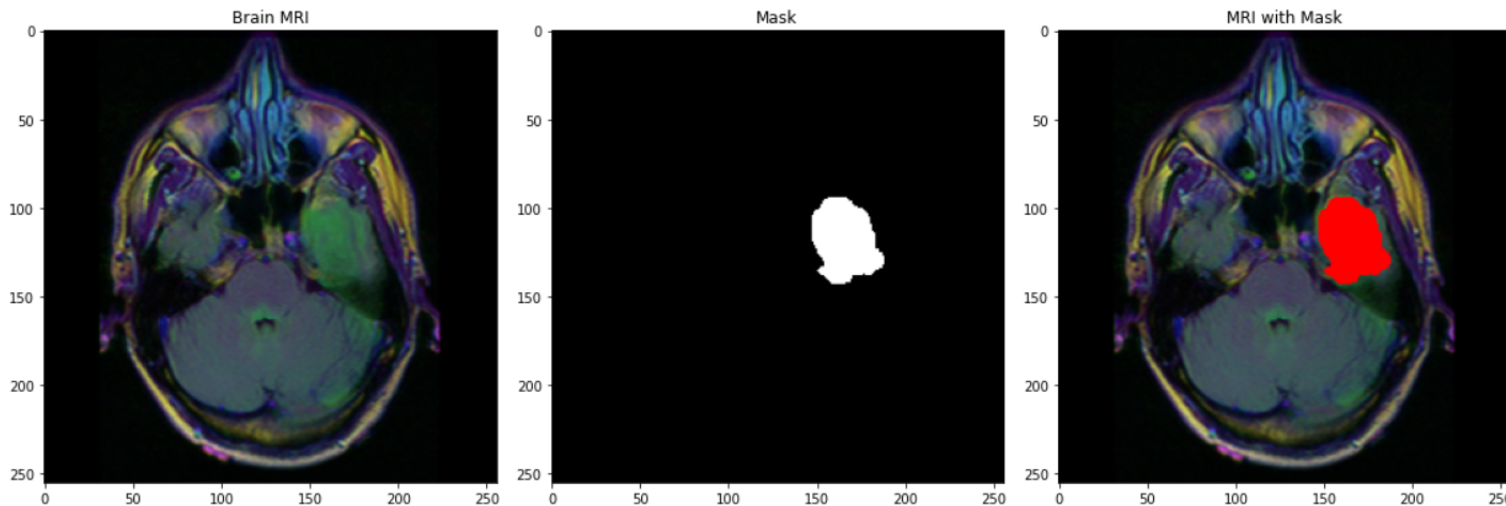


- La inteligencia artificial está revolucionando la atención sanitaria en muchas áreas como:
  - Diagnóstico de enfermedades con imágenes médicas
  - Robots quirúrgicos
  - Maximización de la eficiencia hospitalaria
- Se espera que el mercado de la salud de la IA alcance los \$ 45,2 mil millones de dólares para 2026 a partir de la valoración actual de \$ 4,9 mil millones de dólares.
- Se ha demostrado que el aprendizaje profundo es superior en la detección de enfermedades a partir de rayos X, resonancias magnéticas y tomografías computarizadas, lo que podría mejorar significativamente la velocidad y precisión del diagnóstico.
- Lista completa de nuevas empresas: <https://research.aimultiple.com/looking-for-better-medical-imaging-for-early-diagnostic-and-monitoring-contact-the-leading-vendors-here/>

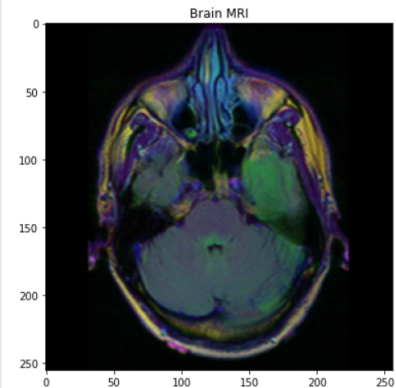


- En este caso de estudio, asumiremos que usted trabaja como consultor de IA / ML y ha sido contratado por una empresa de diagnóstico médico en Nueva York.
- Se te ha encomendado la tarea de mejorar la velocidad y la precisión de la detección y localización de tumores cerebrales en función de las imágenes por resonancia magnética.
- Esto reduciría drásticamente el coste del diagnóstico de cáncer y ayudaría en el diagnóstico temprano de tumores, lo que esencialmente salvaría la vida.
- El equipo ha recopilado imágenes por resonancia magnética del cerebro y se ha acercado a nosotros para desarrollar un modelo que pueda detectar y localizar tumores.
- Se nos han proporcionado 3929 resonancias magnéticas cerebrales junto con la ubicación del tumor cerebral.



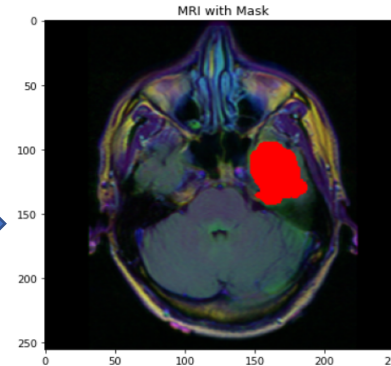
# PIPELINE DE APRENDIZAJE PROFUNDO EN CAPAS PARA REALIZAR CLASIFICACIÓN Y SEGMENTACIÓN

IMÁGENES DE ENTRADA  
MRI CEREBRAL



TUMOR  
DETECTADO

RESUNET  
MODELO DE  
SEGMENTACIÓN



DETECTAR LA  
UBICACIÓN DEL  
TUMOR A NIVEL DE  
PÍXELES

RESNET  
CLASIFICADOR  
DEEP LEARNING

TUMOR NO  
DETECTADO

FIN

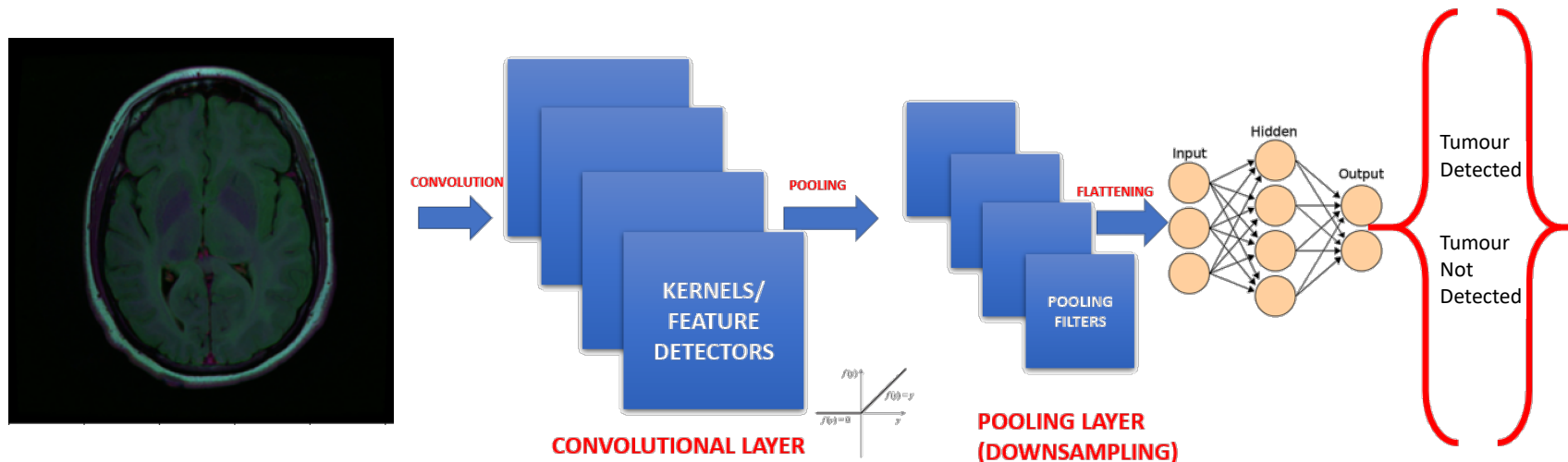
PACIENTE SANO  
(OJALÁ TODOS  
ESTUVIERAN  
AQUÍ)

*"CON IA Y MENTES BRILLANTES COMO LA  
TUYA, CURAREMOS LA ENFERMEDAD"*



# REDES NEURALES CONVOLUCIONALES (REPASO)

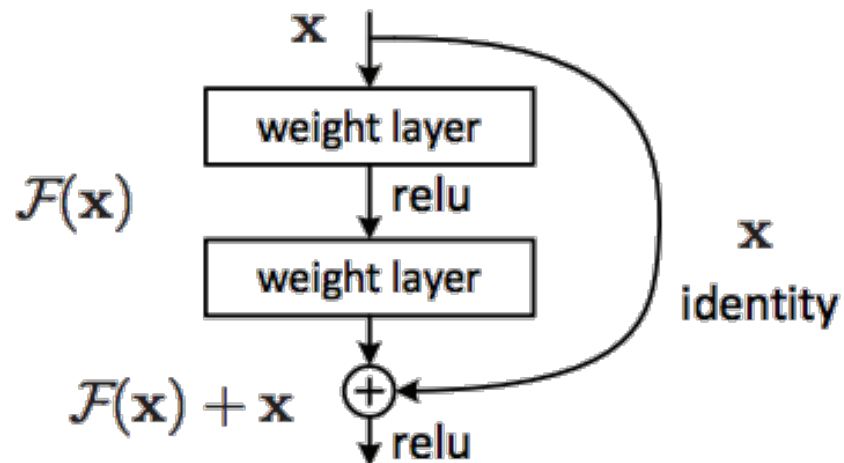
- Las primeras capas de CNN se utilizan para extraer características generales de alto nivel.
- El último par de capas se utilizan para realizar la clasificación (en una tarea específica).
- Los campos respectivos locales escanean la imagen primero en busca de formas simples como bordes / líneas
- Estos bordes son luego recogidos por la capa posterior para formar entidades más complejas.





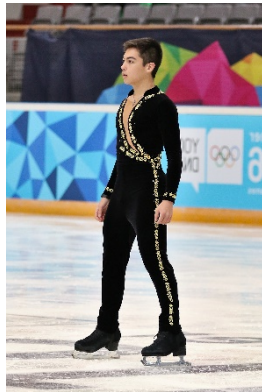
# RESNET (RED RESIDUAL) (REPASO)

- A medida que las RNC se hacen más profundas, tiende a ocurrir el desvanecimiento del gradiente que afectan negativamente el rendimiento de la red.
- El problema del desvanecimiento del gradiente ocurre cuando el gradiente se propaga hacia atrás a capas anteriores, lo que da como resultado un gradiente muy pequeño.
- La red neuronal residual incluye la función de "omisión de conexión" que permite el entrenamiento de 152 capas sin problemas del desvanecimiento del gradiente.
- ResNet funciona agregando "asignaciones de identidad" en la parte superior de la RNC.
- ImageNet contiene 11 millones de imágenes y 11.000 categorías.
- ImageNet se utiliza para entrenar la red profunda ResNet.



# APRENDIZAJE POR TRANSFERENCIA

- El aprendizaje por transferencia es una técnica de aprendizaje automático en la que una red que ha sido entrenada para realizar una tarea específica se reutiliza como punto de partida para otra tarea similar.
- El aprendizaje por transferencia se usa ampliamente ya que partir de modelos previamente entrenados puede reducir drásticamente el tiempo de cálculo requerido si el entrenamiento se realiza desde cero.



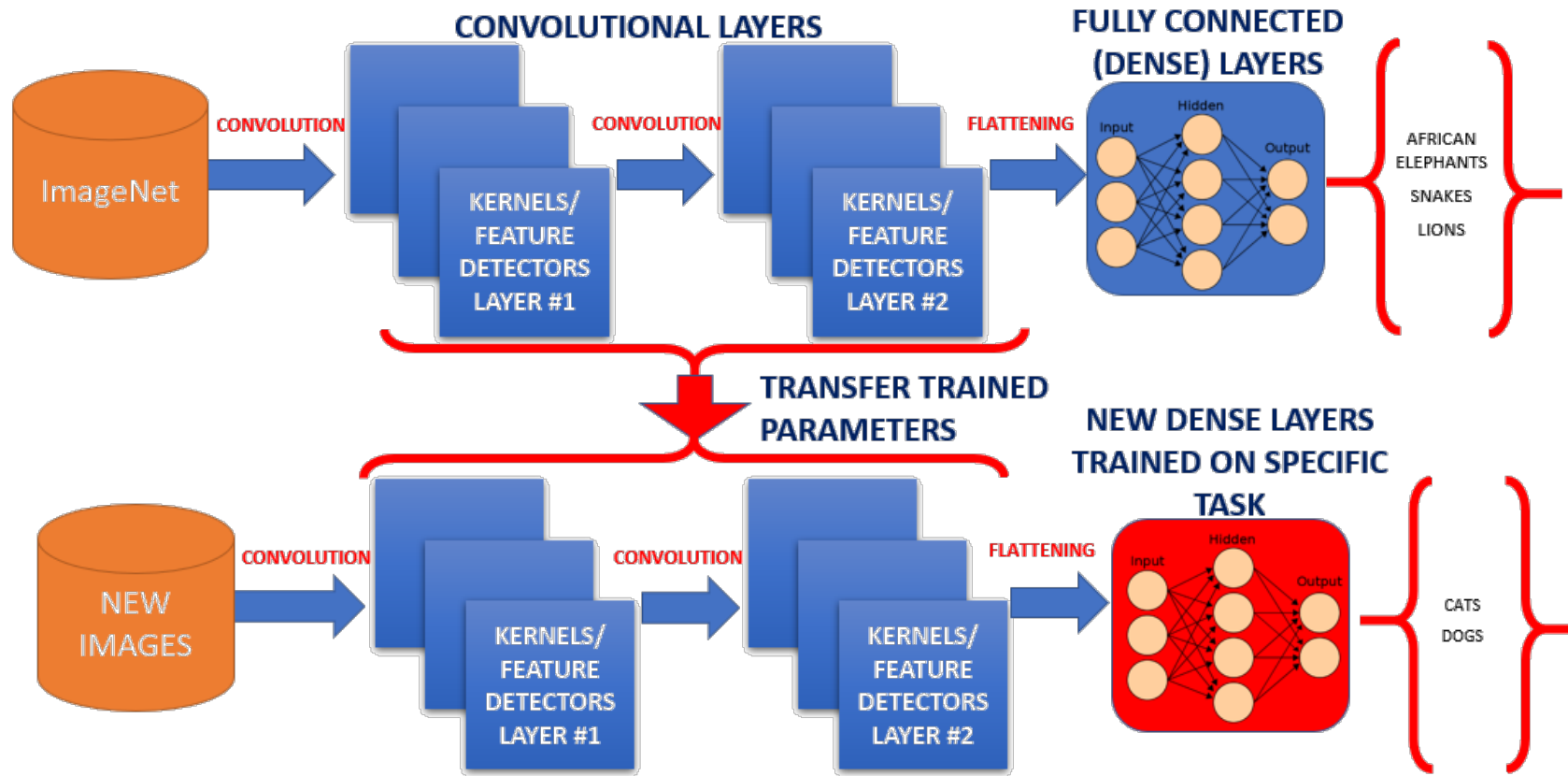
TRANSFERENCIA DEL  
CONOCIMIENTO



- Photo Credit: [https://commons.wikimedia.org/wiki/File:Lillehammer\\_2016\\_-\\_Figure\\_Skating\\_Men\\_Short\\_Program\\_-\\_Camden\\_Pulkinen\\_2.jpg](https://commons.wikimedia.org/wiki/File:Lillehammer_2016_-_Figure_Skating_Men_Short_Program_-_Camden_Pulkinen_2.jpg)
- Photo Credit: [https://commons.wikimedia.org/wiki/Alpine\\_skiing#/media/File:Andrej\\_%C5%A0porn\\_at\\_the\\_2010\\_Winter\\_Olympic\\_downhill.jpg](https://commons.wikimedia.org/wiki/Alpine_skiing#/media/File:Andrej_%C5%A0porn_at_the_2010_Winter_Olympic_downhill.jpg)
- Citations: Olga Russakovsky, *Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei.
- ImageNet Large Scale Visual Recognition Challenge. arXiv:1409.0575, 2014.



# PROCESO DEL APRENDIZAJE POR TRANSFERENCIA



# TRANSFERIR ESTRATEGIAS DE ENTRENAMIENTO APRENDIZAJE

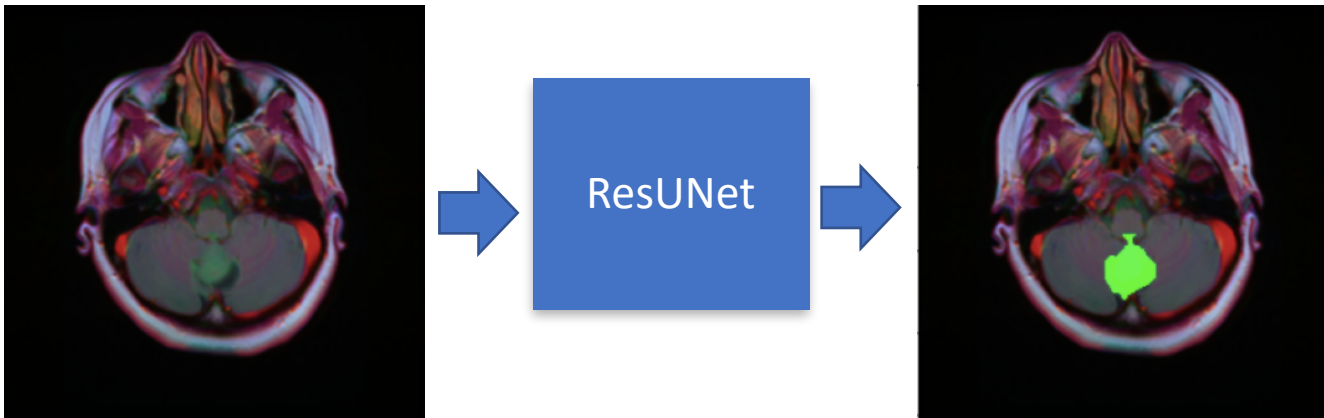
- Pasos de la estrategia n°1:
  - Congelar los pesos de la red RNC entrenada de las primeras capas.
  - Entrenar solo las capas densas recién agregadas (con pesos inicializados aleatoriamente).
- Pasos de la estrategia n°2:
  - Inicializar la red de RNC con los pesos previamente entrenados
  - Volvemos a entrenar a toda la RNC mientras establece la tasa de aprendizaje para que sea muy pequeña, esto es fundamental para garantizar que no cambie agresivamente los pesos entrenados.
- Las ventajas del aprendizaje por transferencia son:
  - Proporciona un progreso de entrenamiento rápido, no tiene que empezar desde cero utilizando pesos inicializados aleatoriamente
  - Se puede utilizar un pequeño conjunto de datos de entrenamiento para lograr resultados increíbles





# ¿QUÉ ES LA SEGMENTACIÓN DE IMÁGENES?

- El objetivo de la segmentación de imágenes es comprender y extraer información de las imágenes a nivel de píxel.
- La segmentación de imágenes se puede utilizar para el reconocimiento y la localización de objetos, lo que ofrece un valor tremendo en muchas aplicaciones, como imágenes médicas y automóviles autónomos, etc.
- El objetivo de la segmentación de imágenes es entrenar una red neuronal para producir una máscara de píxeles de la imagen.
- Las técnicas modernas de segmentación de imágenes se basan en un enfoque de aprendizaje profundo que hace uso de arquitecturas comunes como RNC, RCT (redes de convolución total) y codificadores-decodificadores profundos.
- Utilizamos la arquitectura ResUNet para resolver la tarea actual.



# ¿QUÉ ES LA SEGMENTACIÓN DE IMÁGENES?

- ¿Recuerdas cuándo aplicamos RNC para problemas de clasificación de imágenes? Tuvimos que convertir la imagen en un vector y posiblemente agregar un encabezado de clasificación al final.
- Sin embargo, en el caso de Unet, convertimos (codificamos) la imagen en un vector seguido de un muestreo ascendente (decodificamos) de nuevo en una imagen.
- En el caso de Unet, la entrada y la salida tienen el mismo tamaño, por lo que se conserva el tamaño de la imagen.
- Para RNC clásicas: generalmente se utilizan cuando se necesita clasificar la imagen completa como una etiqueta de clase.
- Para Unet: se realiza una clasificación a nivel de píxeles.
- U-net formula una función de pérdida para cada píxel de la imagen de entrada.
- La función Softmax se aplica a cada píxel, lo que hace que el problema de segmentación funcione como un problema de clasificación donde la clasificación se realiza en cada píxel de la imagen.

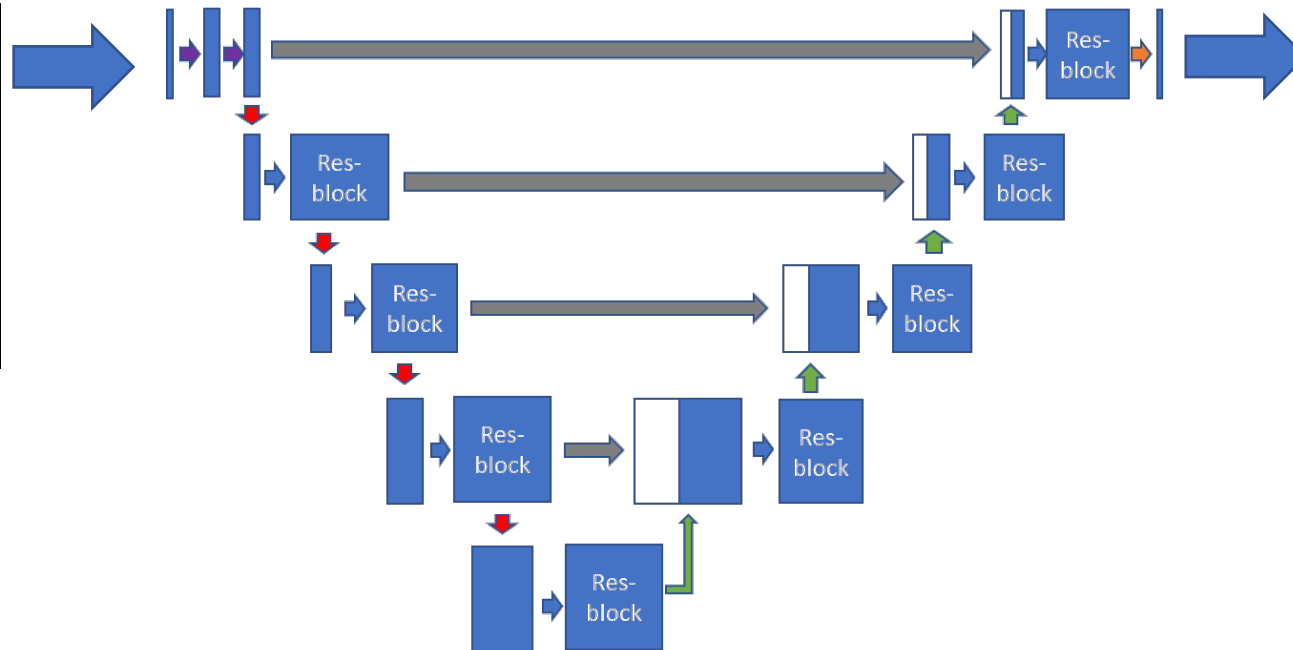
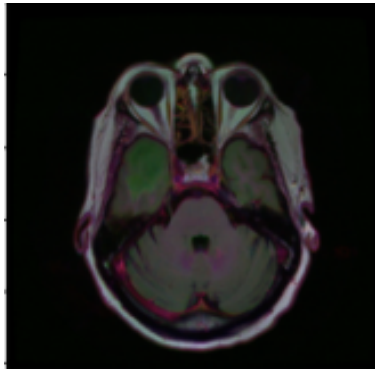
Great article by Aditi Mittal: <https://towardsdatascience.com/introduction-to-u-net-and-res-net-for-image-segmentation-9afcb432ee2f>



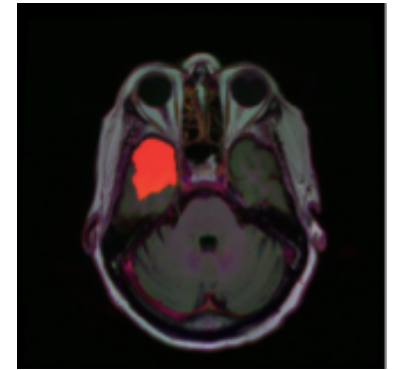
# RESUNET

- La arquitectura ResUNet combina la arquitectura de la red troncal UNet con bloques residuales para superar los problemas de desvanecimiento de gradientes en las arquitecturas profundas.
- La arquitectura de Unet se basa en redes totalmente convolucionales y se modifica de manera que funcione bien en las tareas de segmentación.
- Resunet consta de tres partes:
  - (1) Codificador o ruta de contratación
  - (2) Cuello de botella
  - (3) Decodificador o ruta expansiva

IMAGEN DE ENTRADA



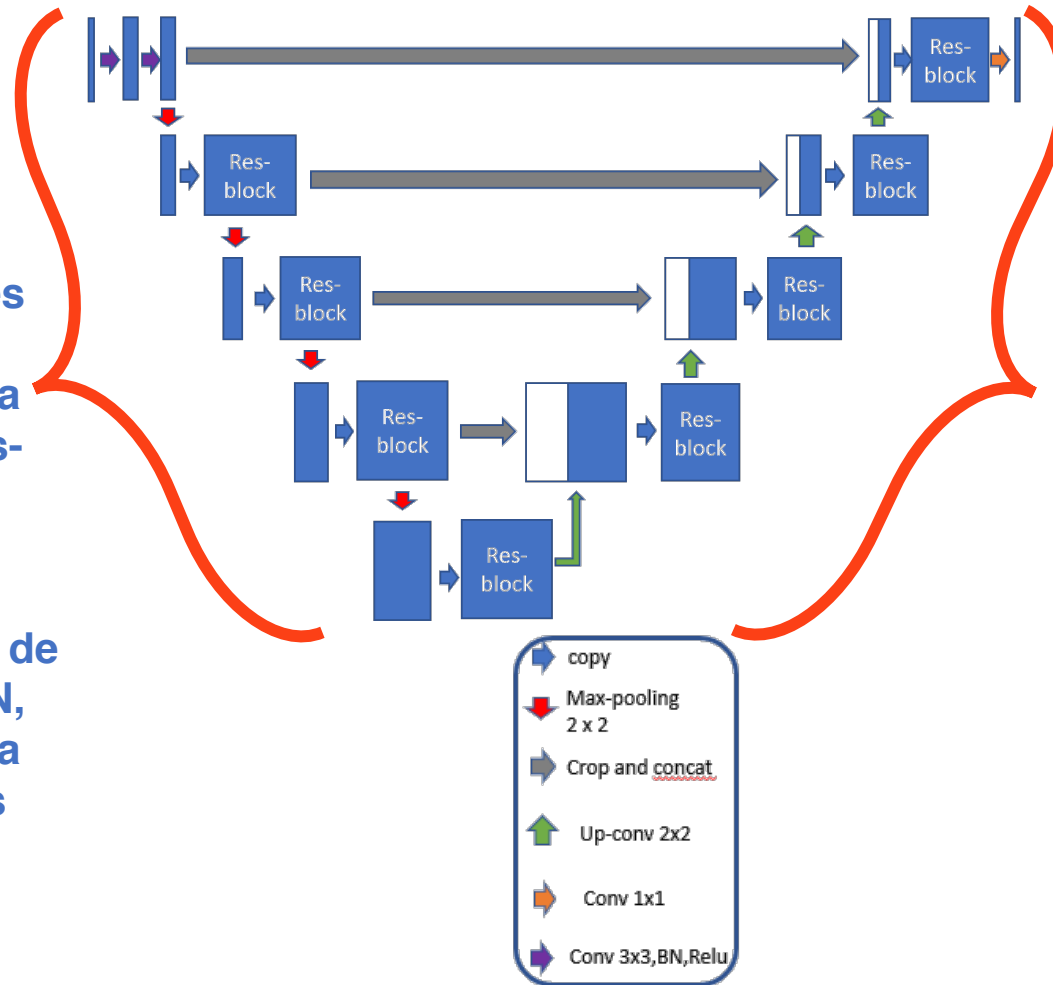
MÁSCARA DE SEGMENTACIÓN



# ARQUITECTURA RESUNET:

## RUTA DE CONTRACCIÓN (ENCODER)

- La ruta de contracción consta de varios bloques de contracción, cada bloque toma una entrada que pasa a través de res-blocks seguido de una agrupación máxima de 2x2. Los mapas de características después de cada bloque se duplican, lo que ayuda al modelo a aprender características complejas de manera efectiva.



## CAMINO DE EXPANSIÓN (DECODIFICADOR)

- La ventaja significativa de esta arquitectura radica en la sección de expansión o decodificador. Cada bloque toma la entrada muestreada de la capa anterior y se concatena con las características de salida correspondientes de los bloques res en la ruta de contracción. Esto luego se pasa nuevamente a través del bloque res seguido por capas de convolución de muestreo ascendente 2x2.
- Esto ayuda a garantizar que las características aprendidas durante la contracción se utilicen al reconstruir la imagen.
- Finalmente, en la última capa de la ruta de expansión, la salida del bloque res pasa a través de la capa de convolución 1x1 para producir la salida deseada con el mismo tamaño que la entrada.

## CUELLO DE BOTELLA

- El bloque de cuello de botella sirve como conexión entre la ruta de contracción y la ruta de expansión.
- El bloque toma la entrada y luego pasa a través de un bloque res seguido de un muestreo de capas de convolución 2x2.



# ARQUITECTURA RESUNET:

## 1. El codificador o ruta de contracción consta de 4 bloques:

- El primer bloque consta de una capa de convolución 3x3 + Relu + Normalización por lotes
- Los tres bloques restantes consisten en bloques Res seguidos de un max pooling de tamaño 2x2.

## 2. Cuello de botella:

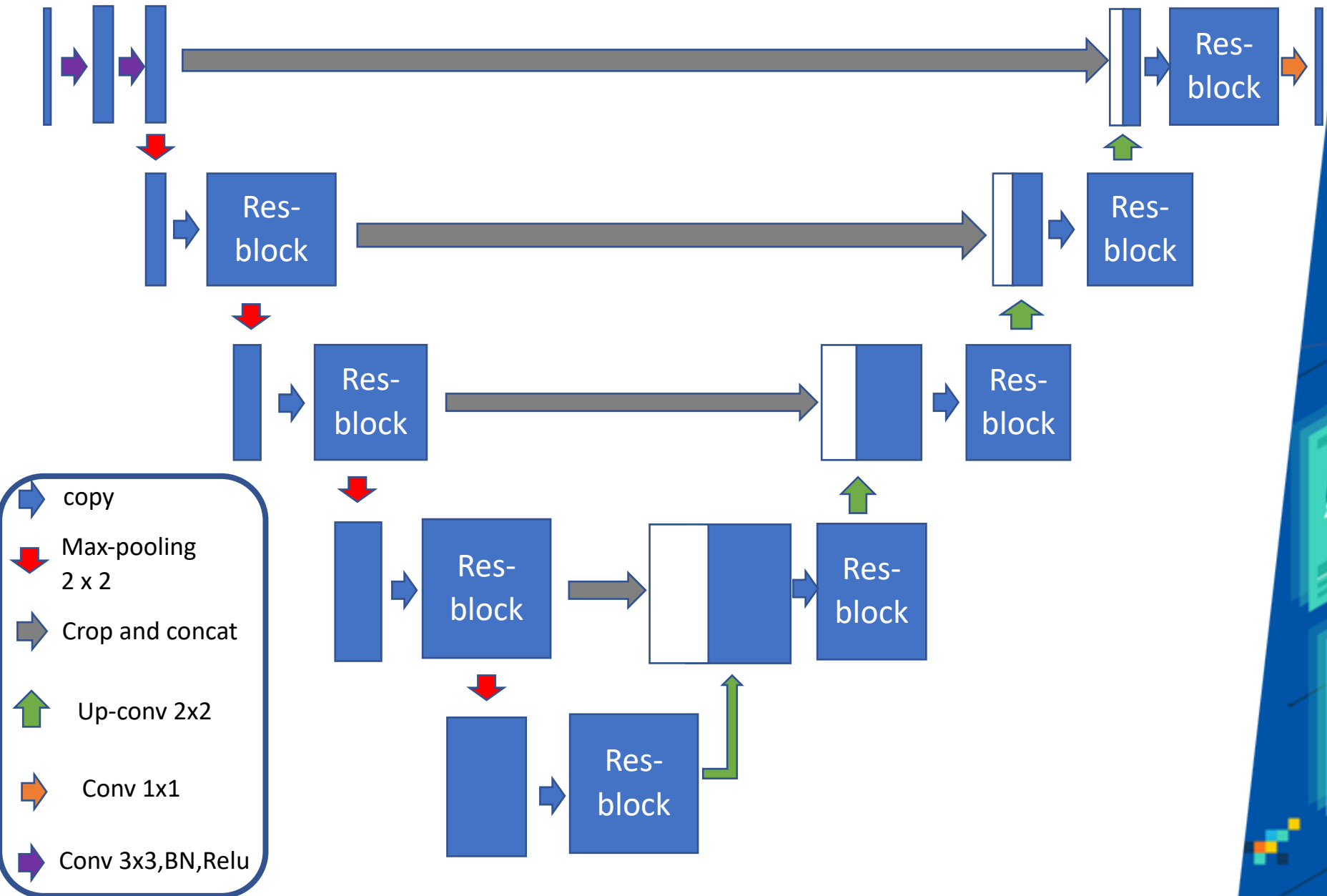
- Está entre el camino de la contracción y la expansión.
- Consiste en un bloque Res seguido de una capa convolucional de upsampling de tamaño 2x2.

## 3. La ruta de expansión o decodificador consta de 4 bloques:

- Los 3 bloques siguientes al cuello de botella consisten en bloques de Res seguidos por una capa convolucional de upsampling de tamaño 2 x 2
- El bloque final consta de un Res-block seguido de una capa de convolucional de 1x1.




# RESUNET ARCHITECTURE:



# RECURSOS ADICIONALES DE RESUNET

Paper #1: <https://arxiv.org/abs/1505.04597>

 Cornell University

arXiv.org > cs > arXiv:1505.04597

Search...  
Help | Advanced

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 18 May 2015]

**U-Net: Convolutional Networks for Biomedical Image Segmentation**

Olaf Ronneberger, Philipp Fischer, Thomas Brox

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. Moreover, the network is fast. Segmentation of a 512x512 image takes less than a second on a recent GPU. The full implementation (based on Caffe) and the trained networks are available at [this http URL](http://this URL).

Comments: conditionally accepted at MICCAI 2015


Subjects: **Computer Vision and Pattern Recognition (cs.CV)**

Cite as: [arXiv:1505.04597 \[cs.CV\]](https://arxiv.org/abs/1505.04597)  
(or [arXiv:1505.04597v1](https://arxiv.org/abs/1505.04597v1) [cs.CV] for this version)

**Bibliographic data**  
[Enable Bibex (What is Bibex?)]

**Submission history**  
From: Olaf Ronneberger [view email]

Paper #2: <https://arxiv.org/abs/1904.00592>

 Cornell University

arXiv.org > cs > arXiv:1904.00592

Search...  
Help | Advanced

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 1 Apr 2019 (v1), last revised 15 Jan 2020 (this version, v3)]

**ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data**

Foivos I. Diakogiannis, François Waldner, Peter Caccetta, Chen Wu

Scene understanding of high resolution aerial images is of great importance for the task of automated monitoring in various remote sensing applications. Due to the large within-class and small between-class variance in pixel values of objects of interest, this remains a challenging task. In recent years, deep convolutional neural networks have started being used in remote sensing applications and demonstrate state of the art performance for pixel level classification of objects. `textcolor{black}`{Here we propose a reliable framework for performant results for the task of semantic segmentation of monosegmental very high resolution aerial images. Our framework consists of a novel deep learning architecture, ResUNet-a, and a novel loss function based on the Dice loss. ResUNet-a uses a U-Net encoder/decoder backbone, in combination with residual connections, atrous convolutions, pyramid scene parsing pooling and multi-tasking inference. ResUNet-a infers sequentially the boundary of the objects, the distance transform of the segmentation mask, the segmentation mask and a colored reconstruction of the input. Each of the tasks is conditioned on the inference of the previous ones, thus establishing a conditioned relationship between the various tasks, as this is described through the architecture's computation graph. We analyse the performance of several flavours of the Generalized Dice loss for semantic segmentation, and we introduce a novel variant loss function for semantic segmentation of objects that has excellent convergence properties and behaves well even under the presence of highly imbalanced classes.} The performance of our modeling framework is evaluated on the ISPRS 2D Potsdam dataset. Results show state-of-the-art performance with an average F1 score of 92.9% over all classes for our best model.

Comments: Accepted for publication to the ISPRS Journal of Photogrammetry and Remote Sensing

Subjects: **Computer Vision and Pattern Recognition (cs.CV)**

DOI: [10.1016/j.isprsjprs.2020.01.013](https://doi.org/10.1016/j.isprsjprs.2020.01.013)

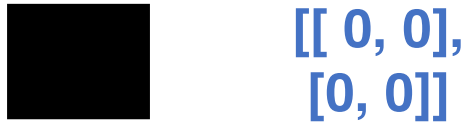
Cite as: [arXiv:1904.00592 \[cs.CV\]](https://arxiv.org/abs/1904.00592)  
(or [arXiv:1904.00592v3](https://arxiv.org/abs/1904.00592v3) [cs.CV] for this version)

Buen artículo: <https://towardsdatascience.com/introduction-to-u-net-and-res-net-for-image-segmentation-9afcb432ee2f>



# MÁSCARAS:

- El objetivo de la segmentación de imágenes es comprender la imagen a nivel de píxel. Asocia cada píxel con una determinada clase. La salida producida por el modelo de segmentación de imágenes se denomina "máscara" de la imagen.
- Las máscaras se pueden representar asociando valores de píxeles con sus coordenadas. Por ejemplo, si tenemos una imagen negra de forma (2,2), esta se puede representar como:



Si nuestra máscara de salida es como sigue:



- Para representar esta máscara, primero debemos aplanar la imagen en una matriz 1-D. Esto resultaría en una máscara temporal similar a [255,0,0,255]. Luego, podemos usar el índice para crear la máscara. Finalmente tendríamos algo como [1,0,0,1] como nuestra máscara.

