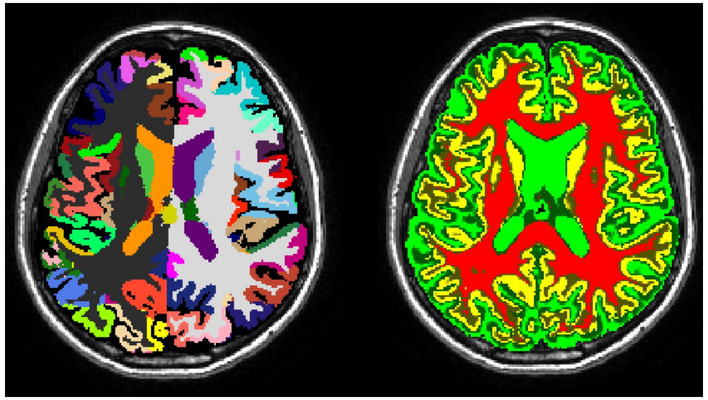
Clase 5

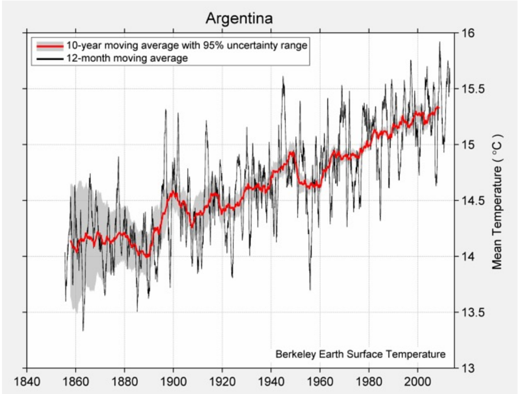
Redes neuronales más allá de la clasificación de imágenes

Enzo Ferrante

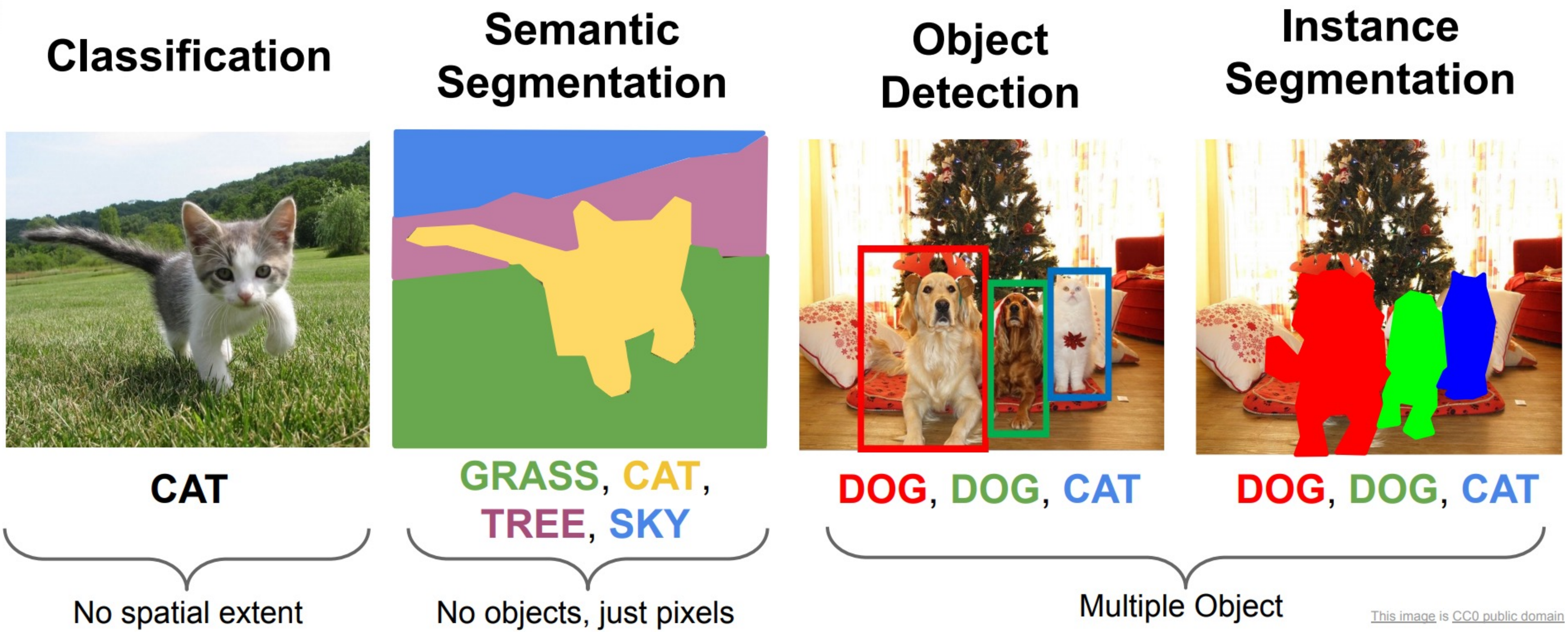
eferrante@sinc.unl.edu.ar

@enzoferrante



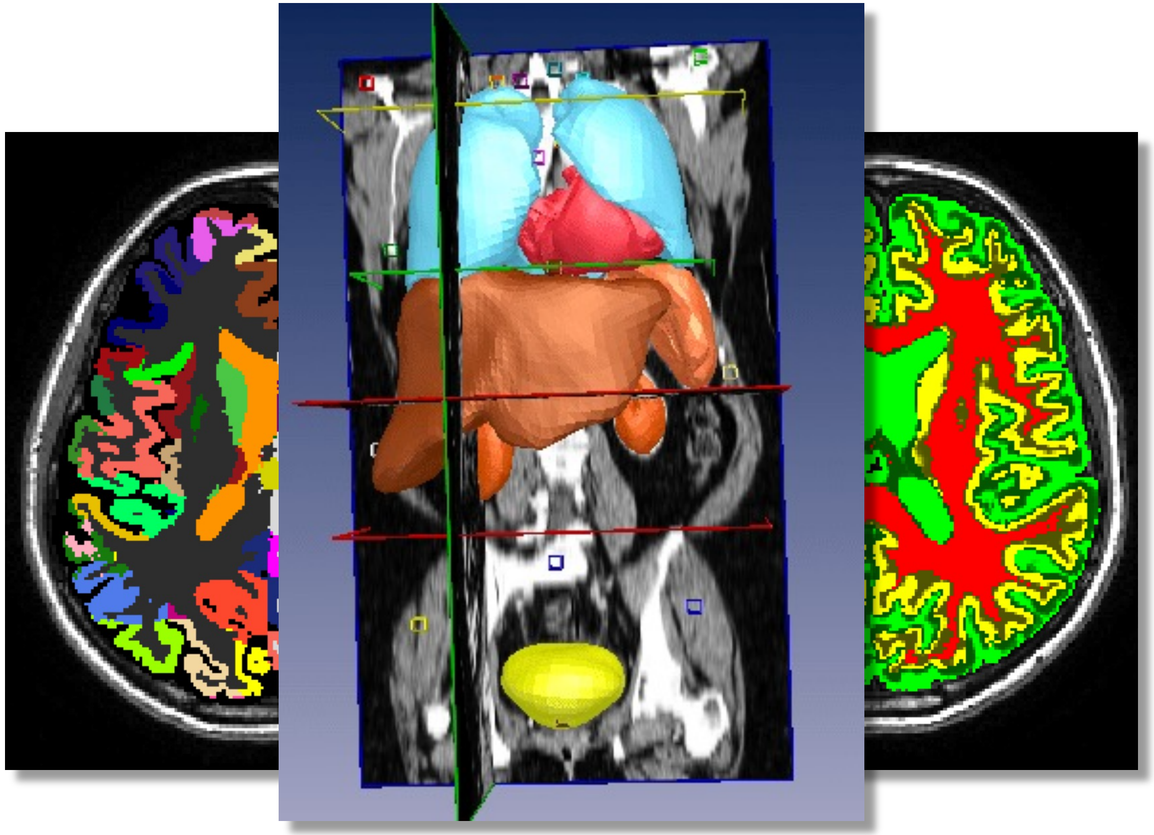


Problemas en computer vision



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

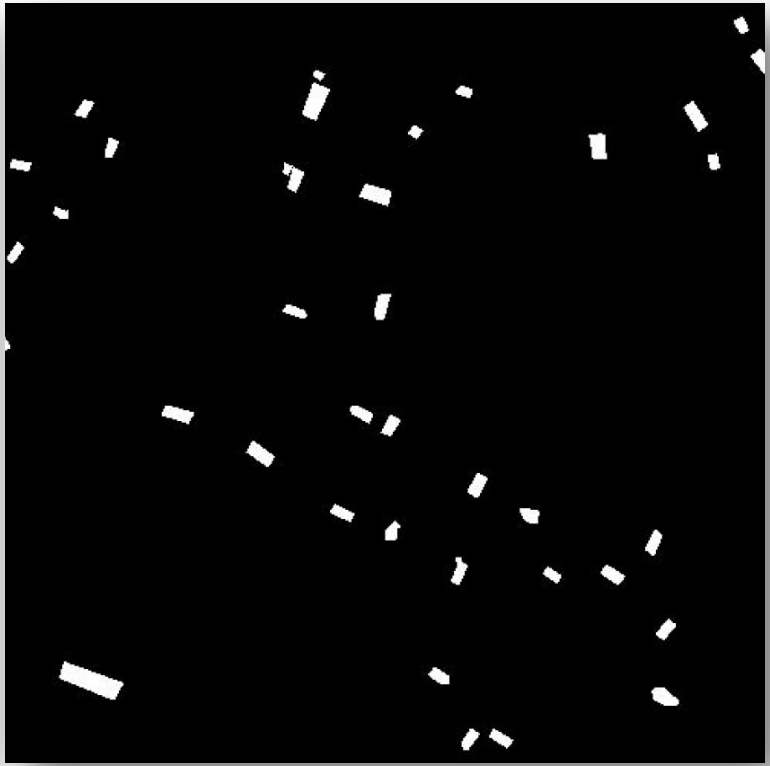
Segmentación de imágenes médicas



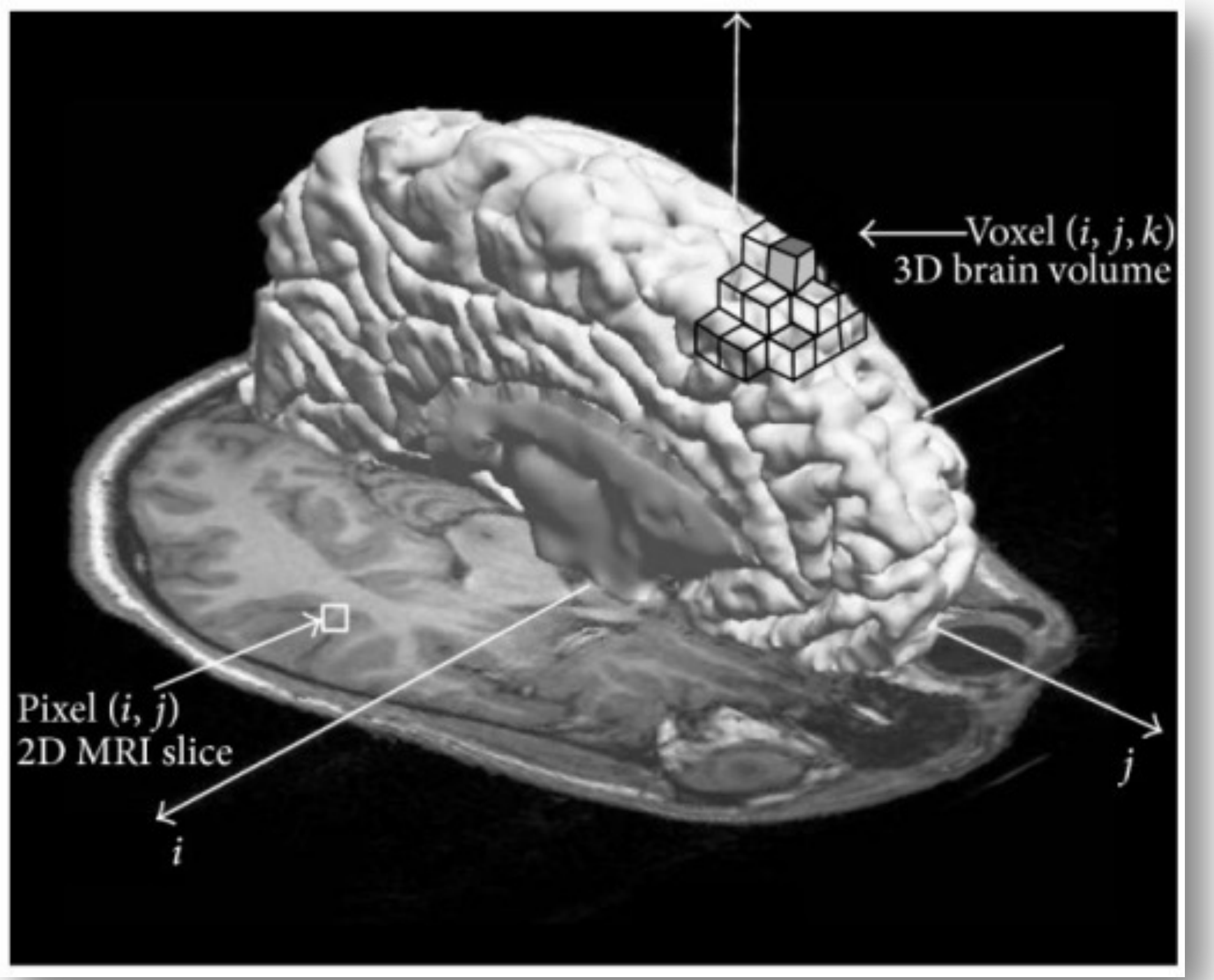
Estructuras anatómicas

Estructuras patológicas

Segmentación de imágenes satelitales



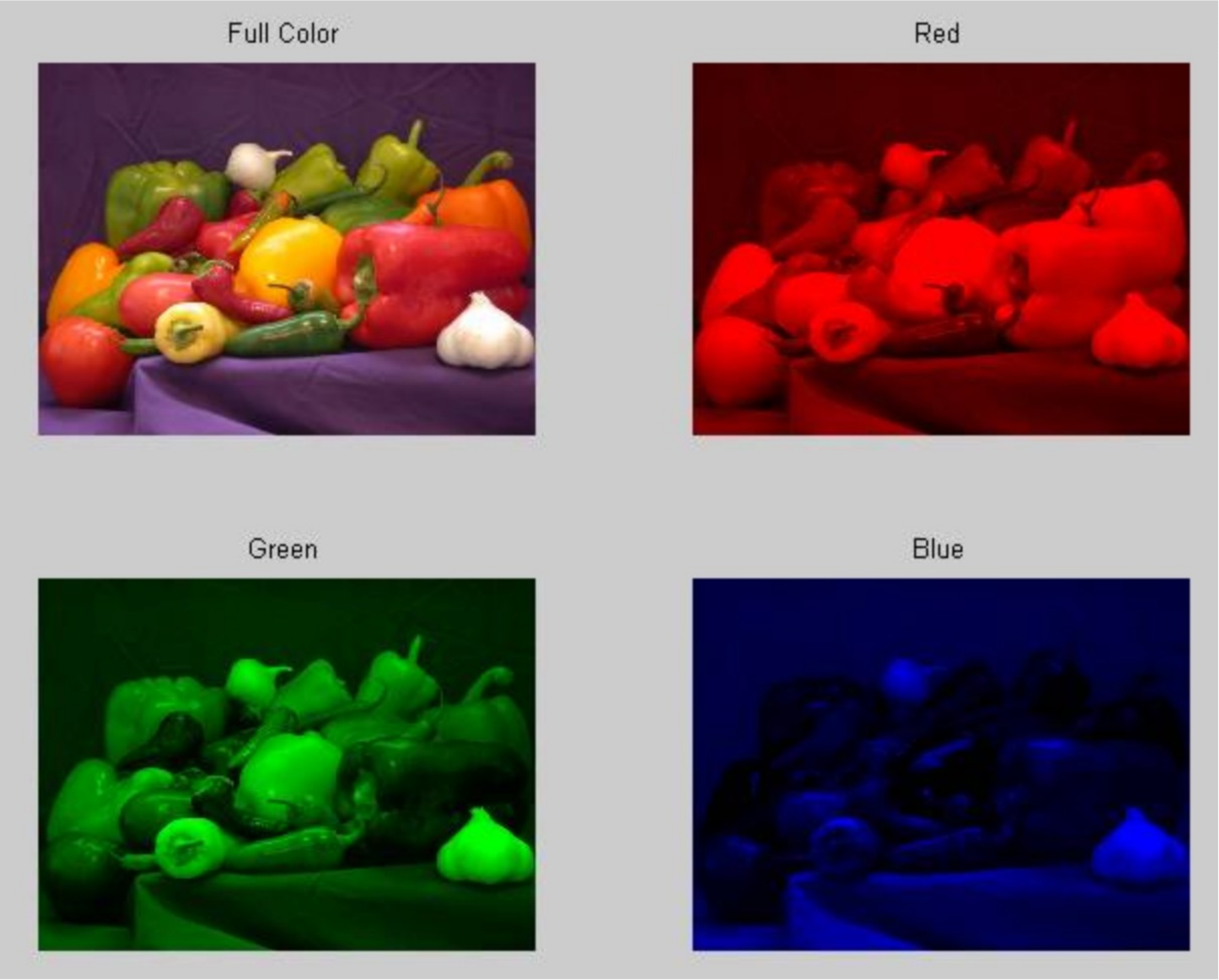
Imágenes 2D vs Imágenes 3D



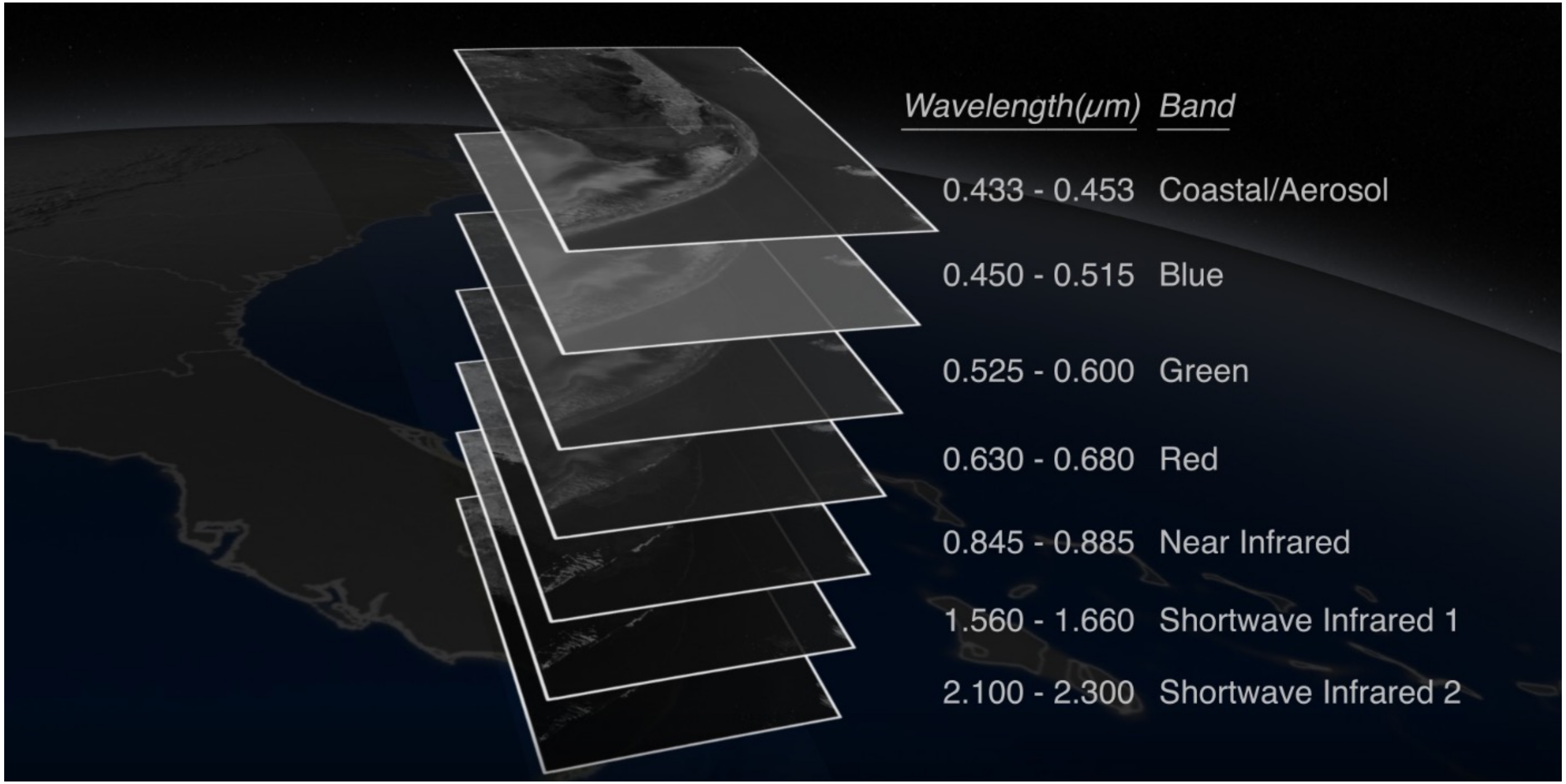
- Las imágenes pueden ser procesadas en 2D o en 3D.

- Diferencia espacio pixel y físico.

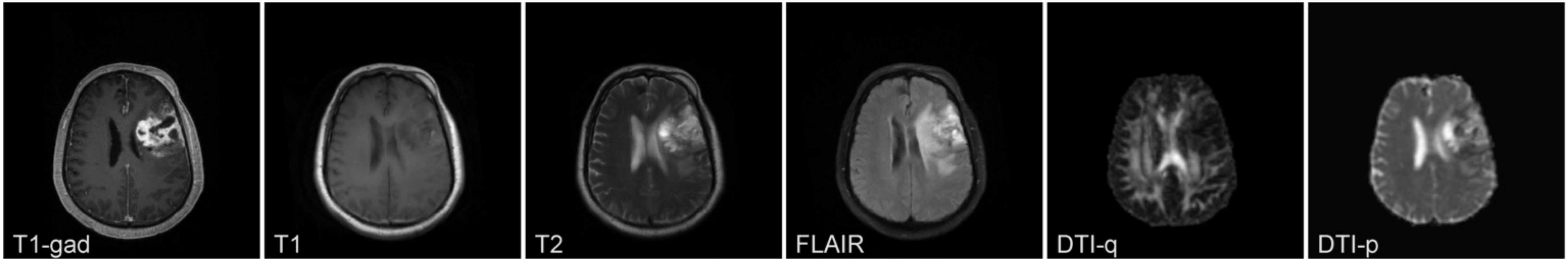
Múltiples canales de entrada



Múltiples canales de entrada

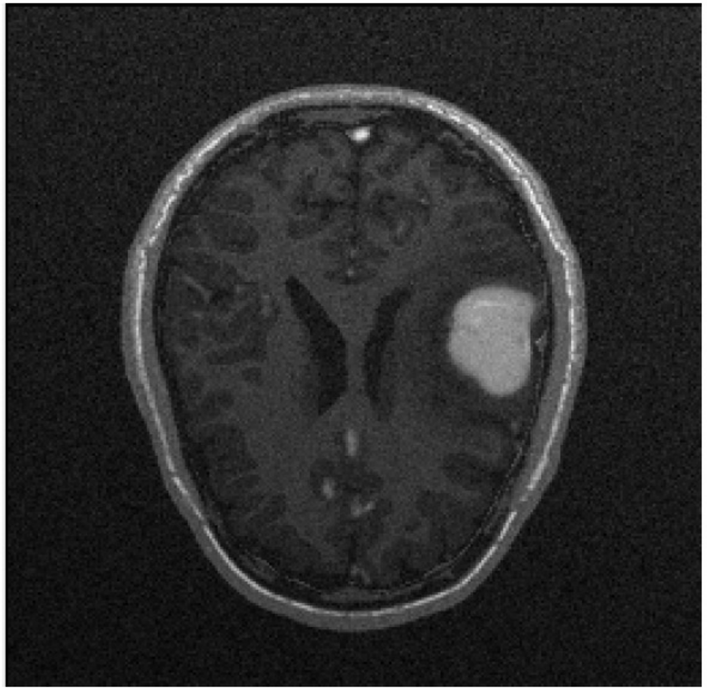


Múltiples canales de entrada



Técnicas clásicas de segmentación de imágenes

Umbralado



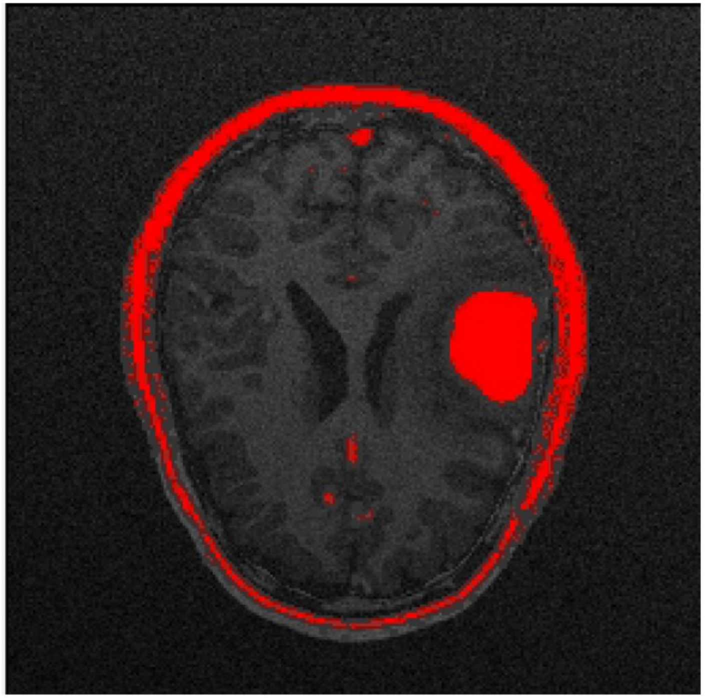


Imagen extraida de Ferrante, 2012

Crecimiento de regiones

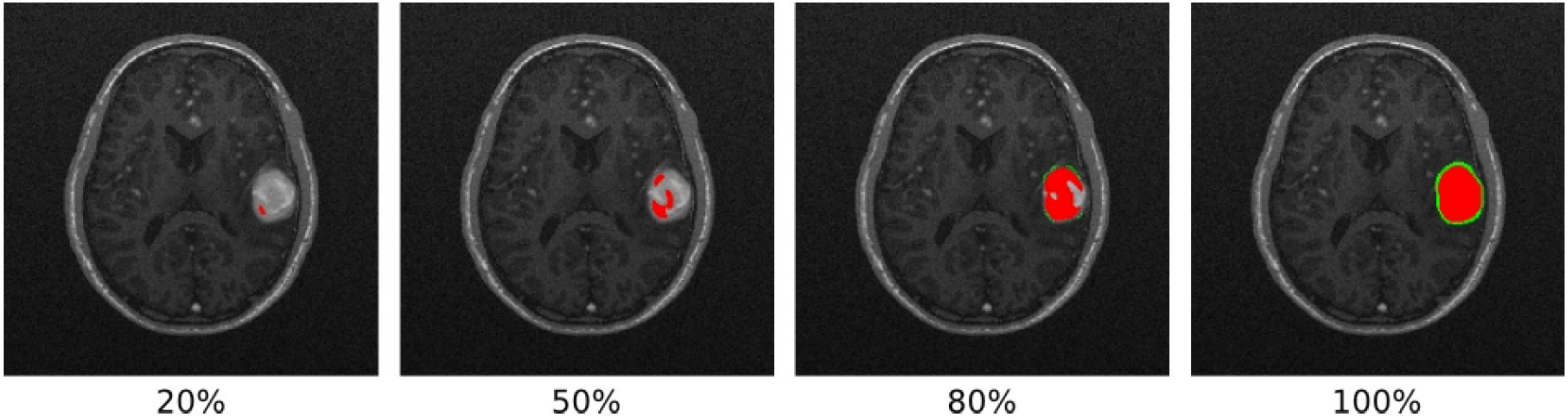


Imagen extraida de Ferrante, 2012

Watershed

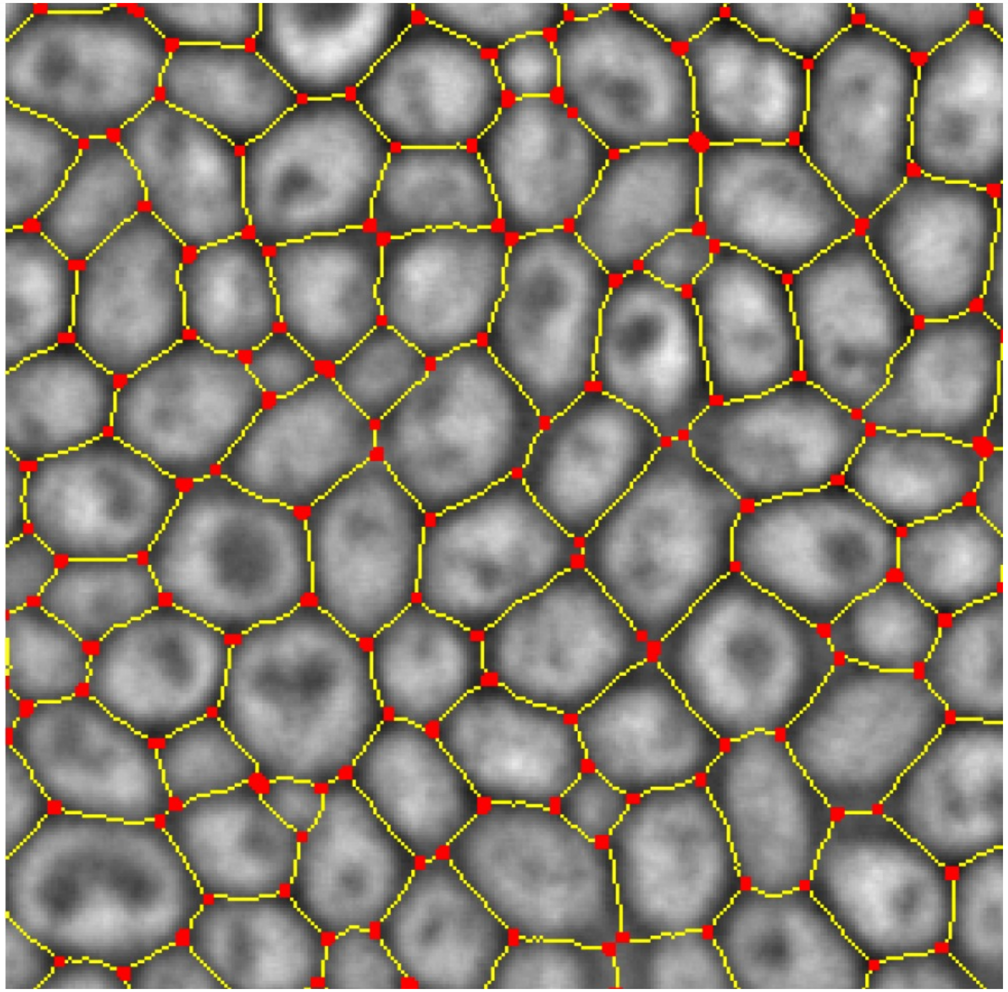
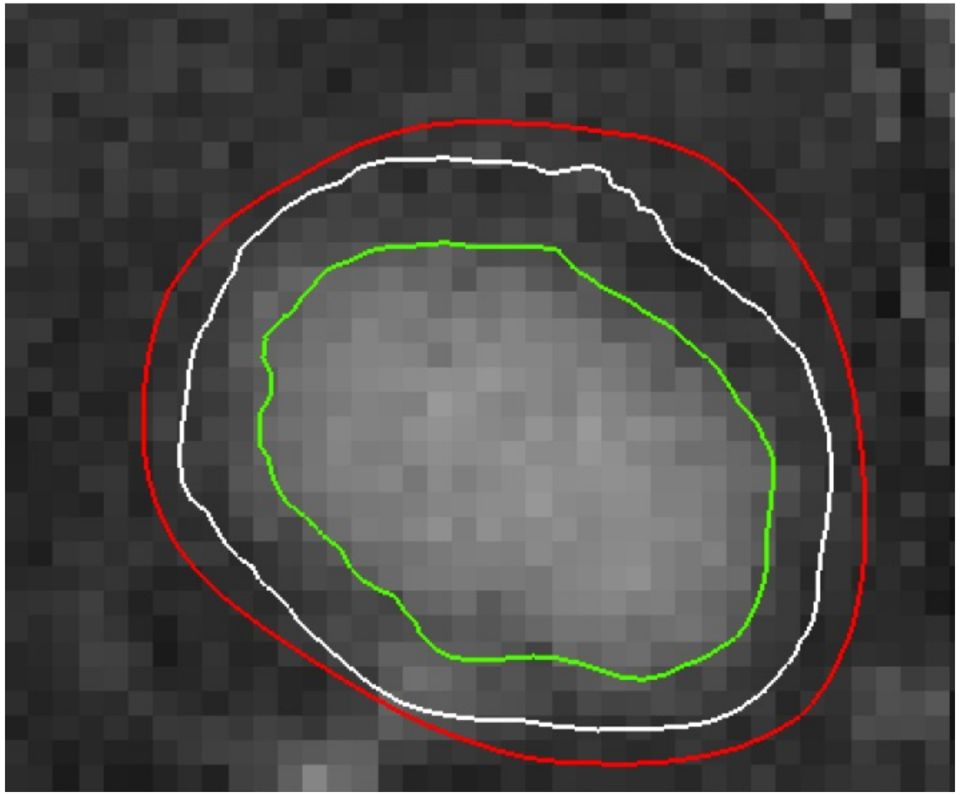
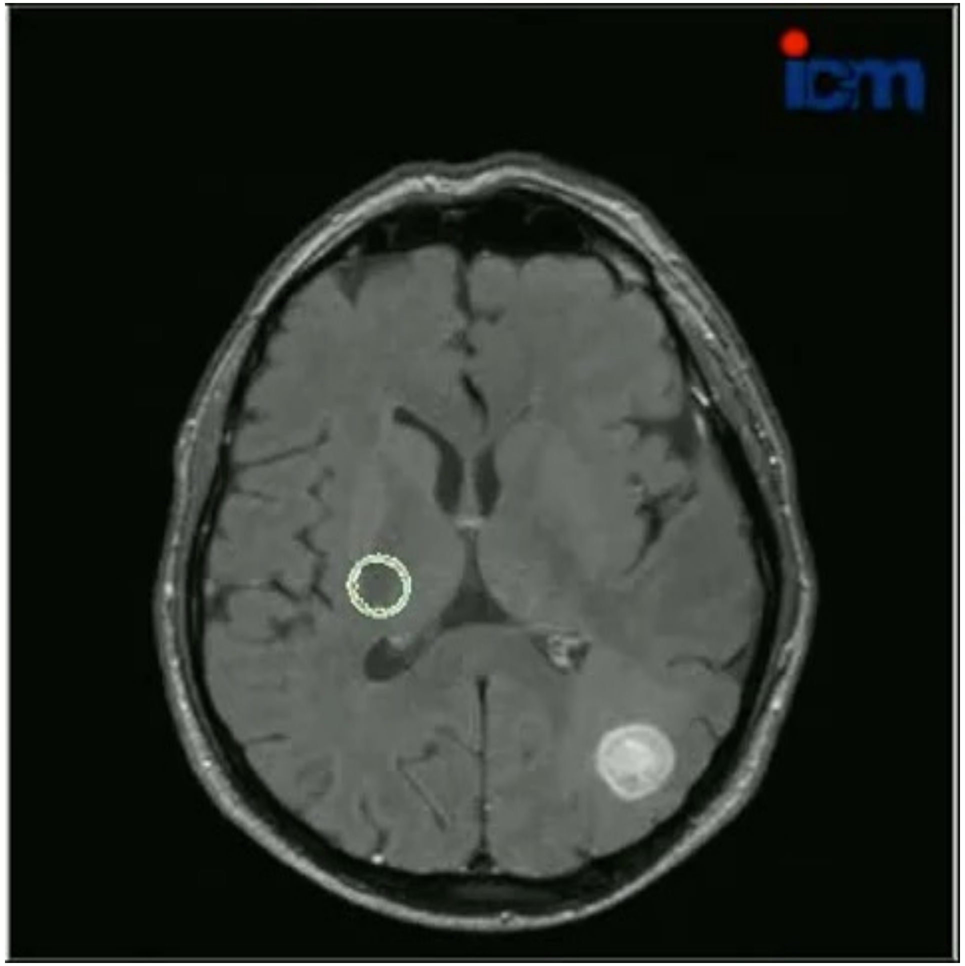


Imagen extraida de http://biodynamics.ucsd.edu/ir

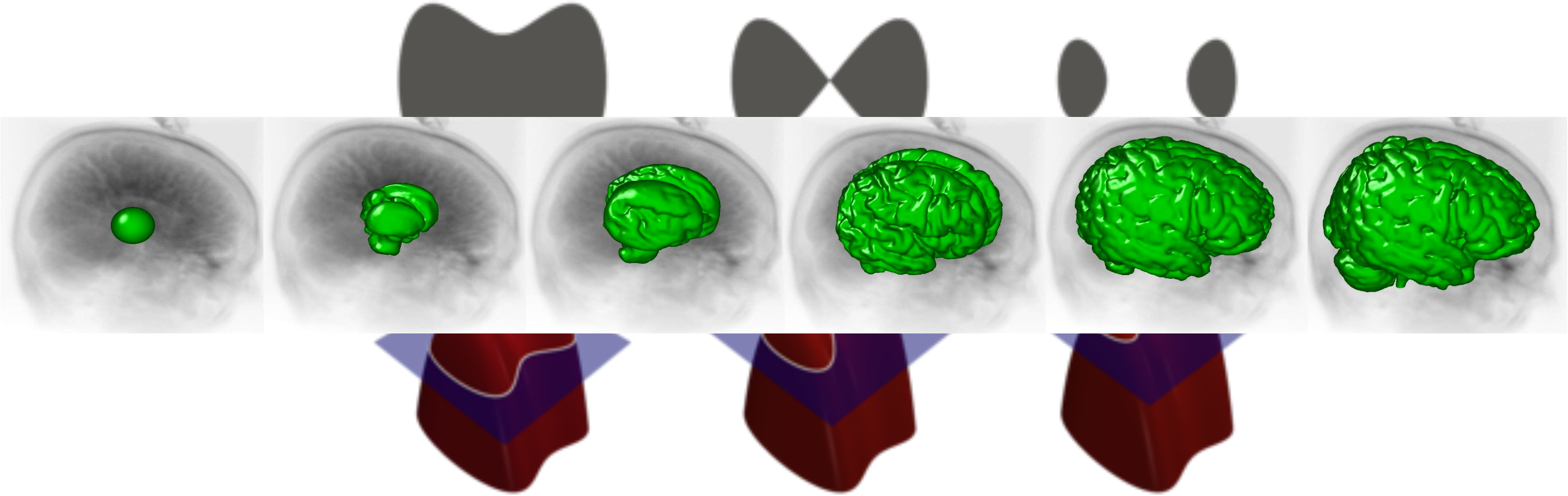
Modelos deformables explícitos

Ej. Contornos activos o Snakes

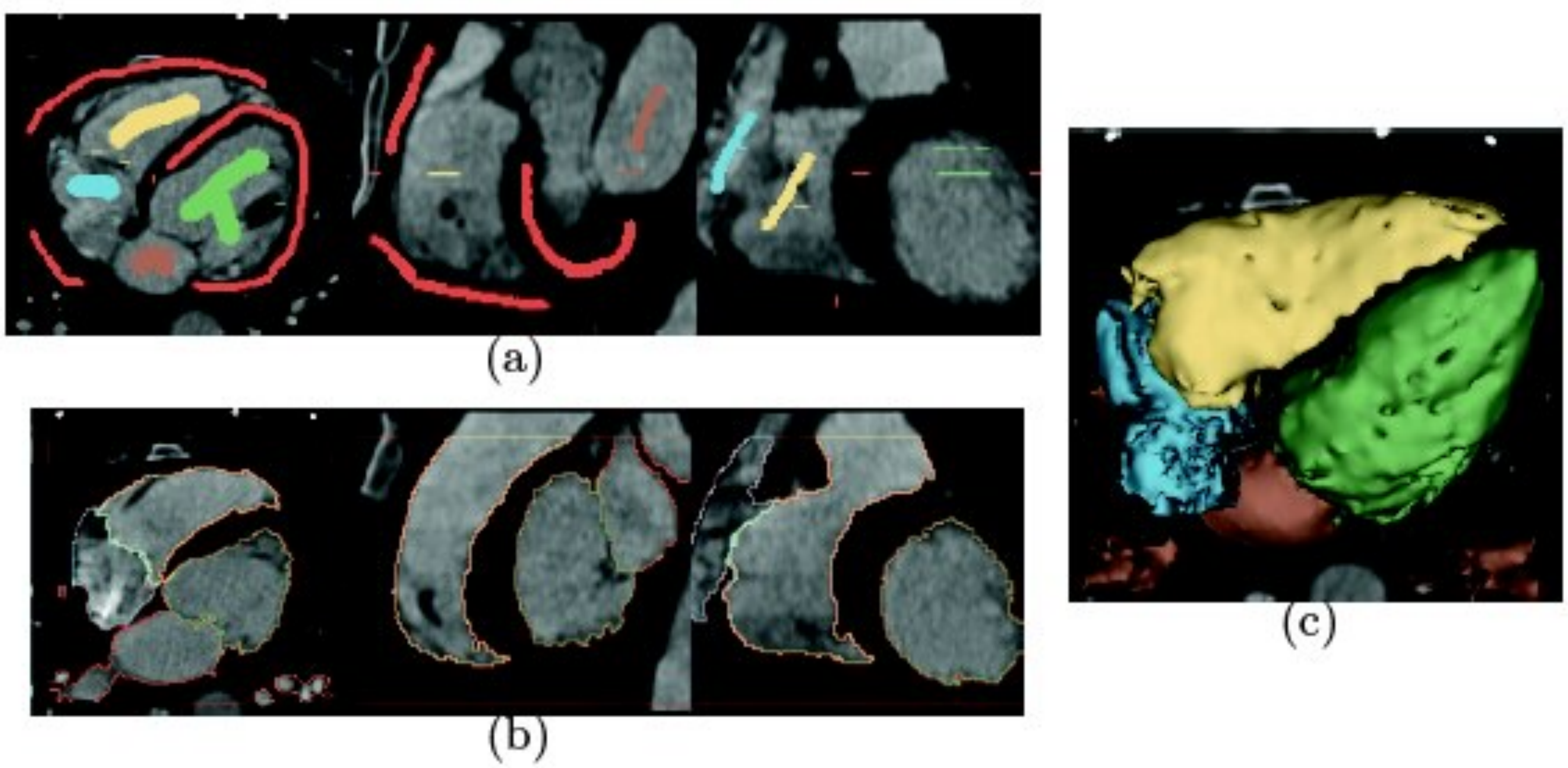




Level-sets



Segmentación interactiva

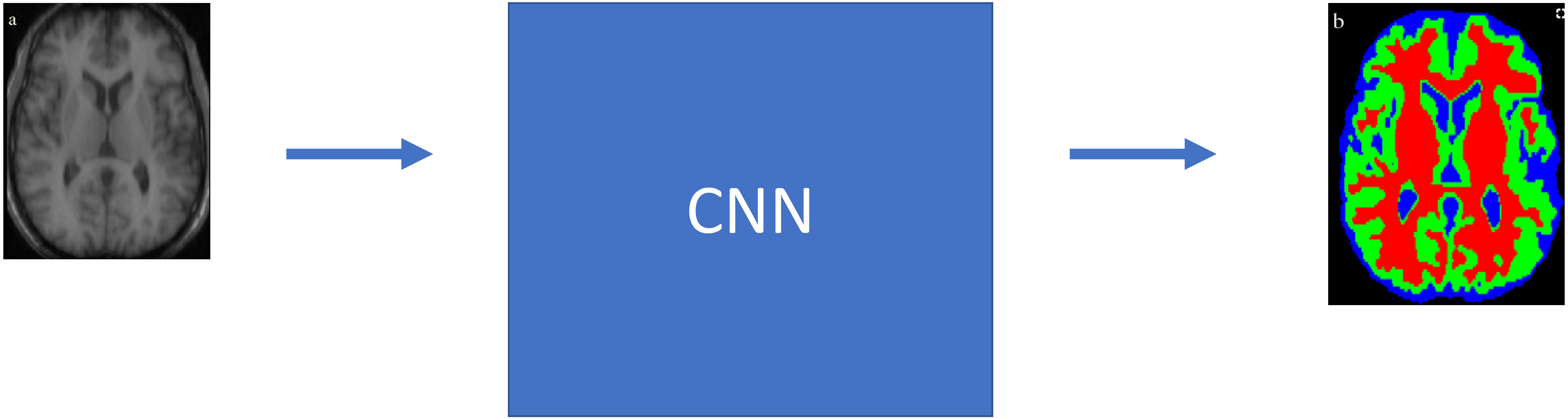


Donde encontrar todos estos métodos implementados?

Redes neuronales convolucionales para la segmentación de imágenes

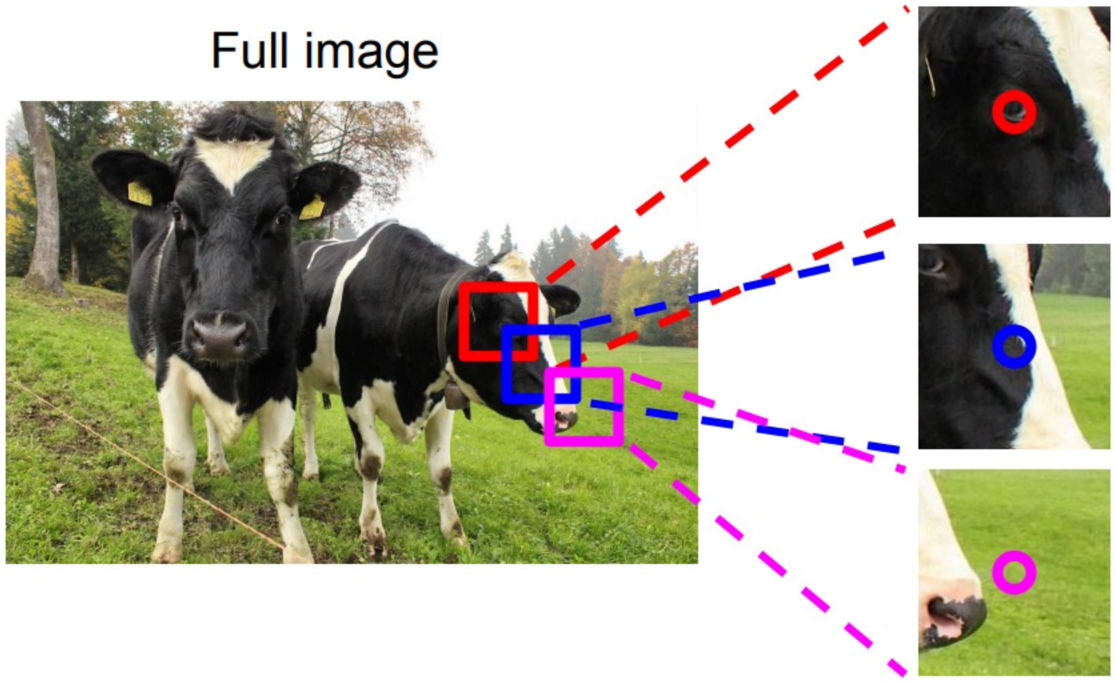
Segmentación de imágenes binaria

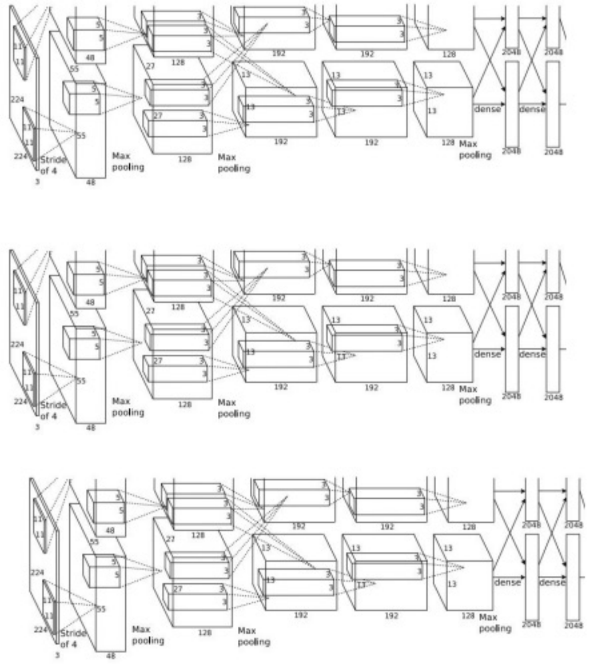
Segmentación de imágenes multiclase



Primera solución: sliding window

AlexNet





Vaca

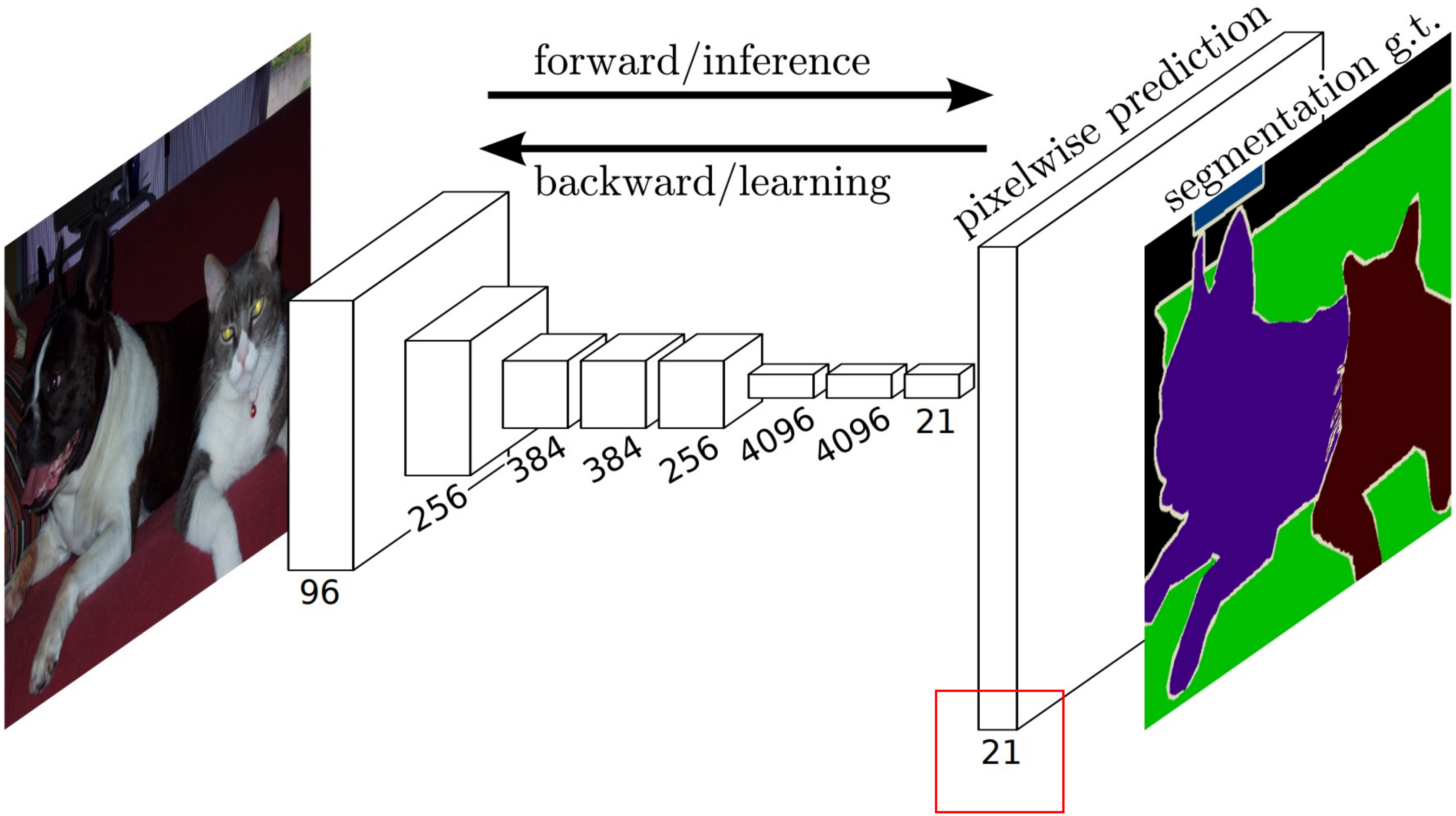
Vaca

Pasto

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

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

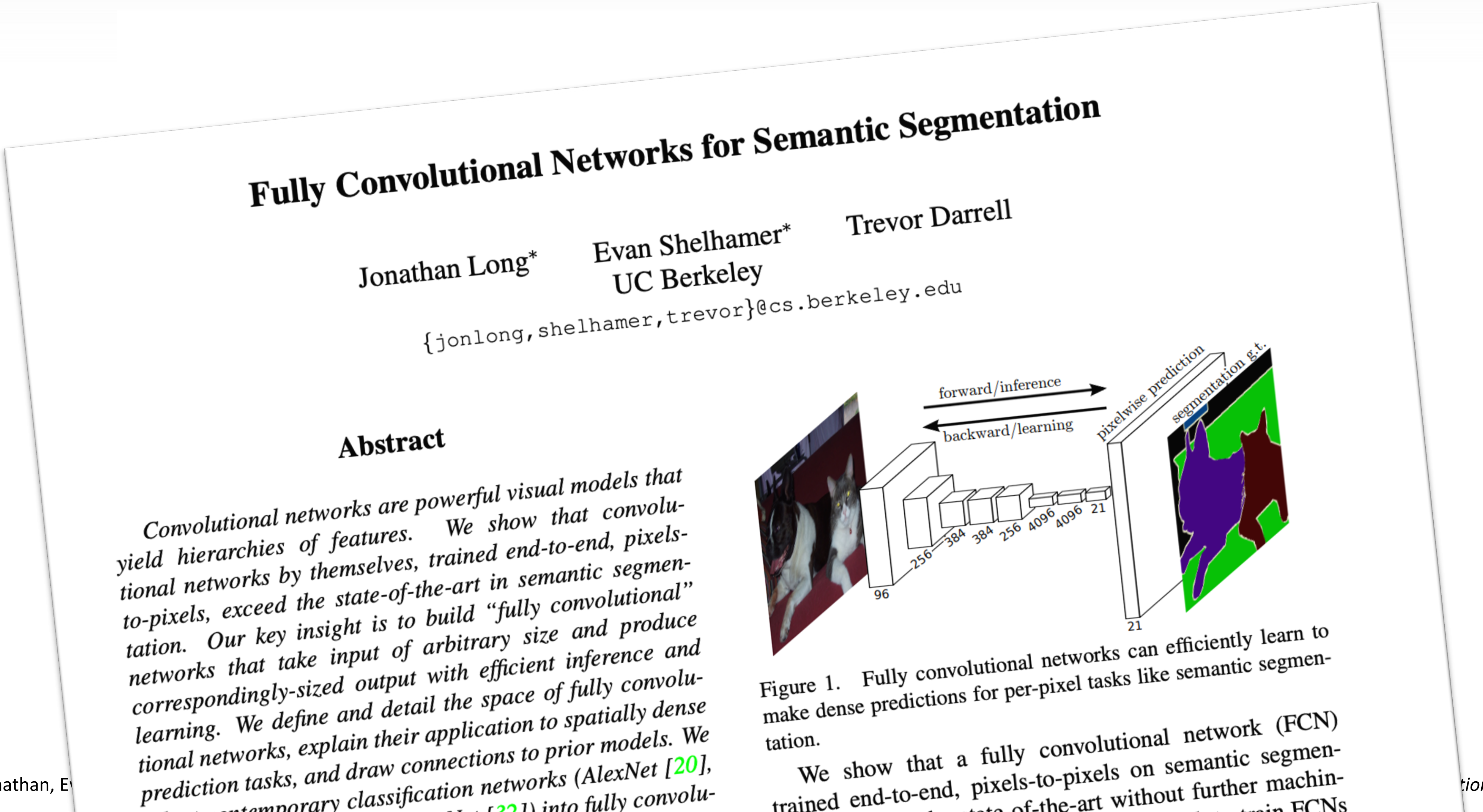
Fully Convolutional Networks



- 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

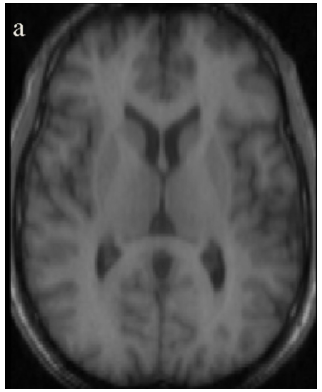
Convolutional networks are powerful visual models that yield hierarchies of features. We show that convolutional networks by themselves, trained end-to-end, pixels-to-pixels, exceed the state-of-the-art in semantic segmentation. Our key insight is to build "fully convolutional" networks that take input 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], ... Net [32]) into fully convolu-

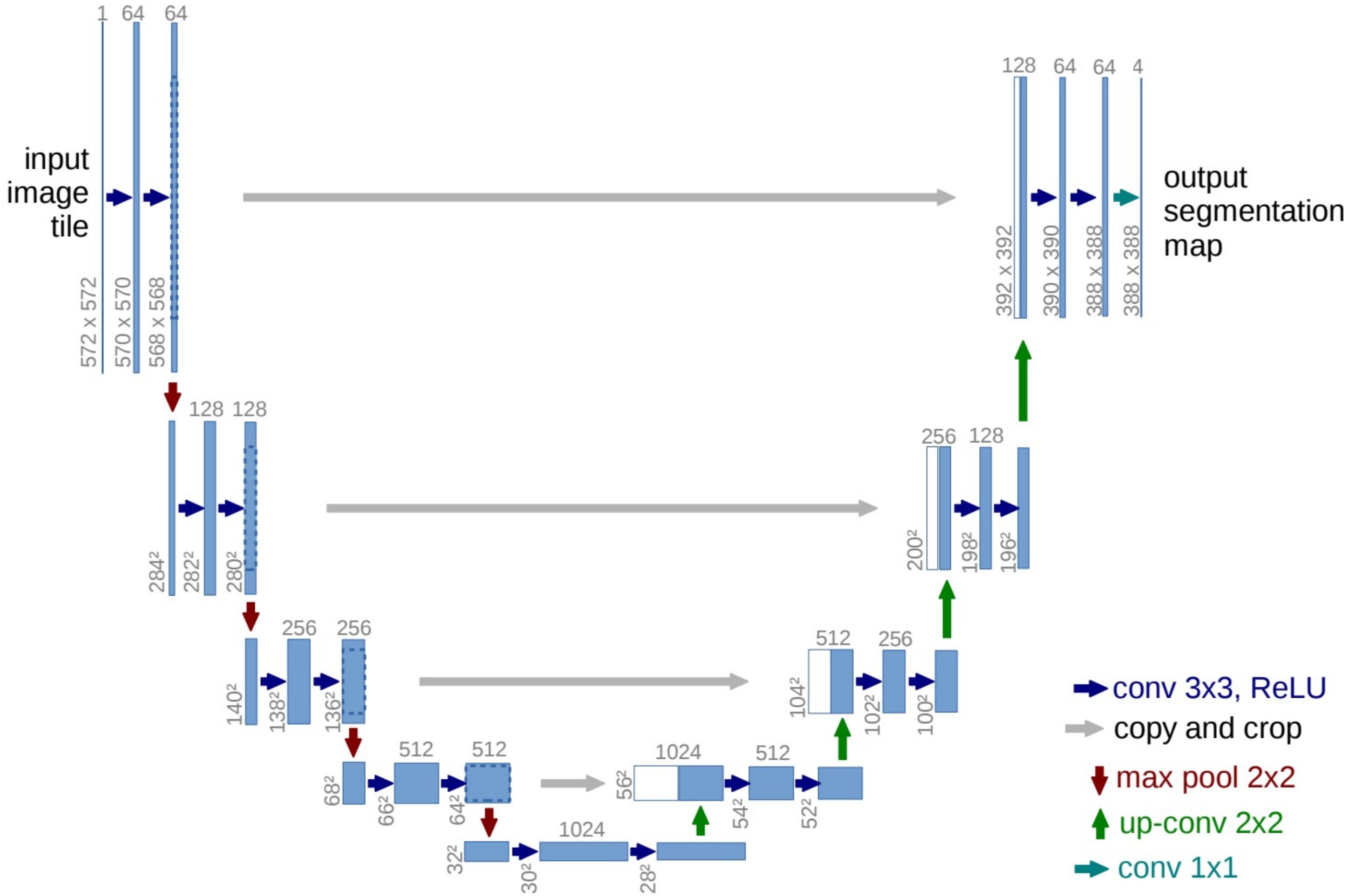
Figure 1. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation.

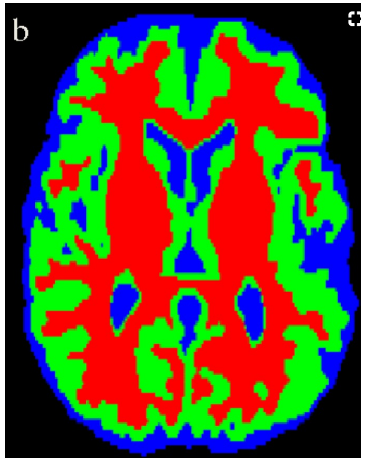
We show that a fully convolutional network (FCN) trained end-to-end, pixels-to-pixels on semantic segmentation ... the state-of-the-art without further machin- ... train FCNs

Long, Jonathan, E... ...*tion*. 2015.

UNet: Arquitectura encoder-decoder para segmentación de imágenes







UNet: Arquitectura encoder-decoder para segmentación de imágenes

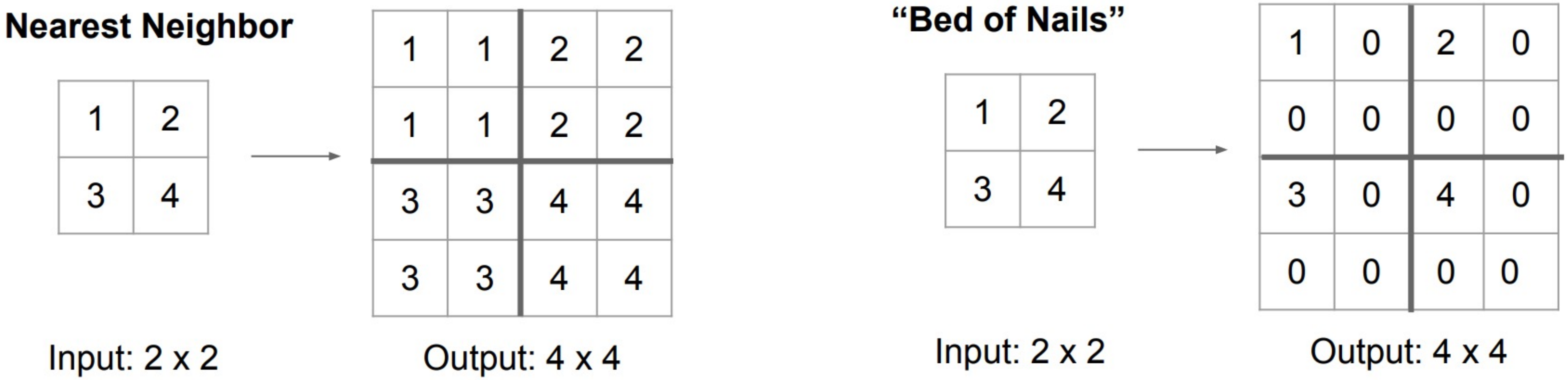
U-Net: Convolutional Networks for Biomedical Image Segmentation

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. More...

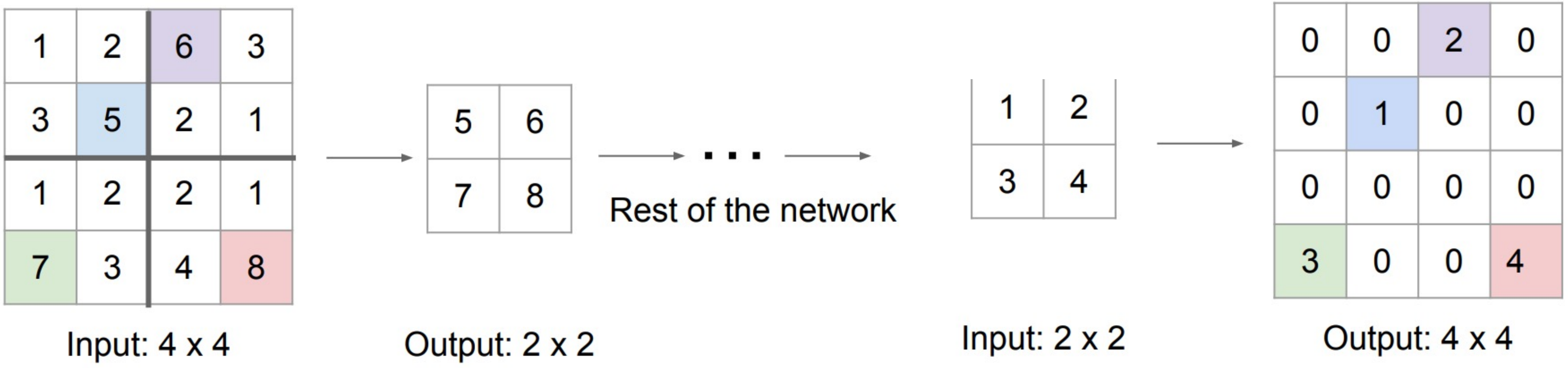
Operaciones de upsampling: Unpooling

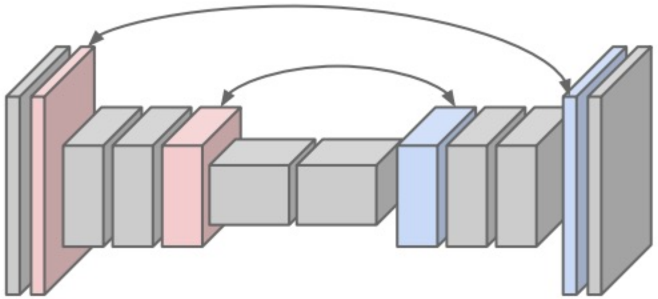


Operaciones de upsampling: Unpooling

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

Durante **MaxUnPooling**,
usamos esa posición

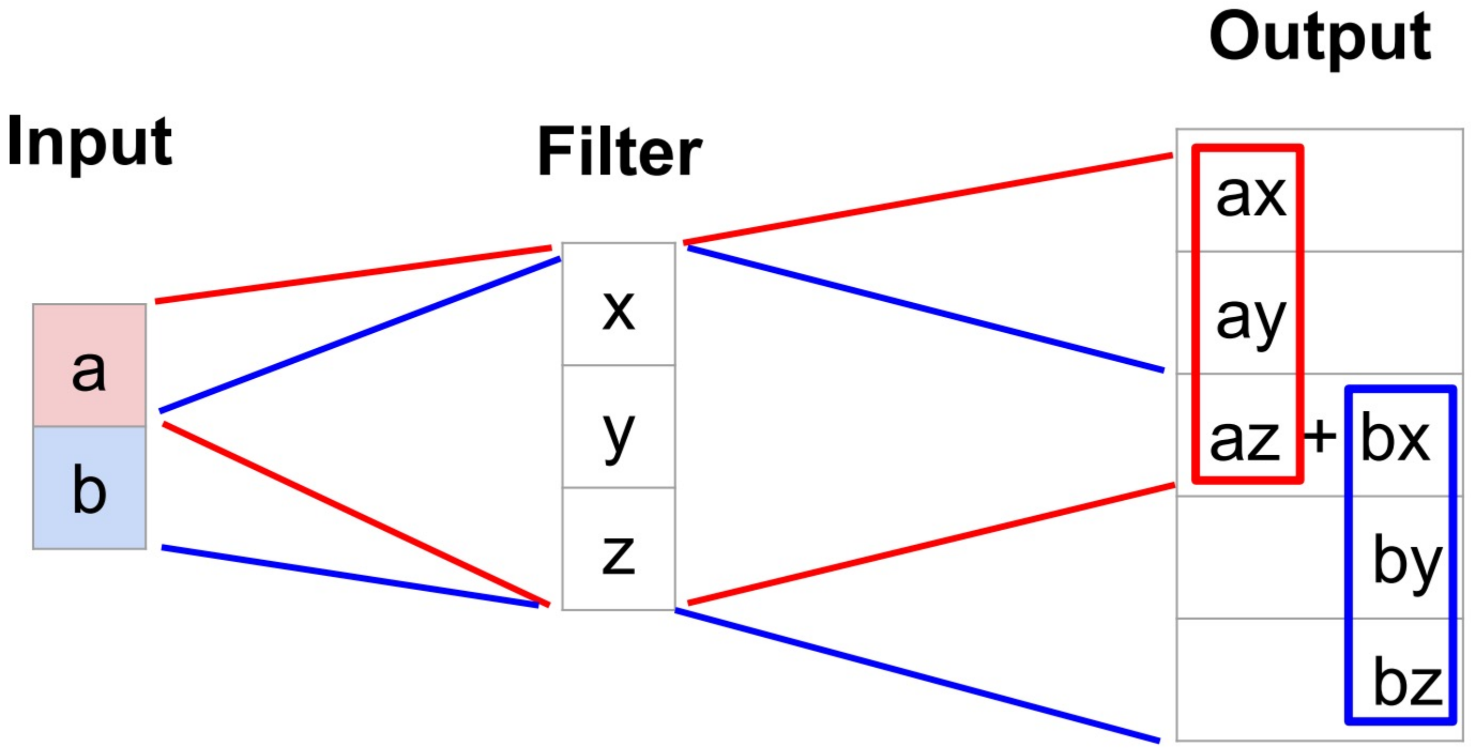




Source: http://cs231n.stanford.edu/slides/2020/lecture_12.pdf

Transpose convolution

Ejemplo para una convolución 1D



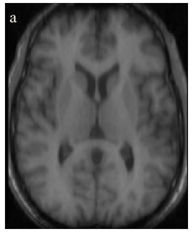
La salida se construye pesando los valores de la entrada con los distintos pesos de un kernel

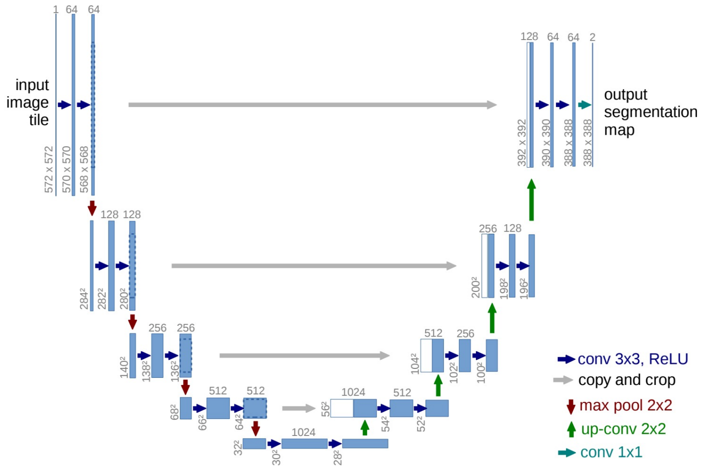
En las zonas de overlap, se suma

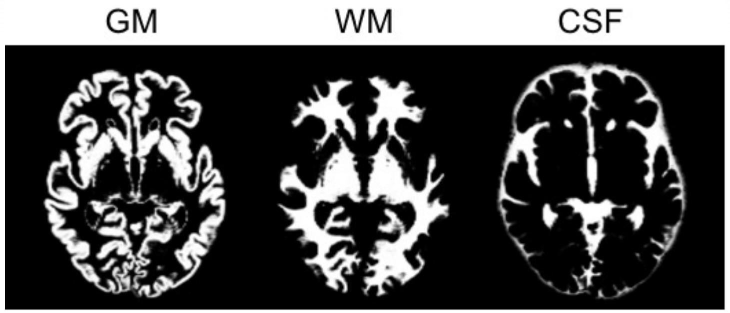
Por cada 1 pixel de la entrada, nos desplazamos 2 en la salida (stride = 2)

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

Cálculo de la función de pérdida para segmentación

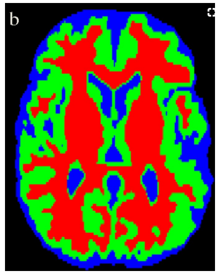






Mapa de probabilidad por clase
(cada voxel suma 1 entre los 3 mapas)

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**

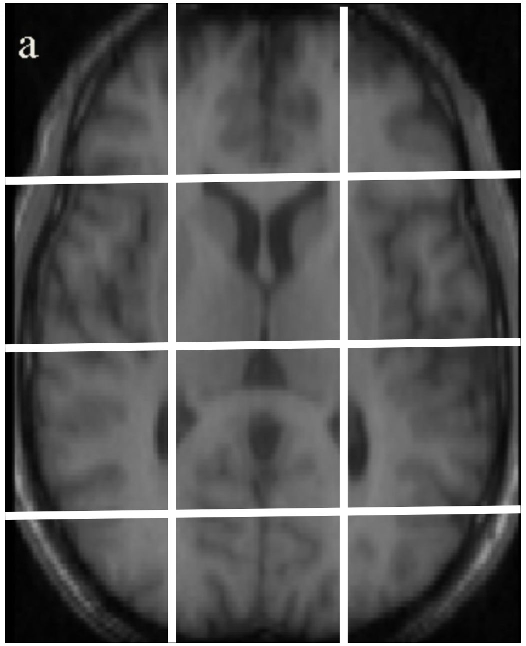
Redes totalmente convolucionales

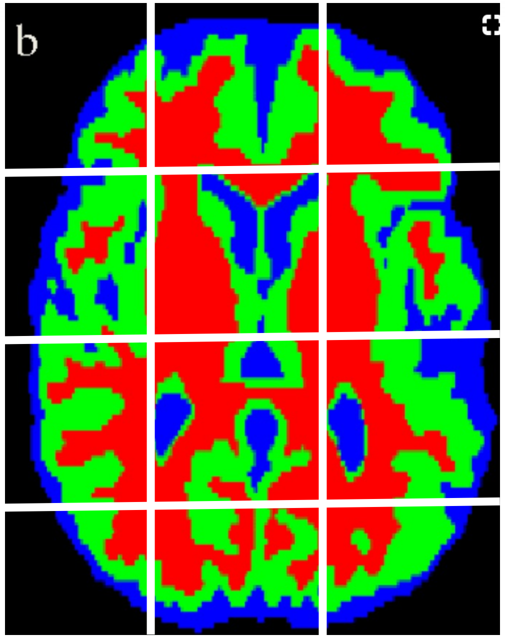
- Si la red no cuenta con ninguna capa totalmente conectada (densa), entonces se dice totalmente convolucional

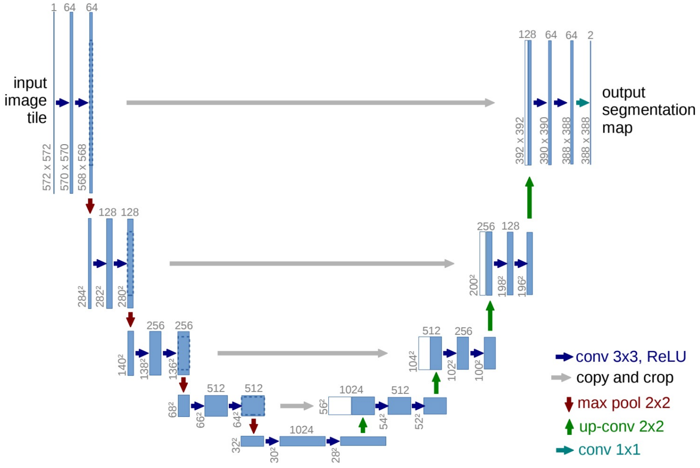
- Las capas densas pueden ser implementadas como capas convolucionales → Cómo?

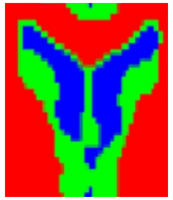
- La gran ventaja de una red totalmente convolucional es que puede procesar imágenes de cualquier tamaño, y puede ser entrenada y evaluada en imágenes de diferentes tamaños.

Entrenamiento por parches





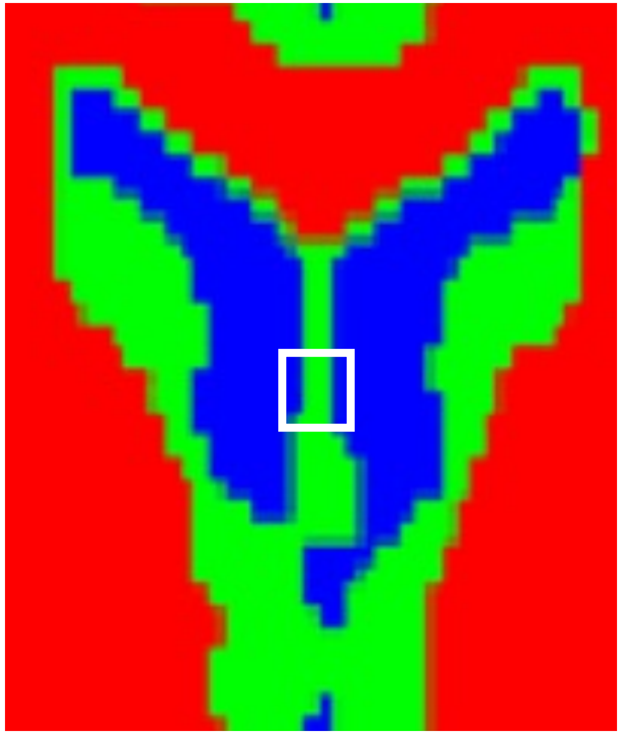




- Los mini-batches se componen de parches, no de imágenes completas
- En test time, si la red es totalmente convolucional, basta con insertar la nueva imagen y la predicción será del tamaño acorde *(cuidado con lo múltiplos de).*
- Si la red no es del tamaño acorde, se puede hacer el tiling fuera de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

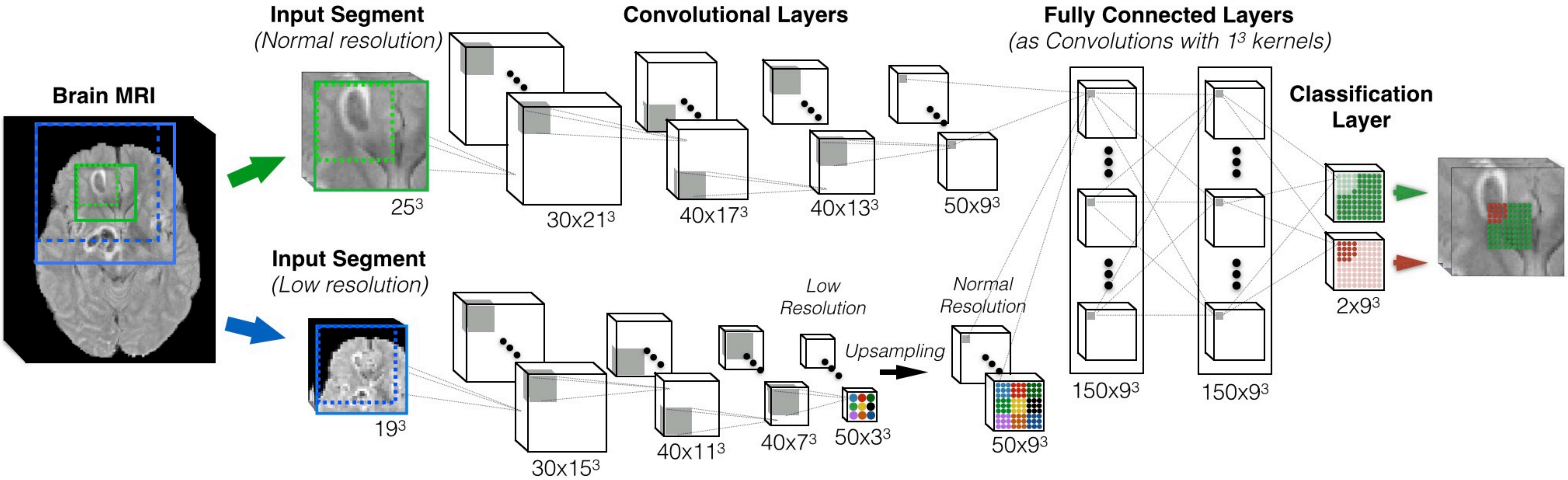
CONV3D

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.

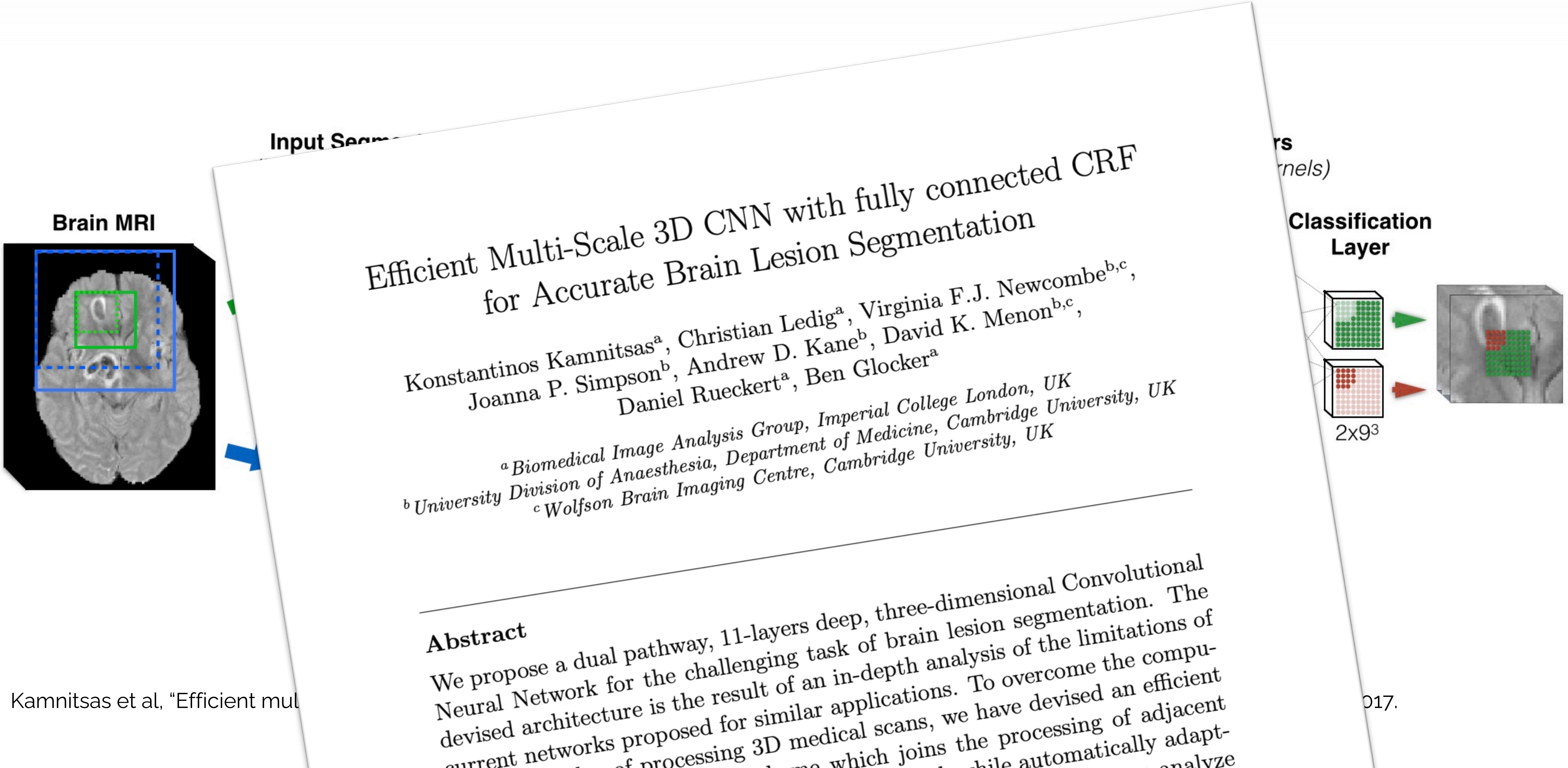
In the simplest case, the output value of the layer with input size (N, C_{in}, D, H, W) and output $(N, C_{out}, D_{out}, H_{out}, W_{out})$ can be precisely described as:

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

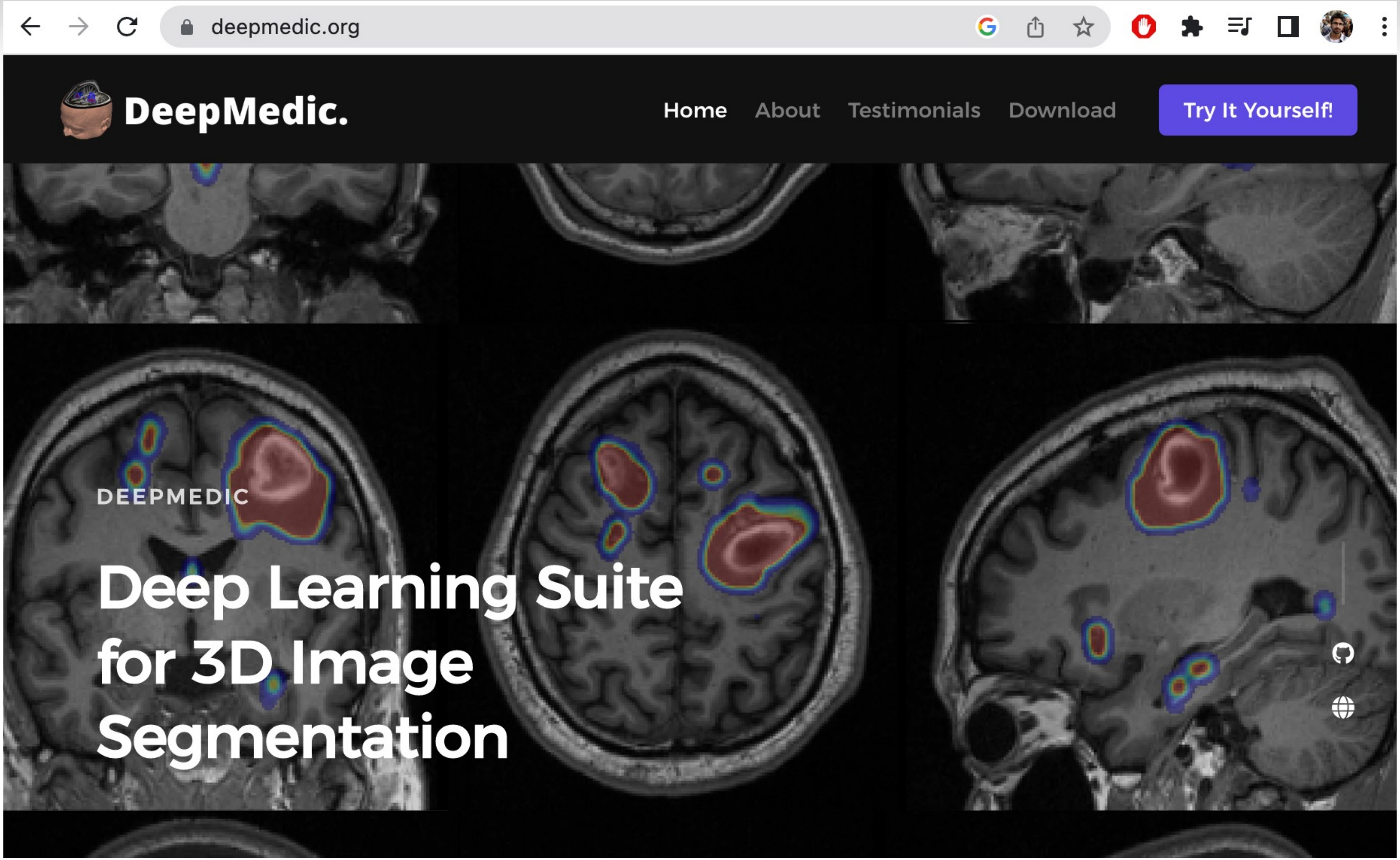
Konstantinos Kamnitsas[a], Christian Ledig[a], Virginia F.J. Newcombe[b,c], Joanna P. Simpson[b], Andrew D. Kane[b], David K. Menon[b,c], Daniel Rueckert[a], Ben Glocker[a]

[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 computational ... of processing 3D medical scans, we have devised an efficient ... which joins the processing of adjacent ... while automatically adapt- ... analyze

Kamnitsas et al, "Efficient mul ... 017.


deepmedic.org
DeepMedic.
Home
About
Testimonials
Download
Try It Yourself!
DEEPMEDIC
Deep Learning Suite for 3D Image Segmentation

nature methods

ARTICLES
https://doi.org/10.1038/s41592-020-01008-z

nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation

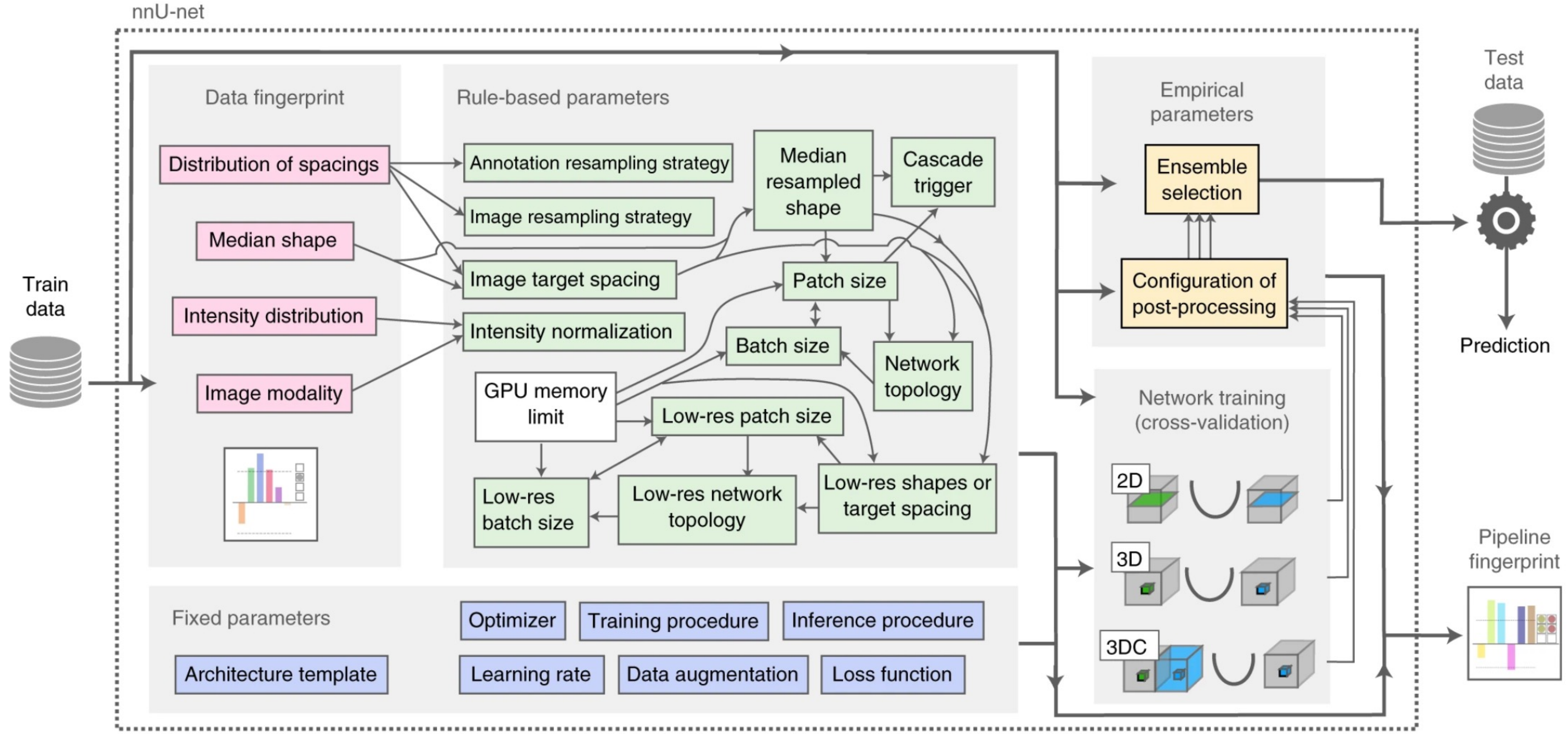
Fabian Isensee[1,2,6], Paul F. Jaeger[1,6], Simon A. A. Kohl[1,3], Jens Petersen[1,4] and Klaus H. Maier-Hein[1,5] ✉

Biomedical imaging is a driver of scientific discovery and a core component of medical care and is being stimulated by the field of deep learning. While semantic segmentation algorithms enable image analysis and quantification in many applications, the design of respective specialized solutions is non-trivial and highly dependent on dataset properties and hardware conditions. We developed nnU-Net, a deep learning-based segmentation method that automatically configures itself, including preprocessing, network architecture, training and post-processing for any new task. The key design choices in this process are modeled as a set of fixed parameters, interdependent rules and empirical decisions. Without manual intervention, nnU-Net surpasses most existing approaches, including highly specialized solutions on 23 public datasets used in international biomedical segmentation competitions. We make nnU-Net publicly available as an out-of-the-box tool, rendering state-of-the-art segmentation accessible to a broad audience by requiring neither expert knowledge nor computing resources beyond standard network training.

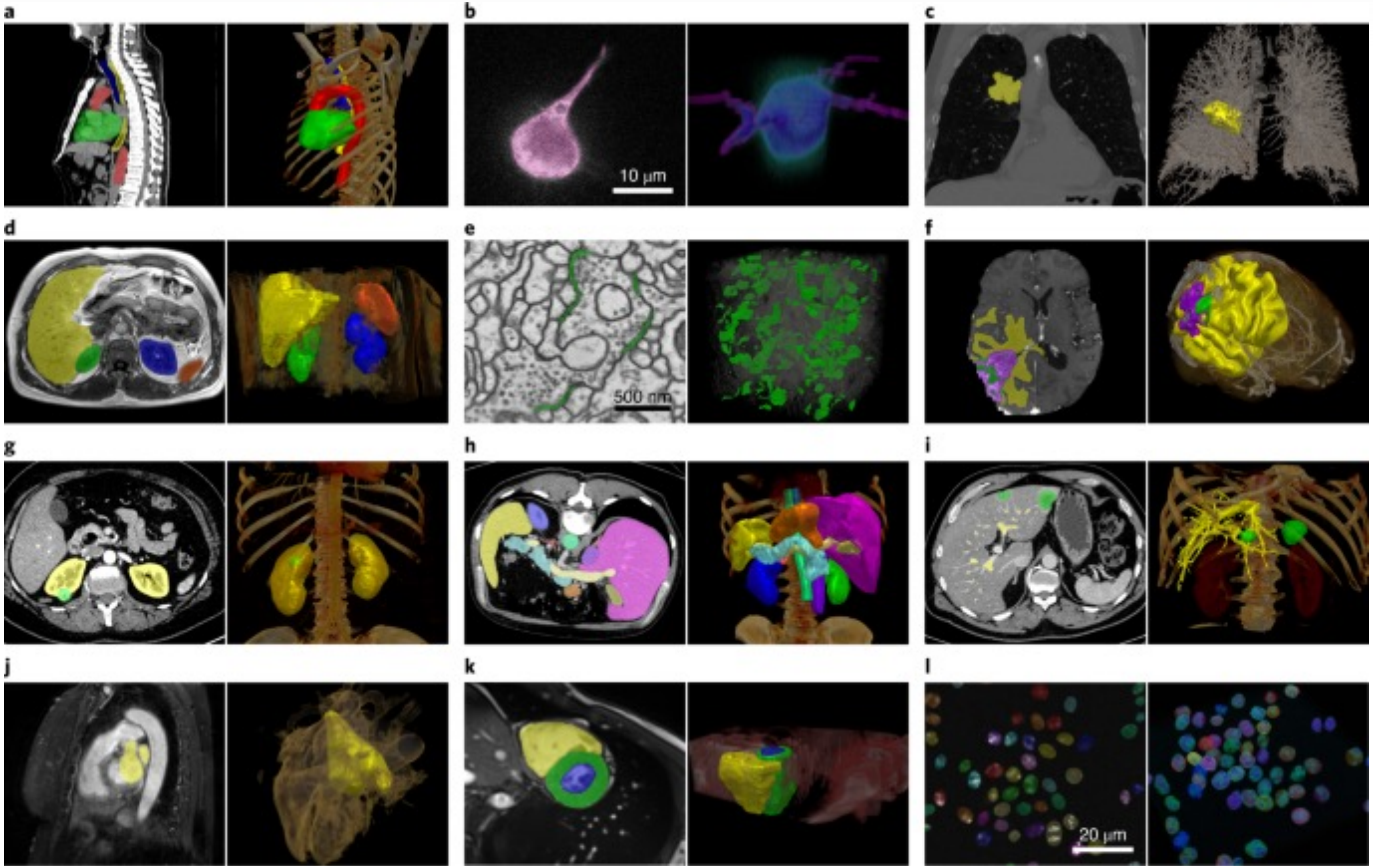
Semantic segmentation transforms raw biomedical image data into meaningful, spatially structured information and thus plays an essential role in scientific discovery[1,2]. At the same time, semantic segmentation is an essential ingredient in numerous clinical applications[3,4], including applications of artificial intelligence in diagnostic support systems[5,6], therapy planning support[7],

expert choices, for example, when considering the construction of a sensible problem-specific search space[12]. As our analysis of the current landscape in international biomedical segmentation challenges indicates (Results), these practical limitations commonly leave users with a manual and an iterative trial-and-error process during method design that is mostly driven by individual experi-

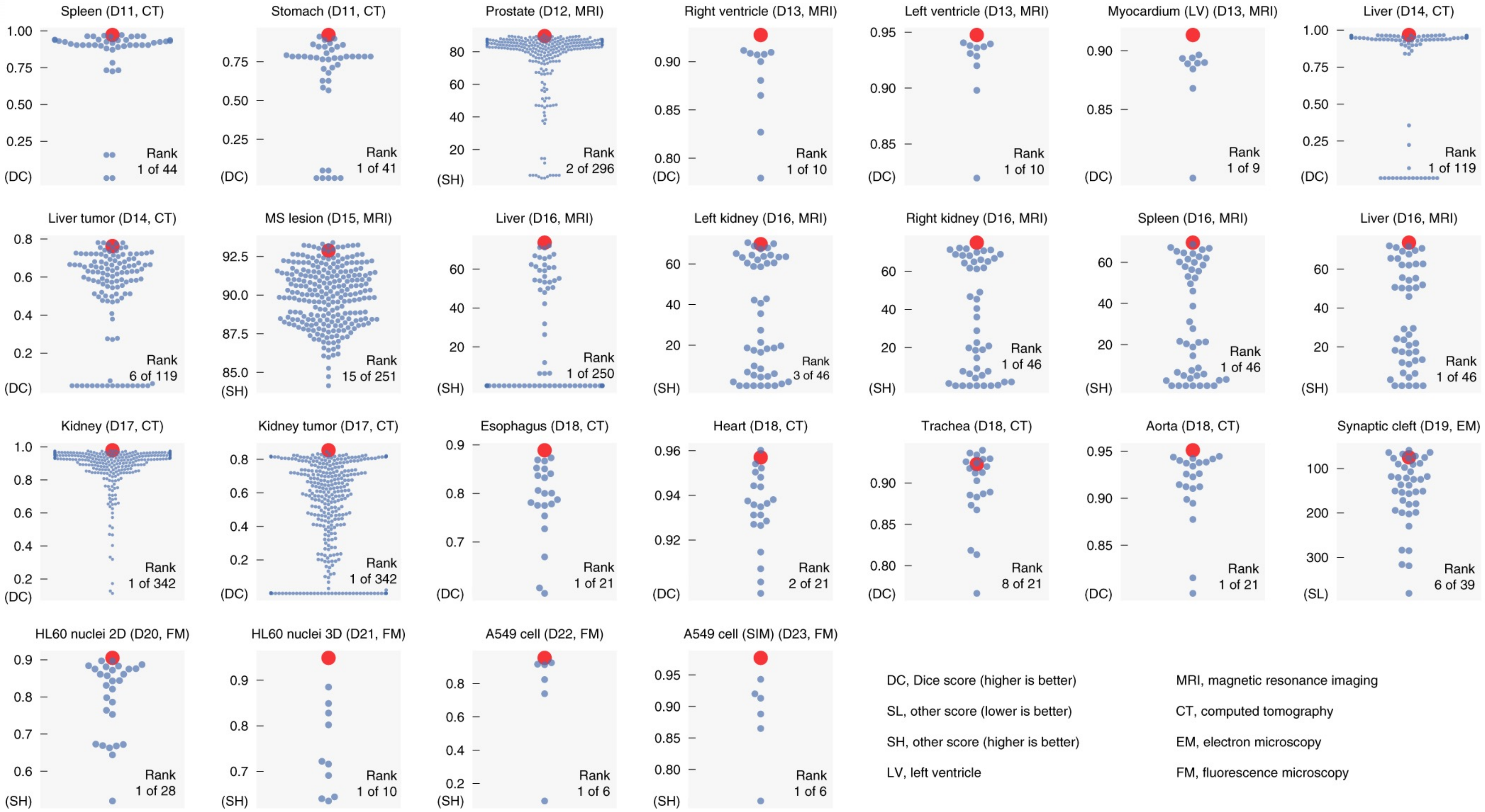
nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation



nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation



nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation



github.com/MIC-DKFZ/nnUNet

FabianIsensee simplify the function for loading pretrained weights — ad081e9 9 hours ago — 1,207 commits

documentation	nothing important	last week
nnunetv2	simplify the function for loading pretrained weights	9 hours ago
.gitignore	initial commit	4 years ago
LICENSE	setup.py	4 years ago
readme.md	update link to v1 branch	last month
setup.cfg	added more setup stuff	4 years ago
setup.py	fix typo in nnUNetv2_plan_experiment entry point	2 weeks ago

readme.md

Welcome to the new nnU-Net!

Click here if you were looking for the old one instead.

Coming from V1? Check out the TLDR Migration Guide. Reading the rest of the documentation is still strongly recommended ;-)

What is nnU-Net?

Image datasets are enormously diverse: image dimensionality (2D, 3D), modalities/input channels (RGB image, CT, MRI, microscopy, ...), image sizes, voxel sizes, class ratio, target structure properties and more change substantially between datasets. Traditionally, given a new problem, a tailored solution needs to be manually designed and optimized - a process that is prone to errors, not scalable and where success is overwhelmingly determined by the skill of the experimenter. Even for experts, this process is anything but simple: there are not only many design choices and data properties that need to be considered, but they are also tightly interconnected,

No description or website provided.

segmentation

Readme

Apache-2.0 license

3.5k stars

76 watching

1.2k forks

Report repository

Releases 2

nnU-Net V2 Latest
last month

+ 1 release

Packages

No packages published

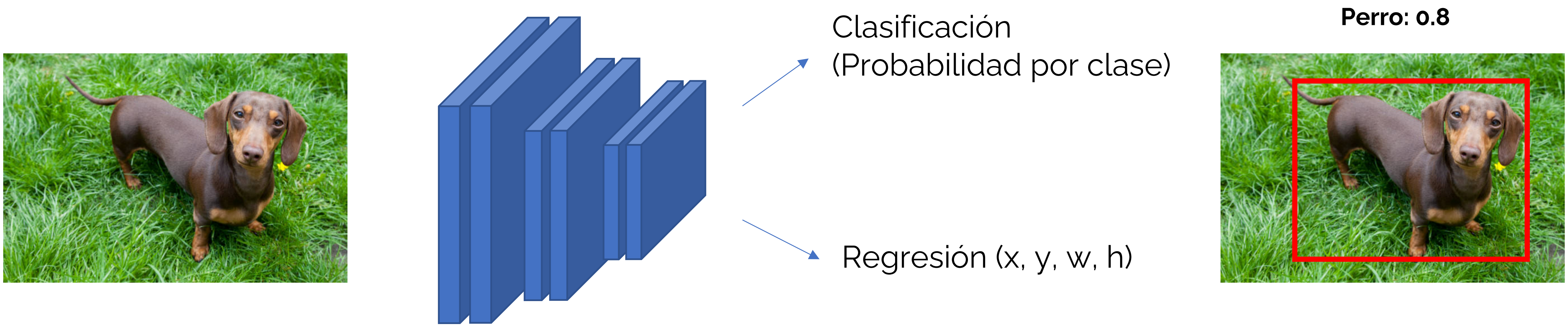
Contributors 38

+ 27 contributors

Languages

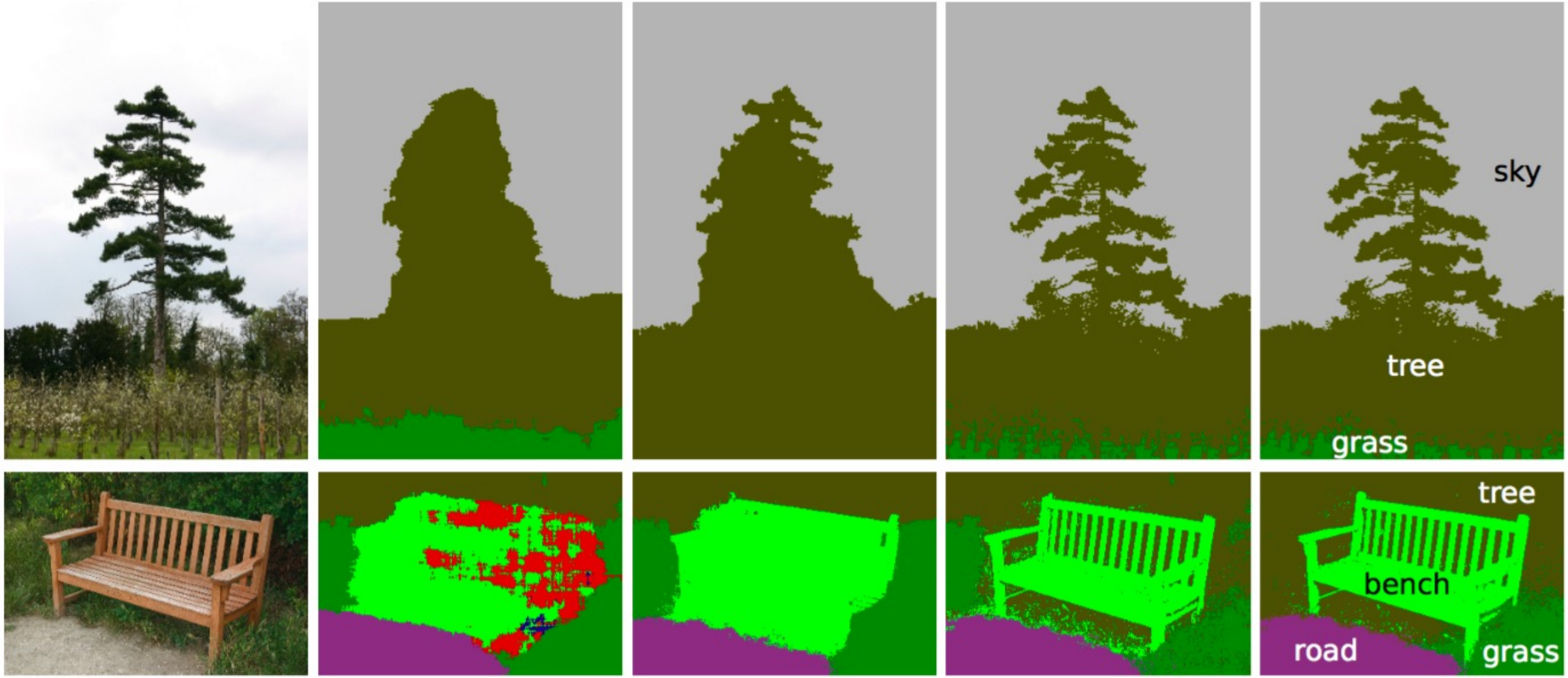
Arquitectura básica para localización

Red con salidas múltiples



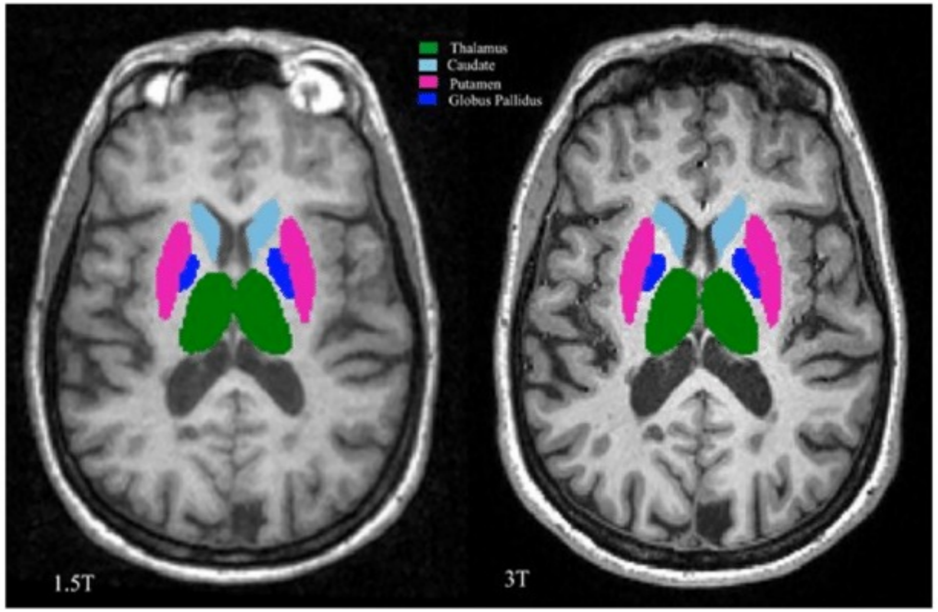
Medidas para evaluar la calidad de la segmentación

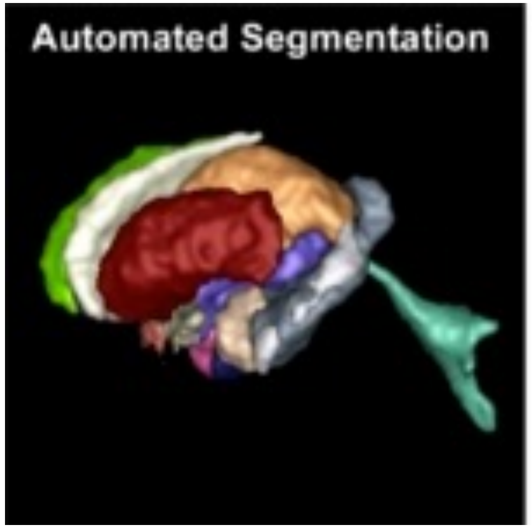
Medidas para evaluar la calidad de la segmentación

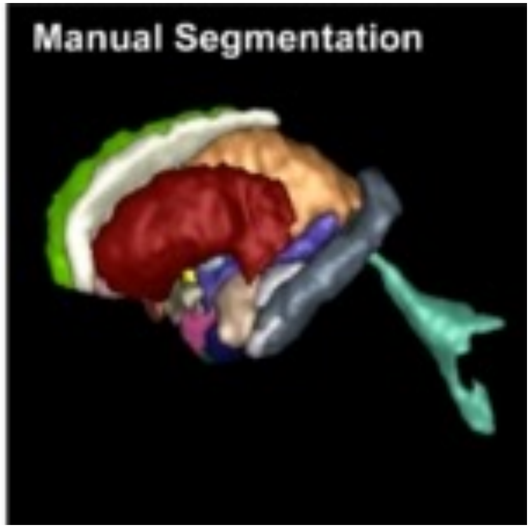


(a) Image (b) Unary classifiers (c) Robust P^n CRF (d) Fully connected CRF, MCMC inference, 36 hrs (e) Fully connected CRF, our approach, 0.2 seconds

Medidas para evaluar la calidad de la segmentación

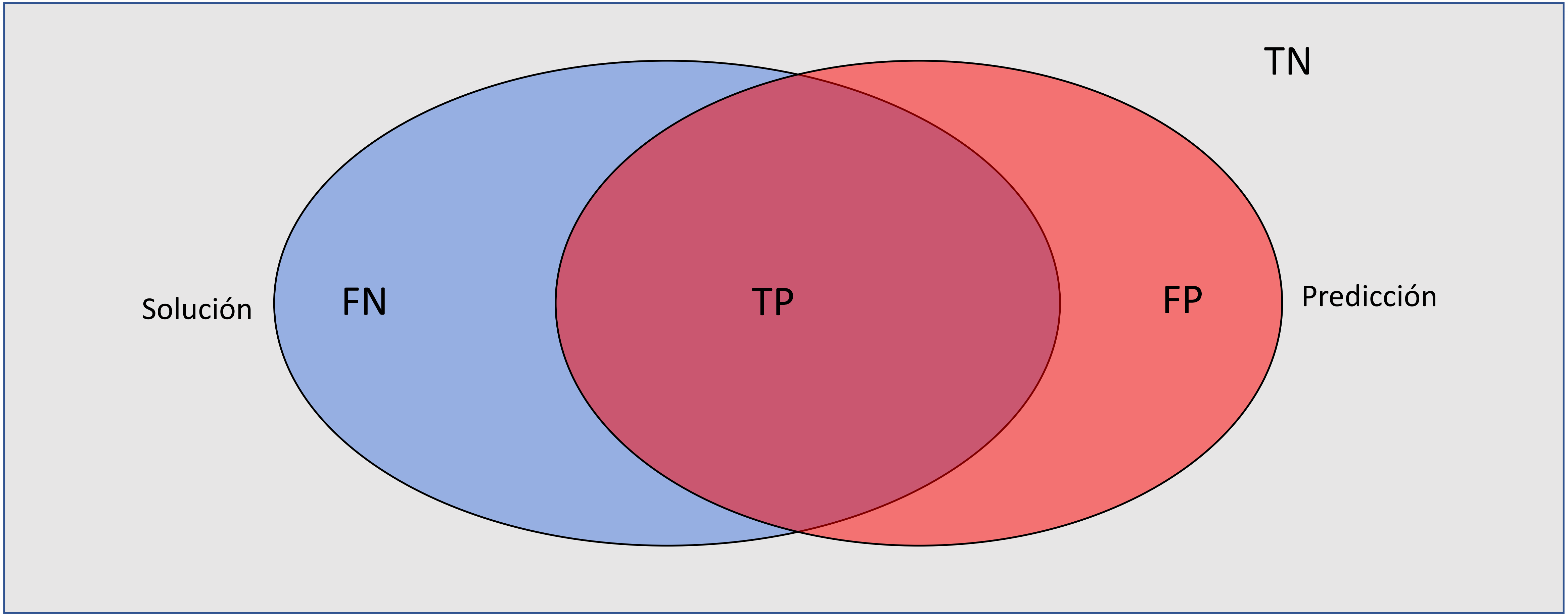






¿Intuición?

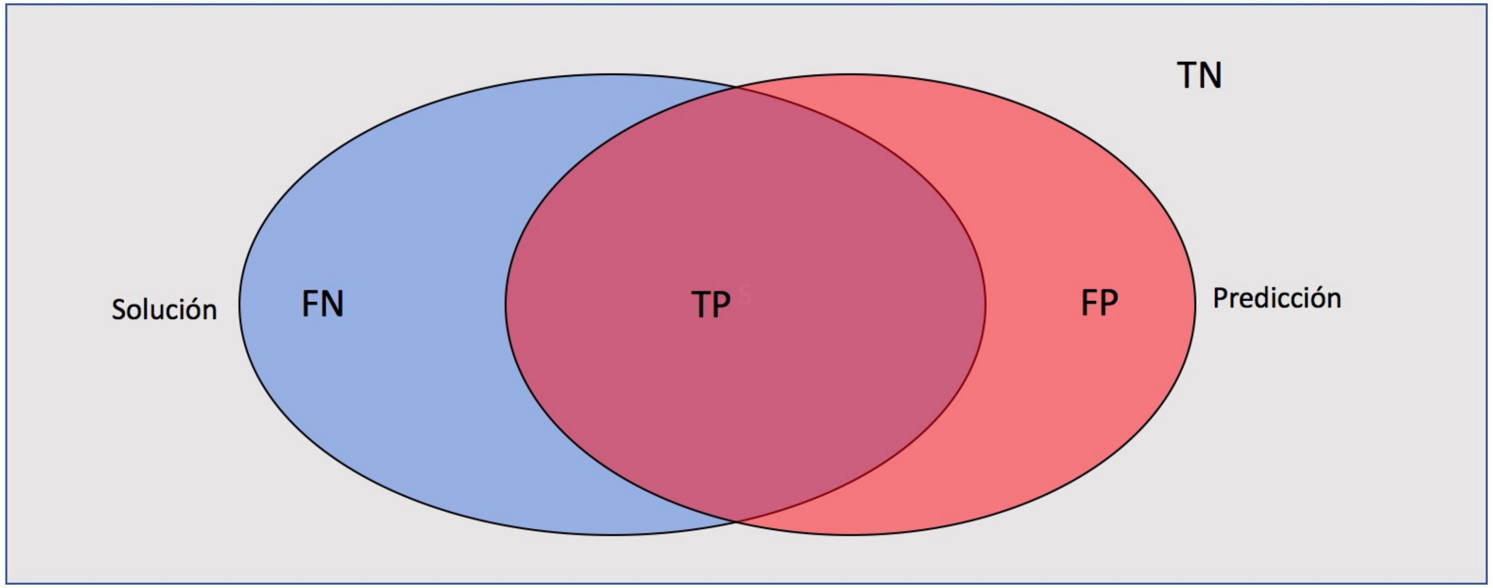
Intuición



Tasa de acierto (accuracy)

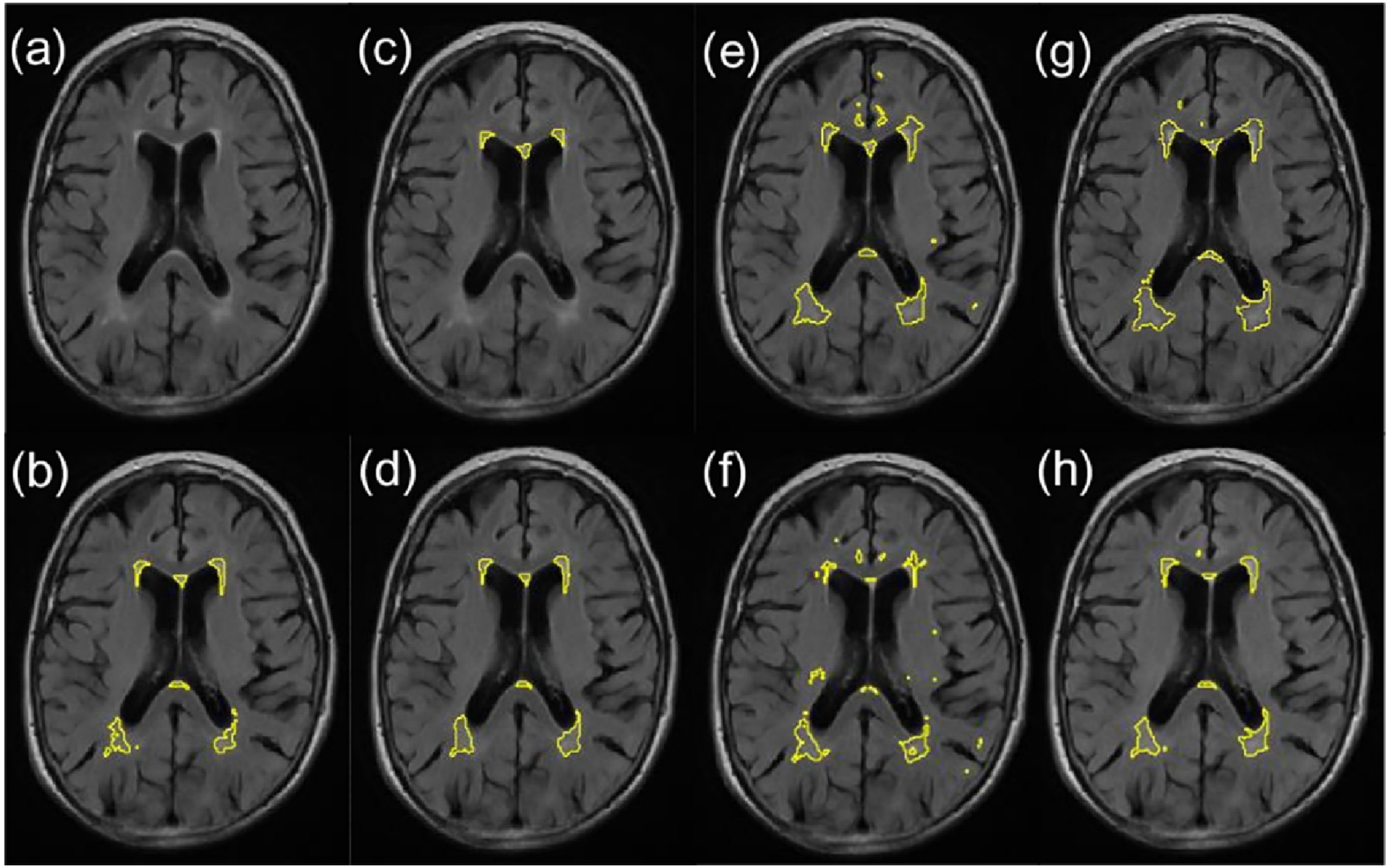
Tasa de acierto sobre el total de los píxeles

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



Tasa de acierto (accuracy)

Problema: Requiere datos balanceados



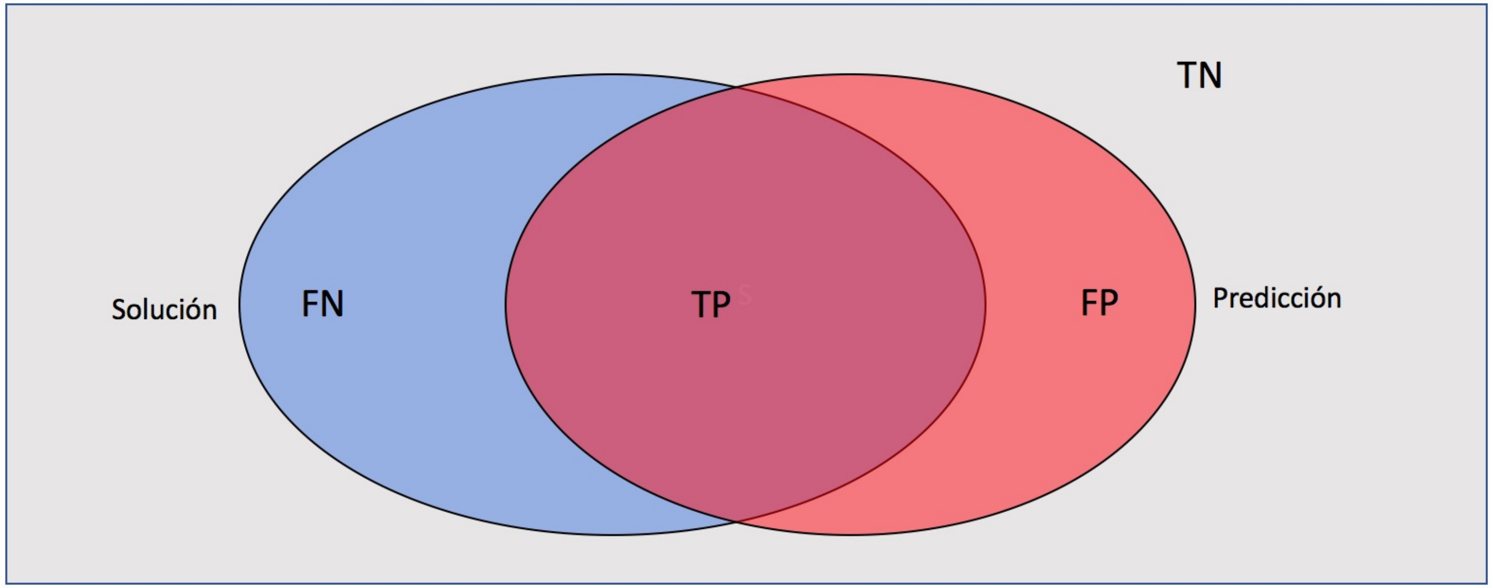
Leucoaraiosis (hiperintesidad en la materia blanca)

Medidas de solapamiento

Coeficiente Dice/ F1-Score

Se aplica en datos binarios

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

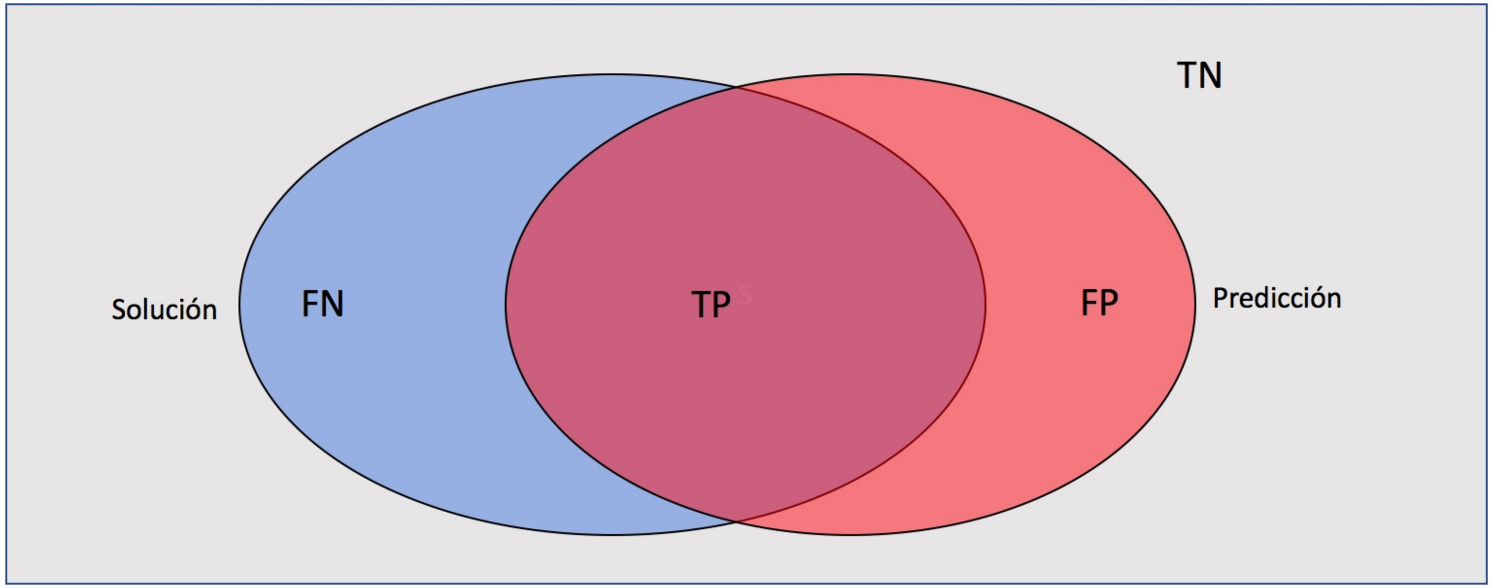


También suele usarse como función de pérdida!

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

Soft Dice as a loss function

(Milletari et al, 2016)

$$Dice = \frac{2\,|A \cap B|}{|A| + |B|}$$

$$Loss\ Soft\ Dice = 1 - Dice$$

$$\begin{bmatrix} 0.01 & 0.03 & 0.02 & 0.02 \\ 0.05 & 0.12 & 0.09 & 0.07 \\ 0.89 & 0.85 & 0.88 & 0.91 \\ 0.99 & 0.97 & 0.95 & 0.97 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

prediction target

$$|A \cap B| = \underbrace{\begin{bmatrix} 0.01 & 0.03 & 0.02 & 0.02 \\ 0.05 & 0.12 & 0.09 & 0.07 \\ 0.89 & 0.85 & 0.88 & 0.91 \\ 0.99 & 0.97 & 0.95 & 0.97 \end{bmatrix}}_{\text{prediction}} * \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}}_{\text{target}} \xrightarrow{\text{element-wise multiply}} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.89 & 0.85 & 0.88 & 0.91 \\ 0.99 & 0.97 & 0.95 & 0.97 \end{bmatrix} \xrightarrow{\text{sum}} 7.41$$

$$|A| = \begin{bmatrix} 0.01 & 0.03 & 0.02 & 0.02 \\ 0.05 & 0.12 & 0.09 & 0.07 \\ 0.89 & 0.85 & 0.88 & 0.91 \\ 0.99 & 0.97 & 0.95 & 0.97 \end{bmatrix}^{2\ (optional)} \xrightarrow{\text{sum}} 7.82$$

$$|B| = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}^{2\ (optional)} \xrightarrow{\text{sum}} 8$$

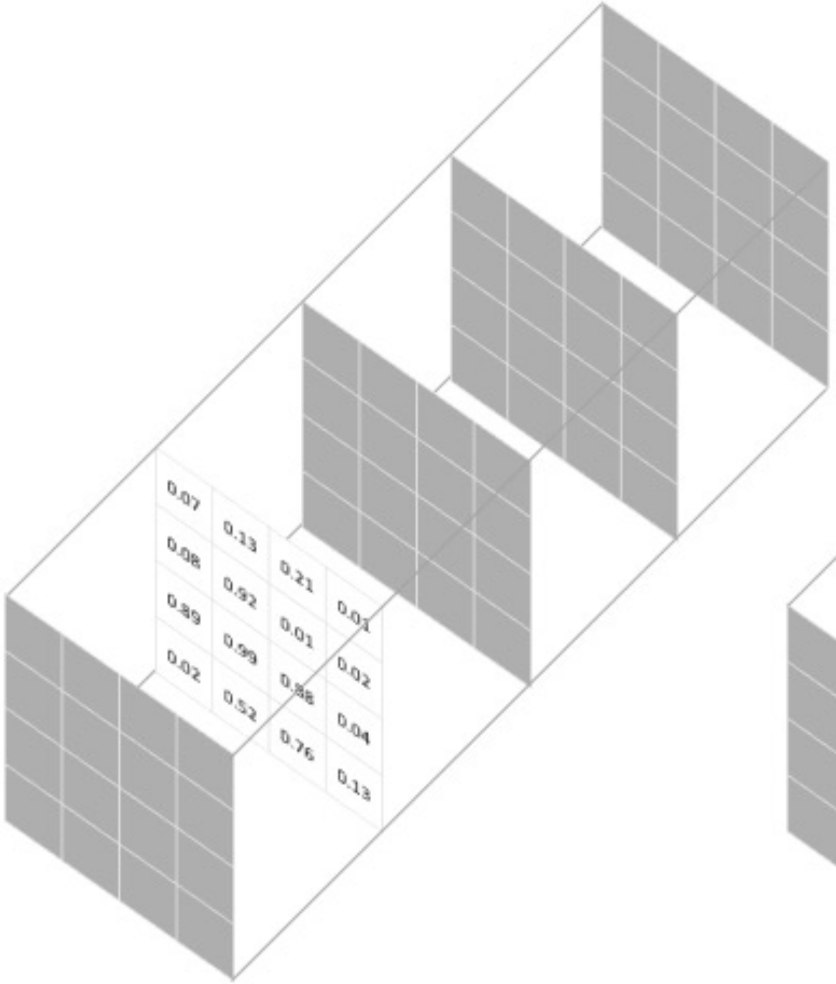
$$Dice = \frac{2 * 7.41}{7.82 + 8} = 0.936$$

$$Loss\ Soft\ Dice = 1 - 0.936 = 0.064$$

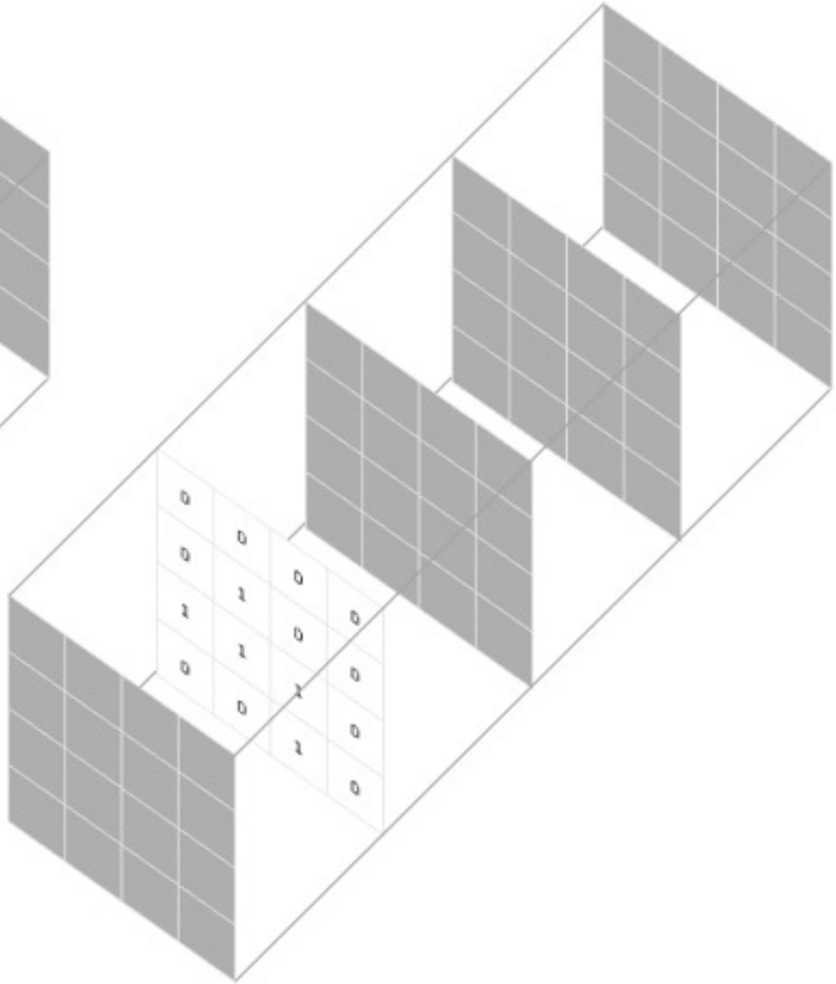
Source: https://www.jeremyjordan.me/semantic-segmentation/

Soft Dice as a loss function

Source: https://www.jeremyjordan.me/semantic-segmentation/



Prediction for a selected class



Target for the corresponding class

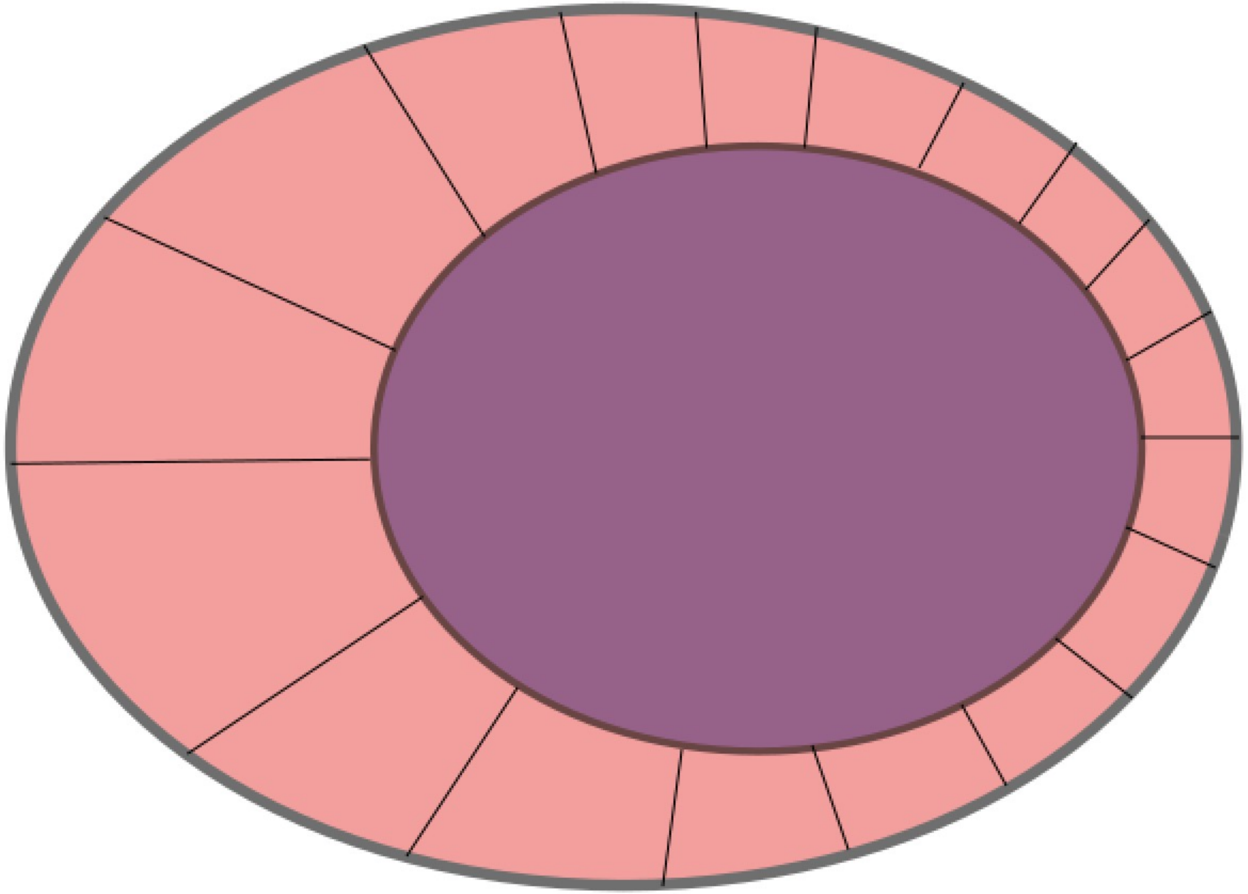
Soft Dice coefficient is calculated for each class mask

$$1-\frac{2\sum_{pixels} y_{true}y_{pred}}{\sum_{pixels} y_{true}^2+\sum_{pixels} y_{pred}^2}$$

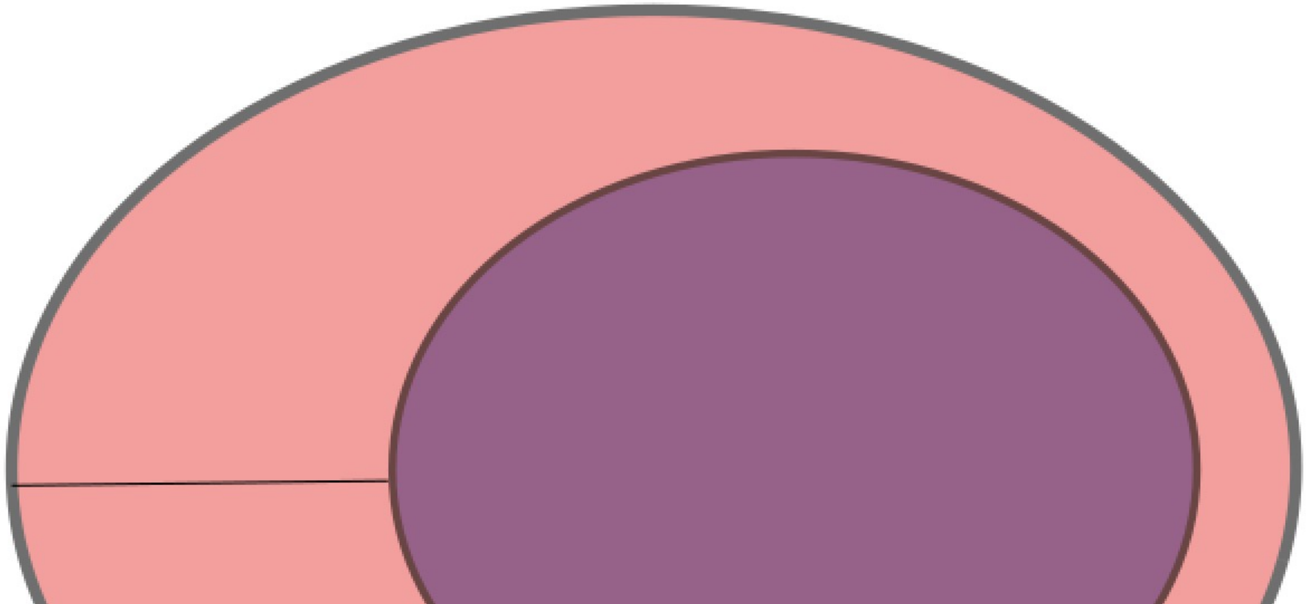
This scoring is repeated over all **classes** and averaged

Medidas distancia entre contornos

Distancia media entre contornos



Distancia de Haussdorf
(máxima distancia entre contornos)



Distancia de Haussdorf

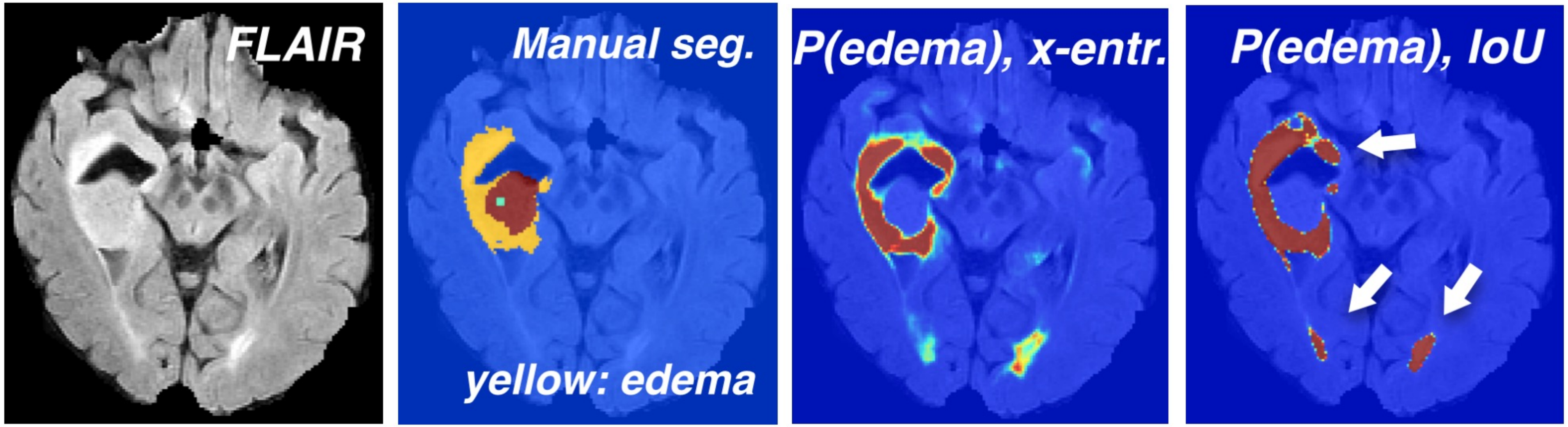
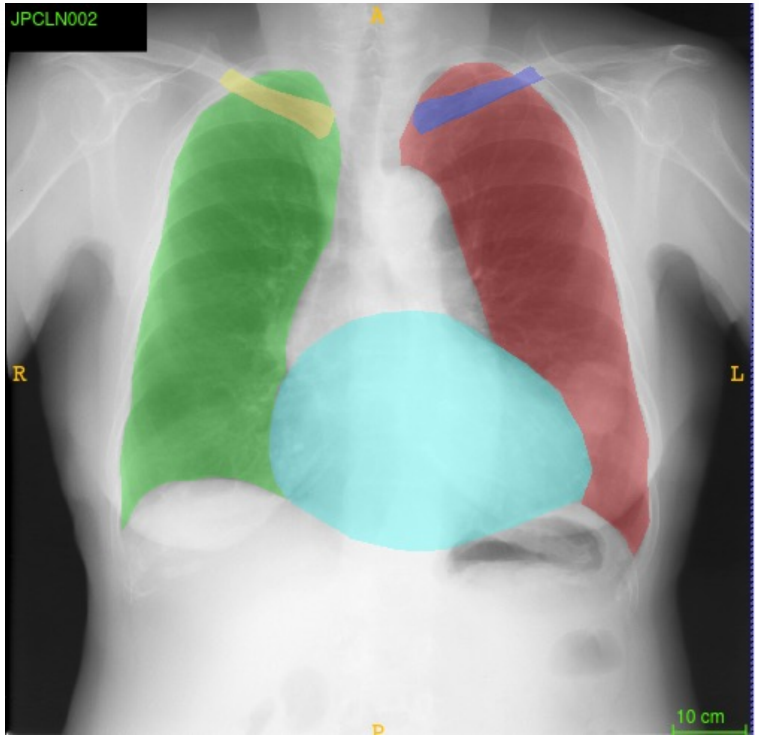
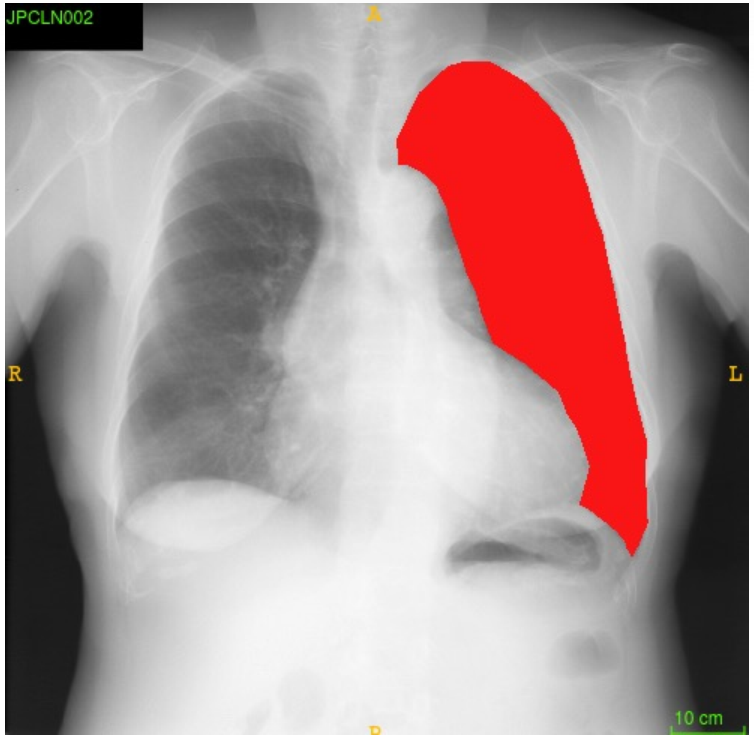


Fig. 1: Left to right: FLAIR; manual annotation of a BRATS'17 subject, where yellow depicts oedema surrounding tumour core; confidence of a CNN predicting oedema, trained with cross-entropy or IoU loss. Although overall performance is similar, training with IoU (or Dice, not shown) loss alters the CNN's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

Qué hacer en los casos multiclase?

Medidas de calidad en segmentación multiclase

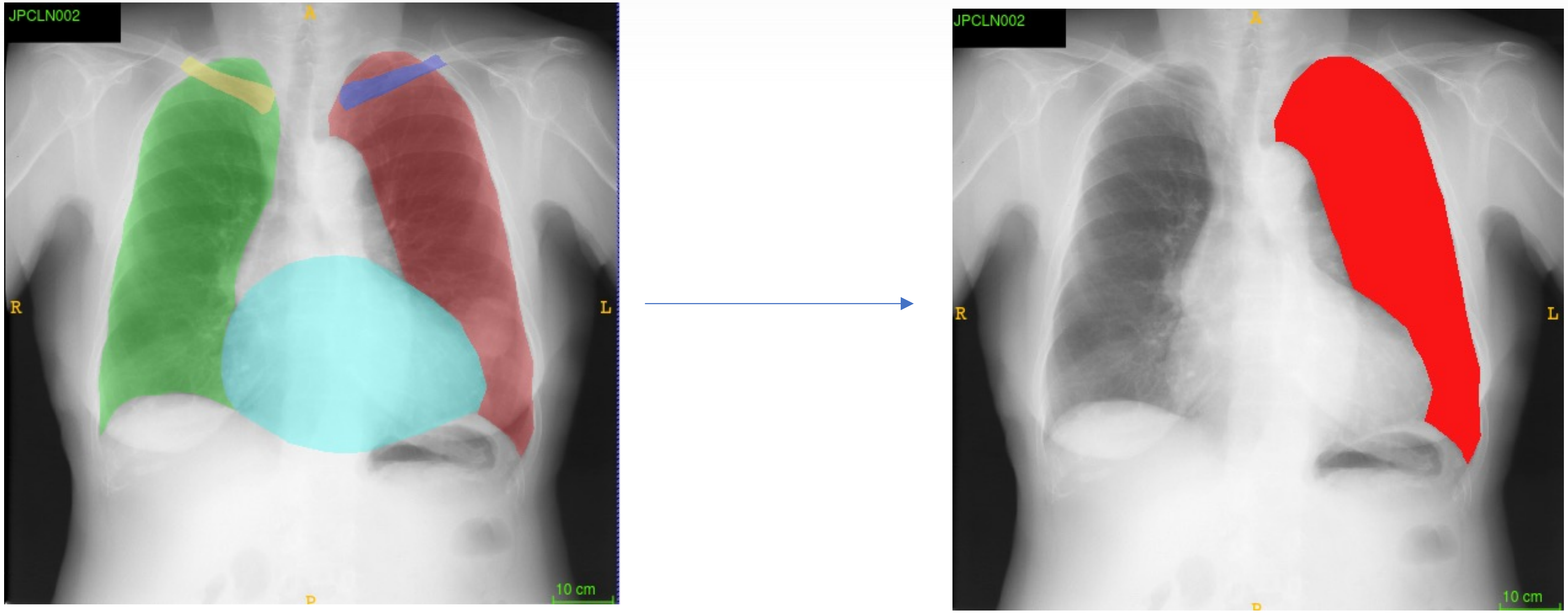




Para cada posible clase c ∈ C, se binariza el problema siguiendo:

- 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 $\in C$ 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.

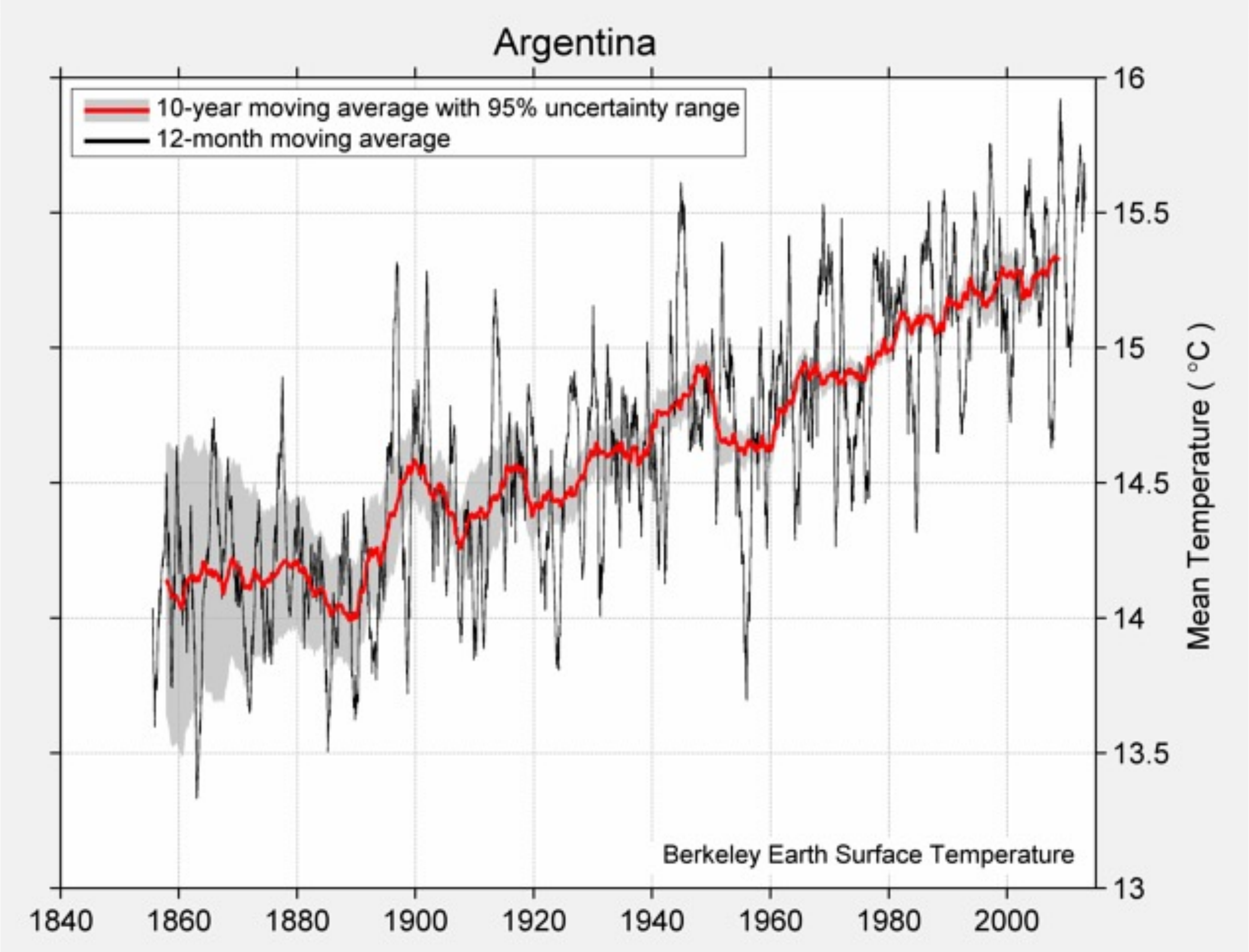
Redes neuronales convolucionales para el procesamiento de series de tiempo

Series de tiempo

Valor del dolar

Series de tiempo

Temperatura media anual en Argentina



Fuente: Berkley Earth

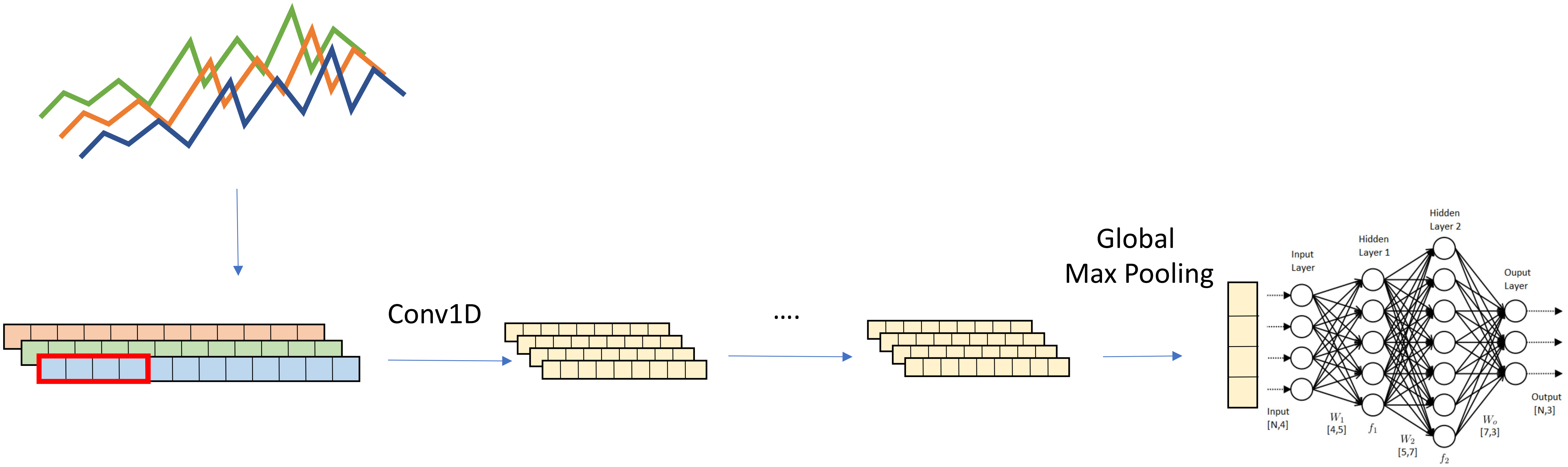
Series de tiempo

Electrocardiograma

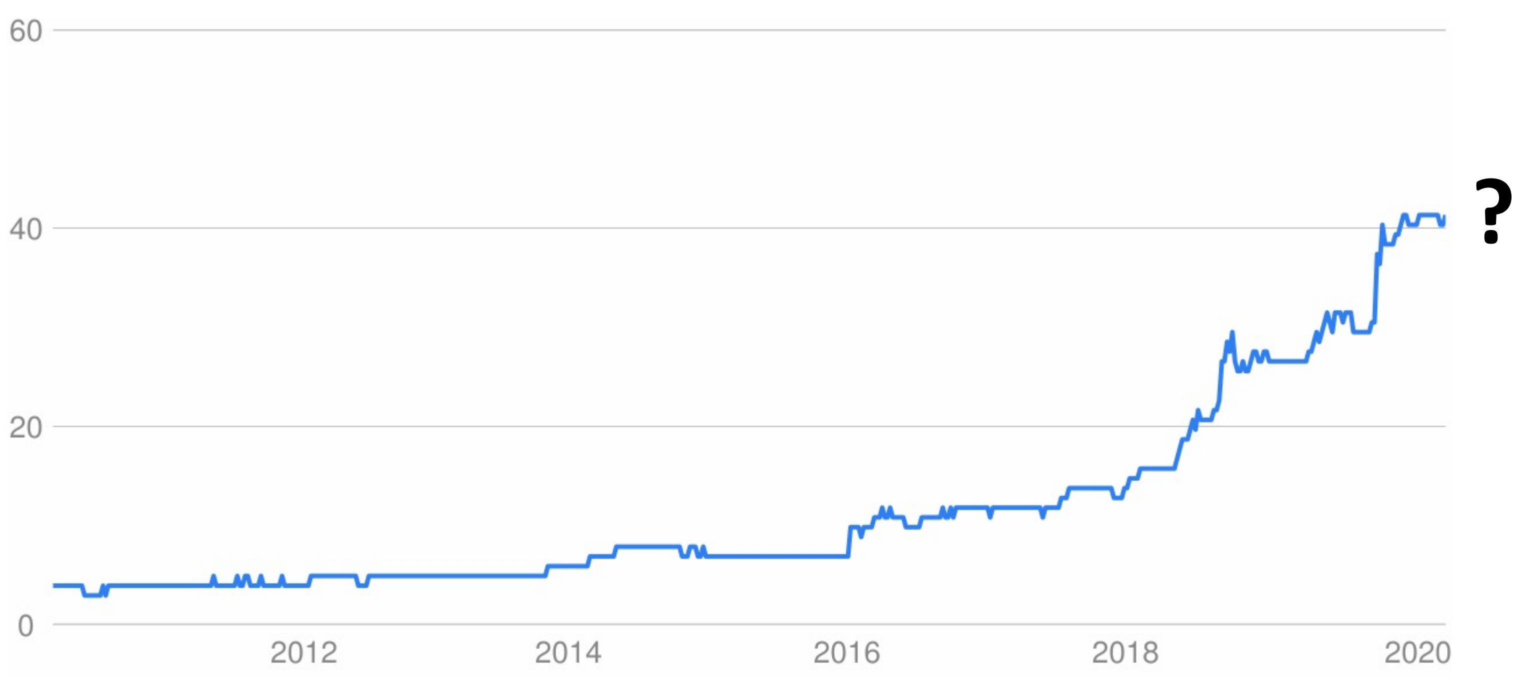
Clasificación de series de tiempo

→ CNN → Contiene un evento anormal?

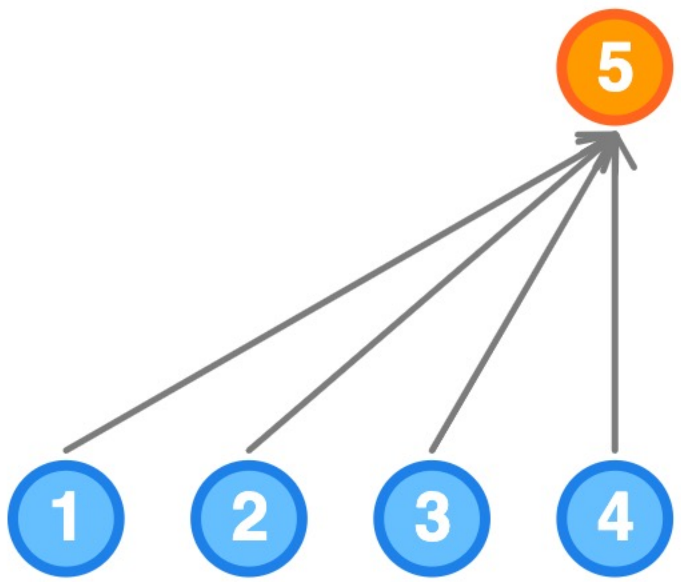
Clasificación de series de tiempo



Predicción en series de tiempo

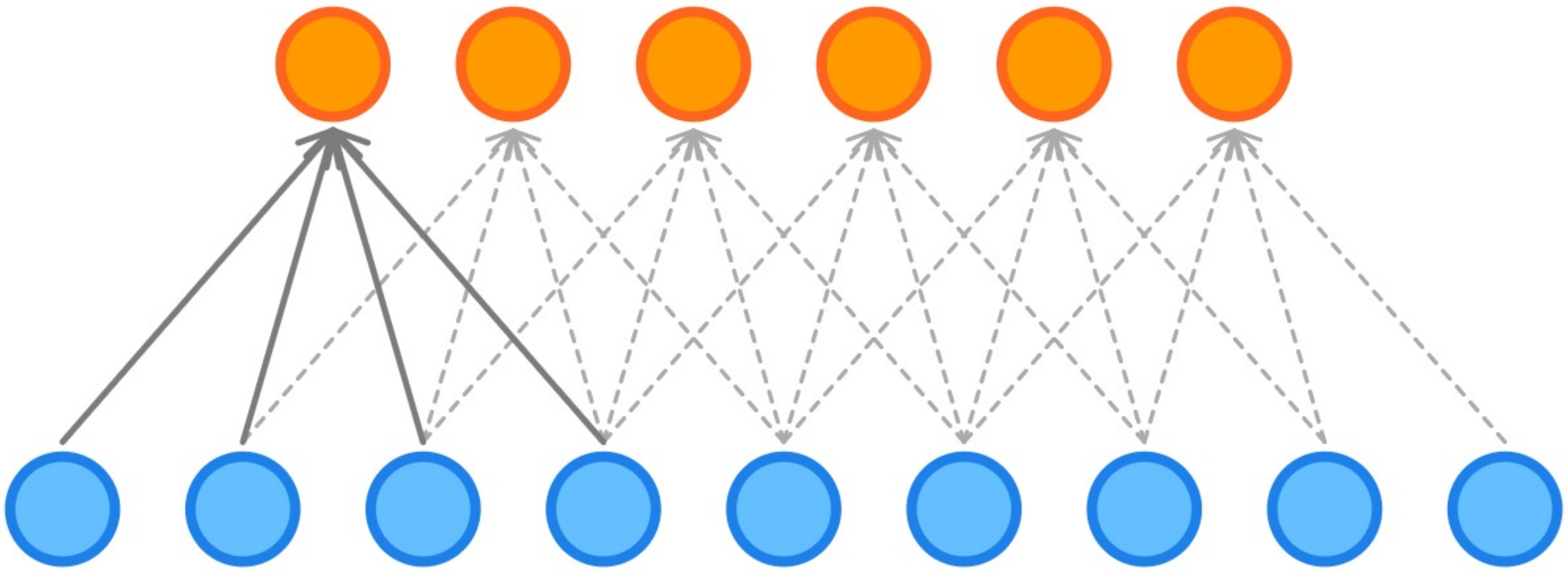


Perceptrón



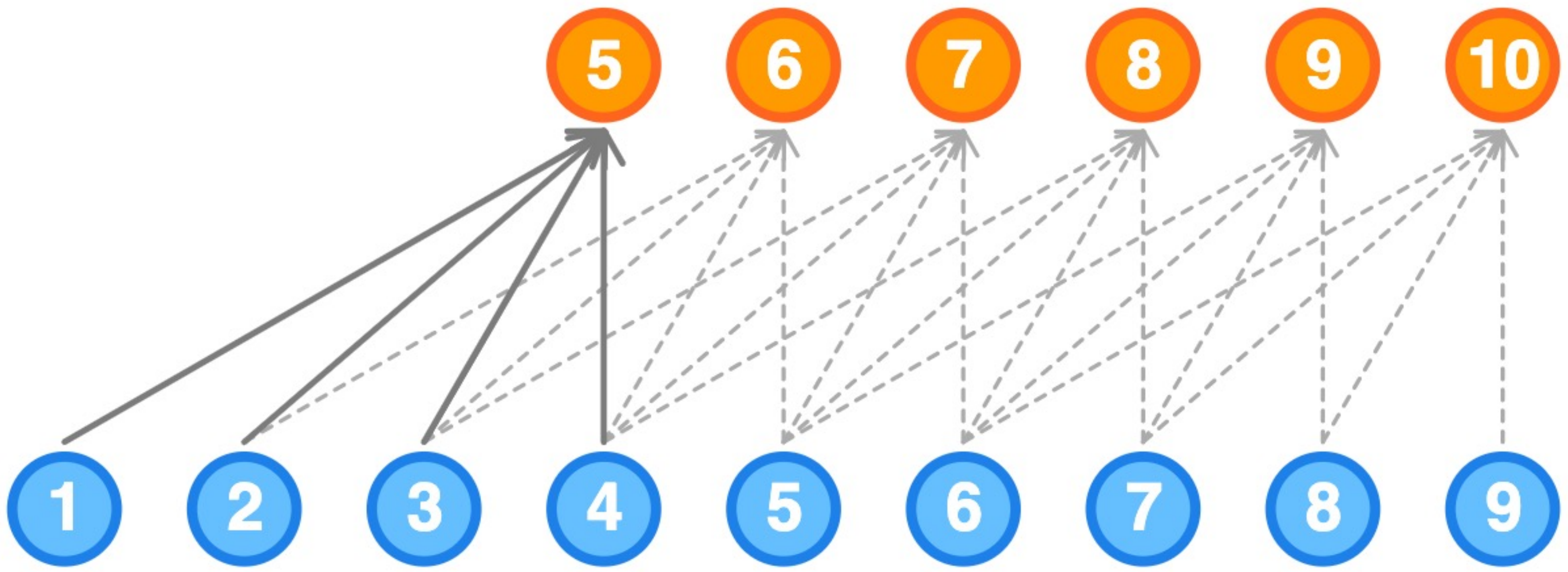
Imágenes de: https://theblog.github.io/post/convolution-in-autoregressive-neural-networks/

Convolución



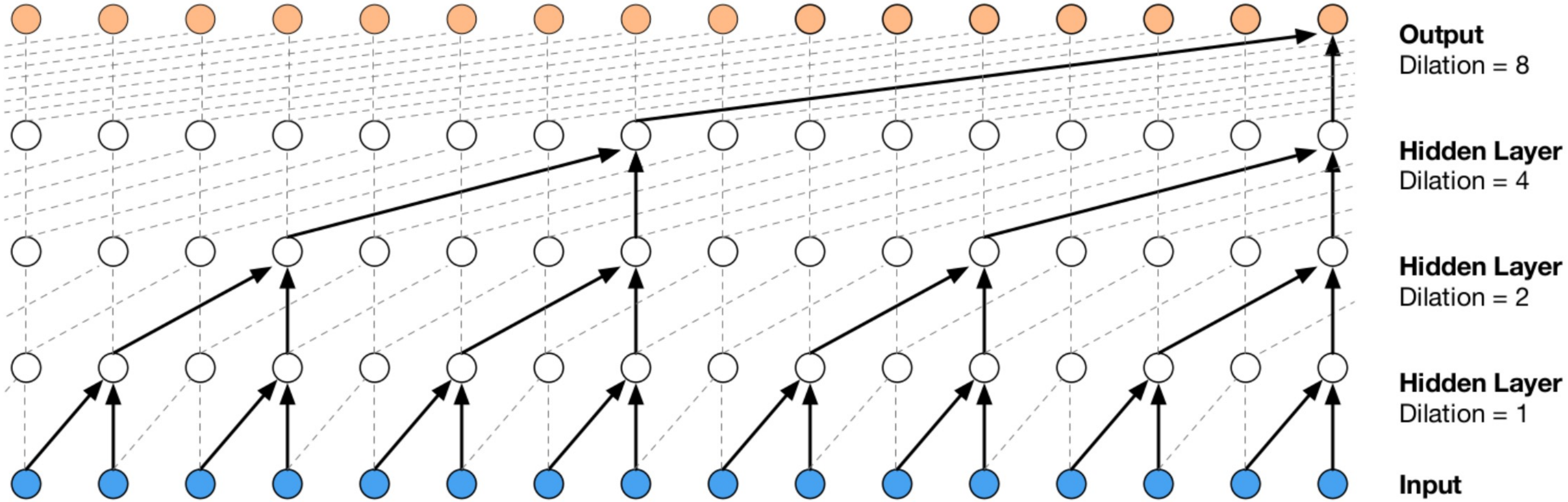
Imágenes de: https://theblog.github.io/post/convolution-in-autoregressive-neural-networks/

Convolución causal



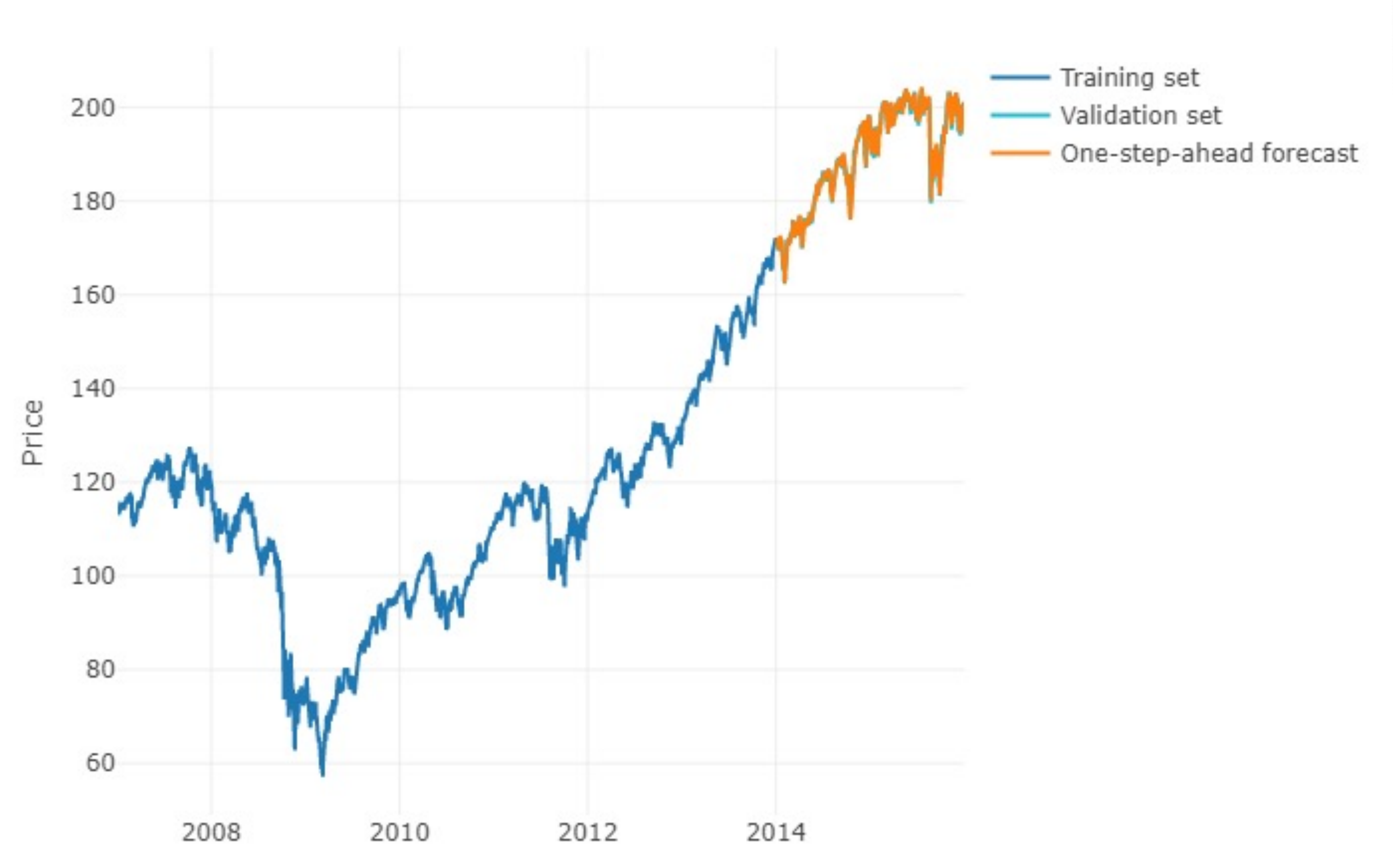
Imágenes de: https://theblog.github.io/post/convolution-in-autoregressive-neural-networks/

Predicción en series de tiempo

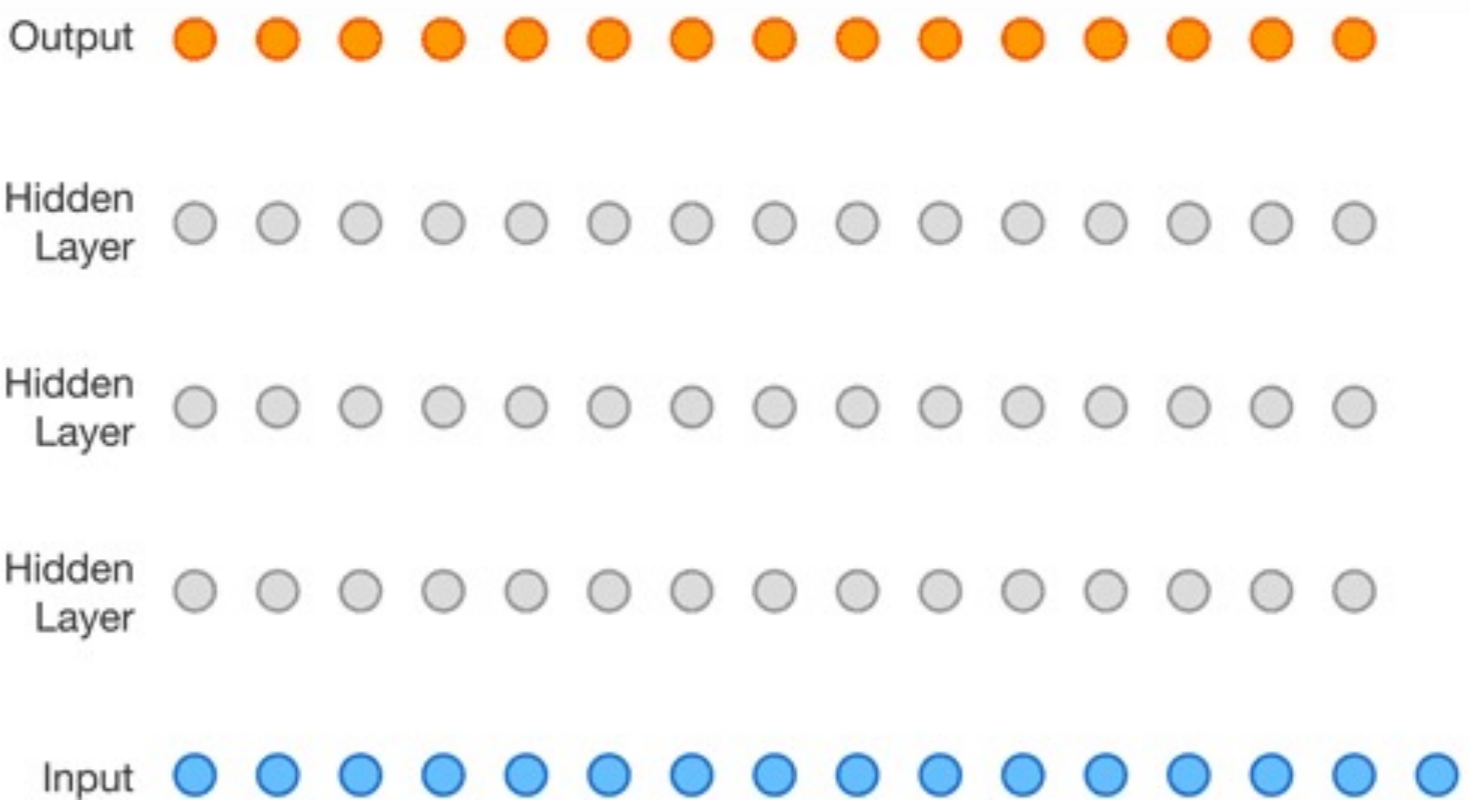


Imágenes de: https://theblog.github.io/post/convolution-in-autoregressive-neural-networks/

Predicción en series de tiempo: más allá de un paso

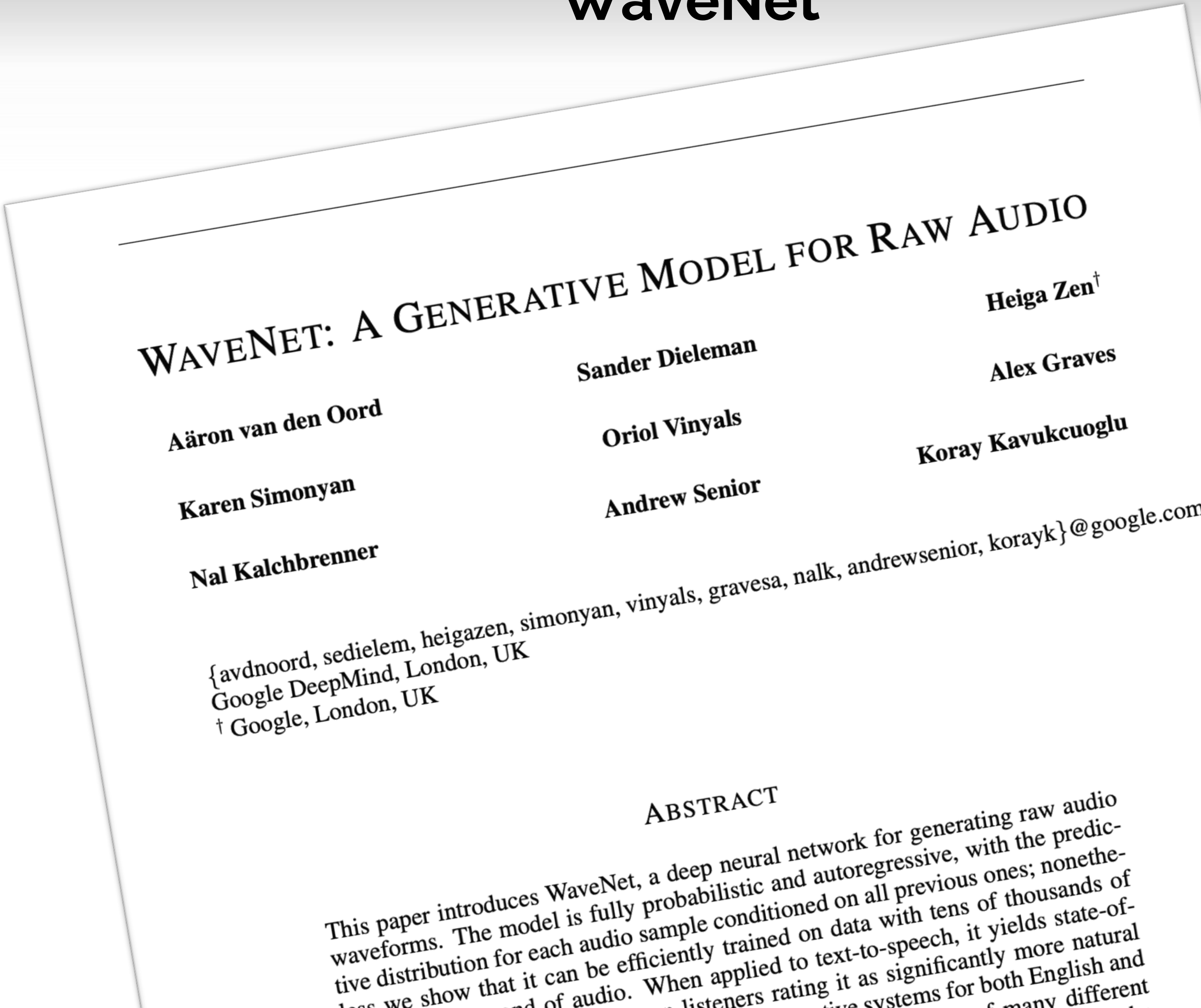


Predicción en series de tiempo: más allá de un paso



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

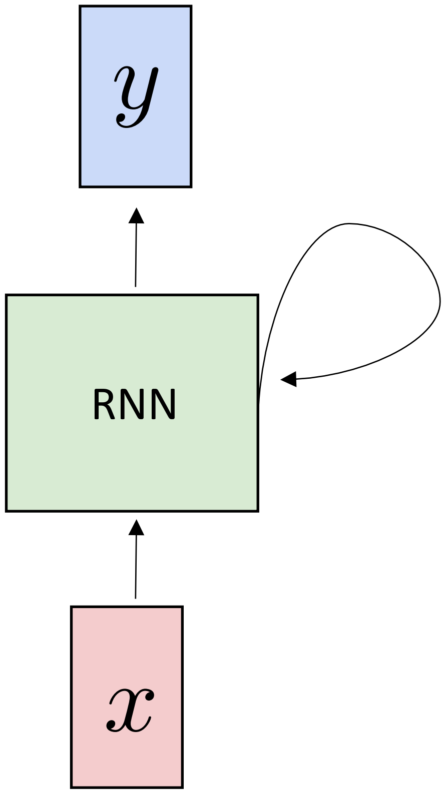
{avdnoord, sedielem, heigazen, simonyan, vinyals, gravesa, nalk, andrewsenior, korayk}@google.com
Google DeepMind, London, UK
† Google, London, UK

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; nonetheless 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 $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(\tau)}$

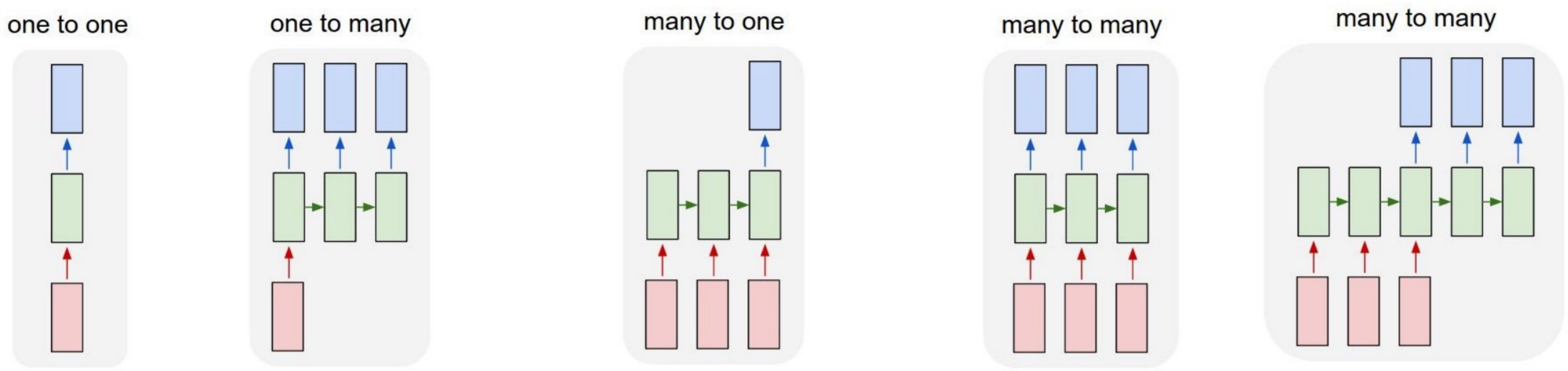
- Así como las redes totalmente convolucionales pueden escalar a cualquier tamaño de imagen, las redes recurrentes pueden en principio procesar cualquier longitud de secuencia.

- La clave son los pesos compartidos (weight sharing) a través del tiempo

Redes neuronales recurrentes

Ej: Sentiment analysis
Secuencia de palabras → Estado sentimental

Ej: Traducción de textos
Secuencia → Secuencia

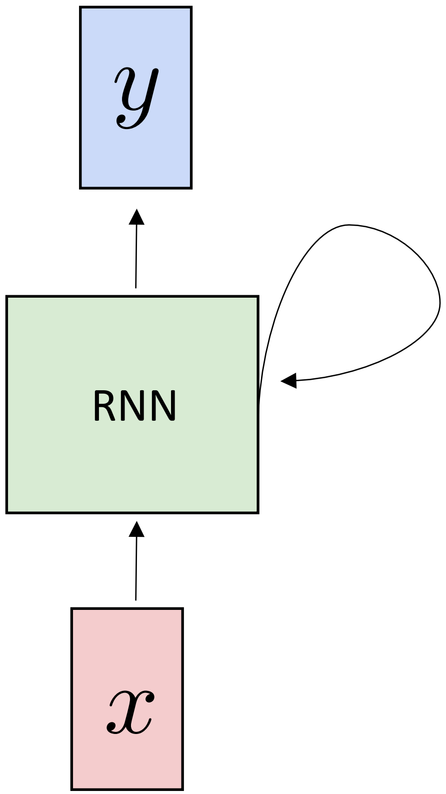


Ej: Image Captioning
Imagen → Secuencia de palabras

Ej: Clasificación de acciones
Video frames → Secuencia de acciones

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

Redes neuronales recurrentes



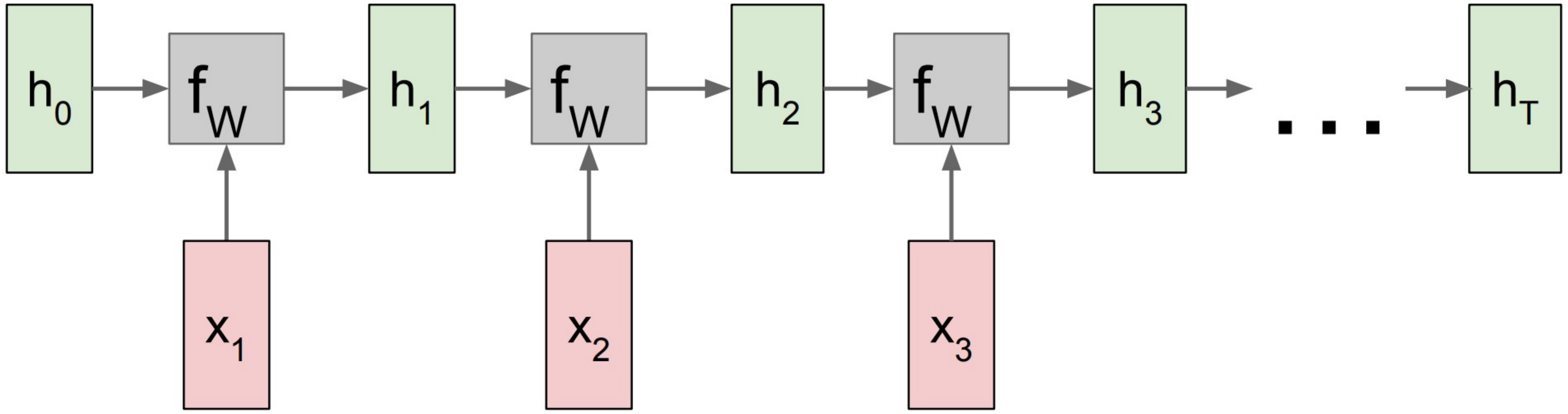
Función parametrizada por W (la misma para todos los instantes *t*)

$$\underbrace{\boxed{h_t}}_{\text{Nuevo Estado}} = \boxed{f_W}(\underbrace{\boxed{h_{t-1}}}_{\text{Antiguo Estado}}, \underbrace{\boxed{x_t}}_{\text{Entrada}})$$

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

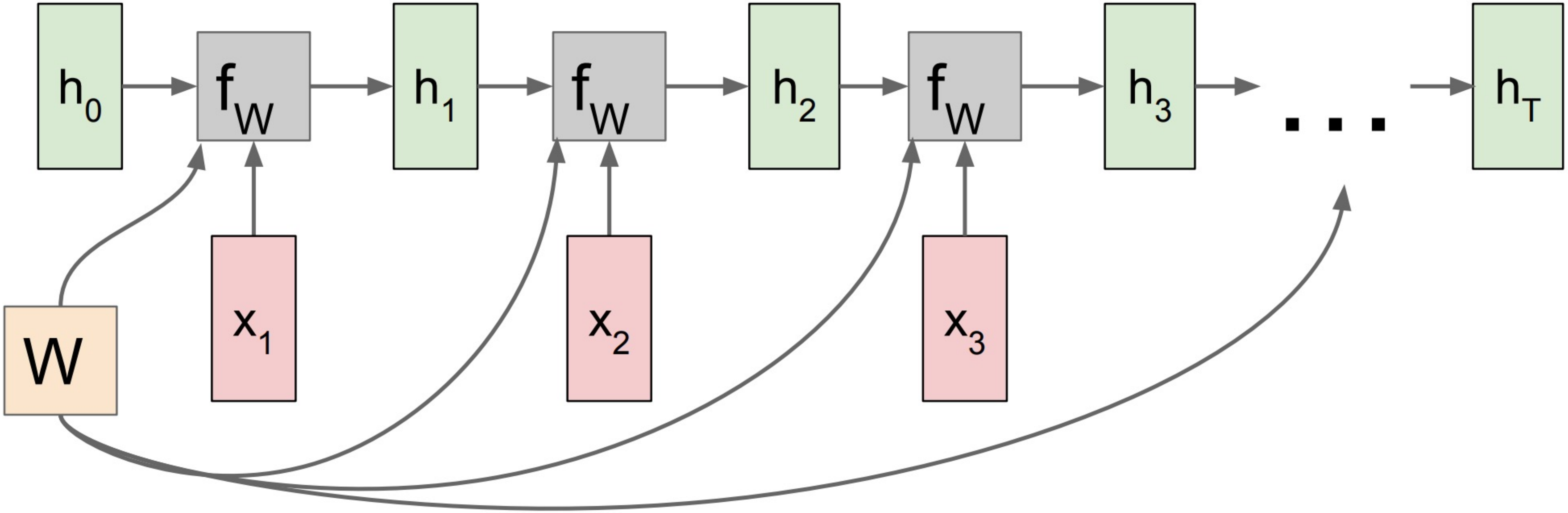
$$y_t = W_{hy}h_t$$

Grafo de cómputo



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

Grafo de cómputo

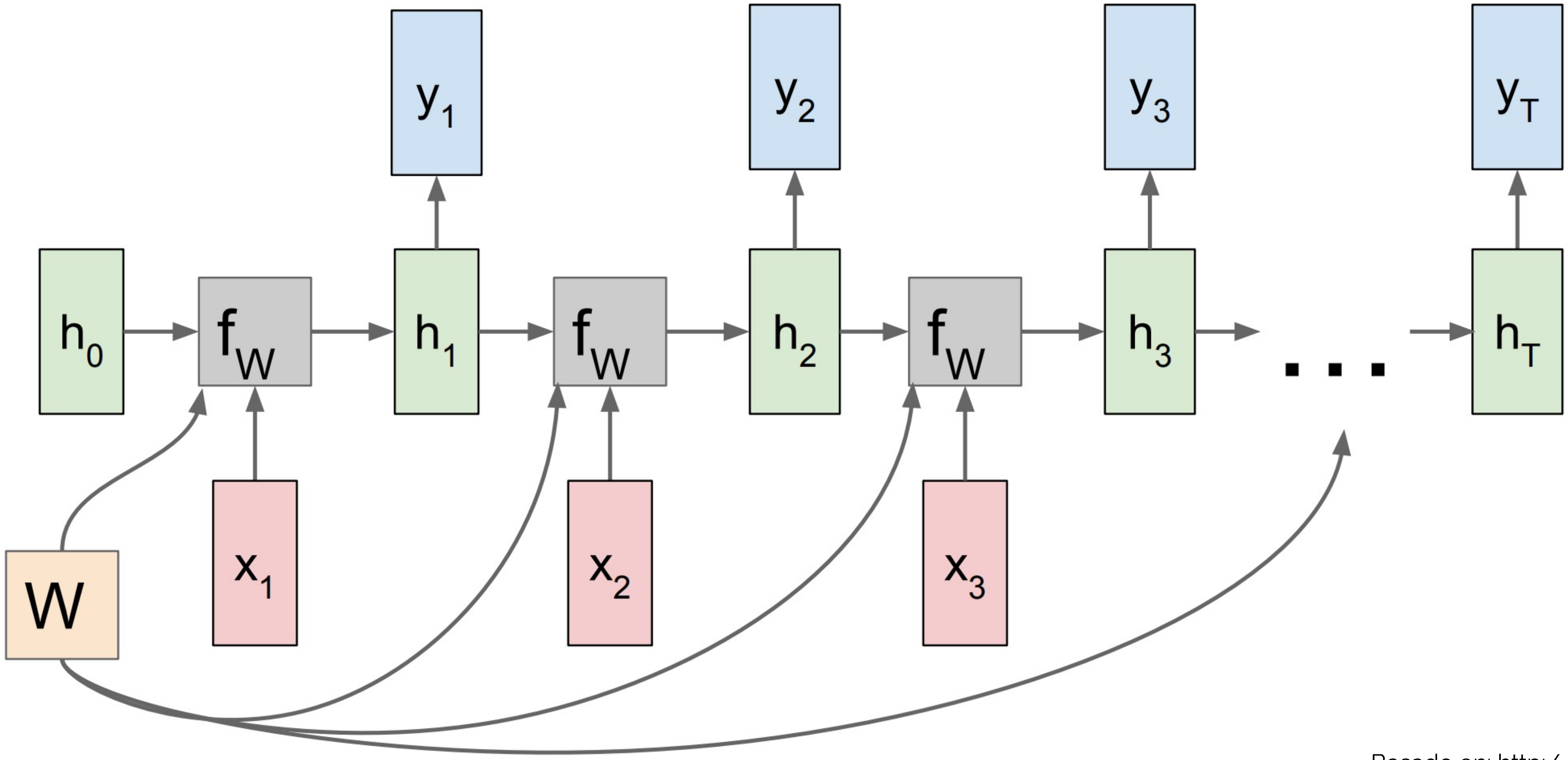


Retropropagación a través del tiempo

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

Grafo de cómputo

Many to many

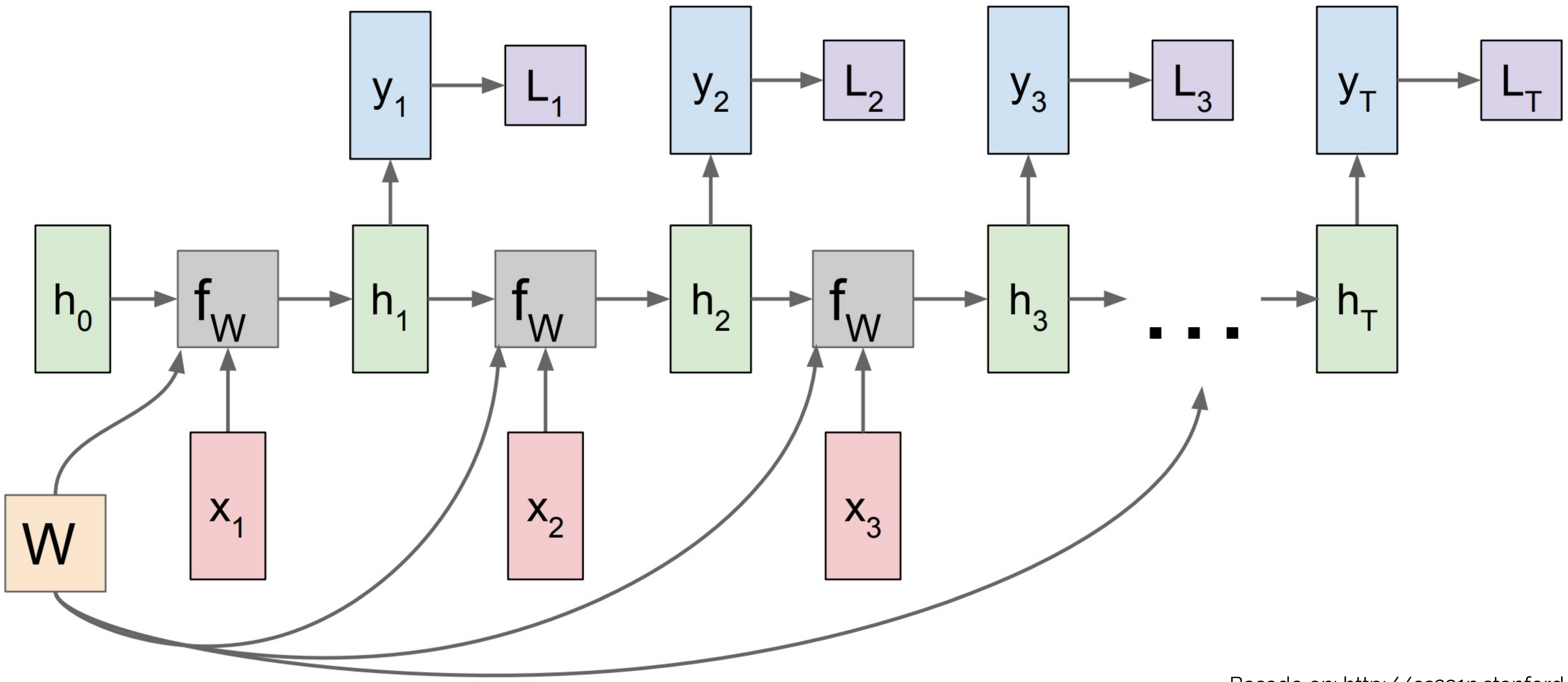


Basado en: http://cs231n.stanford.edu/

Grafo de cómputo

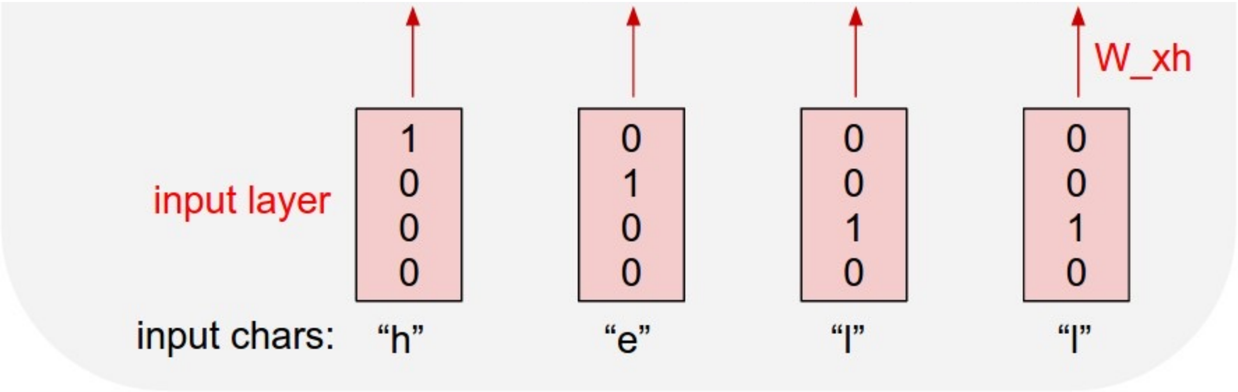
Many to many

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



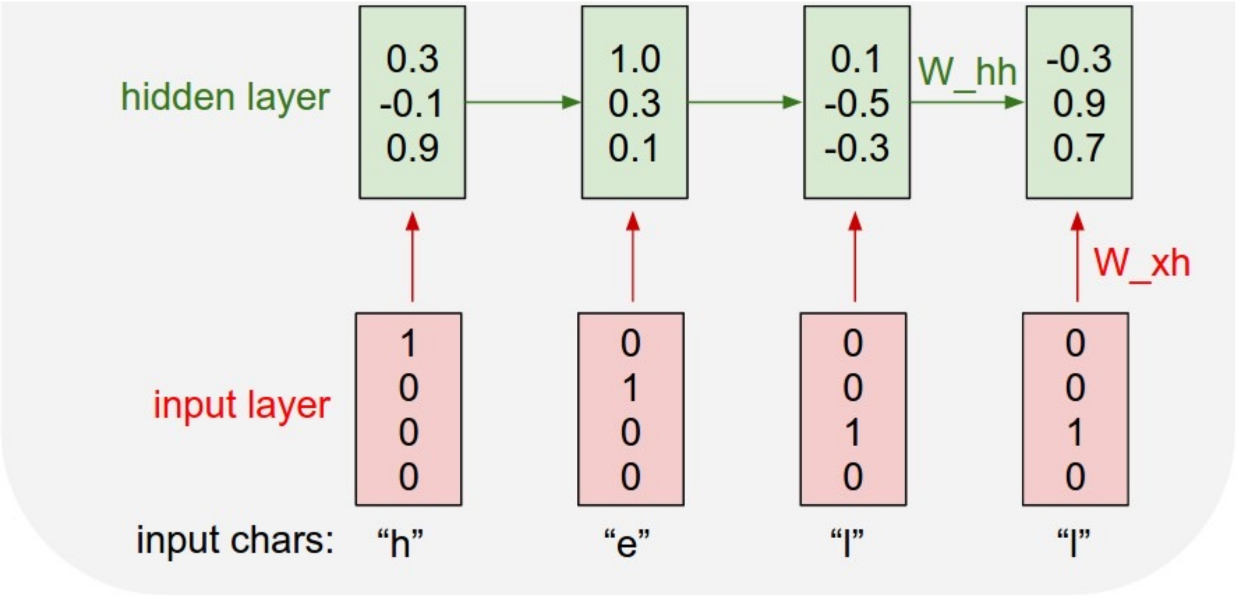
Basado en: http://cs231n.stanford.edu/

Ejemplo: predicción de caracteres



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

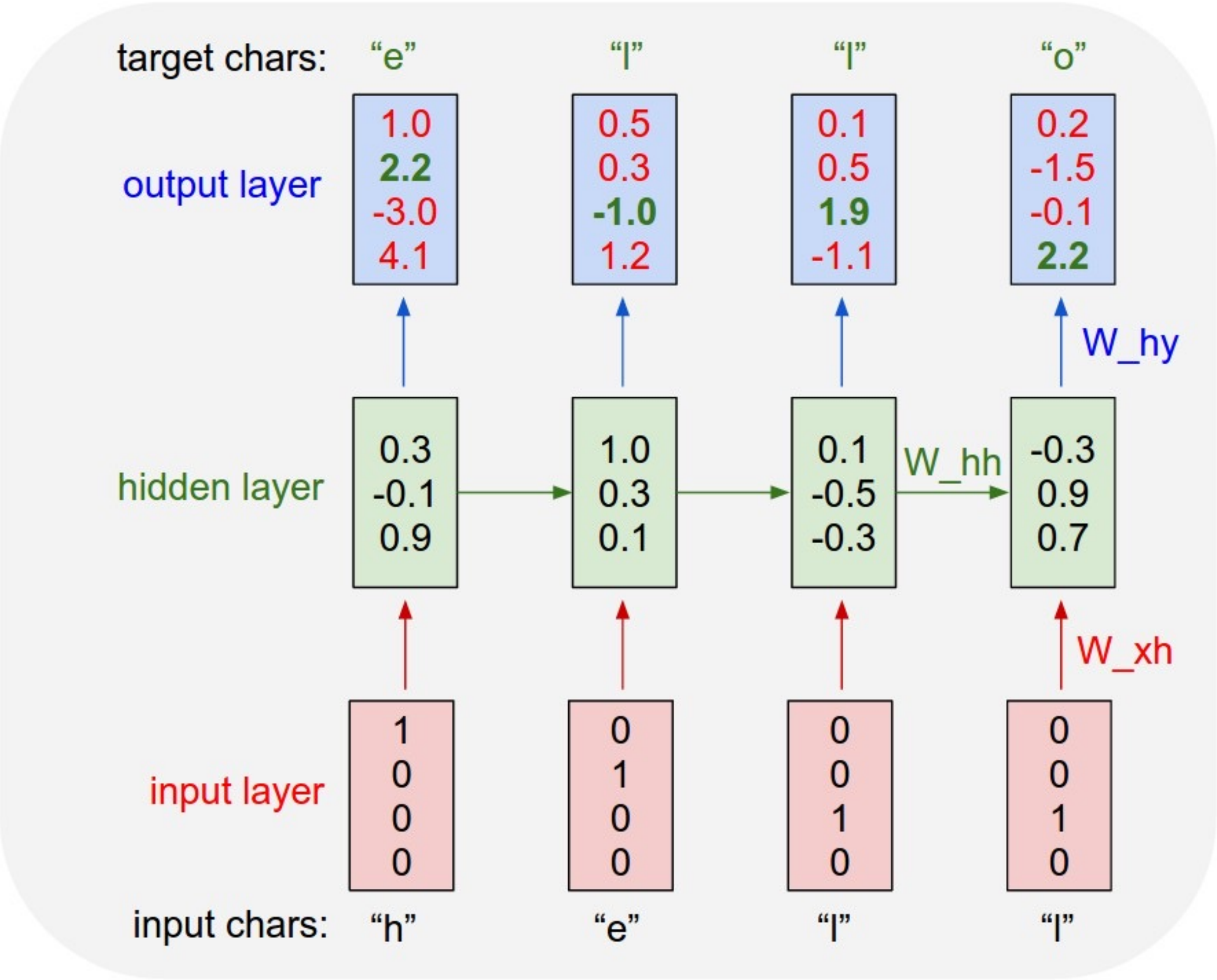
Ejemplo: predicción de caracteres



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

Ejemplo: predicción de caracteres

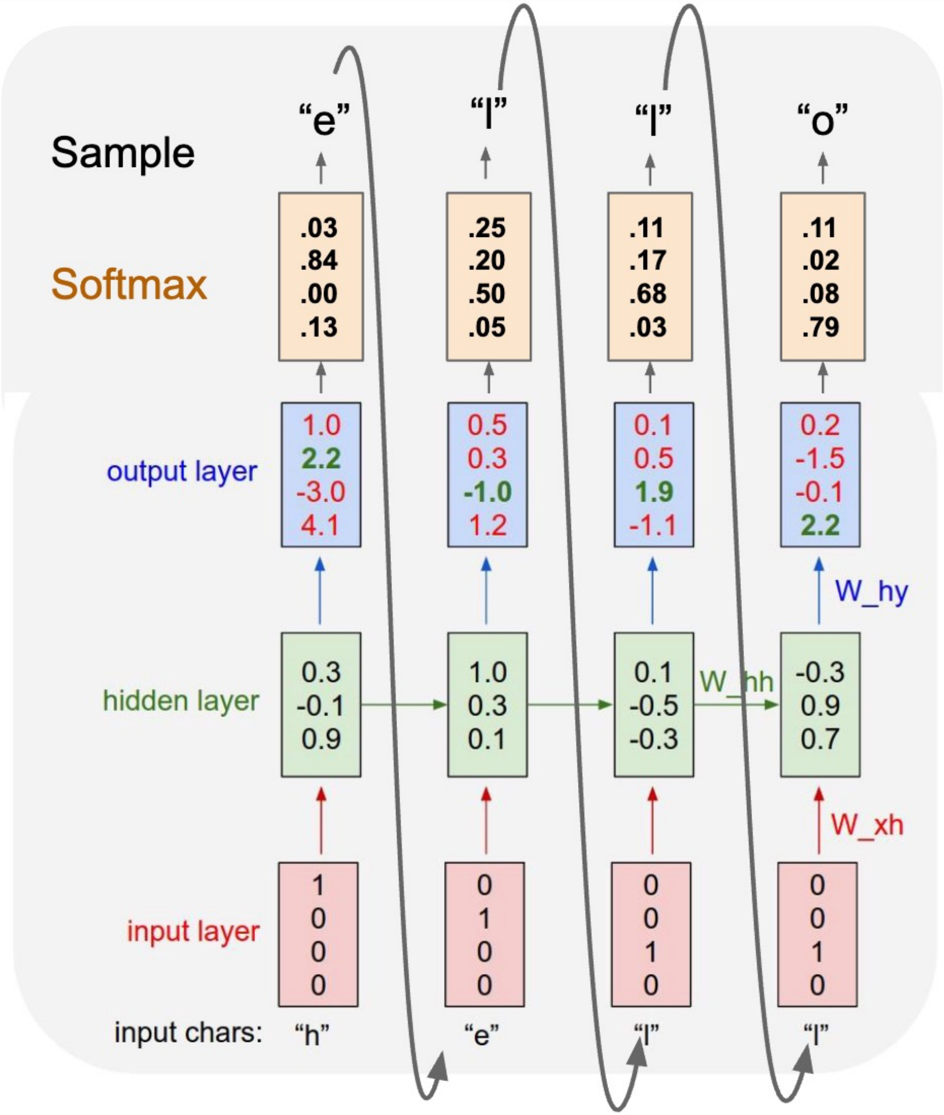
Vocabulario de 4 posibles caracteres



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

Ejemplo: predicción de caracteres

Una vez entrenado el modelo, podemos samplear!



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

Ejemplo: predicción de caracteres

Por ejemplo, un modelo entrenado en Shakespeare

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

Código disponible en: https://github.com/karpathy/char-rnn

Ejemplo: predicción de caracteres

Por ejemplo, un modelo entrenado en papers matemáticos

For $\bigoplus_{n=1,\ldots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, **??** and the fact that any U affine, see Morphisms, Lemma **??**. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma **??** we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example **??**. It may replace S by $X_{spaces,étale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma **??**. Namely, by Lemma **??** we see that R is geometrically regular over S.

Lemma 0.1. *Assume (3) and (3) by the construction in the description.*

Suppose $X = \lim |X|$ (by the formal open covering X and a single map $\underline{Proj}_X(\mathcal{A}) = \mathrm{Spec}(B)$ over U compatible with the complex

$$Set(\mathcal{A}) = \Gamma(X, \mathcal{O}_{X,\mathcal{O}_X}).$$

When in this case of to show that $\mathcal{Q} \to \mathcal{C}_{Z/X}$ is stable under the following result in the second conditions of (1), and (3). This finishes the proof. By Definition **??** *(without element is when the closed subschemes are catenary. If T is surjective we may assume that T is connected with residue fields of S. Moreover there exists a closed subspace $Z \subset X$ of X where U in X' is proper (some defining as a closed subset of the uniqueness it suffices to check the fact that the following theorem*

(1) *f is locally of finite type. Since $S = \mathrm{Spec}(R)$ and $Y = \mathrm{Spec}(R)$.*

Proof. This is form all sheaves of sheaves on X. But given a scheme U and a surjective étale morphism $U \to X$. Let $U \cap U = \coprod_{i=1,\ldots,n} U_i$ be the scheme X over S at the schemes $X_i \to X$ and $U = \lim_i X_i$. □

The following lemma surjective restrocomposes of this implies that $\mathcal{F}_{x_0} = \mathcal{F}_{x_0} = \mathcal{F}_{X,\ldots,0}$.

Lemma 0.2. *Let X be a locally Noetherian scheme over S, $E = \mathcal{F}_{X/S}$. Set $\mathcal{I} = \mathcal{J}_1 \subset \mathcal{I}'_n$. Since $\mathcal{I}^n \subset \mathcal{I}^n$ are nonzero over $i_0 \leq \mathfrak{p}$ is a subset of $\mathcal{J}_{n,0} \circ \overline{A}_2$ works.*

Lemma 0.3. *In Situation* **??**. *Hence we may assume $\mathfrak{q}' = 0$.*

Proof. We will use the property we see that $\mathfrak{p}$ is the mext functor (**??**). On the other hand, by Lemma **??** we see that

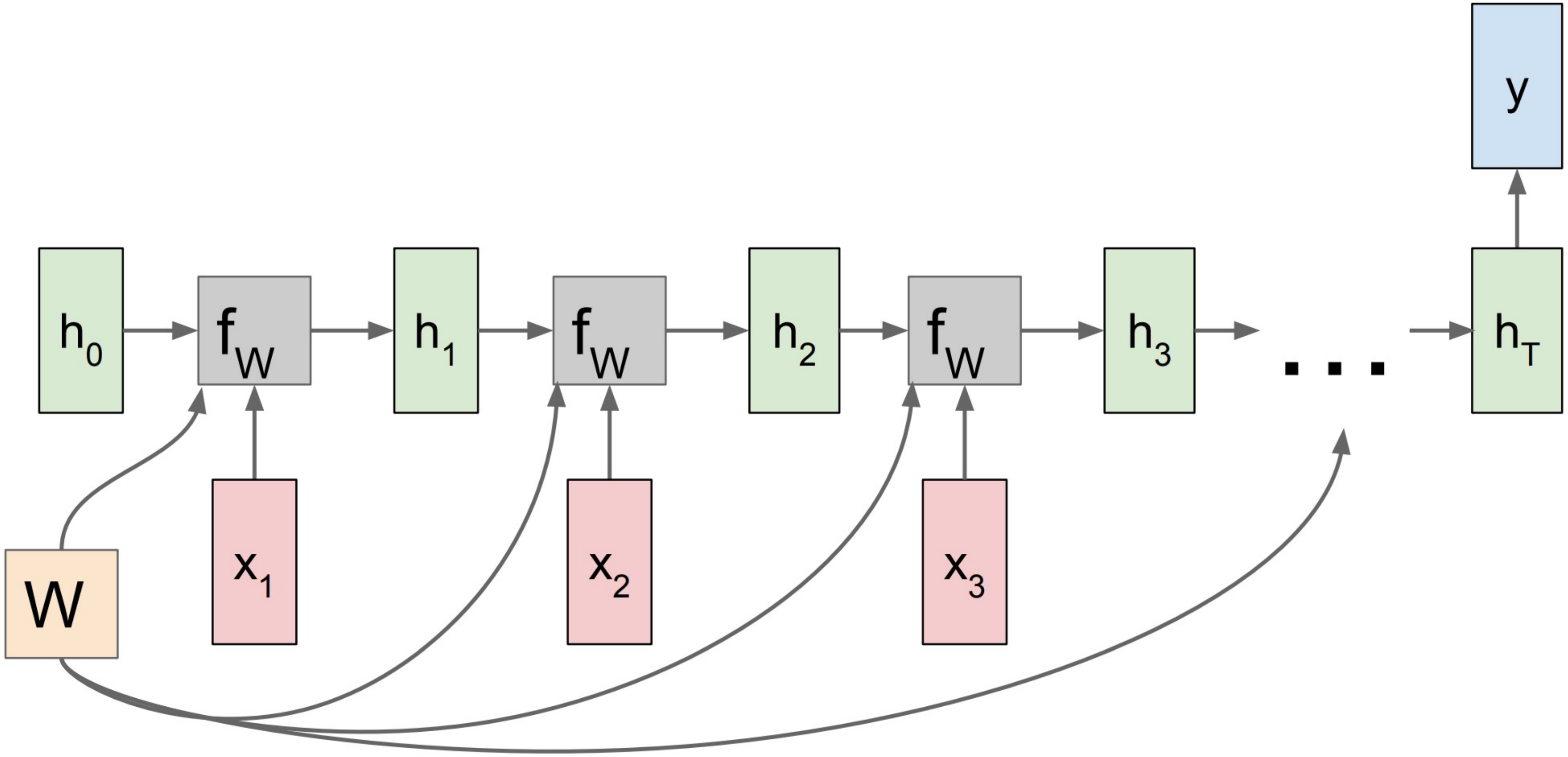
$$D(\mathcal{O}_{X'}) = \mathcal{O}_X(D)$$

where K is an F-algebra where δ_{n+1} is a scheme over S. □

Código disponible en: https://github.com/karpathy/char-rnn

Grafo de cómputo

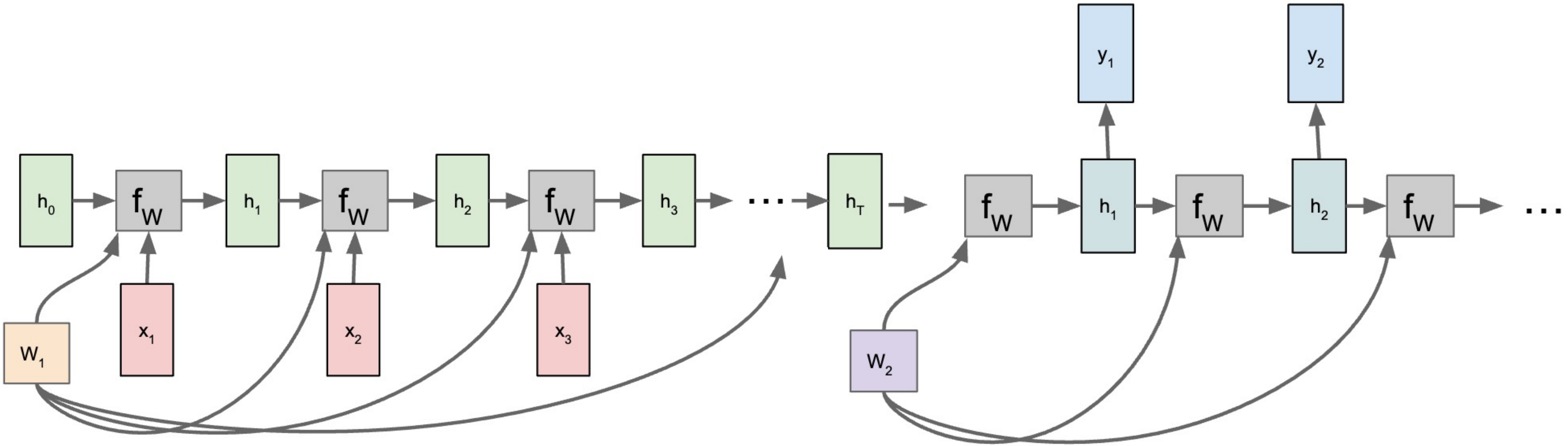
Many to one



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

Grafo de cómputo

Sequence to Sequence: Many to one + one to many



Encoder: codifica la secuencia de entrada en un embedding

Decoder: decodifica la secuencia de salida desde el embedding

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

Ejemplo de uso para image captioning

Deep Visual-Semantic Alignments for Generating Image Descriptions

Andrej Karpathy Li Fei-Fei
Department of Computer Science, Stanford University
{karpathy,feifeili}@cs.stanford.edu

Abstract

We present a model that generates natural language descriptions of images and their regions. Our approach leverages datasets of images and their sentence descriptions to learn about the inter-modal correspondences between language and visual data. Our alignment model is based on a novel combination of Convolutional Neural Networks over image regions, bidirectional Recurrent Neural Networks over sentences, and a structured objective that aligns the two modalities through a multimodal embedding. We then describe a Multimodal Recurrent Neural Network architecture that uses the inferred alignments to learn to generate novel descriptions of image regions. We demonstrate that our alignment model produces state of the art results in retrieval experiments on Flickr8K, Flickr30K and MSCOCO datasets. We then show that the generated descriptions significantly outperform retrieval baselines on both full images and on a new dataset of region-level annotations.

1. Introduction

A quick glance at an image is sufficient for a human to point out and describe an immense amount of details about

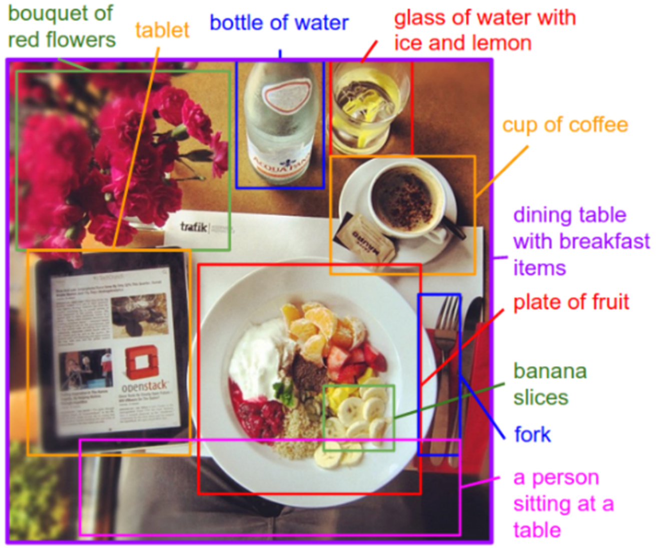


Figure 1. Motivation/Concept Figure: Our model treats language as a rich label space and generates descriptions of image regions.

generating dense descriptions of images (Figure 1). The primary challenge towards this goal is in the design of a model that is rich enough to simultaneously reason about contents of images and their representation in the domain

Dataset of images and sentence descriptions

training image

"A Tabby cat is leaning on a wooden table, with one paw on a laser mouse and the other on a black laptop"

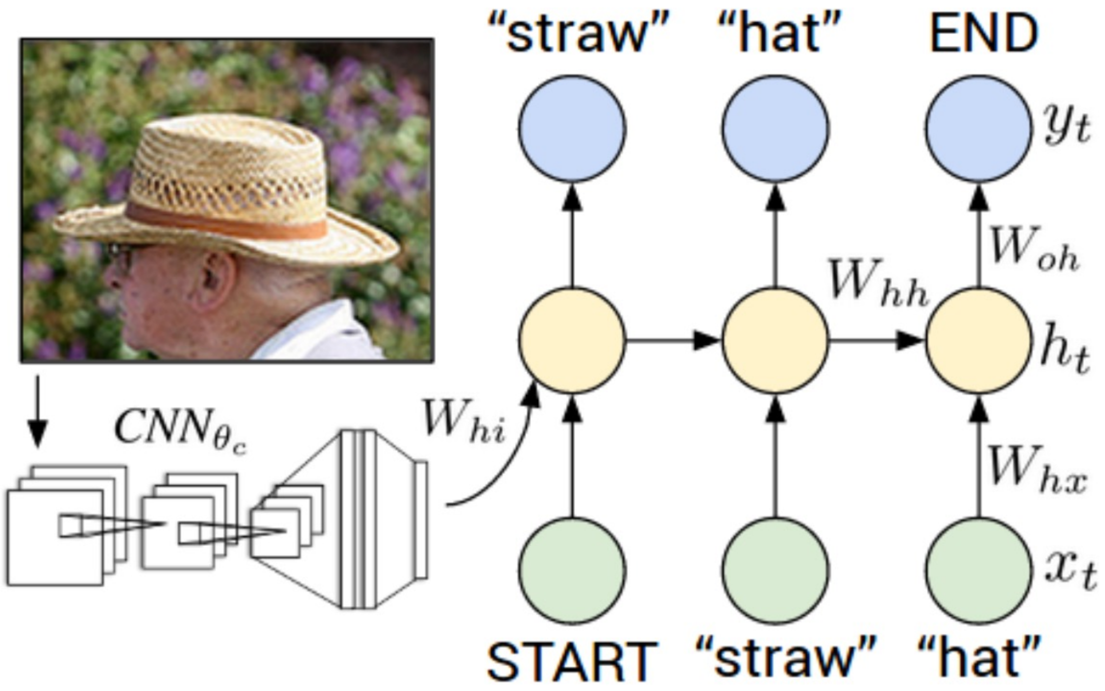
Ejemplo de uso para image captioning

before:

$h = \tanh(W_{xh} * x + W_{hh} * h)$

now:

$h = \tanh(W_{xh} * x + W_{hh} * h + W_{ih} * v)$



Código disponible en: https://github.com/karpathy/char-rnn

Clase 5

Redes neuronales más allá de la clasificación de imágenes

Enzo Ferrante

eferrante@sinc.unl.edu.ar

@enzoferrante