

- La inteligencia artificial y el aprendizaje automático están transformando la industria de la manufacturación. Según el informe publicado por el Foro Económico Mundial, estas tecnologías jugarán un papel importante en la cuarta revolución industrial. Las principales áreas que pueden beneficiarse de esto son:
 - Departamento de mantenimiento
 - Departamento de producción
 - Departamento de Cadena de Suministro
- Se ha demostrado que el aprendizaje profundo es superior en la detección y localización de defectos utilizando datos de imágenes que podrían mejorar significativamente la eficiencia de producción en la industria de la manufacturación.
- Gran ejemplo de LandingAI: <https://landing.ai/defect-detection/>



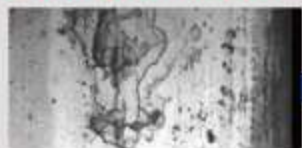
- **En este caso de estudio, asumiremos que trabajas como consultor de IA / ML.**
- **Has sido contratado por una empresa de fabricación de acero en San Diego y se le te encomendado la tarea de automatizar el proceso de detección y localización de defectos encontrados en la fabricación del acero.**
- **La detección de defectos ayudaría a mejorar la calidad de fabricación, así como a reducir el desperdicio debido a defectos de producción.**
- **El equipo ha recopilado imágenes de superficies de acero y se ha acercado a ti para desarrollar un modelo que pueda detectar y localizar defectos en tiempo real.**
- **Se te han proporcionado 12600 imágenes que contienen 4 tipos de defectos, junto con su ubicación en la superficie de acero.**



METODOLOGÍA DE APRENDIZAJE PROFUNDO POR CAPAS PARA REALIZAR LA CLASIFICACIÓN Y SEGMENTACIÓN

ENTRADA DE
IMÁGENES

DE LA LÍNEA DE
FABRICACIÓN O DEL
DEPARTAMENTO DE
MANTENIMIENTO



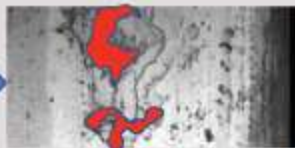
MODELO DE
CLASIFICADOR
DE APRENDIZAJE
PROFUNDO
RESNET

LIBRE DE
DEFECTOS

FIN

DEFECTO
DETECTADO

MODELO DE
SEGMENTACIÓN
RESUNET

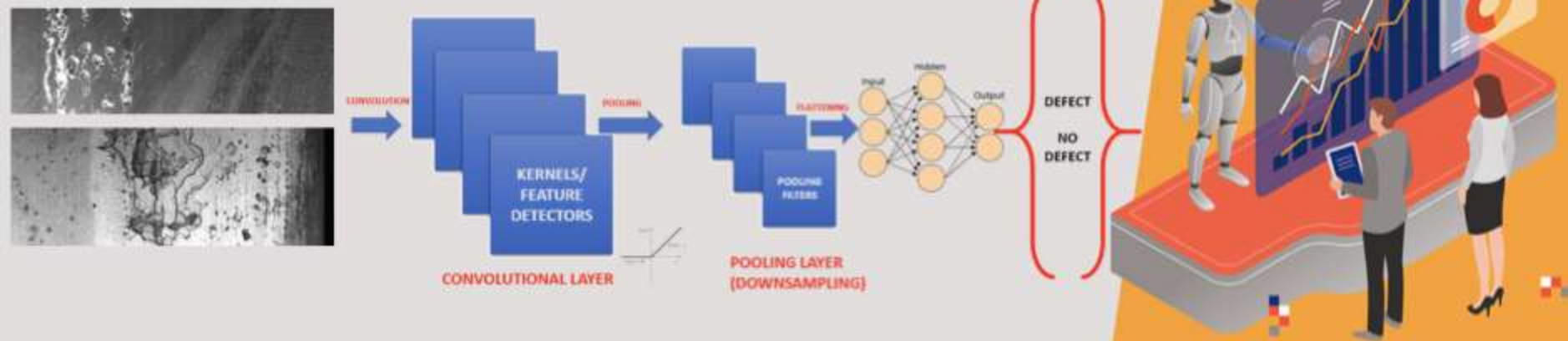


DETECTAR LA
UBICACIÓN DEL
DEFECTO A NIVEL DE
PÍXEL



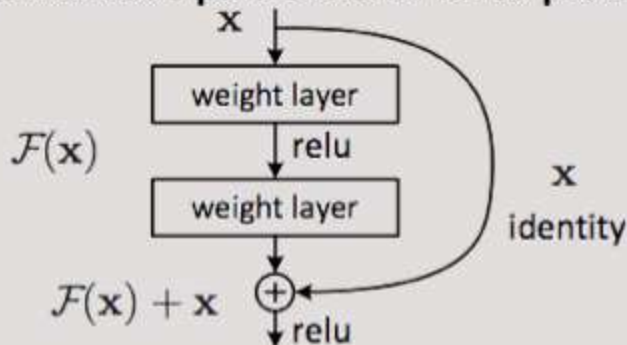
REDES NEURALES CONVOLUCIONALES (REPASO)

- Las primeras capas de una RNC 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 detectores de características 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, tienden a ocurrir gradientes de desvanecimiento que impactan negativamente en 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 sufrir los problemas de 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 es el dataset utilizado para entrenar la red profunda ResNet.



¿APRENDIZAJE POR TRANSFERENCIA? (REPASO)

- 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 se puede reducir drásticamente el tiempo de cálculo requerido comparado con realizar el entrenamiento desde cero.



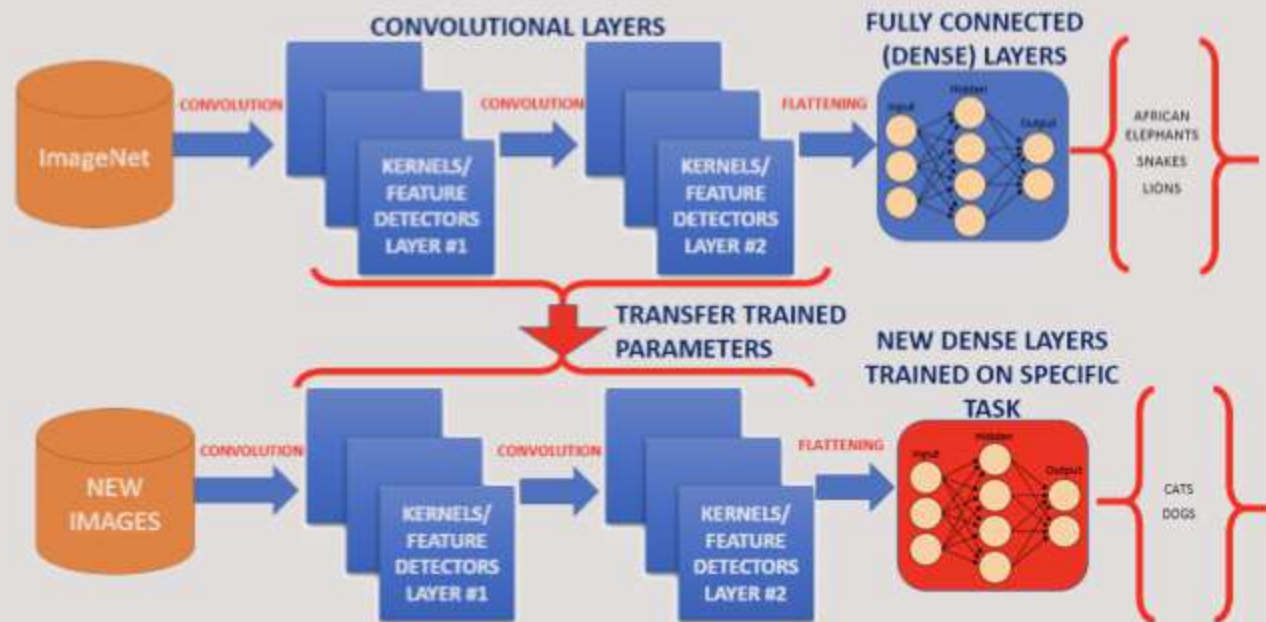
KNOWLEDGE TRANSFER



- Photo Credit: 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
- 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.



EL PROCESO DE APRENDIZAJE POR TRANSFERENCIA (REPASO)



ESTRATEGIAS DE ENTRENAMIENTO EN APRENDIZAJE POR TRANSFERENCIA

- **Pasos de la estrategia n. ° 1:**
 - Congelar los pesos de las primeras capas de la red RNC entrenada.
 - Entrenar solo las capas densas recién agregadas al final (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 red de RNC mientras establecemos que la tasa de aprendizaje sea muy pequeña, esto es fundamental para garantizar que no cambie agresivamente los pesos ya entrenados.
- Las ventajas del aprendizaje por transferencia son:
- Proporciona un progreso de entrenamiento rápido, no tiene que empezar desde cero utilizando pesos inicializados aleatoriamente
- 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, FCN (redes de convolución total) y encoder-decoders profundos.
- Utilizaremos la arquitectura ResUNet para resolver la tarea actual.



¿QUÉ ES LA SEGMENTACIÓN DE IMÁGENES?

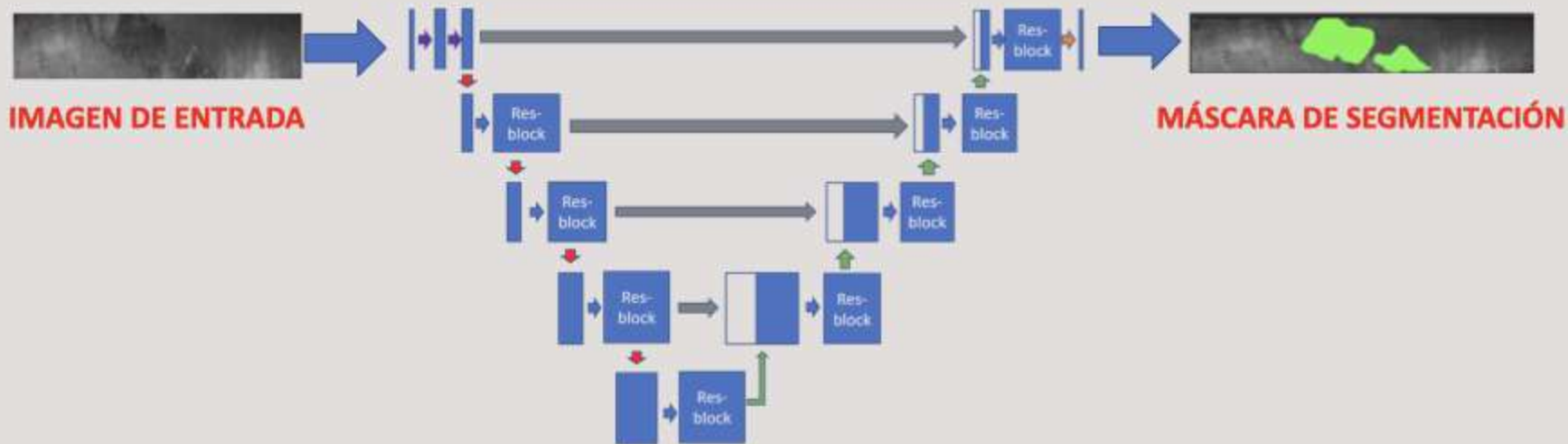
- ¿Recuerdas cuándo aplicamos RNC para problemas de clasificación de imágenes? Tuvimos que convertir la imagen en un vector y además agregar un encabezado de clasificación al final.
- Sin embargo, en el caso de Unet, convertimos (codificamos) la imagen en un vector seguido de un muestreo ascendente (decodificamos) 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 CNN clásicas: generalmente se utilizan cuando se necesita clasificar la imagen completa otorgando una etiqueta de clase.
- Para Unet: se realiza una clasificación a nivel de píxeles.
- U-net construye 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.

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



RESUNET

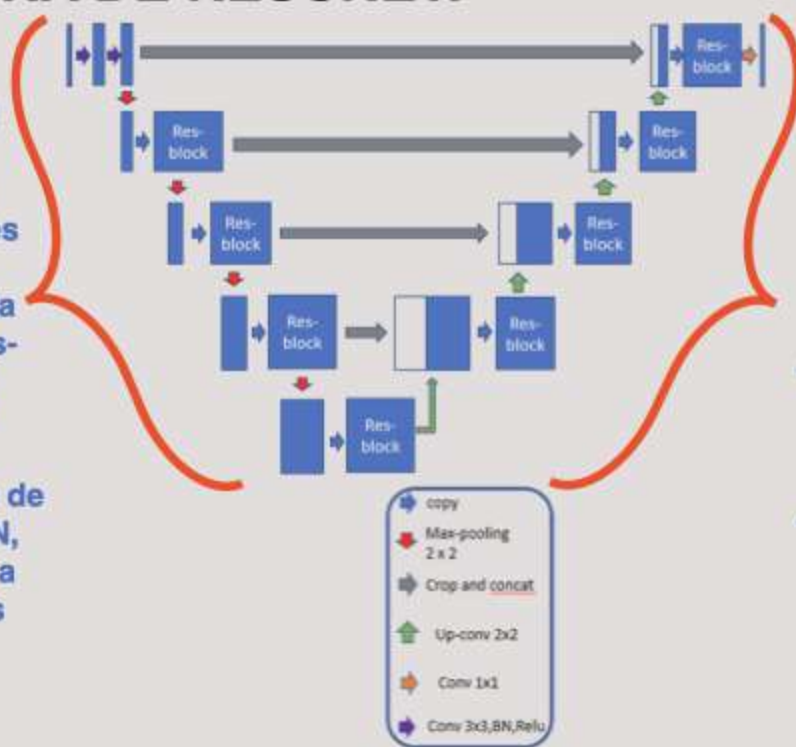
- La arquitectura ResUNet combina la arquitectura troncal de UNet con bloques residuales para superar los problemas del desvanecimiento del gradiente presentes en las arquitecturas profundas.
- La arquitectura de Unet se basa en redes totalmente convolucionales y está modificada de manera que funcione 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



ARQUITECTURA DE 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 DE RESUNET:

1. El codificador o ruta de contratación consta de 4 bloques:

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

2. Cuello de botella:

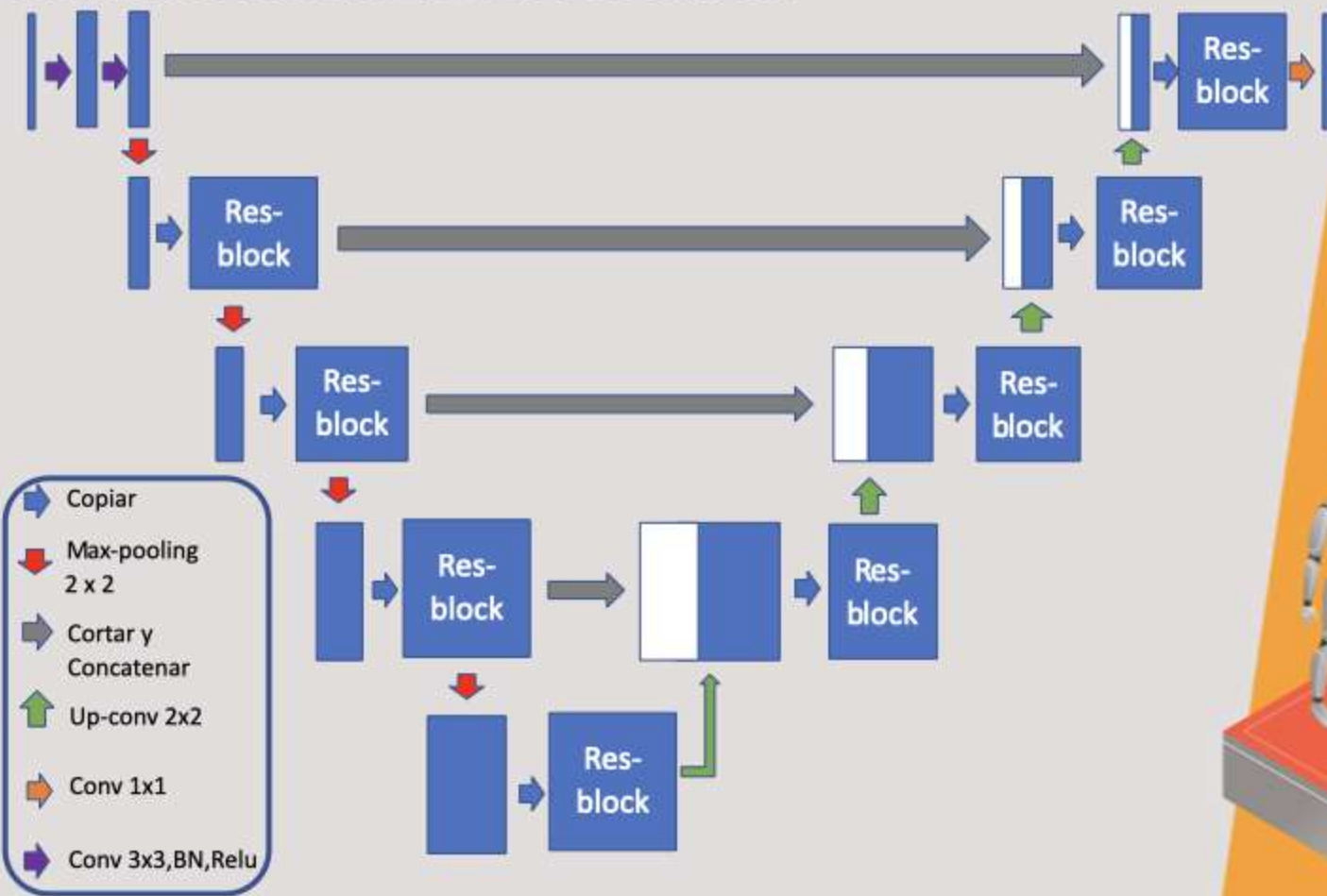
- Está entre el camino de la contracción y la expansión.
- Consiste en un bloque Res seguido de una capa convolucional de upsampling de tamaño 2×2 .

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 2×2
- El bloque final consta de un Res-block seguido de una capa de convolucional de 1×1 .

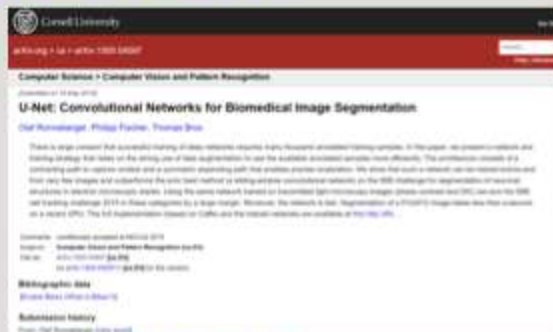


ARQUITECTURA DE RESUNET:



RECURSOS ADICIONALES DE RESUNET

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



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



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.



RUN LENGTH ENCODING (RLE):

- A veces es difícil representar la máscara usando el índice, ya que haría que la longitud de la máscara fuera igual al producto de la altura y el ancho de la imagen.
- Para superar esto, utilizamos una técnica de compresión de datos sin pérdidas llamada codificación de longitud de ejecución (RLE), que almacena secuencias que contienen muchos elementos de datos consecutivos como un valor de datos único seguido del recuento.
- Por ejemplo, supongamos que tenemos una imagen (una sola fila) que contiene texto negro sin formato sobre un fondo blanco sólido. B representa píxel negro y W representa blanco:

WWWWWWWWWWWWBWWWWWWWWWW
WWBWWWWWWWWWWWWWWWWWWWWWW
WWWWWWBWWWWWWWWWWWWWWWWWW

- Run-length encoding (RLE):

12W1B12W3B24W1B14W

- *Esto se puede interpretar como una secuencia de doce W, una B, doce W, tres B, etc.*

