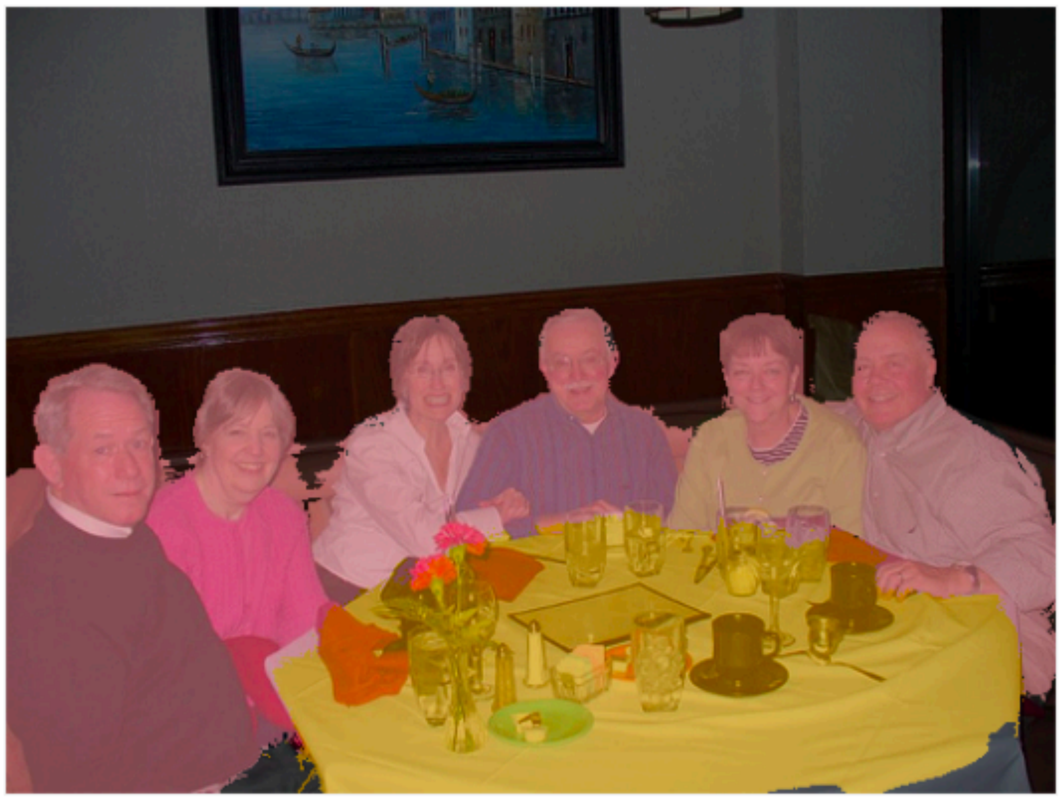
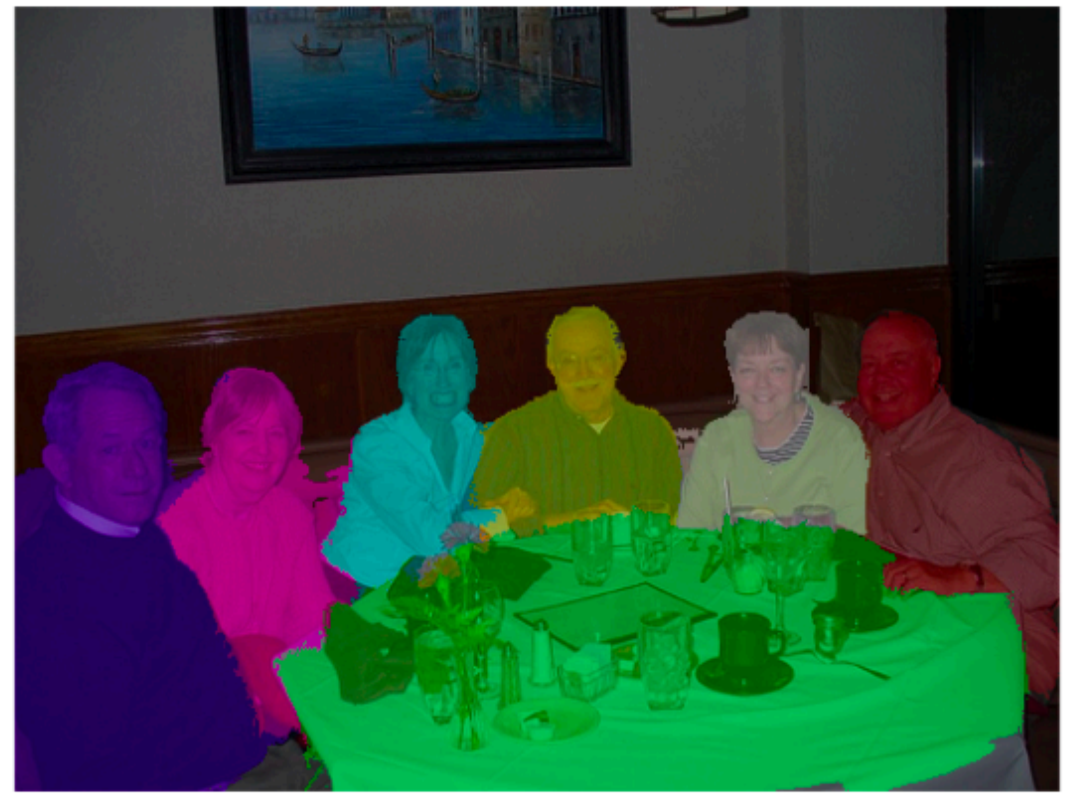
Segmentación



Segmentación semántica

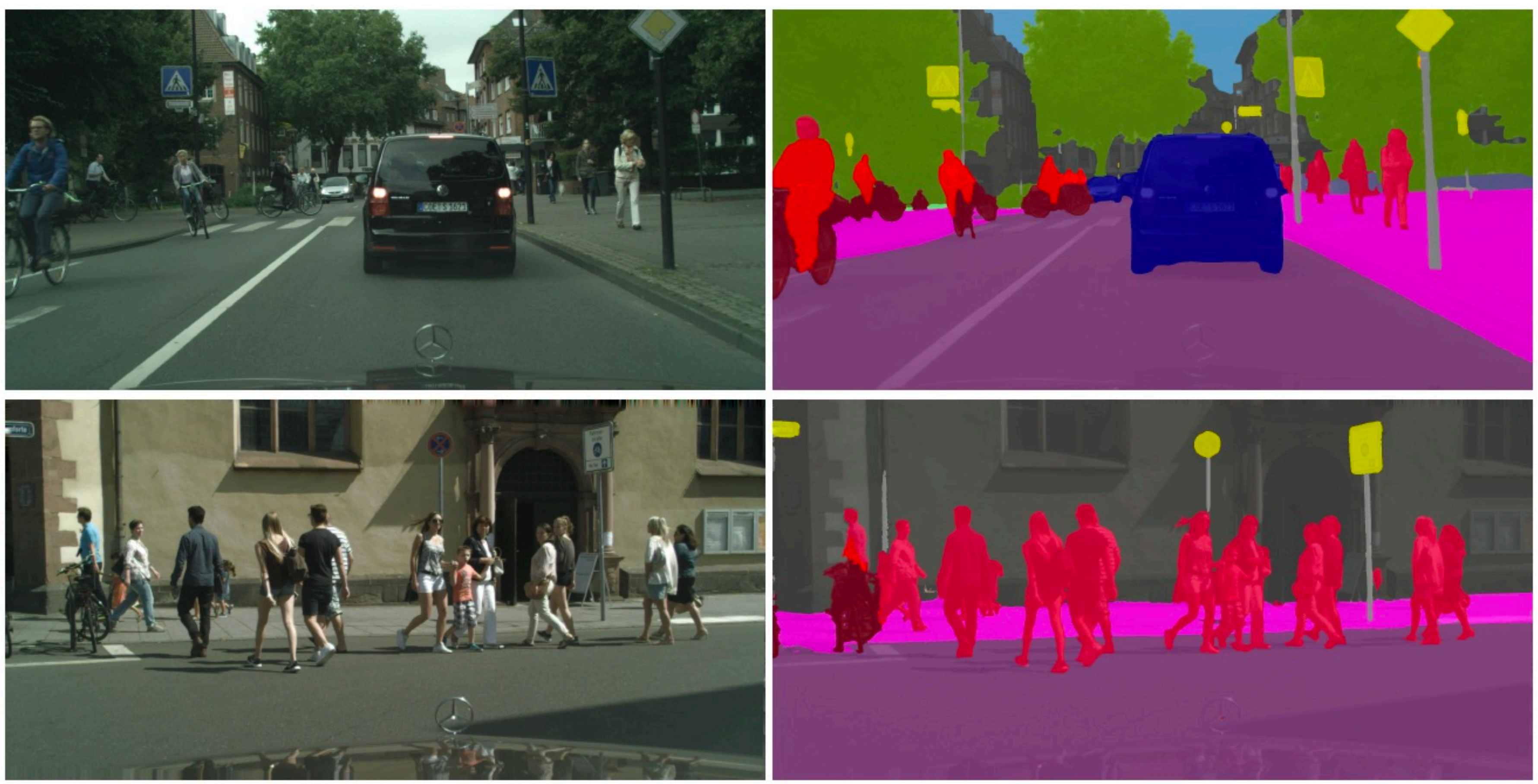


Segmentación de instancias

Segmentación semántica
+
detección de objetos

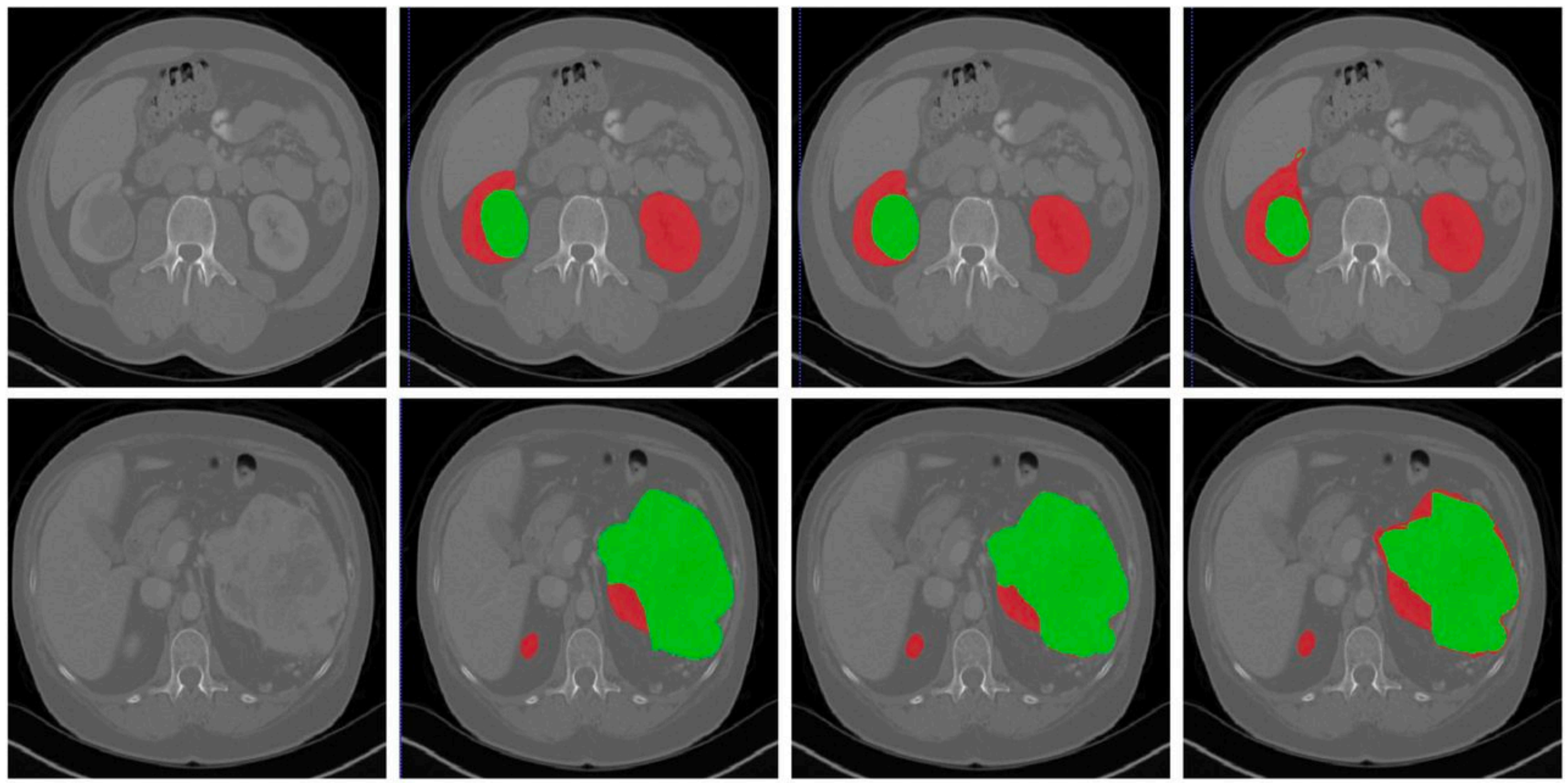
Usos

Conducción autónoma

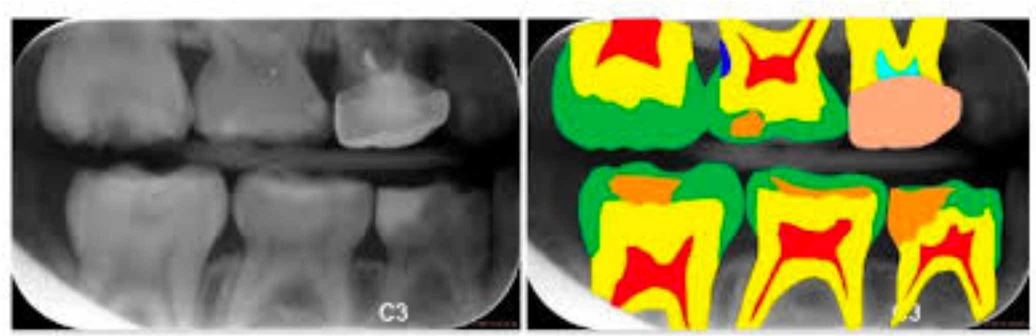


Usos

Medicina

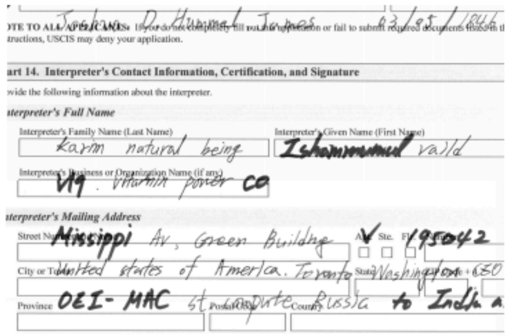


(a) Input images (b) Semantic Labels (c) Seg-Net+EG-CNN (d) Seg-Net

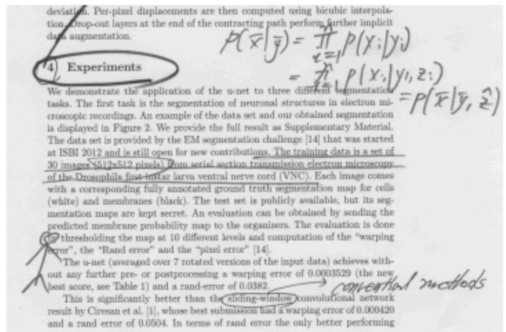


Usos

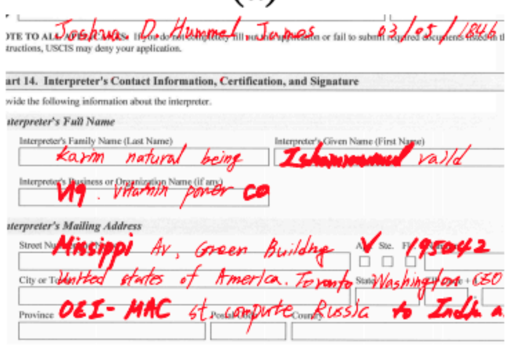
Detección de texto y OCR



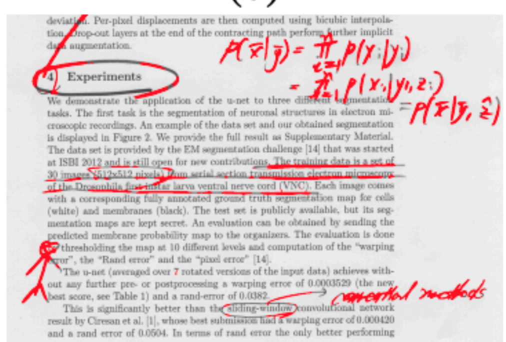
(a)



(b)



(c)



(d)

Modo retrato

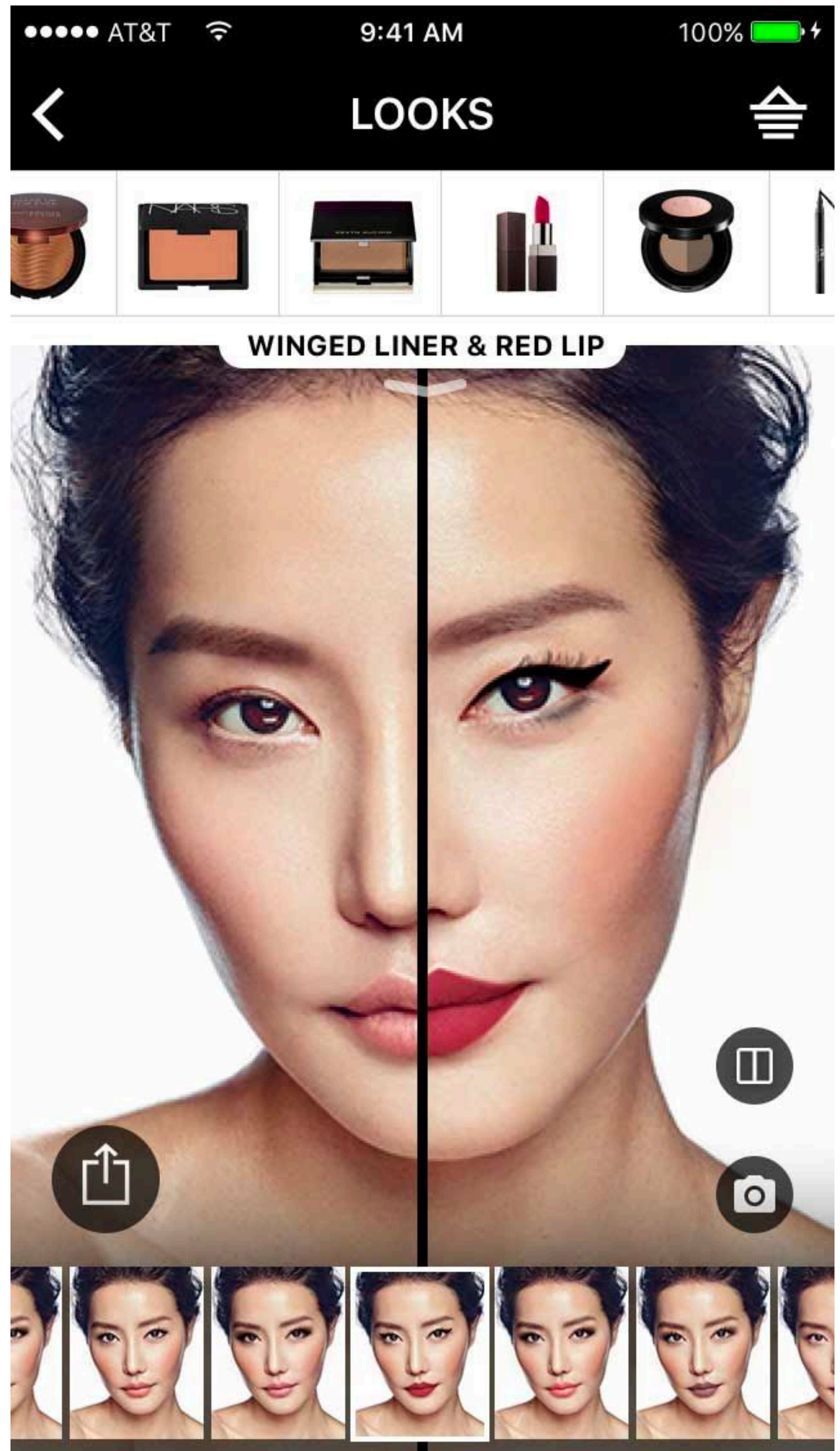
Cambios de Fondo

Usos

Detección de texto y OCR

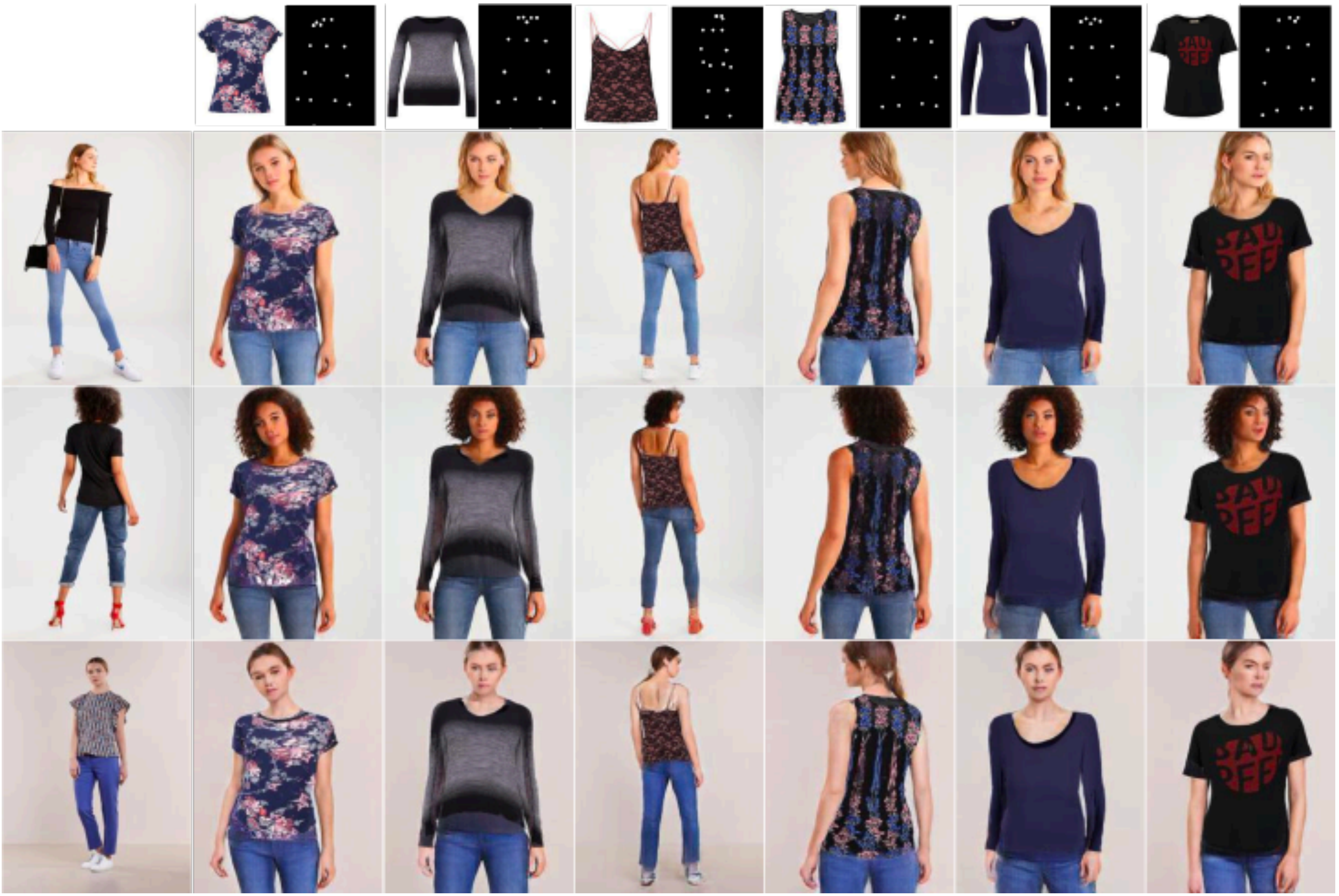
Usos

Maquillaje virtual



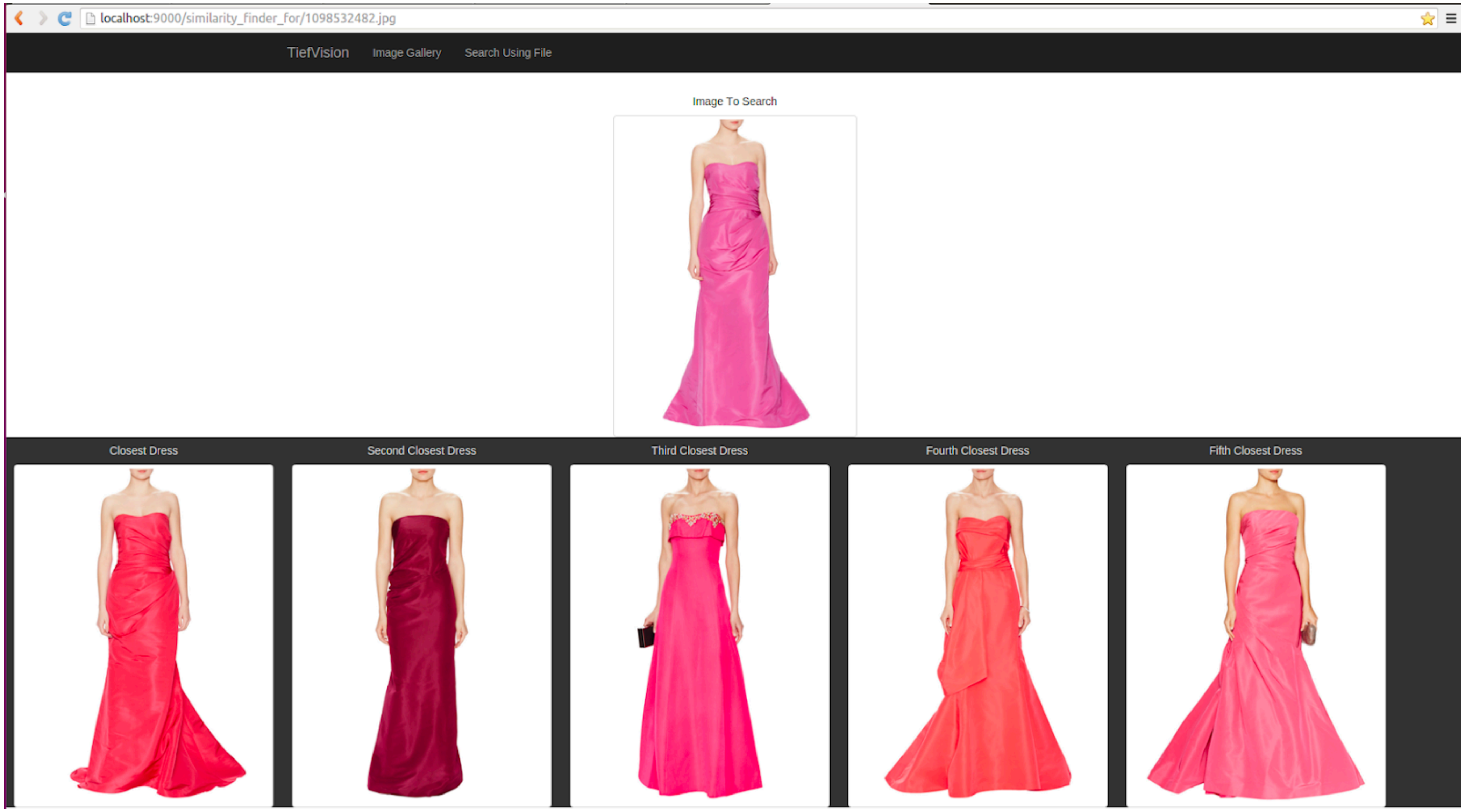
Usos

Probador virtual



Usos

Buscador de imágenes inteligente



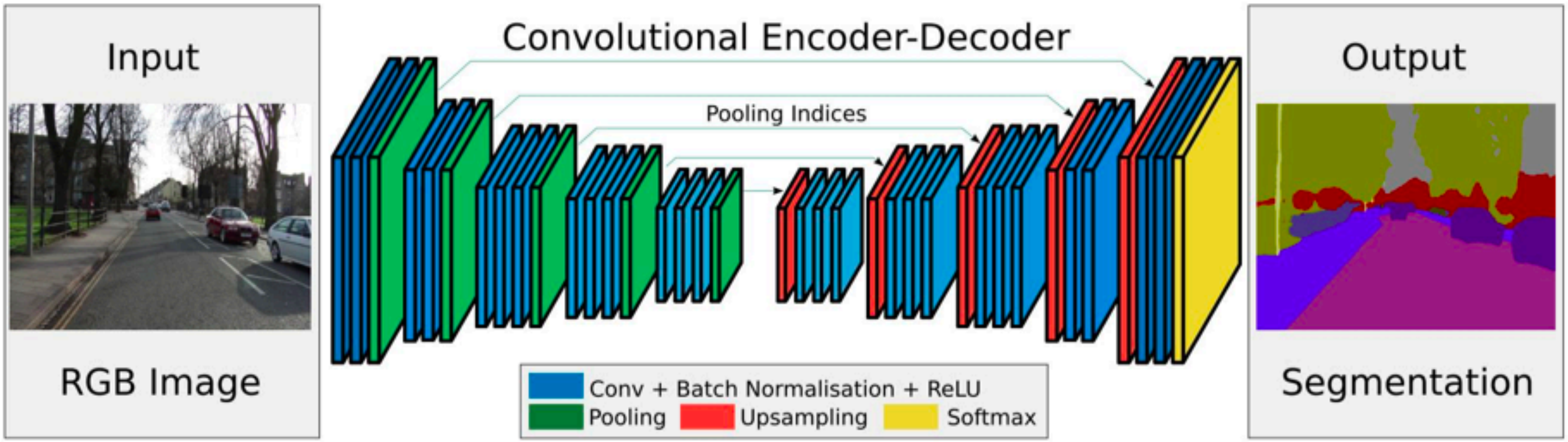
Usos

Discriminación de objetos en cámaras seguridad

Usos

Webcams con auto tracking

Segmentación



En segmentación, la mayoría de las arquitecturas constan de un codificador y un decodificador. Normalmente el decodificador genera una máscara que contiene la forma del objeto y objetos.

U-Net

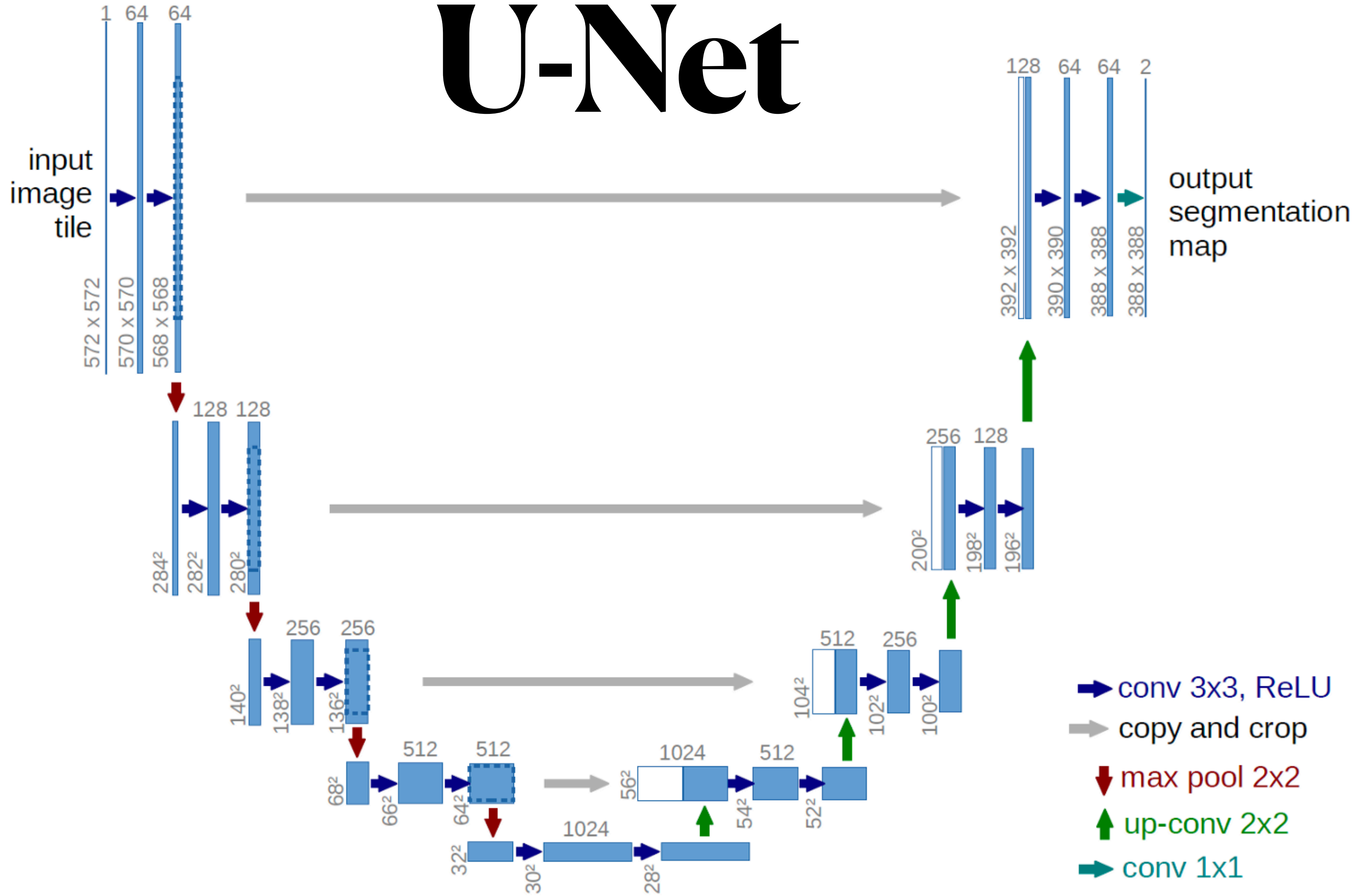


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

FastFCN —Fast Fully Convolutional Network

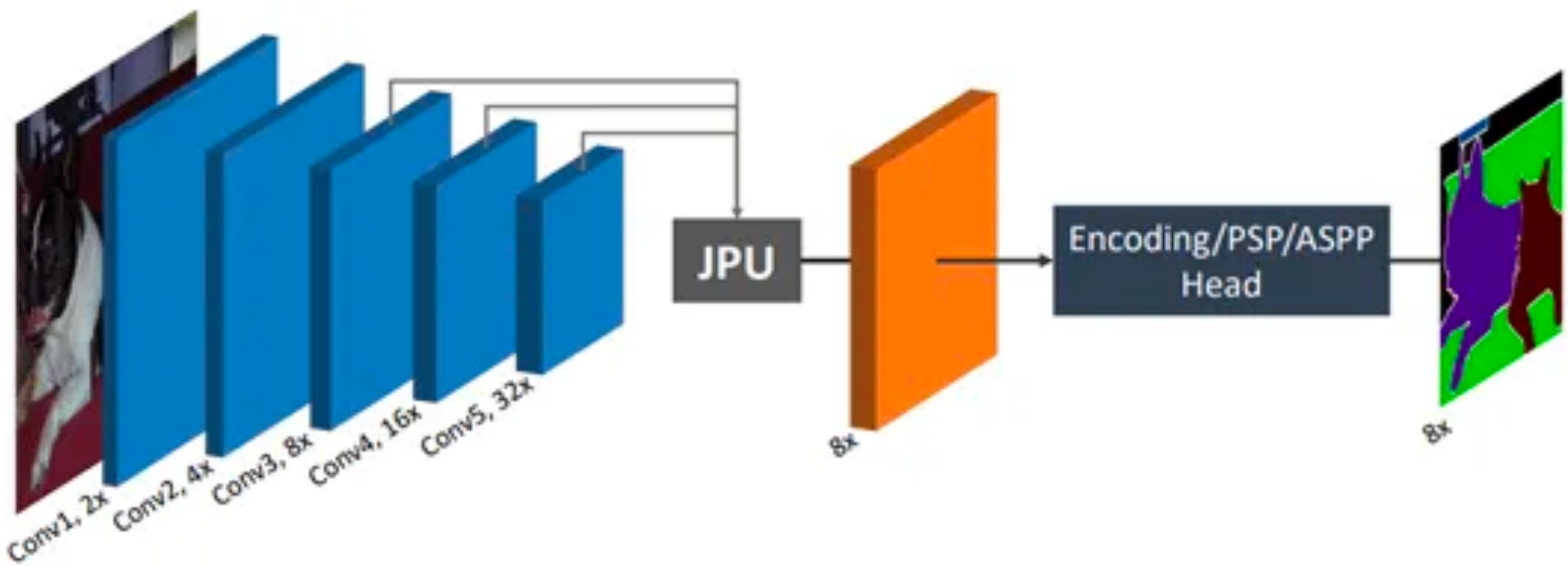


Figure 2: **Framework Overview of Our Method.** Our method employs the same backbone as the original FCN. After the backbone, a novel upsampling module named Joint Pyramid Upsampling (JPU) is proposed, which takes the last three feature maps as the inputs and generates a high-resolution feature map. A multi-scale/global context module is then employed to produce the final label map. Best viewed in color.

Gated-SCNN

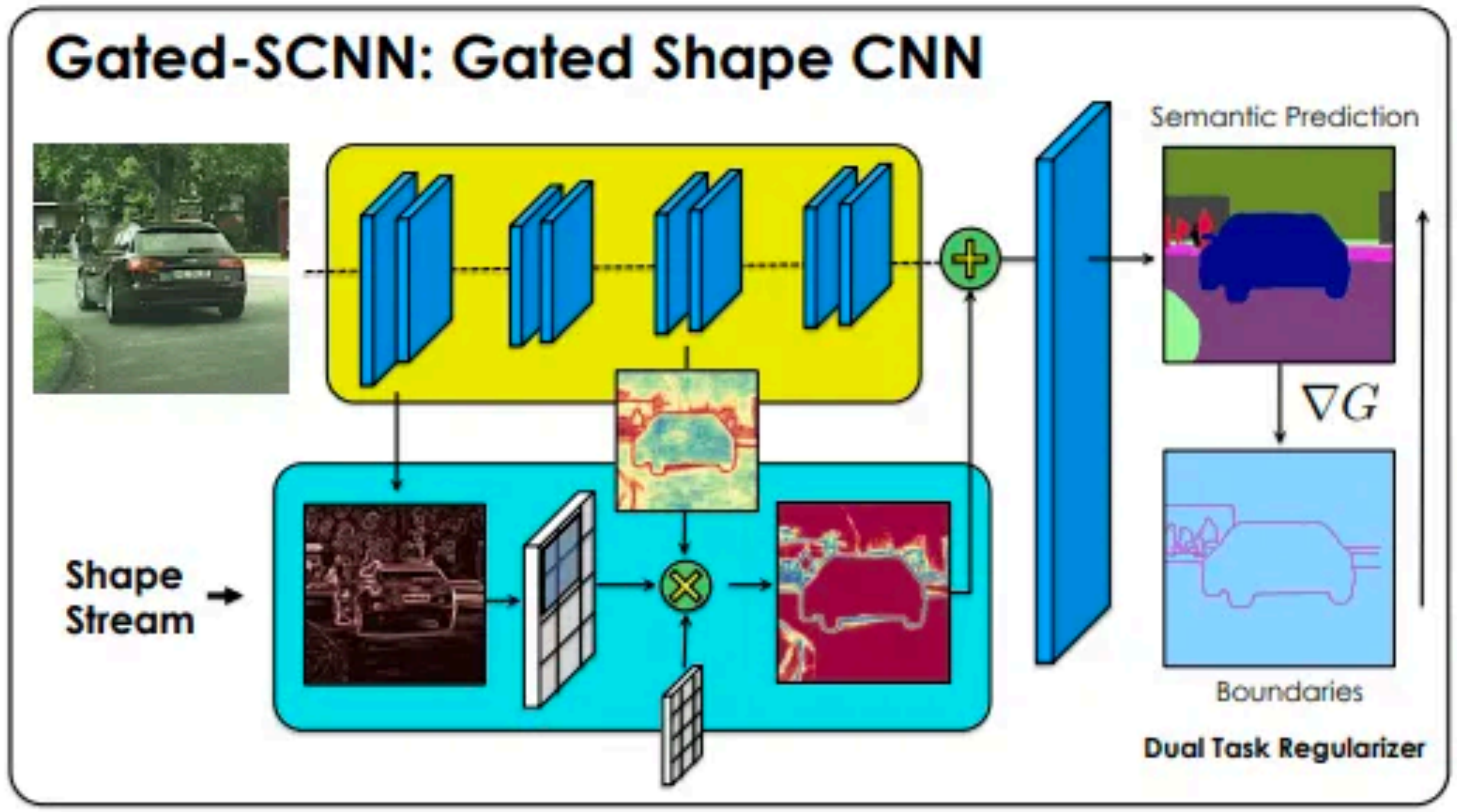


Figure 1: We introduce *Gated-SCNN* (GSCNN), a new two-stream CNN architecture for semantic segmentation that explicitly wires shape information as a separate processing stream. GSCNN uses a new gating mechanism to connect the intermediate layers. Fusion of information between streams is done at the very end through a fusion module. To predict high-quality boundaries, we exploit a new loss function that encourages the predicted semantic segmentation masks to align with ground-truth boundaries.

DeepLab

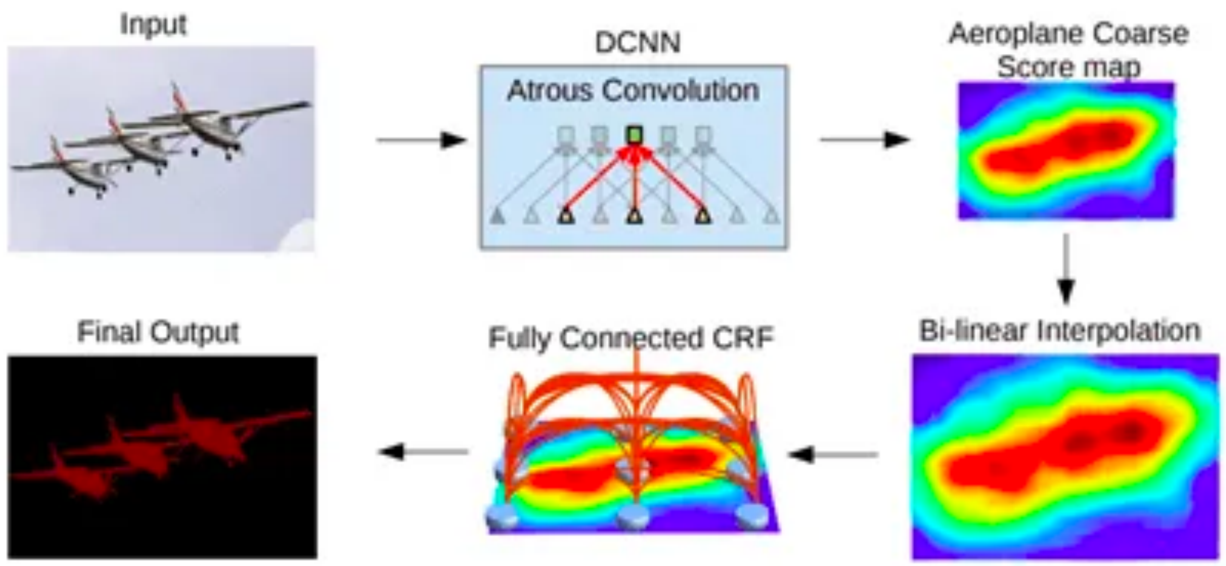


Fig. 1: Model Illustration. A Deep Convolutional Neural Network such as VGG-16 or ResNet-101 is employed in a fully convolutional fashion, using atrous convolution to reduce the degree of signal downsampling (from 32x down 8x). A bilinear interpolation stage enlarges the feature maps to the original image resolution. A fully connected CRF is then applied to refine the segmentation result and better capture the object boundaries.

Mask R-CNN

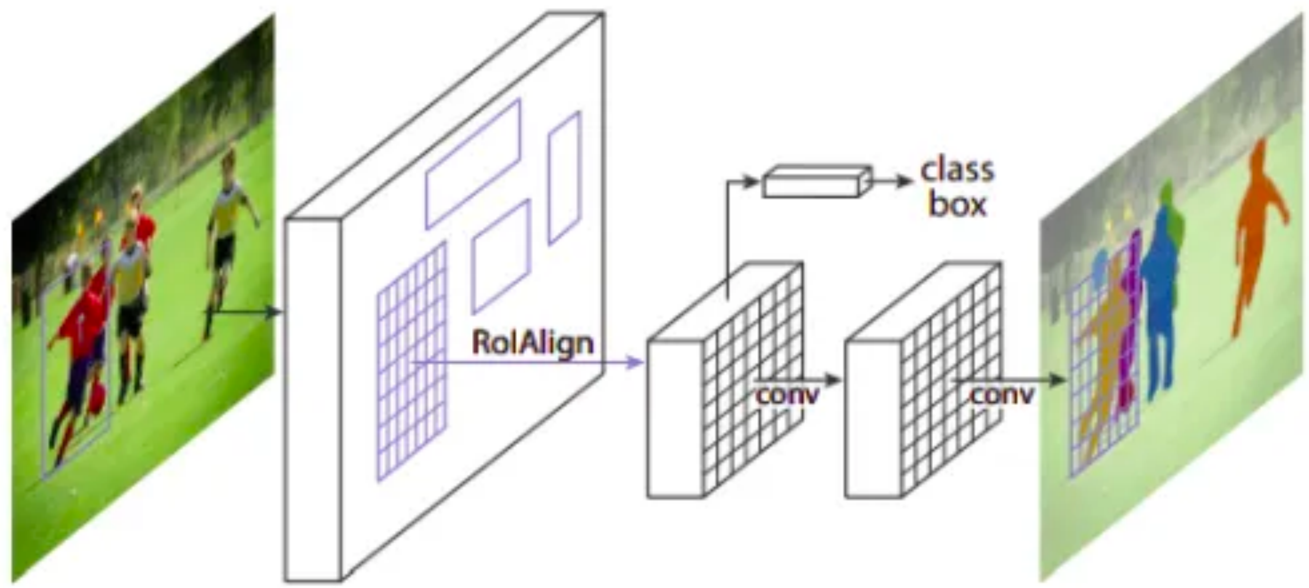
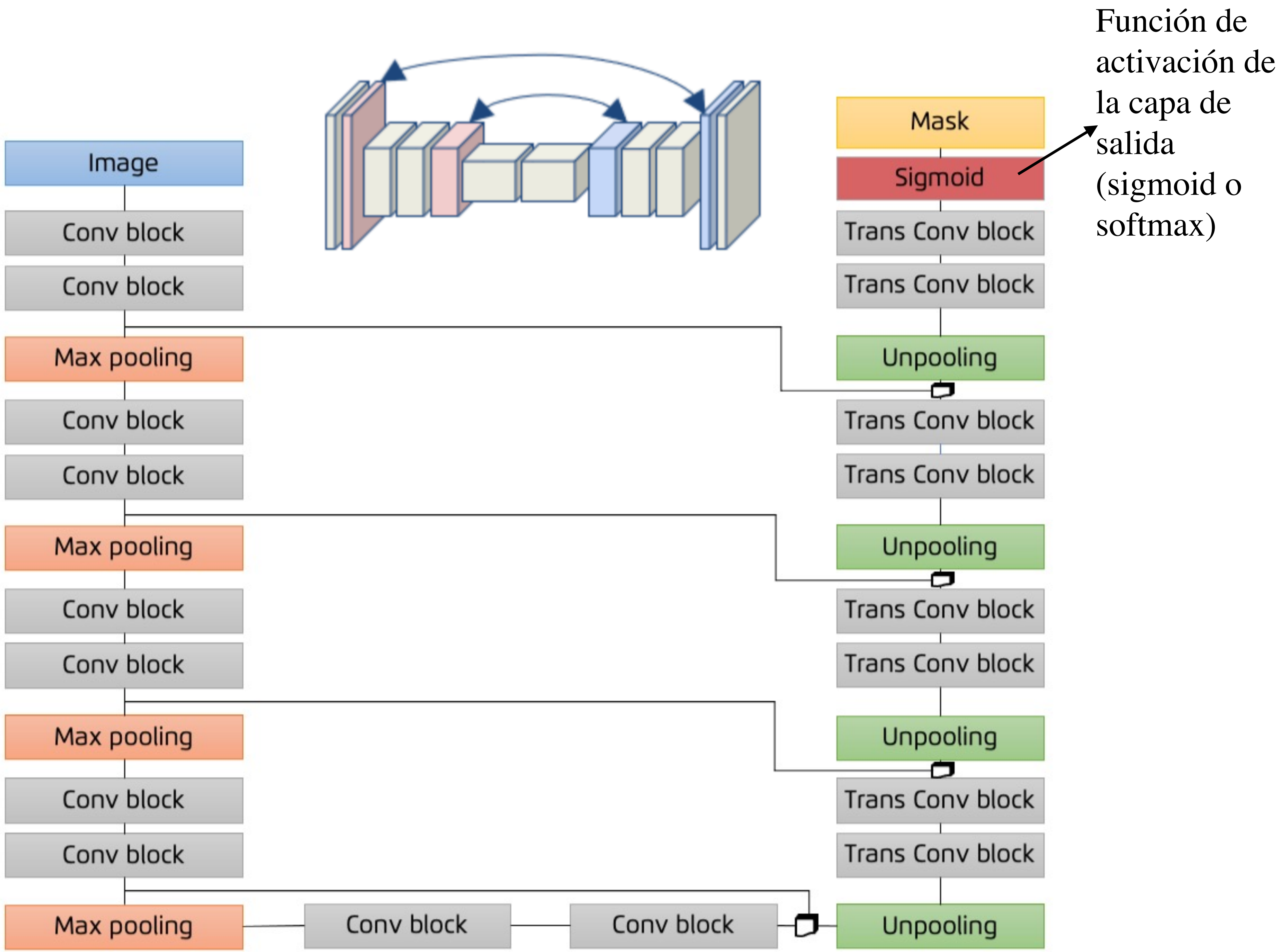


Figure 1. The **Mask R-CNN** framework for instance segmentation.



Función de activación de la capa de salida (sigmoid o softmax)
Image
Conv block
Conv block
Max pooling
Conv block
Conv block
Max pooling
Conv block
Conv block
Max pooling
Conv block
Conv block
Max pooling
Conv block
Conv block
Unpooling
Trans Conv block
Trans Conv block
Unpooling
Trans Conv block
Trans Conv block
Unpooling
Trans Conv block
Trans Conv block
Unpooling
Trans Conv block
Trans Conv block
Sigmoid
Mask

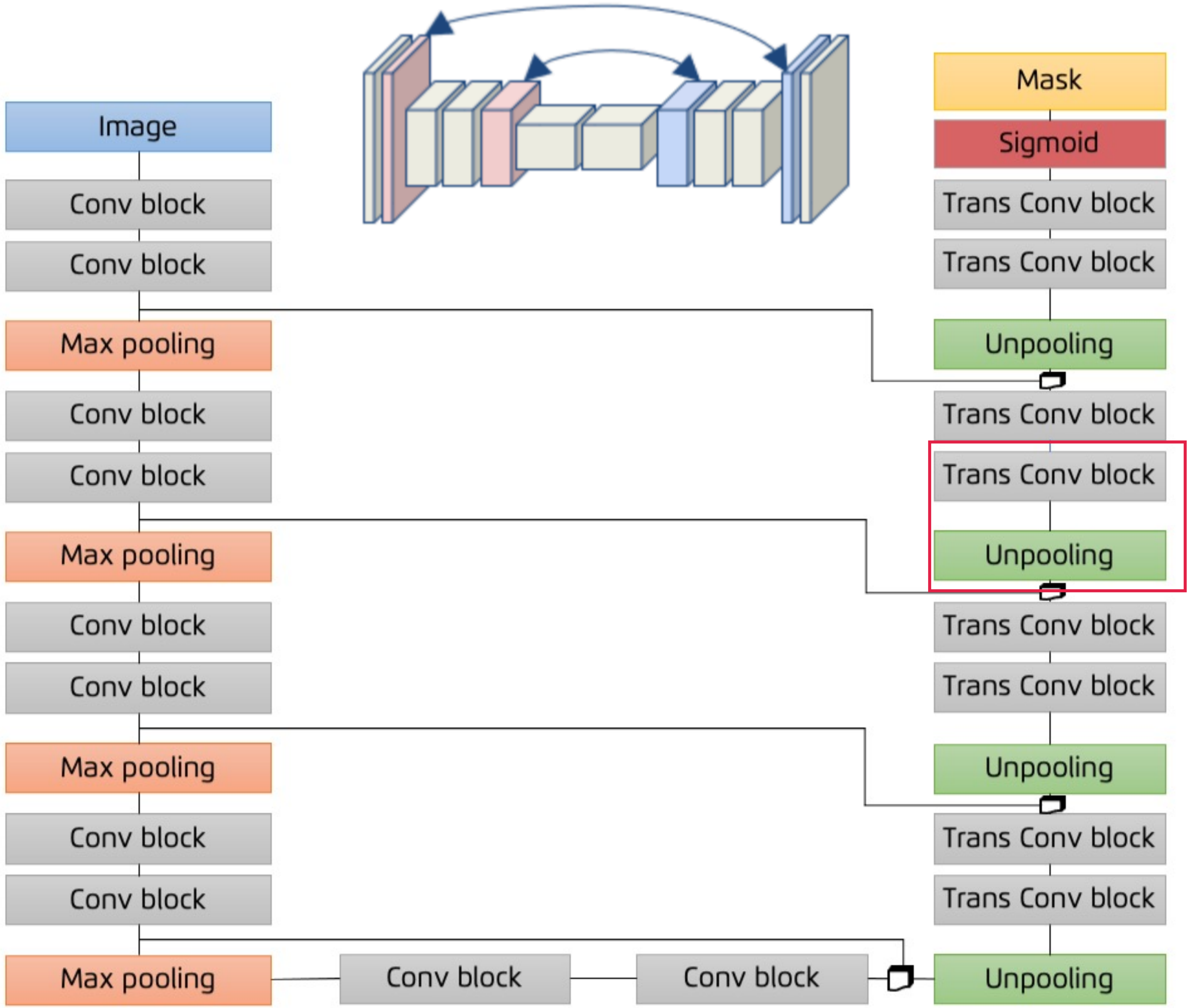
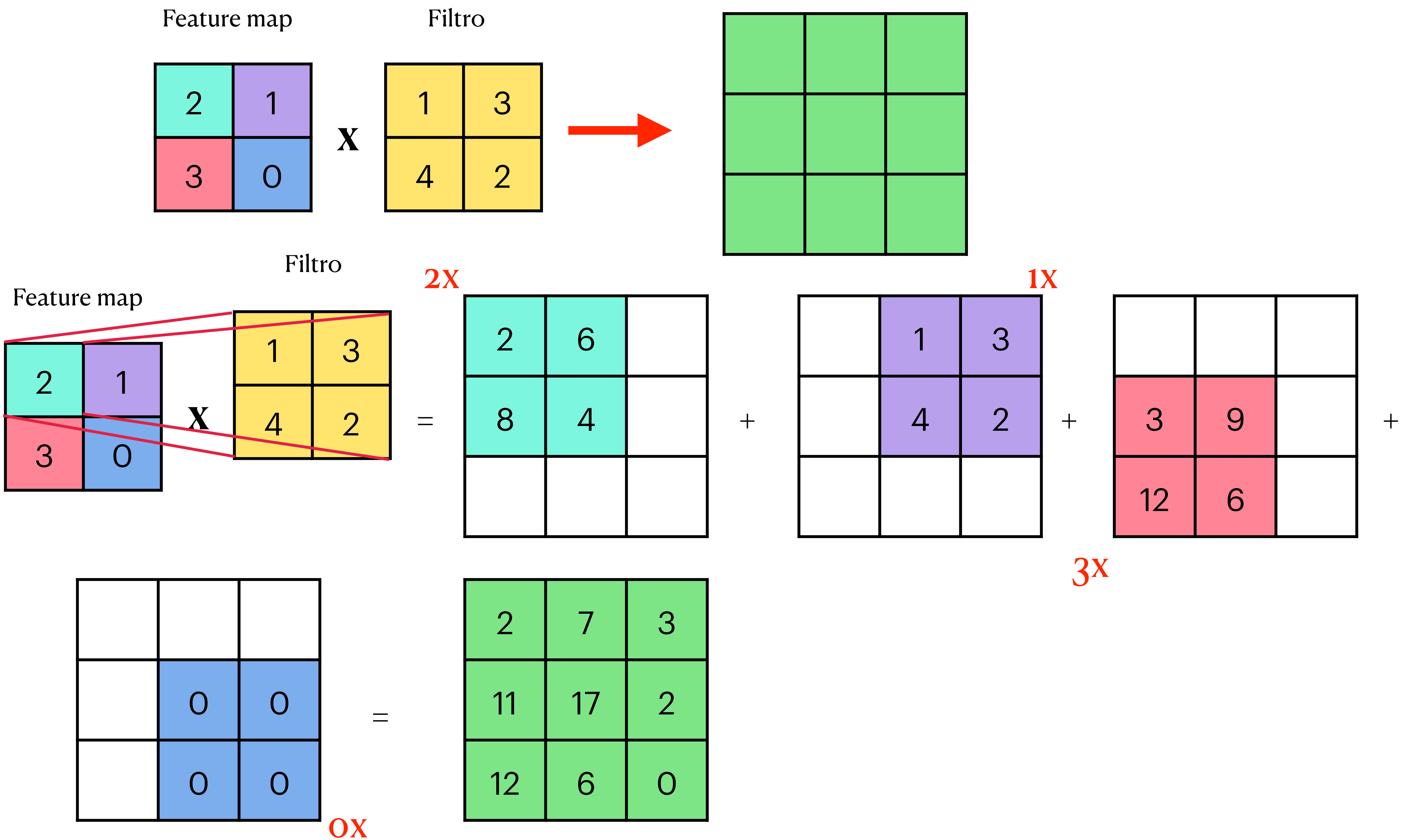


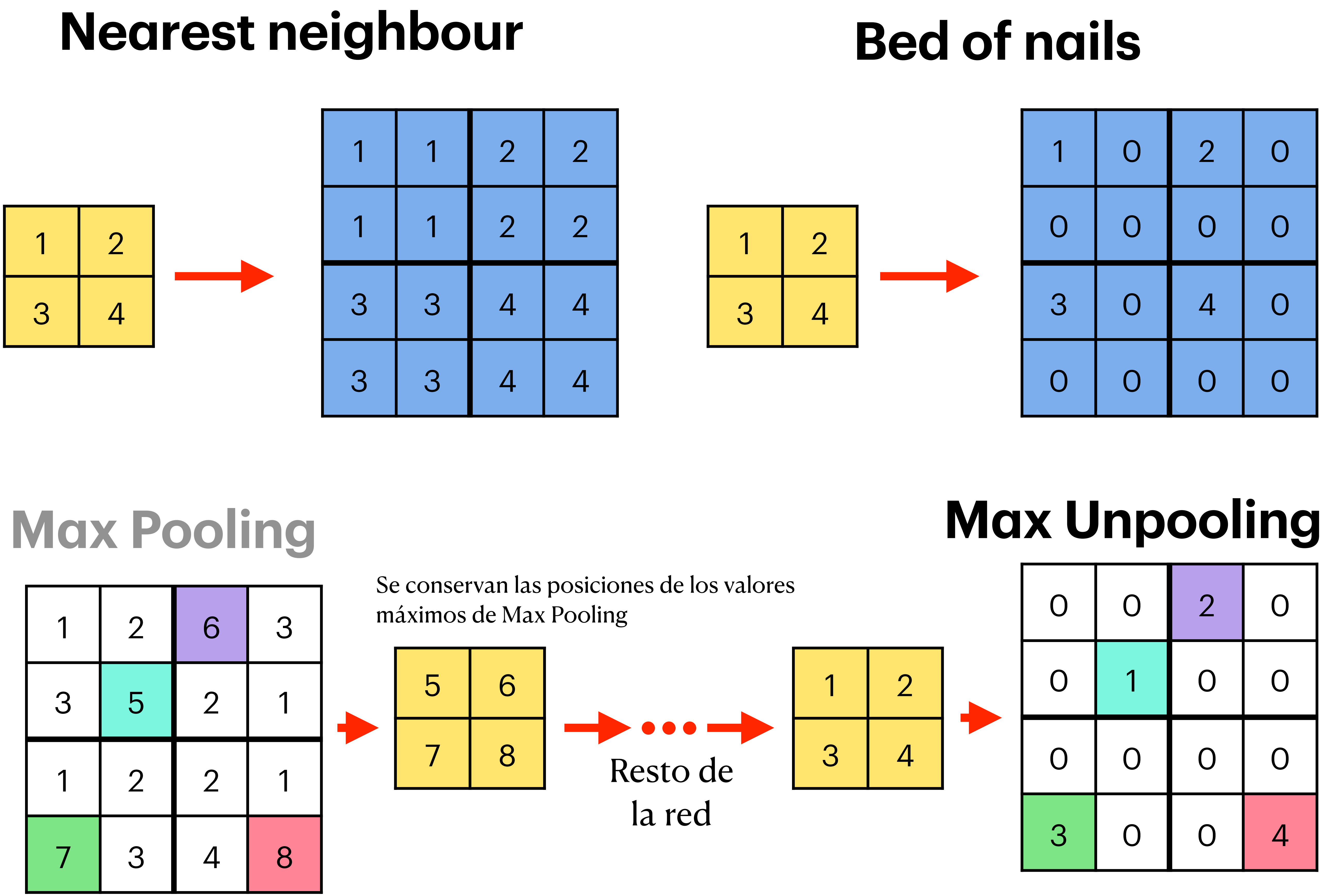
Image
Conv block
Conv block
Max pooling
Conv block
Conv block
Max pooling
Conv block
Conv block
Max pooling
Conv block
Conv block
Max pooling
Conv block
Conv block
Unpooling
Trans Conv block
Trans Conv block
Unpooling
Trans Conv block
Trans Conv block
Unpooling
Trans Conv block
Trans Conv block
Unpooling
Trans Conv block
Trans Conv block
Sigmoid
Mask

Capa de convolución transpuesta

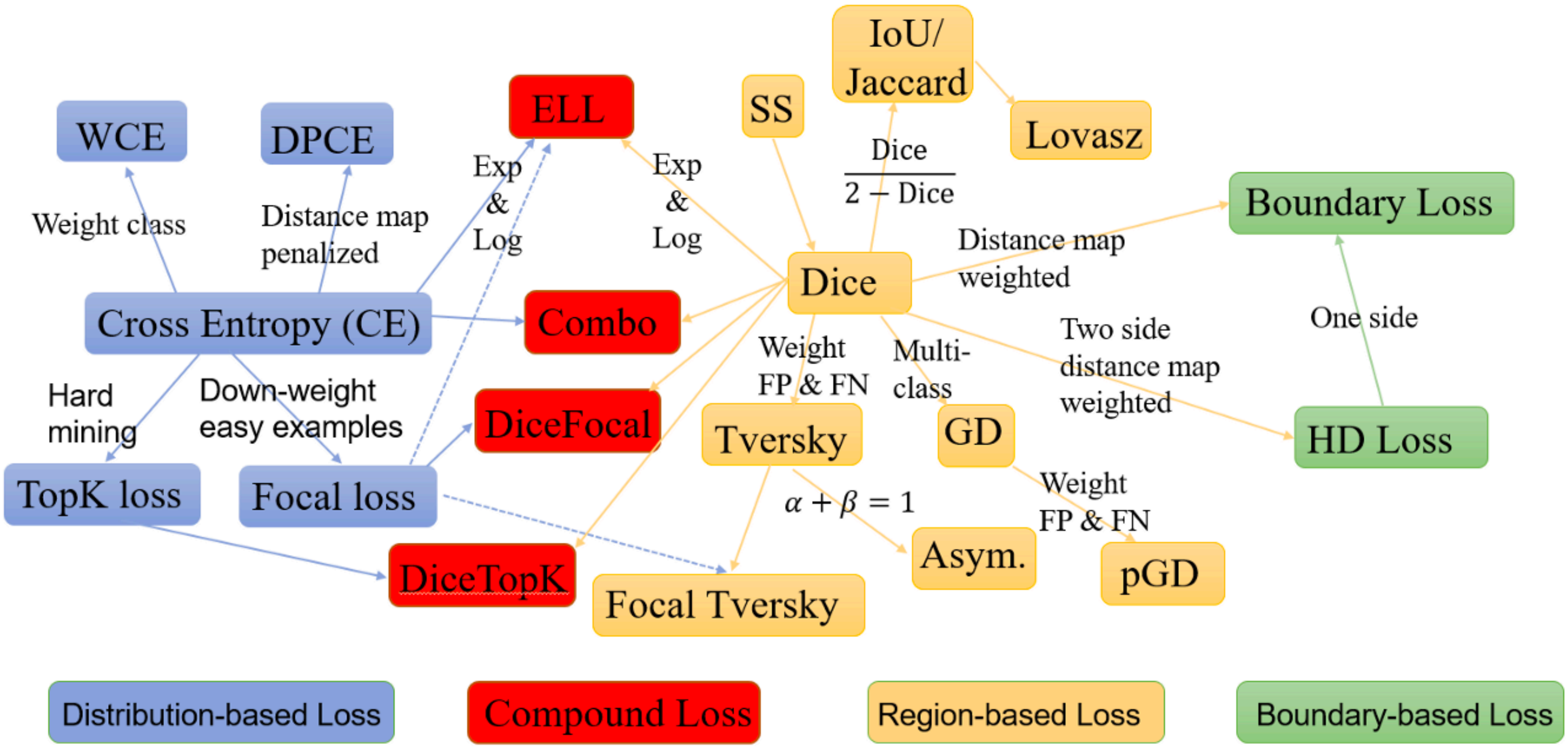
Ejemplo: filtro de 2x2, stride 1 y cero padding



Capa de upsampling o unpooling



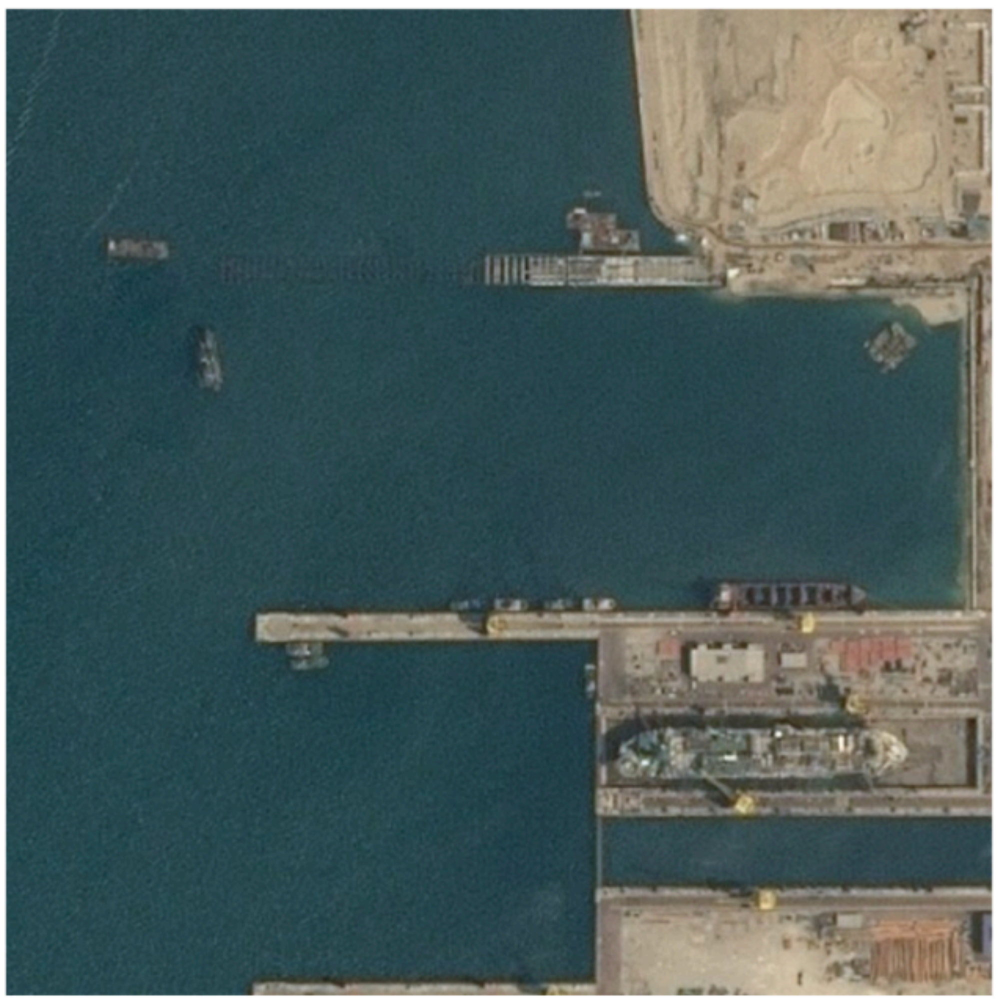
Función de pérdida (loss function)

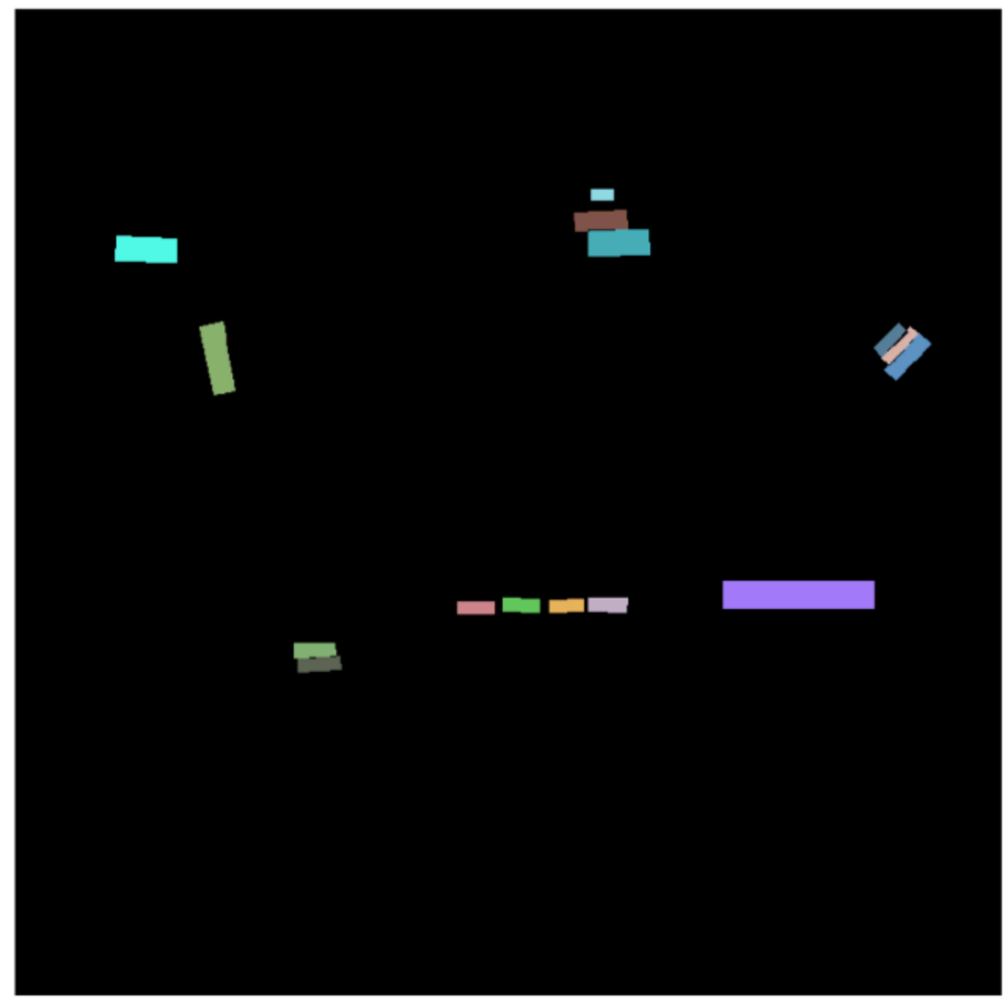


Las más usadas son **Dice**, o **Tversky** para clases poco balanceadas

Métricas

- Pixel Accuracy
- Intersection-Over-Union (Jaccard Index)
- Dice Coefficient (F1 Score)



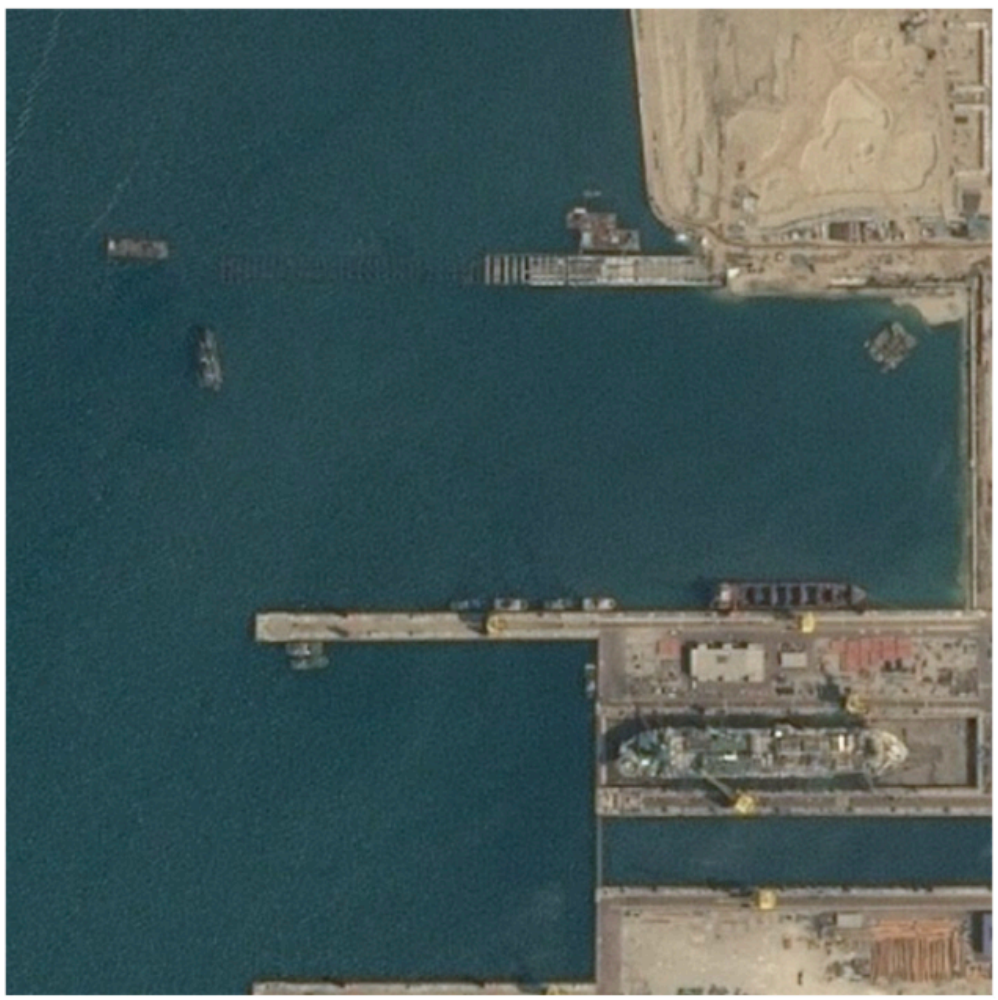


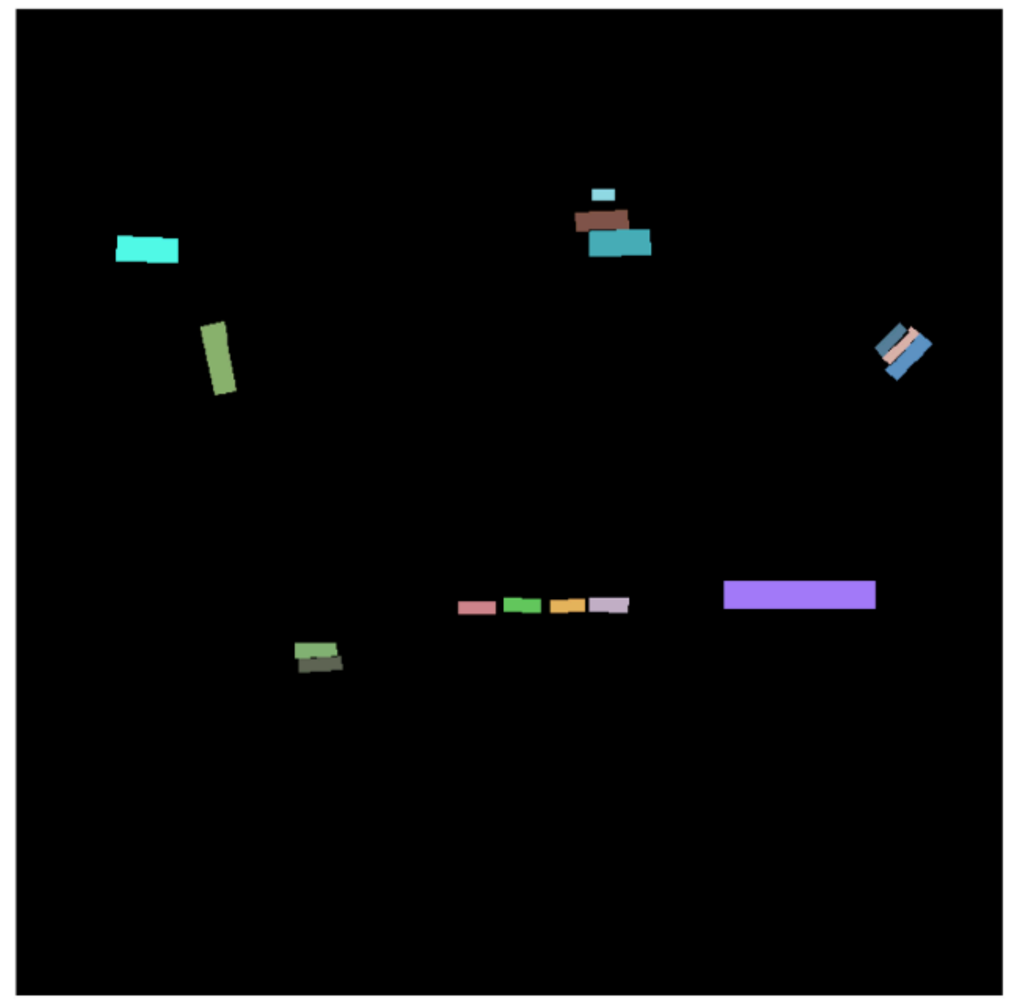
Métricas

- Pixel Accuracy

Es el porcentaje de píxels correctamente clasificado

Es la masa sencilla de entender pero no la mejor



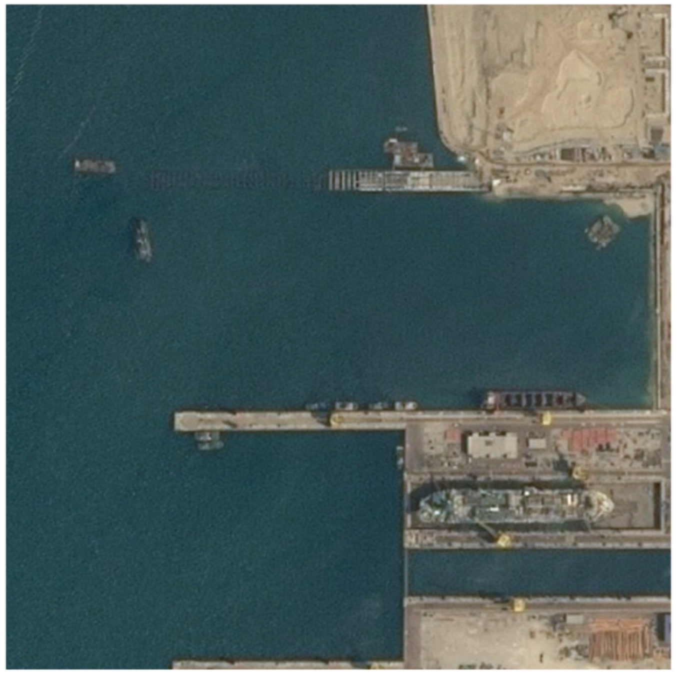


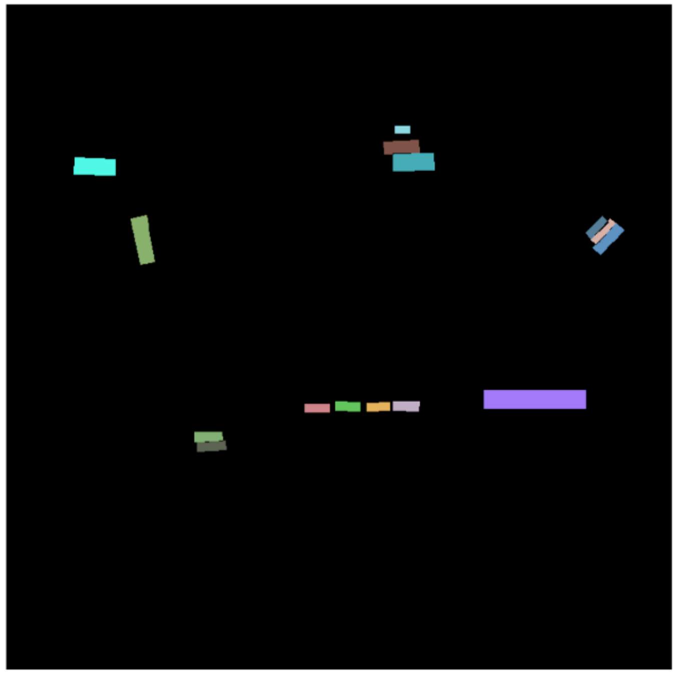
Métricas

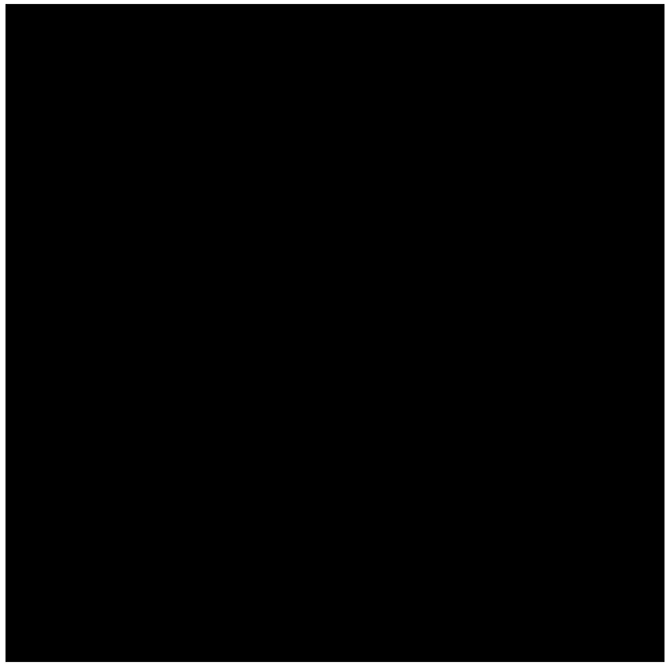
- Pixel Accuracy

Es el porcentaje de píxels correctamente clasificados

Es la más sencilla de entender pero no la mejor





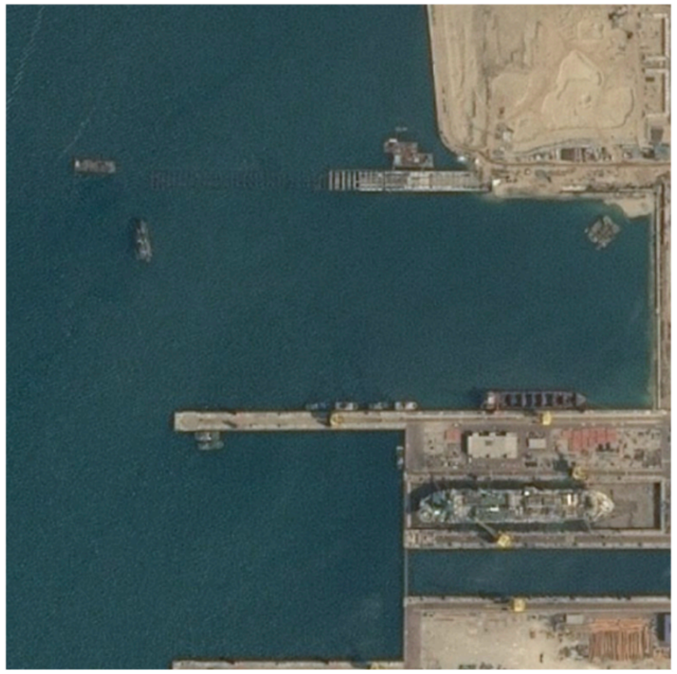


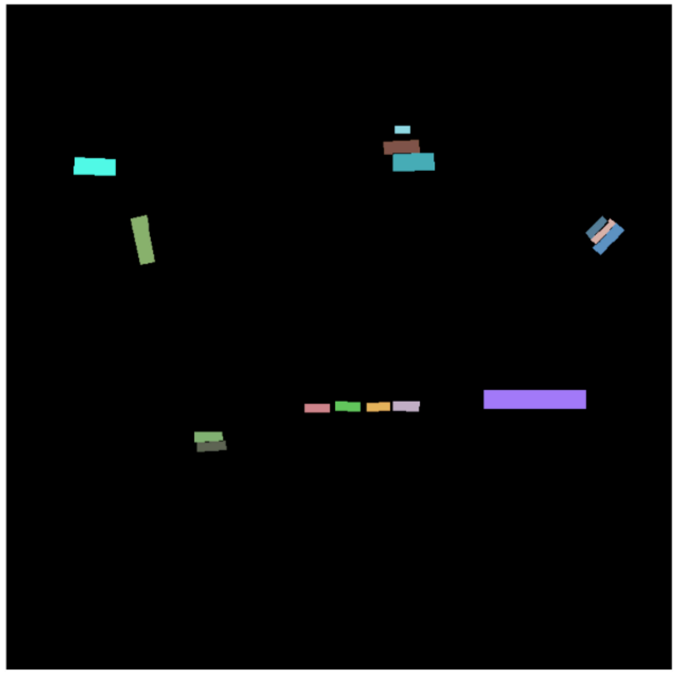
Esta máscara conseguiría el 95% de exactitud.
El culpable es el desbalanceo de clases

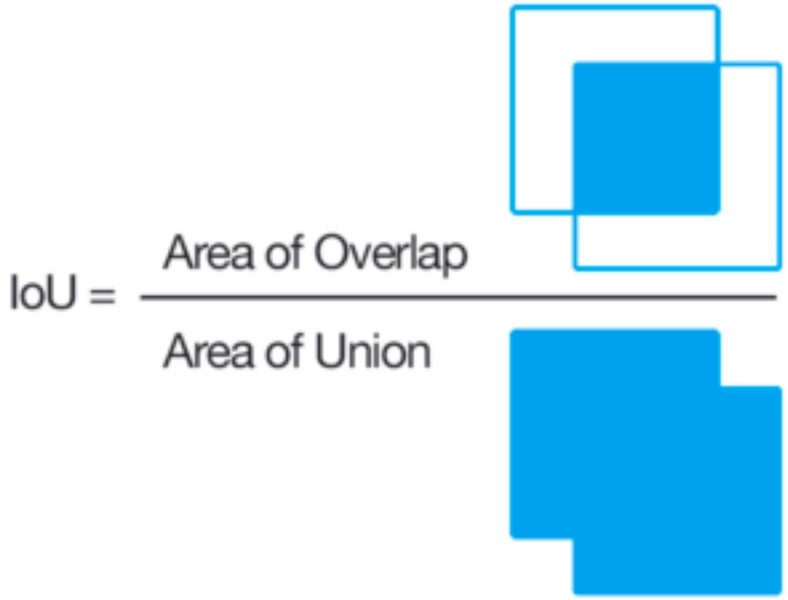
Métricas

- Intersection-Over-Union (Jaccard Index)

Es el área de superposición entre la segmentación predicha y la real dividida por el área de la unión entre la segmentación predicha y la real



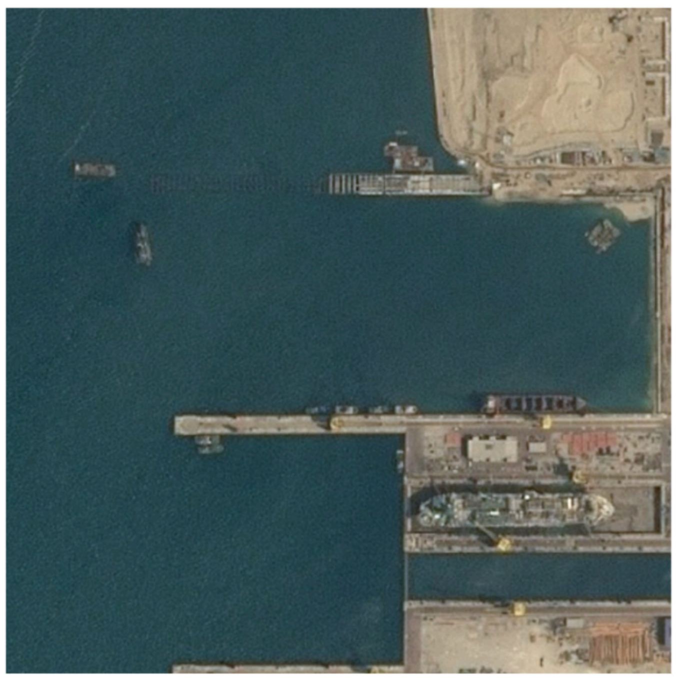


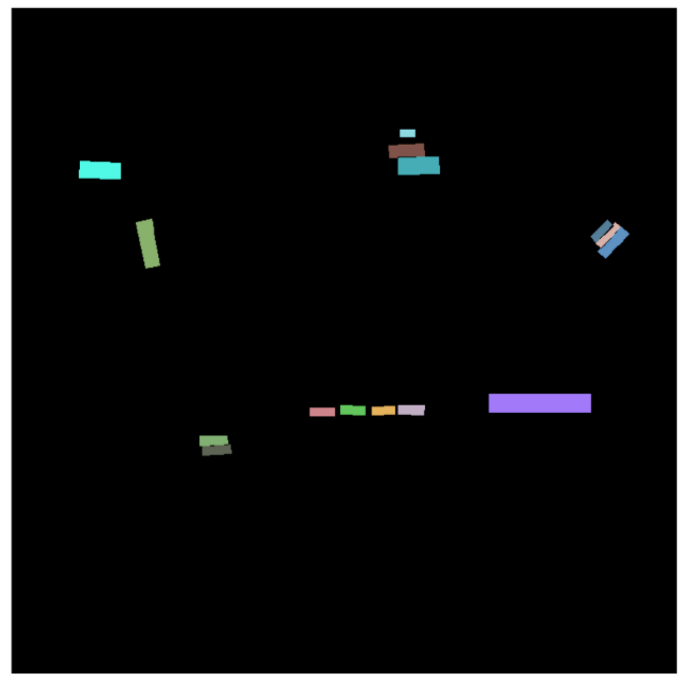


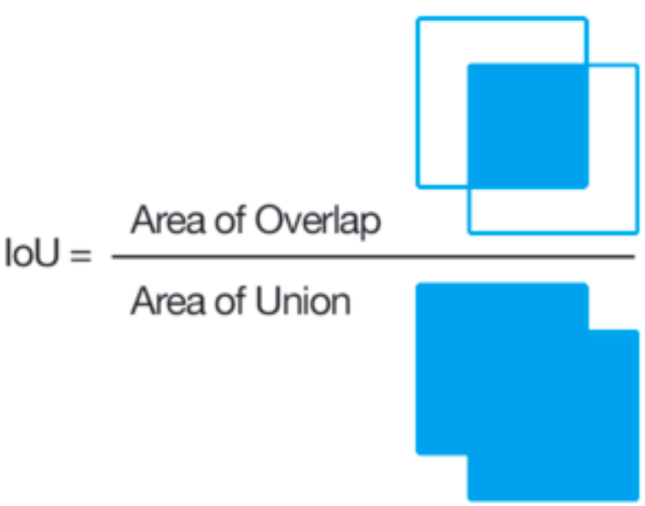
Métricas

- Intersection-Over-Union (Jaccard Index)

Es el área de superposición entre la segmentación predicha dividida por el área de la unión entre la segmentación predicha y el realidad







Supongamos área total de la imagen = 100

Area superpuesta para barcos = 0
Area unión para barcos = 5

IoU = 0 / 5 = 0%

Area superpuesta para fondo = 95
Area unión para fondo = 100

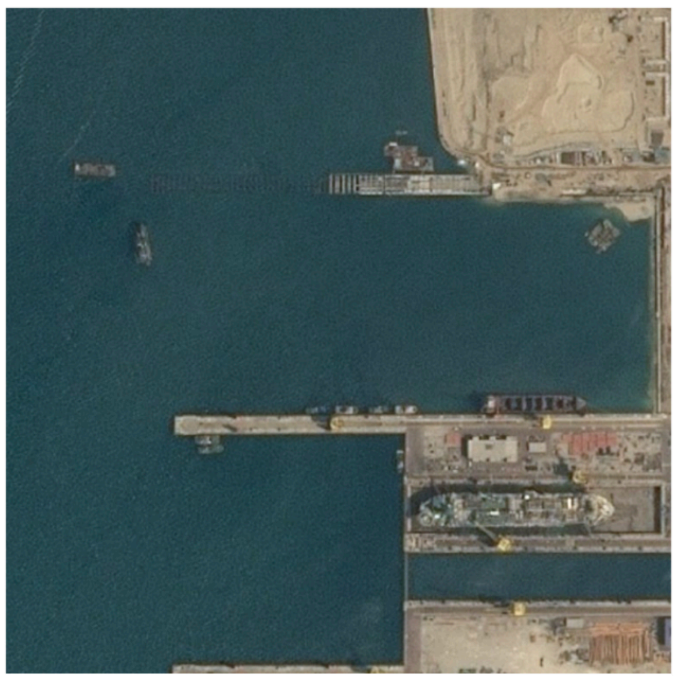
IoU = 95 / 100 = 95%

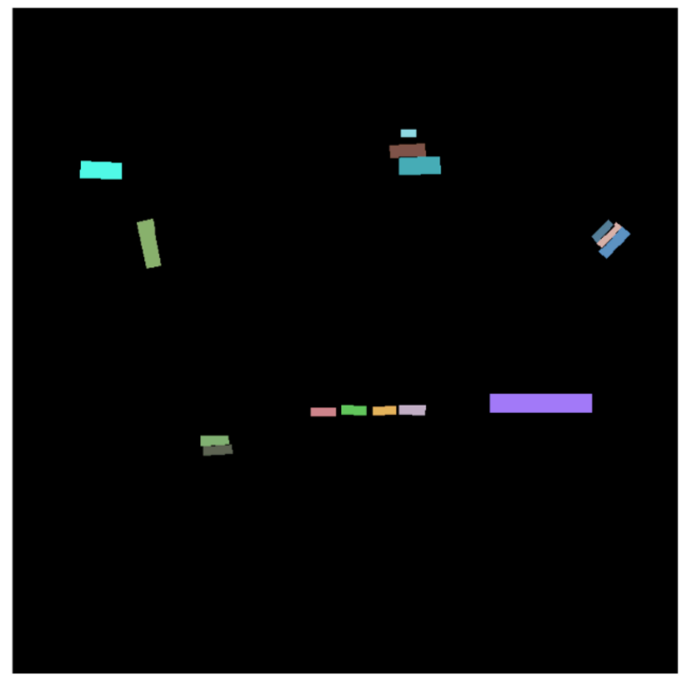
IoU media= (0 + 95) / 2 = 47.5%

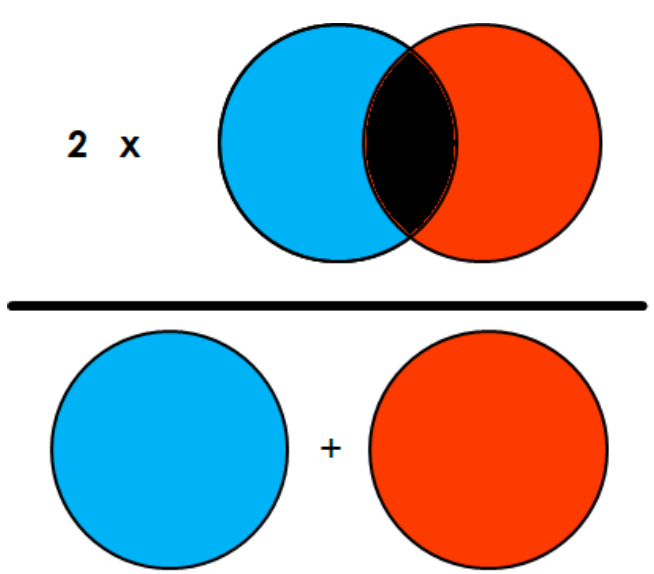
Métricas

- Dice Coefficient (F1 Score)

Es 2 * el área superpuesta dividido por el número total de píxels en ambas imágenes







Nº total de píxels en ambas imágenes = 200

Área de superposición para barcos = 0

2 * 0 / 200 = 0

Área de superposición para fondo = 95

2 * 95 / 200 = 0.95

Dice = (barcos + fondo) / = (0 + 0.95) / 2 = 47.5%

Mismo valor que IoU, pero no siempre será así