada posible clase c ∈ C, se binariza el problema siguiendo:

- PredicciónBinarizada(x) = 1 si Predicción(x) = c
- PredicciónBinarizada(x) = 0 si Predicción(x) != c

Medidas de calidad en segmentación multiclase





Luego se computa una medida para cada etiqueta c ∈ C y se analiza el comportamiento por etiqueta y en promedio

Medidas de calidad en Python

```
$ pip install medpy
```

Metric measures (`medpy.metric`)

This package provides a number of metric measures that e.g. can be used for testing and/or evaluation purposes on two binary masks (i.e. measuring their similarity) or distance between histograms.

Binary metrics (`medpy.metric.binary`)

Metrics to compare binary objects and classification results.

Compare two binary objects

`dc` (result, reference)	Dice coefficient
`jc` (result, reference)	Jaccard coefficient
`hd` (result, reference[, voxelspacing, ...])	Hausdorff Distance.
`asd` (result, reference[, voxelspacing, ...])	Average surface distance metric.
`assd` (result, reference[, voxelspacing, ...])	Average symmetric surface distance.
`precision` (result, reference)	Precison.
`recall` (result, reference)	Recall.
`sensitivity` (result, reference)	Sensitivity.
`specificity` (result, reference)	Specificity.
`true_positive_rate` (result, reference)	True positive rate.
`true_negative_rate` (result, reference)	True negative rate.
`positive_predictive_value` (result, reference)	Positive predictive value.
`ravd` (result, reference)	Relative absolute volume difference.

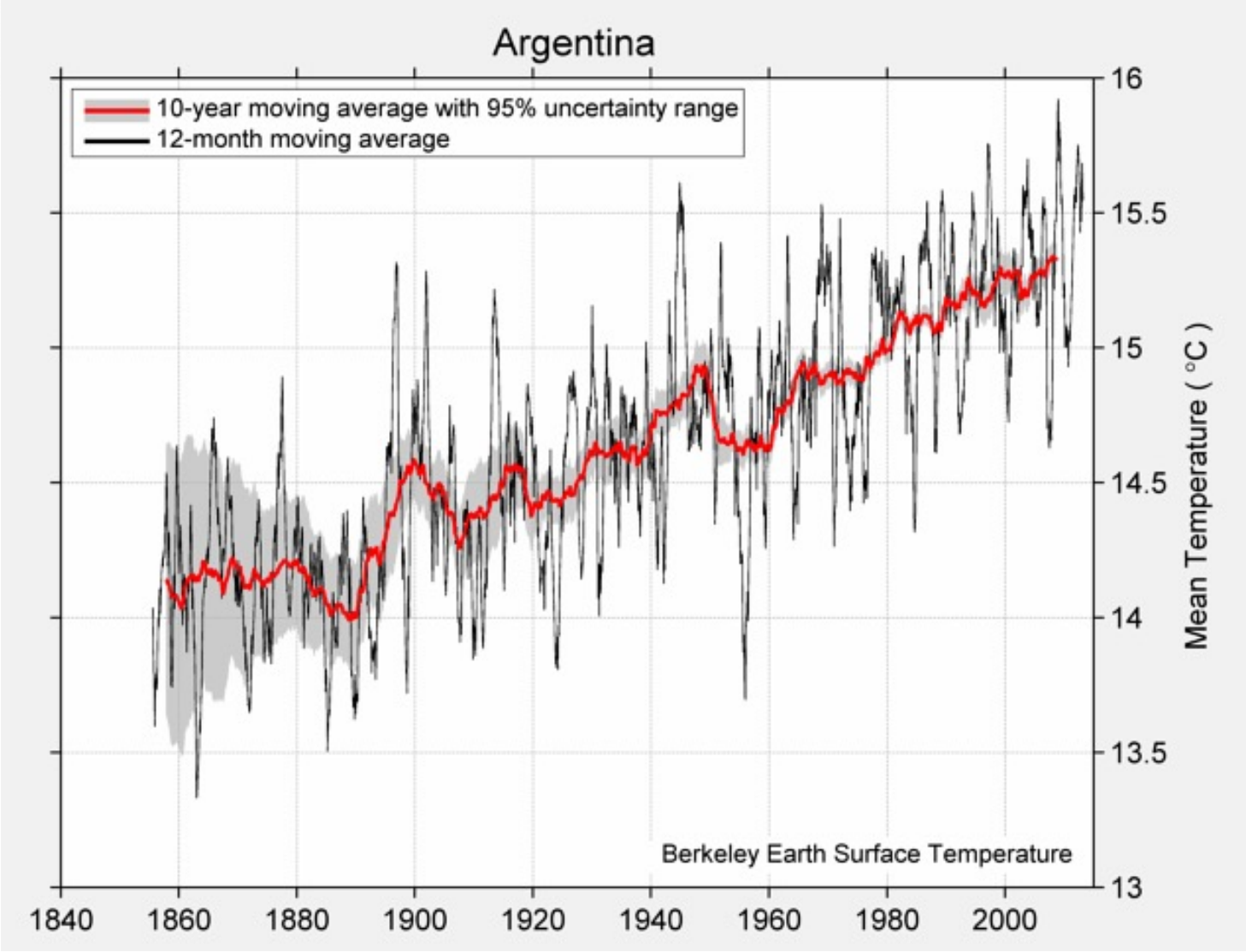
Redes neuronales convolucionales para el procesamiento de series de tiempo

Series de tiempo

Valor del dolar

Series de tiempo

Temperatura media anual en Argentina



Fuente: Berkley Earth

Series de tiempo

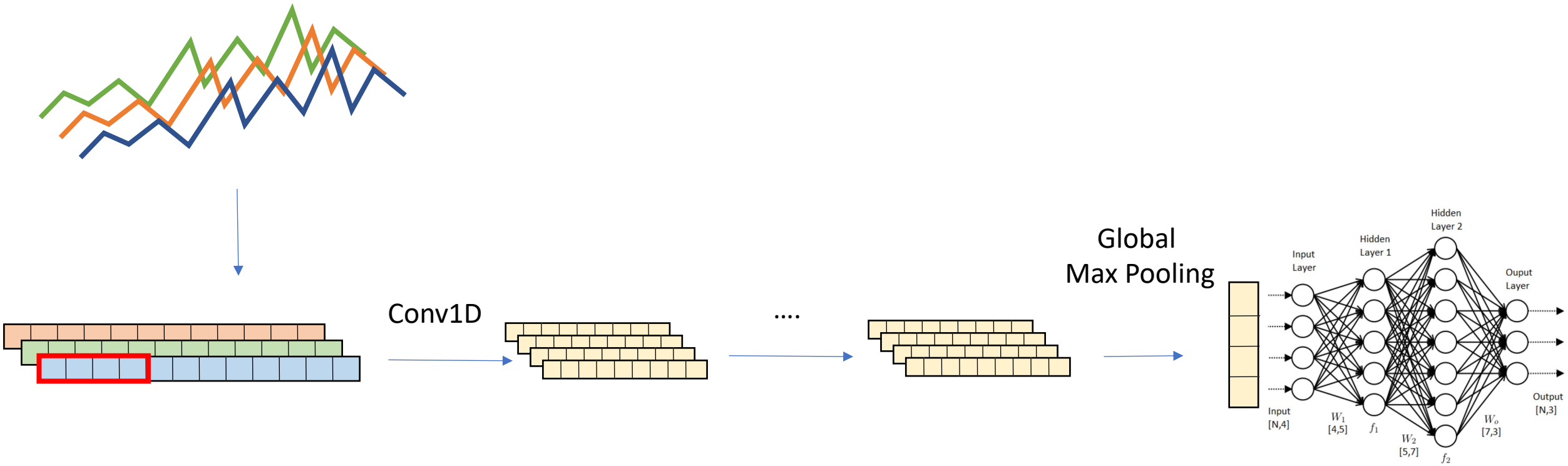
Electrocardiograma

Clasificación de series de tiempo

CNN

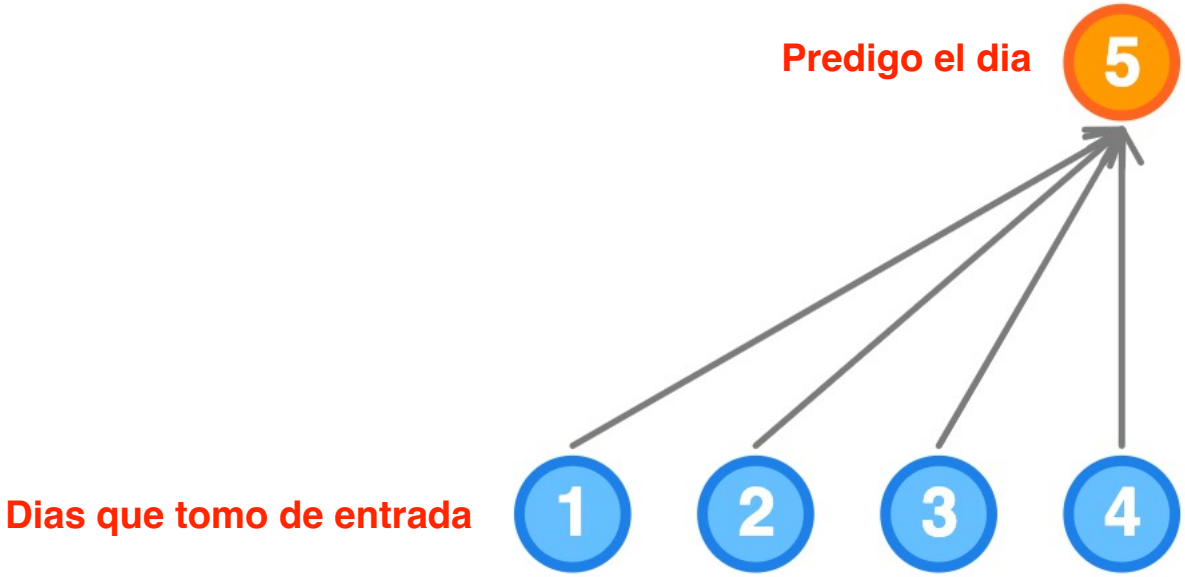
Contiene un evento anormal?

Clasificación de series de tiempo



Predicción en series de tiempo

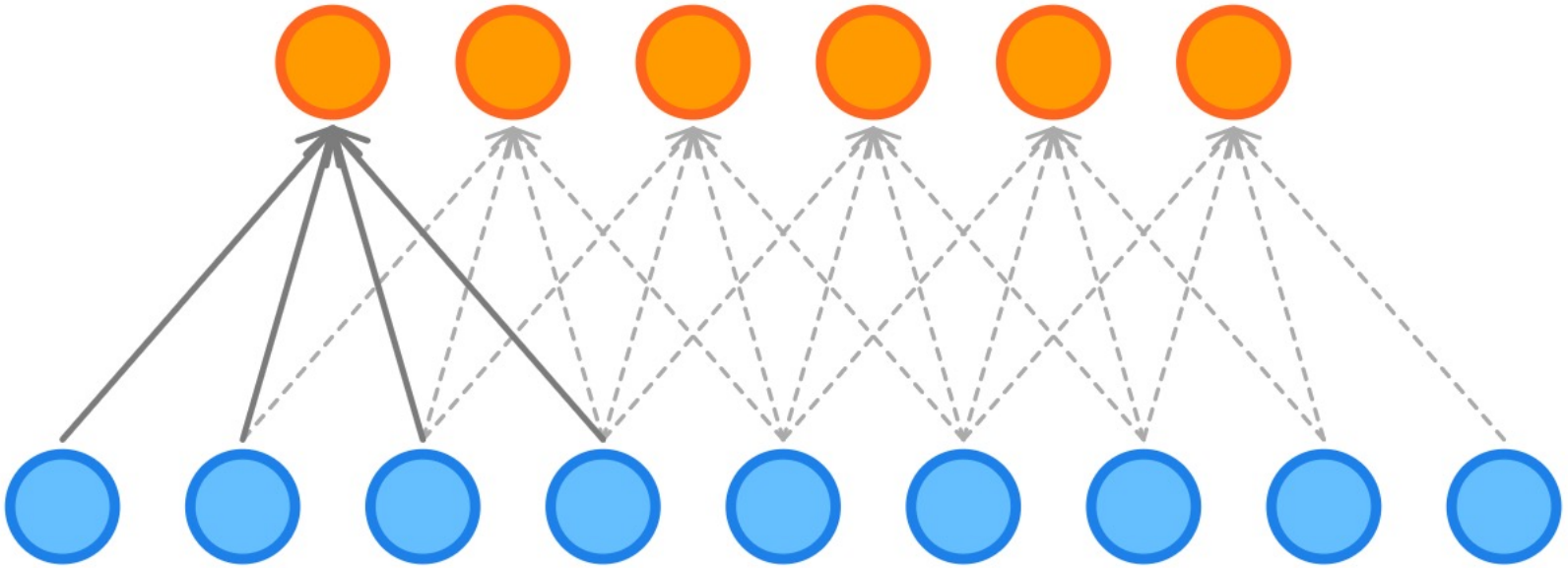
Perceptrón



CONTRAS:
- Debo utilizar la idea del sliding window, para los datos que utilizo de feature para predecir el siguiente.

Imágenes de: https://theblog.github.io/post/convolution-in-autoregressive-neural-networks/

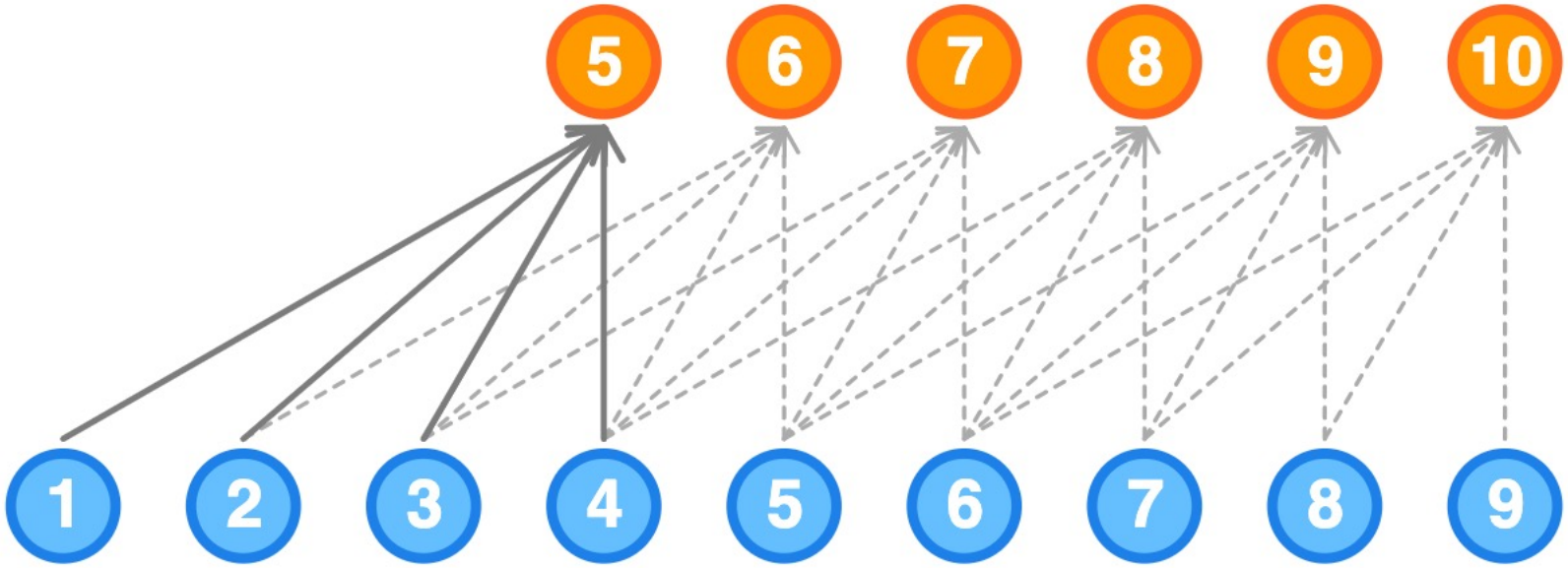
Convolución



Imágenes de: https://theblog.github.io/post/convolution-in-autoregressive-neural-networks/

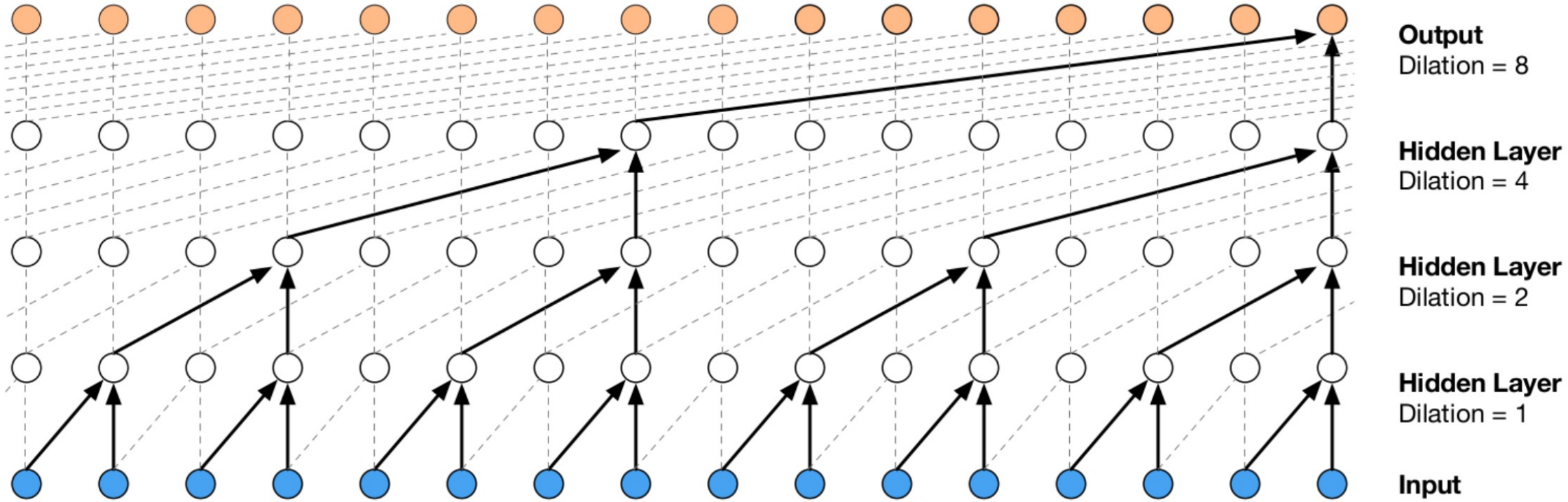
Convolución causal

- Una convolucion causal es aquella que me asegura que el dato que predigo es un dato futuro, que aun no ocurrio (“una convencion”).
- Al utilizar una convolucion, el “sliding window” o ventaneo que necesito, lo obtengo “gratis”.
- me voy moviendo y prediciendo dia a dia



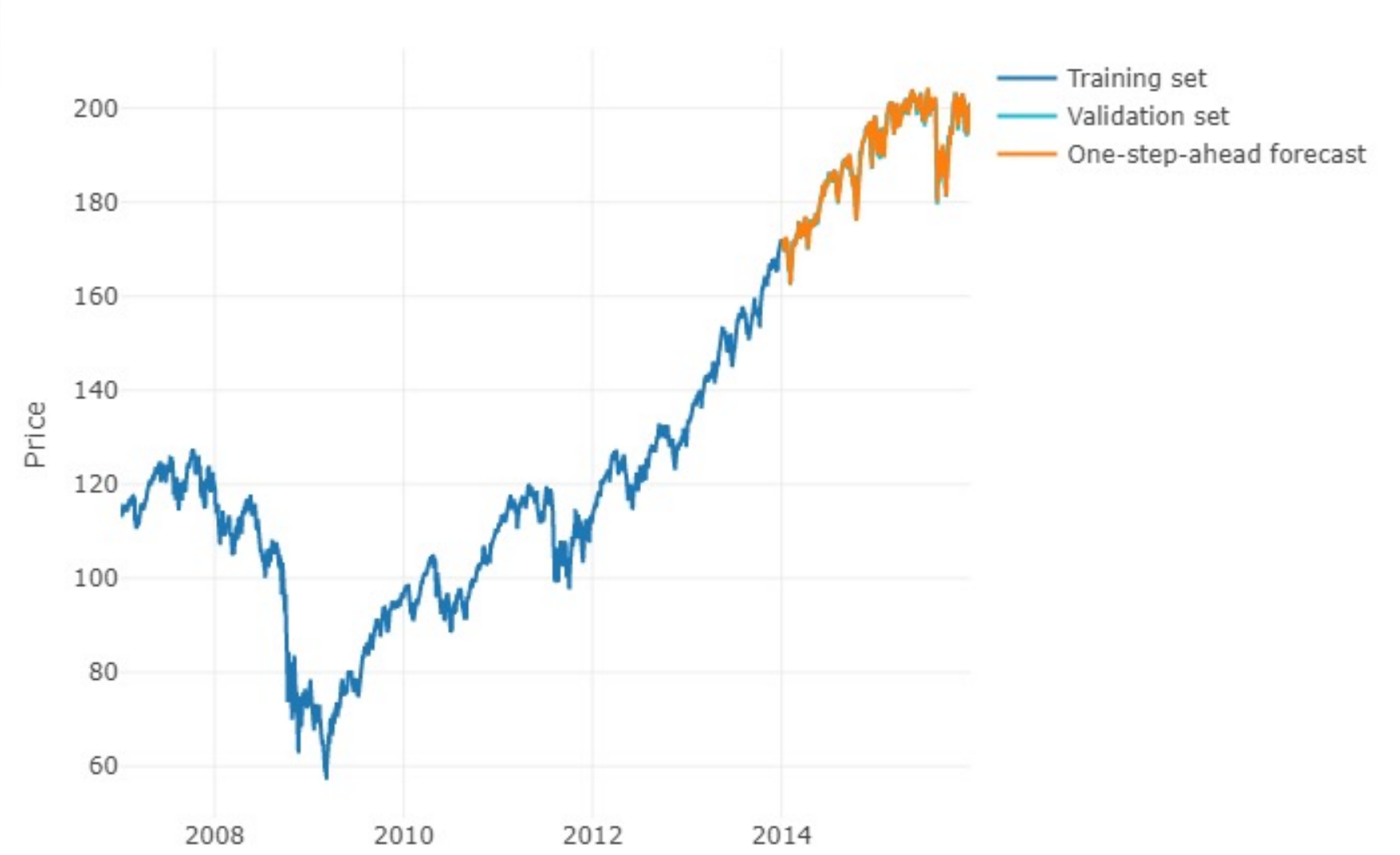
Imágenes de: https://theblog.github.io/post/convolution-in-autoregressive-neural-networks/

Predicción en series de tiempo



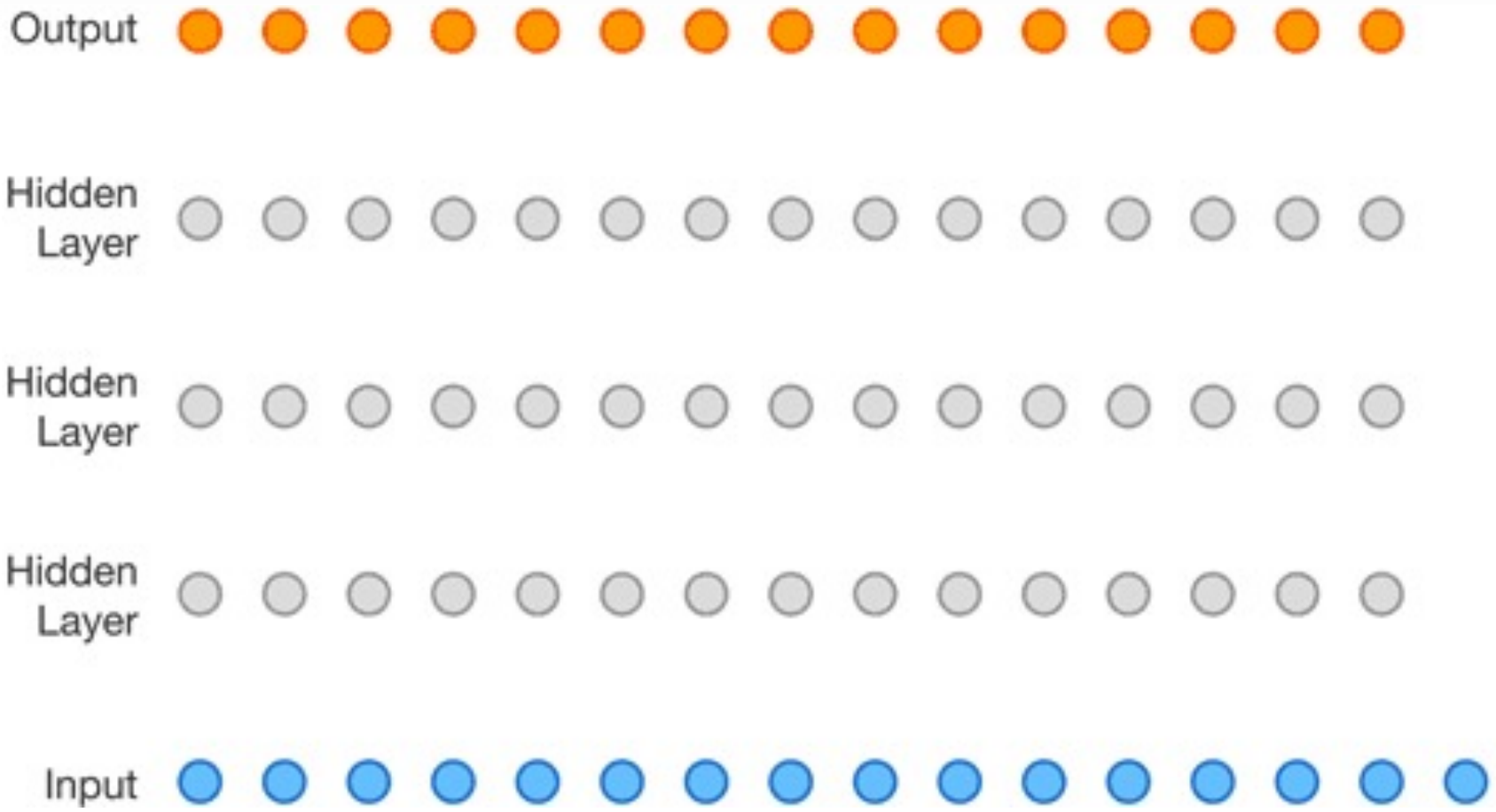
Imágenes de: https://theblog.github.io/post/convolution-in-autoregressive-neural-networks/

Predicción en series de tiempo: más allá de un paso



Predicción en series de tiempo: más allá de un paso

WaveNET, se usa para predecir los acordes siguientes y predecir asi musica



Imágenes de: https://theblog.github.io/post/convolution-in-autoregressive-neural-networks/

WaveNet

WaveNet: A Generative Model for Raw Audio

Aäron van den Oord
Sander Dieleman
Heiga Zen[†]
Karen Simonyan
Oriol Vinyals
Alex Graves
Nal Kalchbrenner
Andrew Senior
Koray Kavukcuoglu

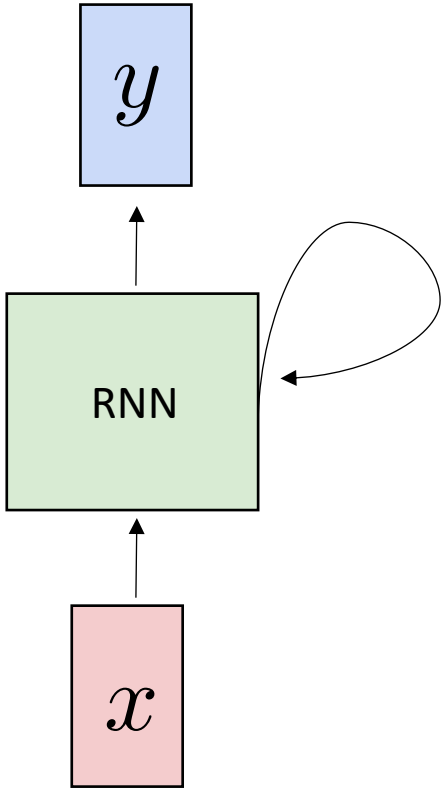
{avdnoord, sedielem, heigazen, simonyan, vinyals, gravesa, nalk, andrewsenior, korayk}@google.com
Google DeepMind, London, UK
† Google, London, UK

Abstract

This paper introduces WaveNet, a deep neural network for generating raw audio waveforms. The model is fully probabilistic and autoregressive, with the predictive distribution for each audio sample conditioned on all previous ones; nonethe-
less we show that it can be efficiently trained on data with tens of thousands of
nd of audio. When applied to text-to-speech, it yields state-of-
listeners rating it as significantly more natural
tive systems for both English and
f many different

Redes recurrentes

Redes neuronales recurrentes

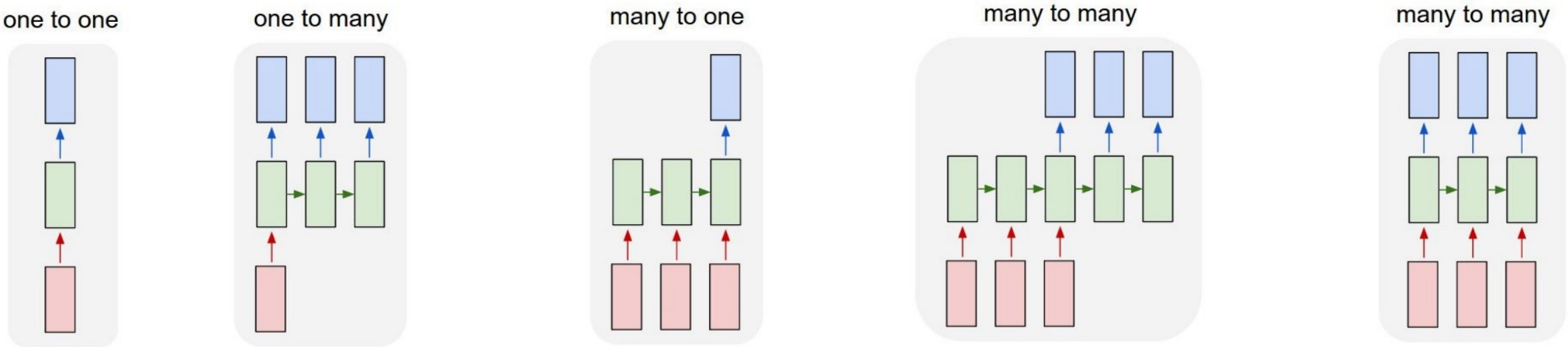


- Redes neuronales diseñadas para procesar datos secuenciales $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(\tau)}$
- Así como las redes totalmente convolucionales pueden escalar a cualquier tamaño de imagen, las redes recurrentes pueden en principio procesar cualquier longitud de secuencia.
- La clave son los pesos compartidos (weight sharing) a través del tiempo

Redes neuronales recurrentes

Ej: Sentiment analysis
Secuencia de palabras → Estado sentimental

Ej: Clasificación de acciones
Video frames → Secuencia de acciones

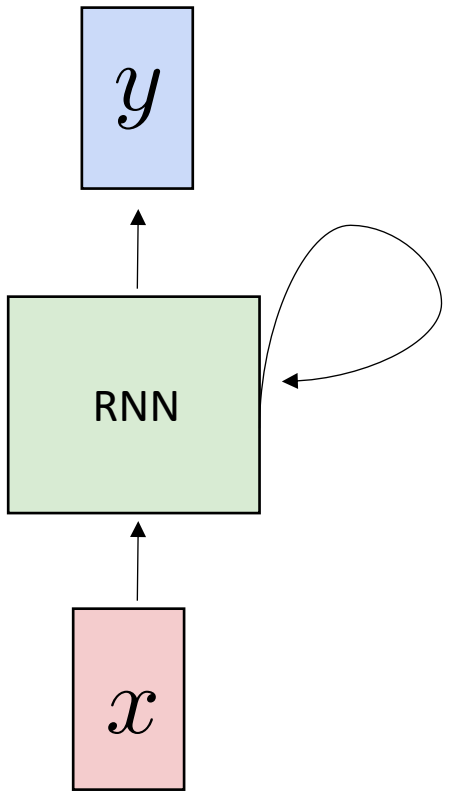


Ej: Image Captioning
Imagen → Secuencia de palabras

Ej: Traducción de textos
Secuencia de palabras → Secuencia de palabras

Basado en: http://cs231n.stanford.edu/

Redes neuronales recurrentes



Función parametrizada por W (la misma para todos los instantes *t*)

$$\boxed{h_t} = \boxed{f_W}(\boxed{h_{t-1}}, \boxed{x_t})$$

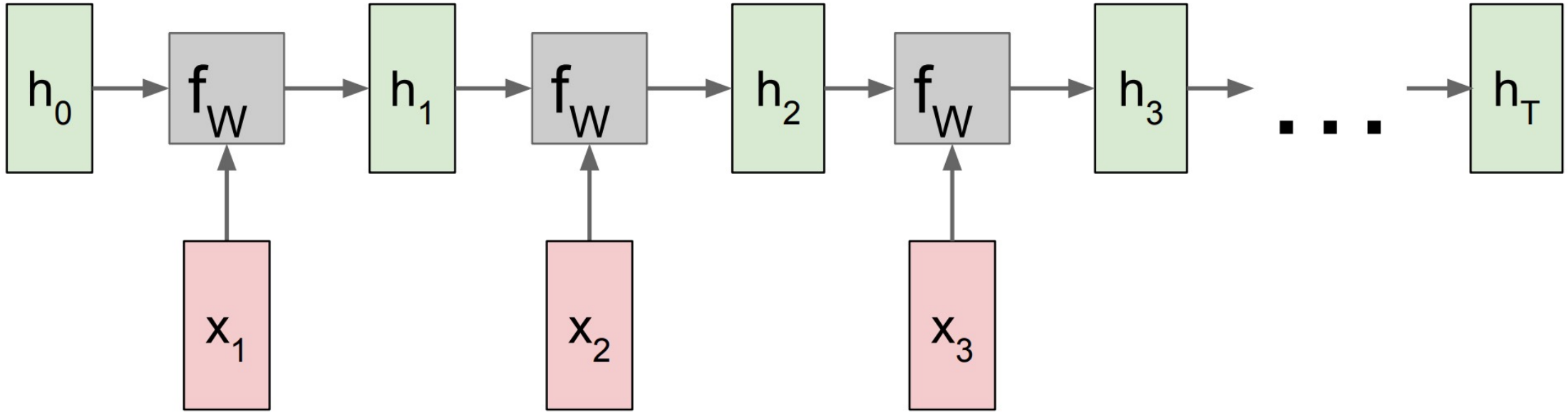
Nuevo Estado — Antiguo Estado — Entrada

$$h_t = tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

$$y_t = W_{hy}h_t$$

Basado en: http://cs231n.stanford.edu/

Grafo de cómputo

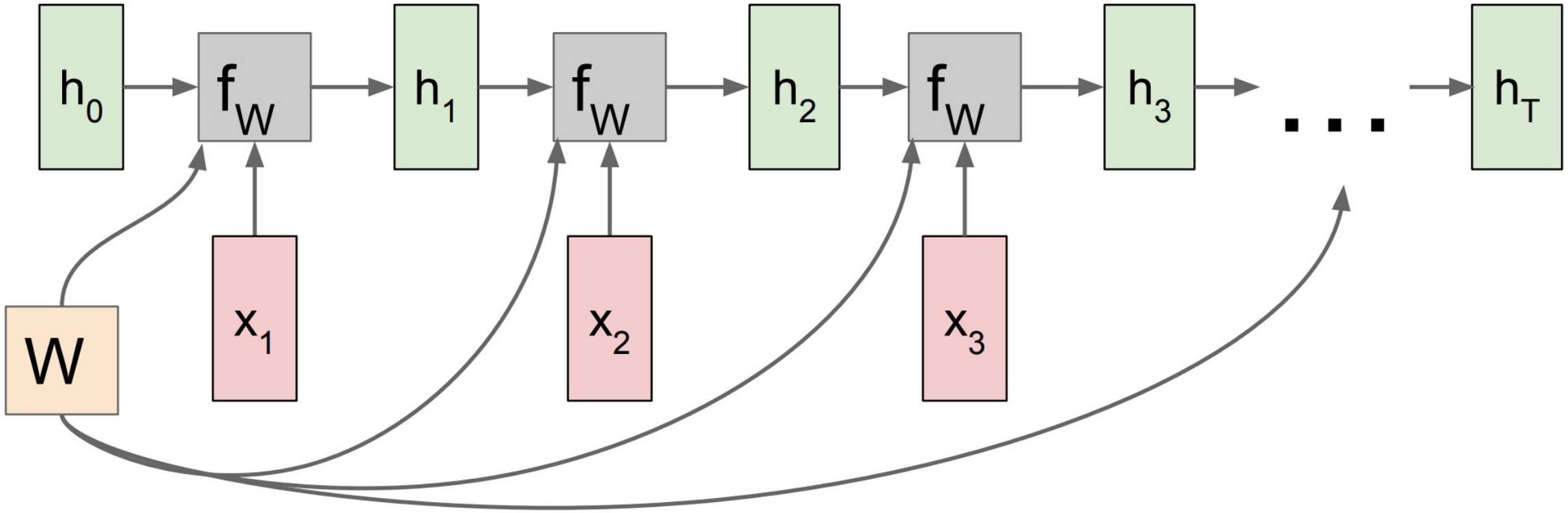


Recurrent Neural Network is a generalization of feed-forward neural network that has an internal memory. RNN is recurrent in nature as it performs the same function for every input of data while the output of the current input depends on the past one computation. After producing the output, it is copied and sent back into the recurrent network. For making a decision, it considers the current input and the output that it has learned from the previous input.

Unlike feed-forward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition. In other neural networks, all the inputs are independent of each other. But in RNN, all the inputs are related to each other.

Basado en: http://cs231n.stanford.edu/

Grafo de cómputo

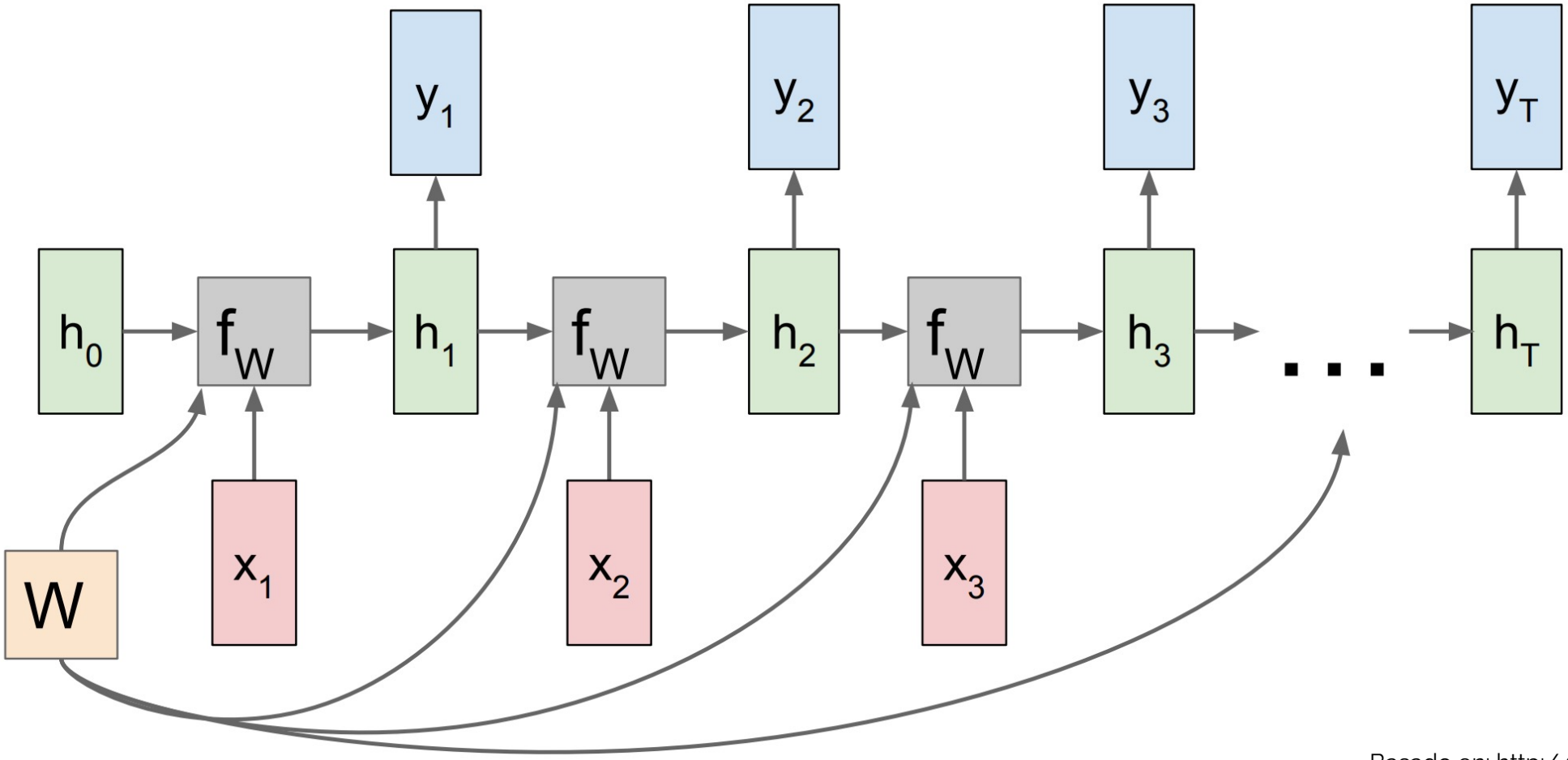


Retropropagación a través del tiempo

Basado en: http://cs231n.stanford.edu/

Grafo de cómputo

Many to many

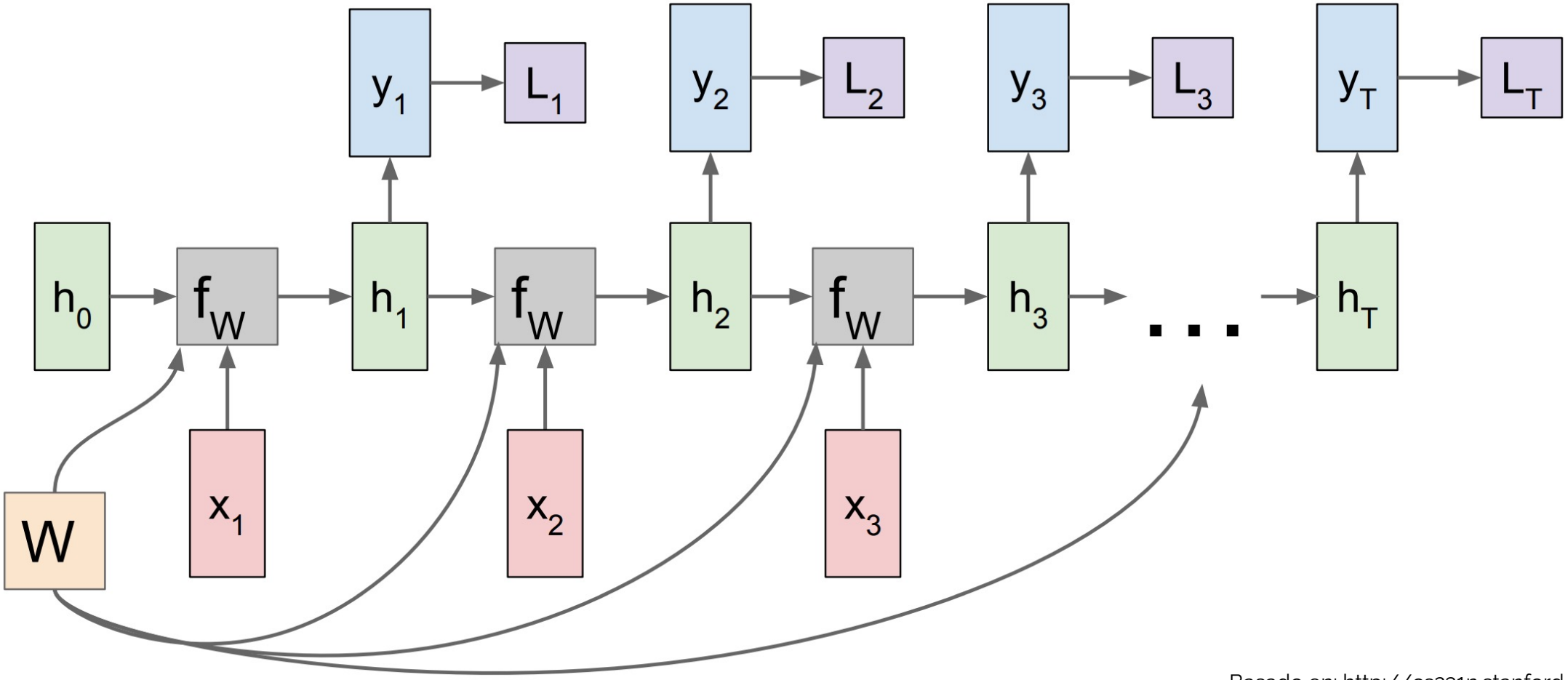


Basado en: http://cs231n.stanford.edu/

Grafo de cómputo

Many to many

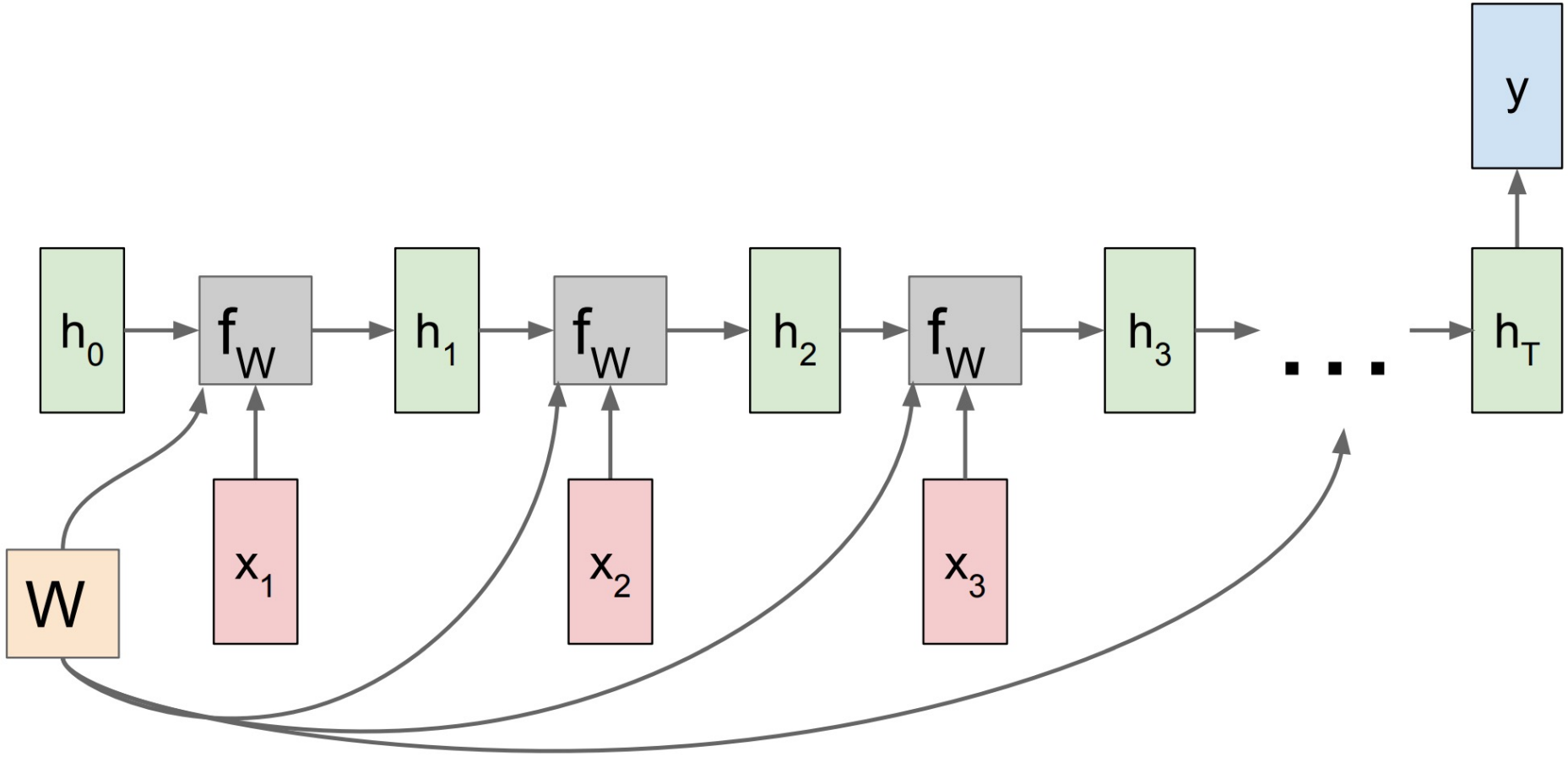
La función de pérdida la computamos como la suma de las Li



Basado en: http://cs231n.stanford.edu/

Grafo de cómputo

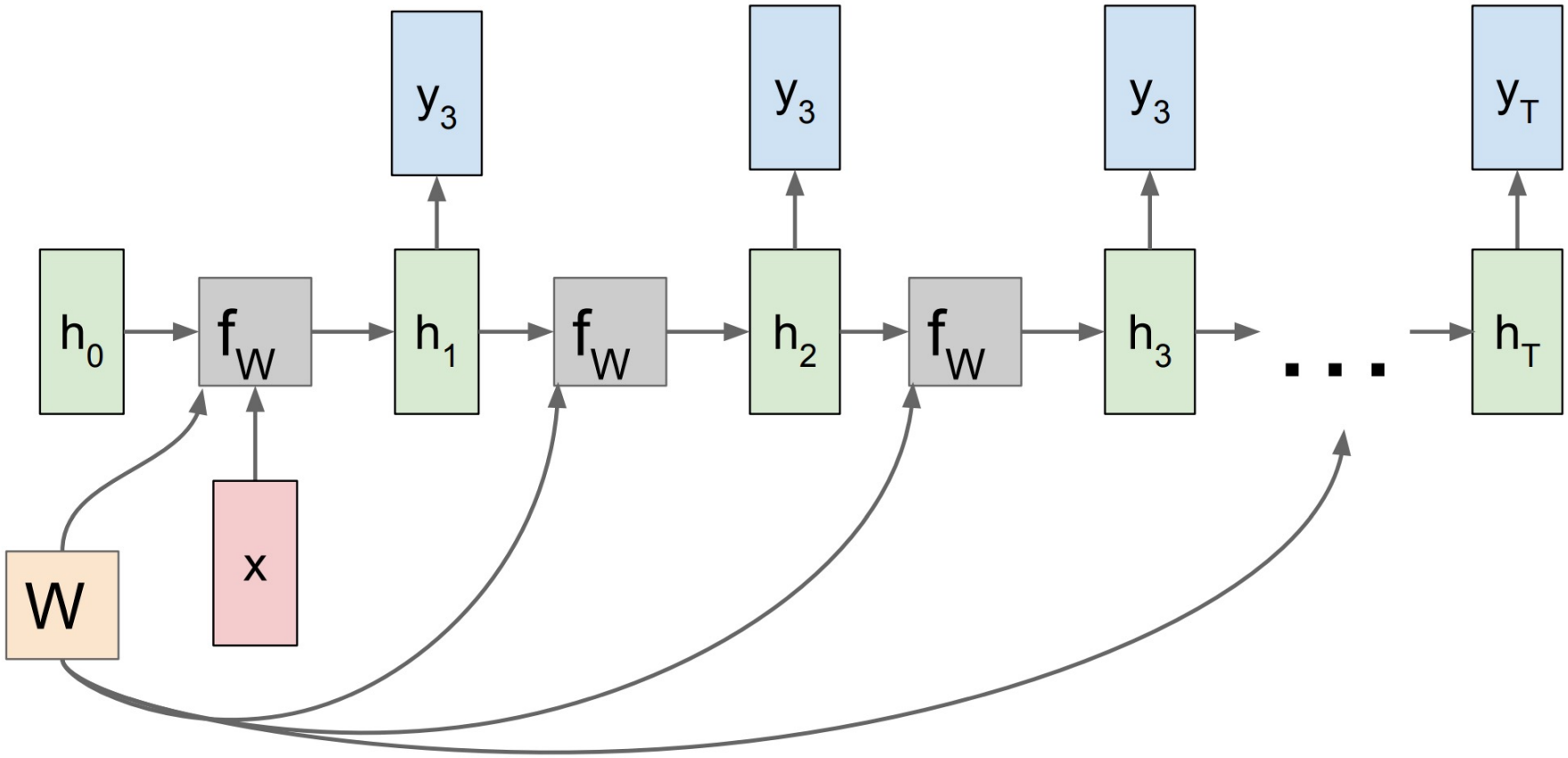
Many to one



Basado en: http://cs231n.stanford.edu/

Grafo de cómputo

One to many



Basado en: http://cs231n.stanford.edu/

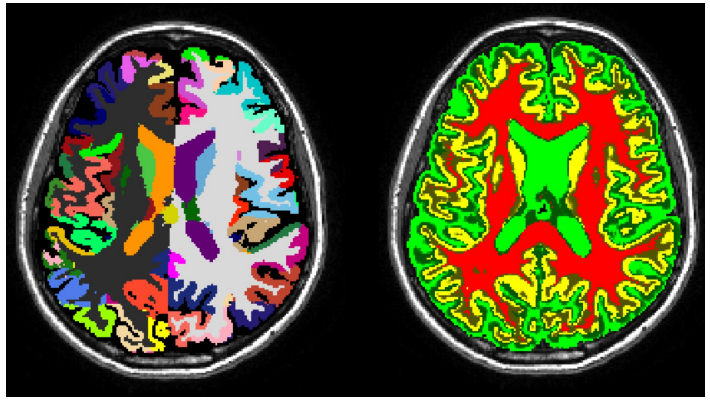
Clase 6

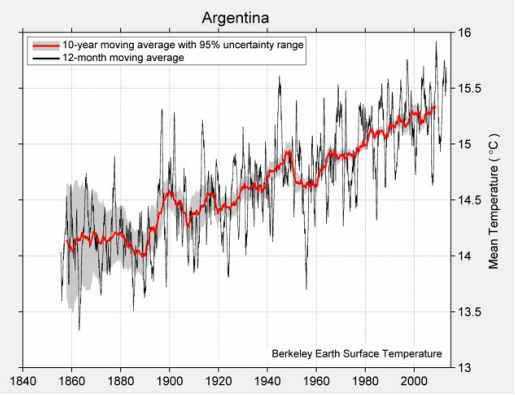
Redes neuronales más allá de la clasificación de imágenes

Enzo Ferrante

eferrante@sinc.unl.edu.ar

@enzoferrante





Universidad de SanAndrés

UNL CONICET

s i n c (i)