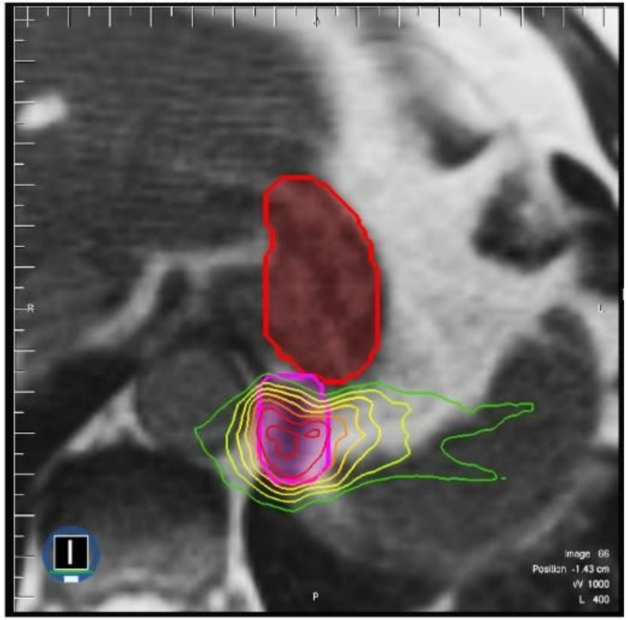
Segmentación semántica y de instancias de órganos del tracto gastrointestinal

Visión por Computadora II - CEIA
Diego Tomás Gomila Molina
Trinidad Monreal

Descripción del problema

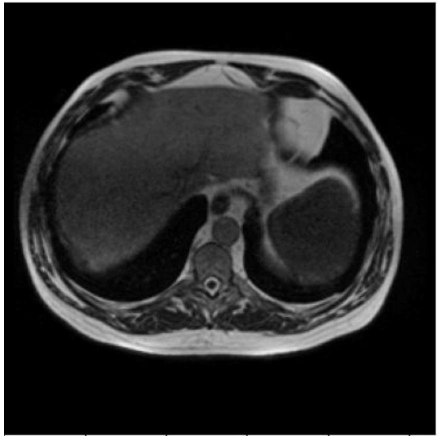
[UW-Madison GI Tract Image Segmentation]

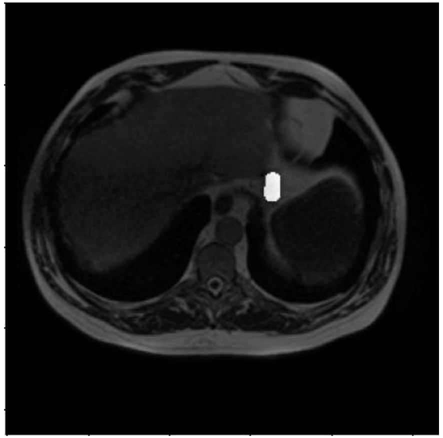
- 5 millones de personas diagnosticadas con cáncer del tracto gastrointestinal (2019).
  - 2.5 millones → elegibles para radioterapia.
- Radiación en altas dosis al tumor evitando el estómago y los intestinos.
  - Tecnología MR-Linacs → oncólogo debe delinear manualmente la posición de los órganos (15 - 60 min).
- **Objetivo**: segmentar intestino grueso, intestino delgado y estómago empleando técnicas de visión por computadora.



Dataset UW-Madison GI Tract Image Segmentation

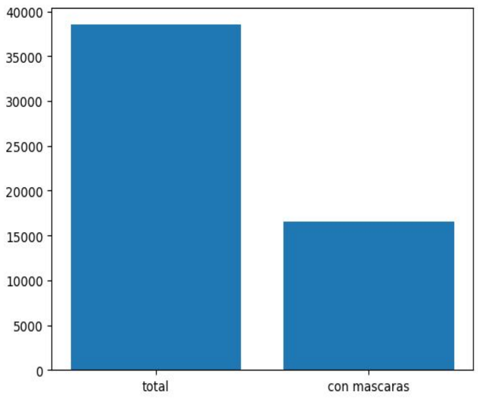
- Dividido por casos y días de muestra.
  - En general ⟶ 140 *slices* cada día.



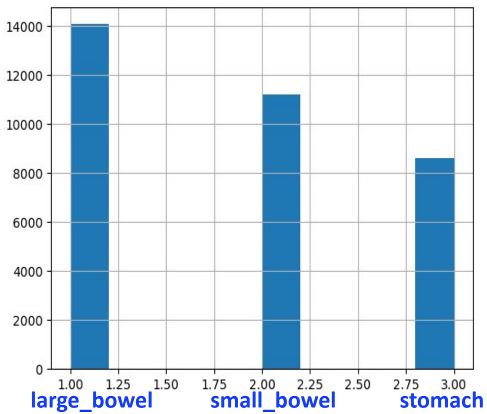


Dataset [UW-Madison GI Tract Image Segmentation]

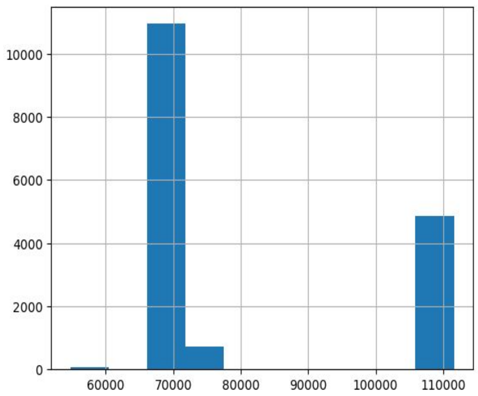
- 38.496 imágenes PNG en escala de grises de 16 bits. Resolución promedio: 293 x 279 px.
    - Imágenes con al menos una máscara: 16.590.
    - En total: 33.913 máscaras de 3 clases diferentes: “large_bowel”, “small_bowel” y “stomach”.



Cantidad de imágenes



Máscaras por clases



img con máscaras por área

Modelos a entrenar

Segmentación semántica: U-Net

1) Modelo clásico usando PyTorch.
2) Modelo con backbone pre-entrenado usando Segmentation Models.

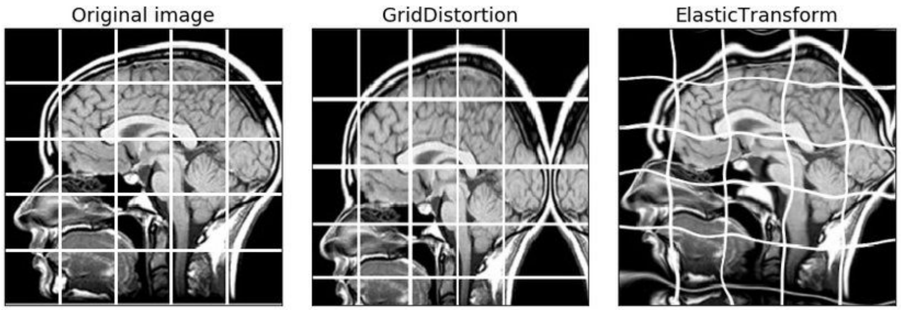
Segmentación de instancias: MASK R-CNN

1) Modelo clásico.
2) Modelo con backbone actualizado y aumento de datos.

Aumento de datos

U-Net

- Horizontal flip
- Vertical flip
- Grid Distortion
- Elastic Transform
- Random Brightness



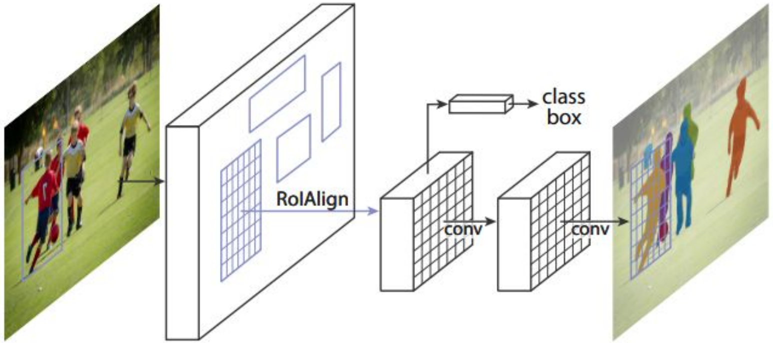
MASK R-CNN

- Random flip: 50%
- Random zoom out: 30%
  - Rango: 1 a 1.2
- Random Photometric Distort: 30%
  - Contrast: 0.875 a 1.125
  - Brightness: 0.875 a 1.125
- Random IoU Crop: 100%
  - Escala: 0.9 a 1.1
  - Relación de aspecto: 0.9 a 1.1

Mask R-CNN

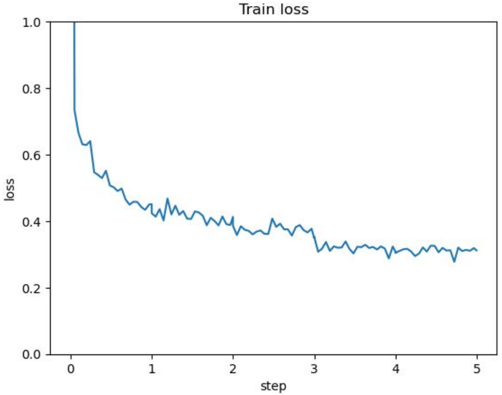
Modelo clásico

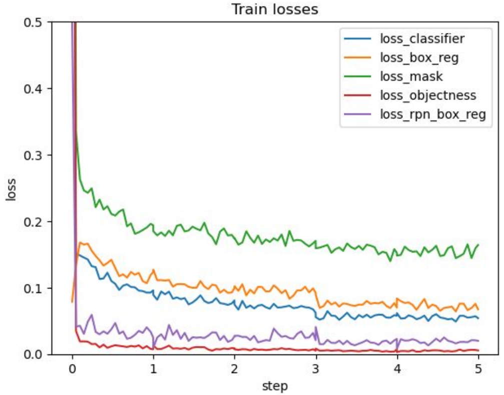
- Backbone: ResNet50
- Parámetros: 43.933.159
- lr_scheduler: 0.0005, disminuye cada 3 épocas
- Pesos pre entrenados:
  - Database: COCO2017
  - categorías 90
  - box_map (on COCO-val2017) 37.9
  - mask_map (on COCO-val2017) 34.6

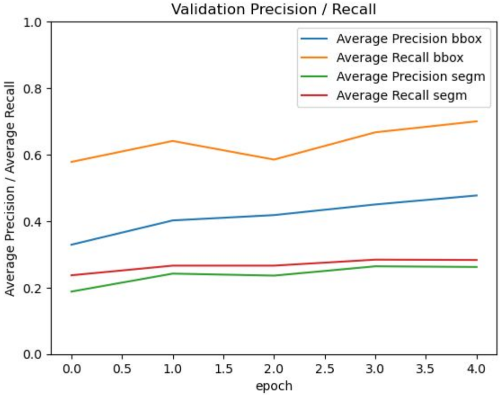


Mask R-CNN

Entrenamiento



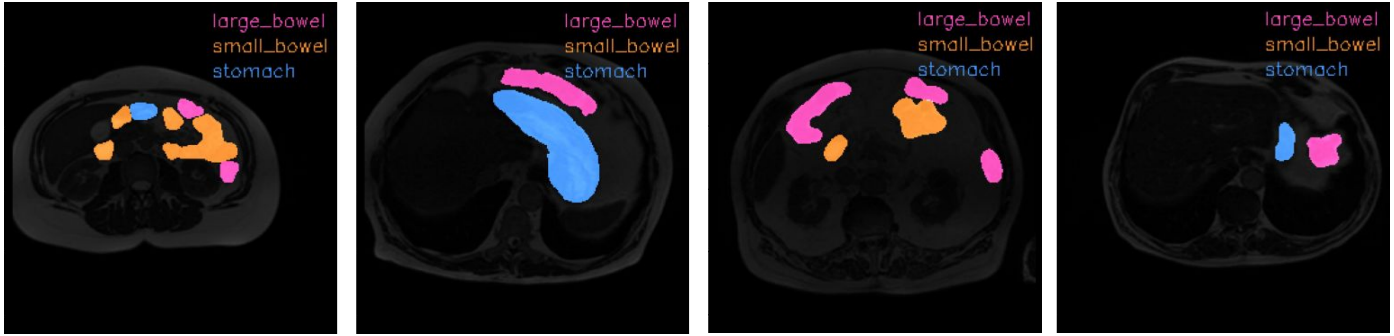




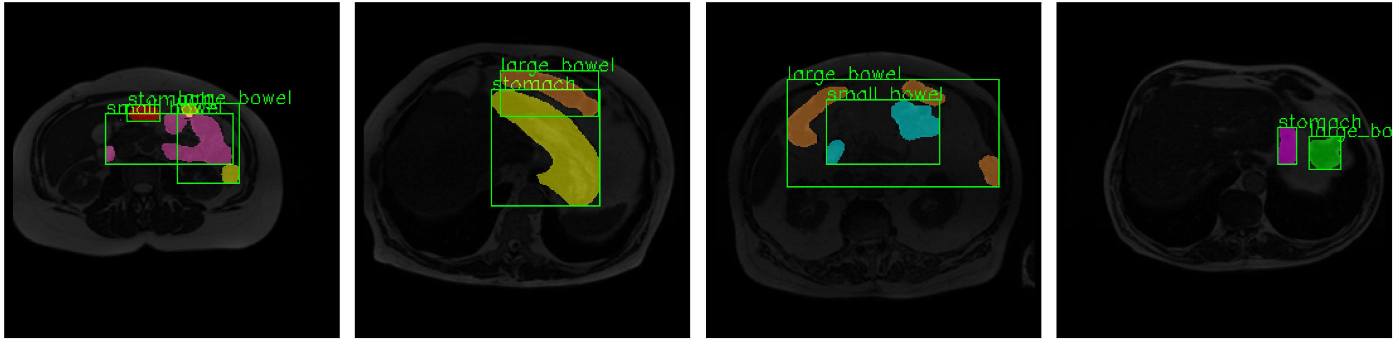
Mask R-CNN

Resultados

Imágenes + máscaras



Predicciones



Mask R-CNN v2

Modelo modificado

- Backbone: ResNet50
- Parámetros: 46.359.409
- Pesos pre entrenados:
  - Database: COCO2017
  - categorías 90
  - box_map (on COCO-val2017)   47.4
  - mask_map (on COCO-val2017)   41.8

Mask R-CNN v2

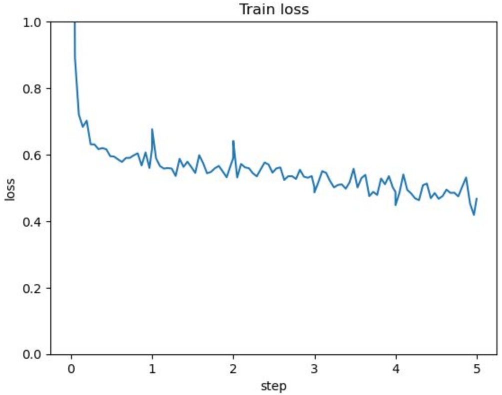
Modelo modificado

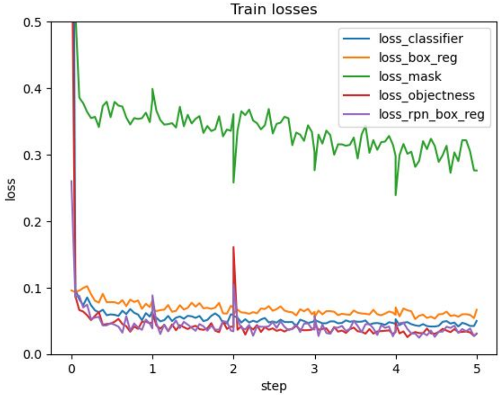
├─RoIHeads: 1-4	[100, 4]	--
└─MultiScaleRoIAlign: 2-5	[1000, 256, 7, 7]	--
└─TwoMLPHead: 2-6	[1000, 1024]	--
└─Linear: 3-33	[1000, 1024]	12,846,080
└─Linear: 3-34	[1000, 1024]	1,049,600
└─FastRCNNPredictor: 2-7	[1000, 4]	--
└─Linear: 3-35	[1000, 4]	4,100
└─Linear: 3-36	[1000, 16]	16,400
└─MultiScaleRoIAlign: 2-8	[100, 256, 14, 14]	--
└─MaskRCNNHeads: 2-9	[100, 256, 14, 14]	--
└─Conv2dNormActivation: 3-37	[100, 256, 14, 14]	590,080
└─Conv2dNormActivation: 3-38	[100, 256, 14, 14]	590,080
└─Conv2dNormActivation: 3-39	[100, 256, 14, 14]	590,080
└─Conv2dNormActivation: 3-40	[100, 256, 14, 14]	590,080
└─MaskRCNNPredictor: 2-10	[100, 4, 28, 28]	--
└─ConvTranspose2d: 3-41	[100, 256, 28, 28]	262,400
└─ReLU: 3-42	[100, 256, 28, 28]	--
└─Conv2d: 3-43	[100, 4, 28, 28]	1,028

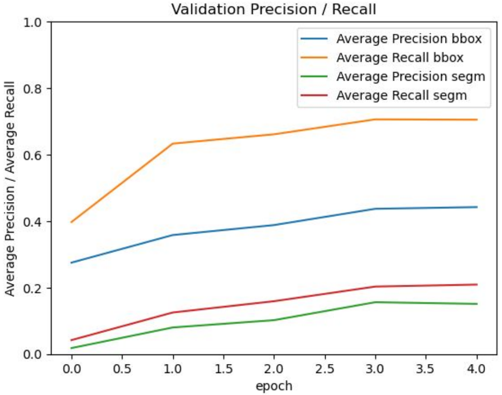
├─RoIHeads: 1-4	[1, 4]	--
└─MultiScaleRoIAlign: 2-5	[1000, 256, 7, 7]	--
└─FastRCNNConvFCHead: 2-6	[1000, 1024]	--
└─Conv2dNormActivation: 3-33	[1000, 256, 7, 7]	590,336
└─Conv2dNormActivation: 3-34	[1000, 256, 7, 7]	590,336
└─Conv2dNormActivation: 3-35	[1000, 256, 7, 7]	590,336
└─Conv2dNormActivation: 3-36	[1000, 256, 7, 7]	590,336
└─Flatten: 3-37	[1000, 12544]	--
└─Linear: 3-38	[1000, 1024]	12,846,080
└─ReLU: 3-39	[1000, 1024]	--
└─FastRCNNPredictor: 2-7	[1000, 91]	--
└─Linear: 3-40	[1000, 91]	93,275
└─Linear: 3-41	[1000, 364]	373,100
└─MultiScaleRoIAlign: 2-8	[1, 256, 14, 14]	--
└─MaskRCNNHeads: 2-9	[1, 256, 14, 14]	--
└─Conv2dNormActivation: 3-42	[1, 256, 14, 14]	590,336
└─Conv2dNormActivation: 3-43	[1, 256, 14, 14]	590,336
└─Conv2dNormActivation: 3-44	[1, 256, 14, 14]	590,336
└─Conv2dNormActivation: 3-45	[1, 256, 14, 14]	590,336
└─MaskRCNNPredictor: 2-10	[1, 91, 28, 28]	--
└─ConvTranspose2d: 3-46	[1, 256, 28, 28]	262,400
└─ReLU: 3-47	[1, 256, 28, 28]	--
└─Conv2d: 3-48	[1, 91, 28, 28]	23,387

Mask R-CNN

Entrenamiento



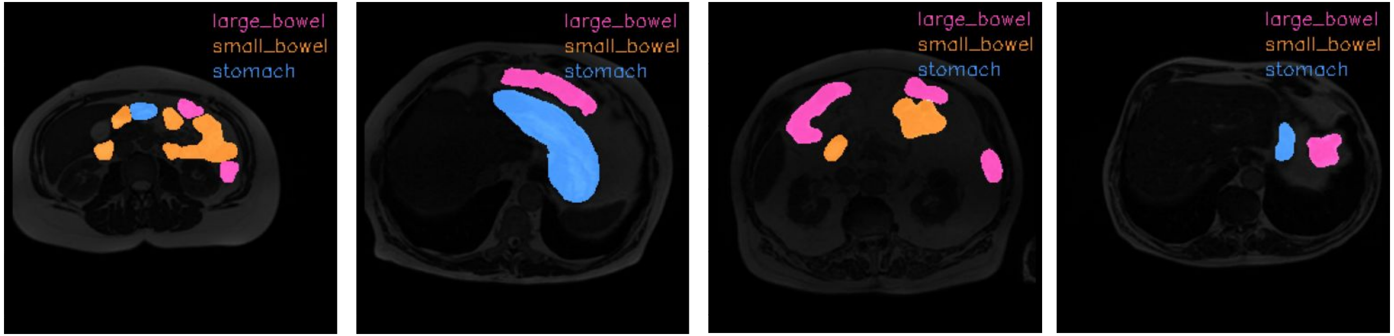




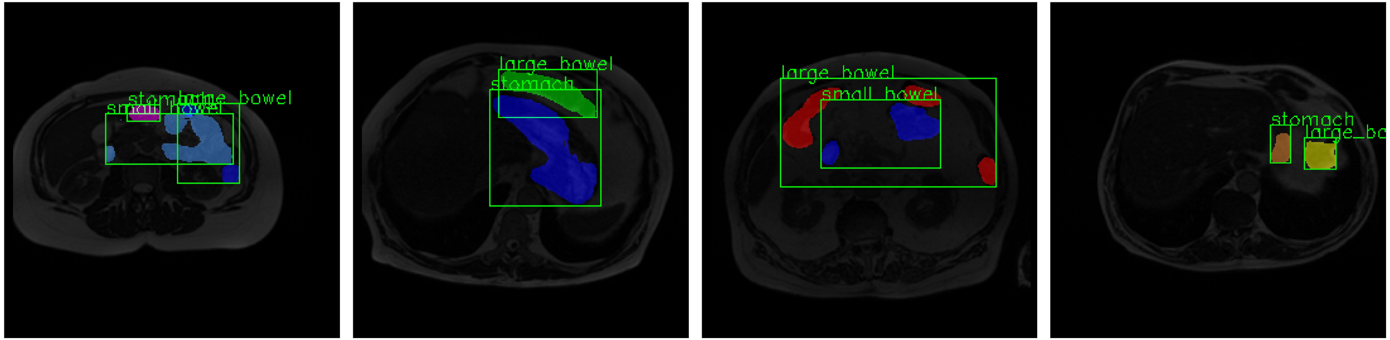
Mask R-CNN

Resultados

Imágenes + máscaras



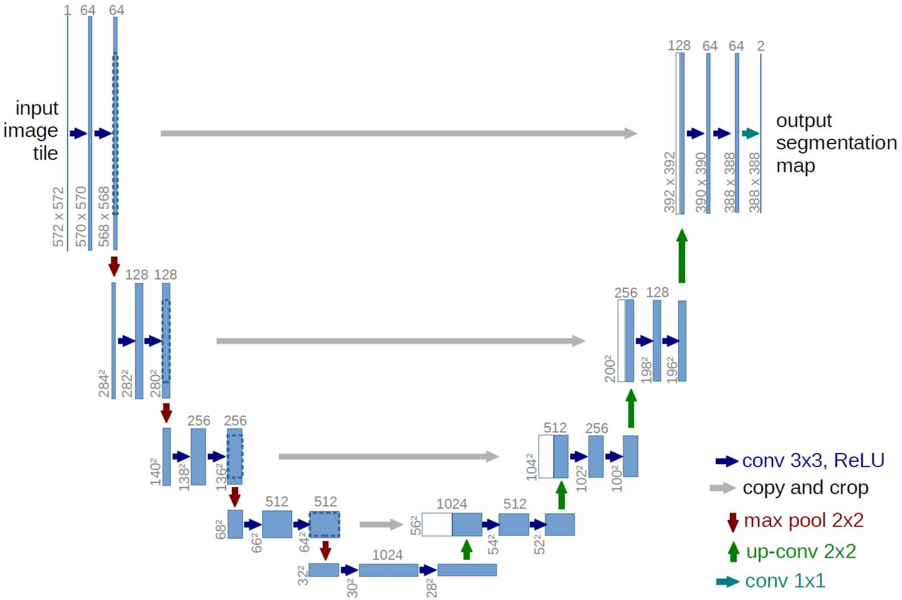
Predicciones



U-Net

Modelo 1

- Learning rate scheduler
- Parámetros: 30,101,475
- Dice Loss Function

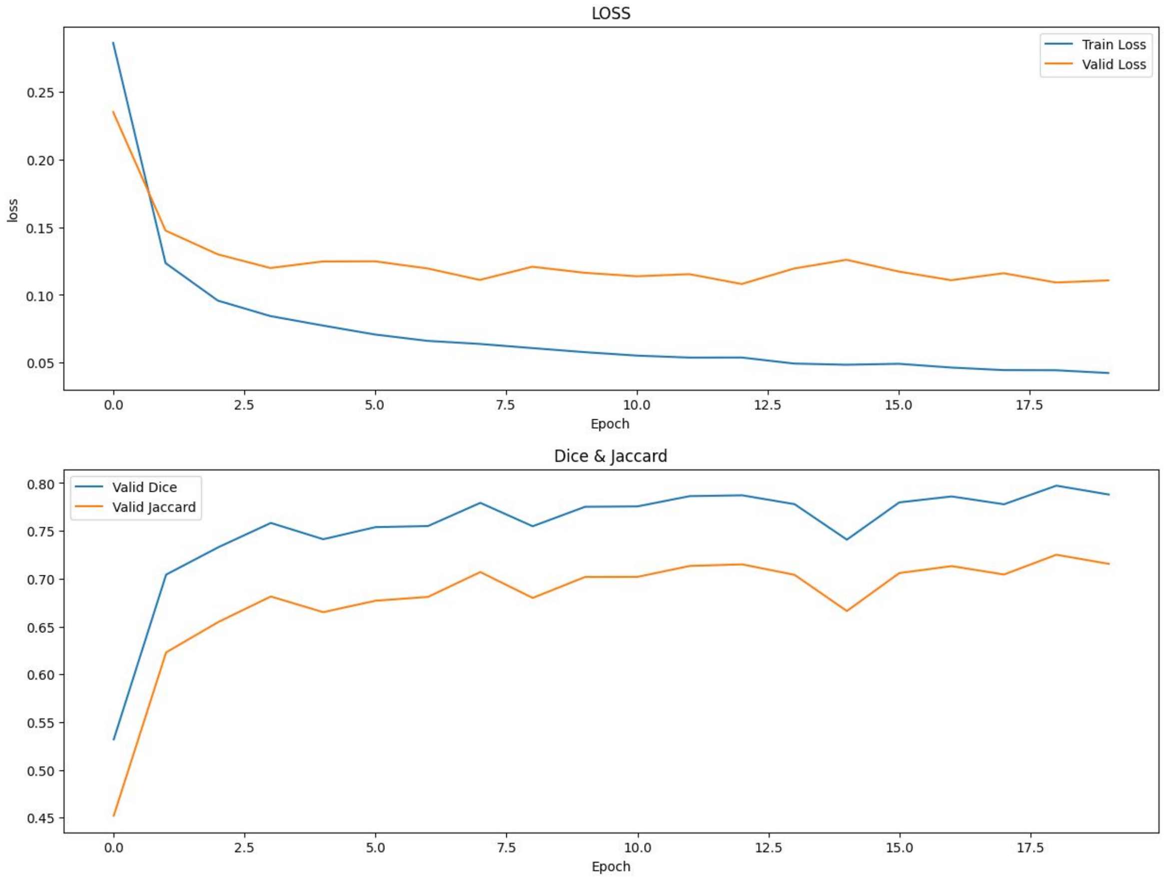


Ronnemberg et al, U-Net: Convolutional Networks for Biomedical Image Segmentation

U-Net

Entrenamiento

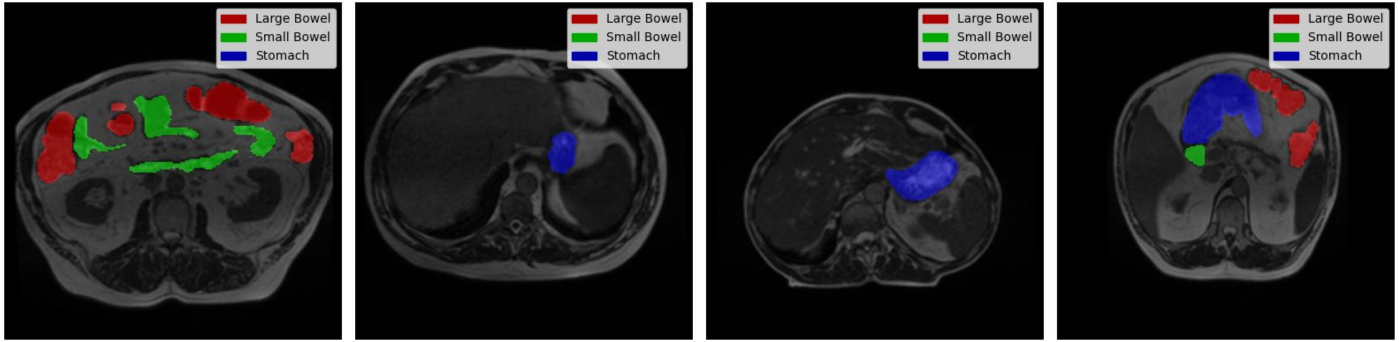
- Valid Dice: 0.7973
- Valid IoU: 0.7251



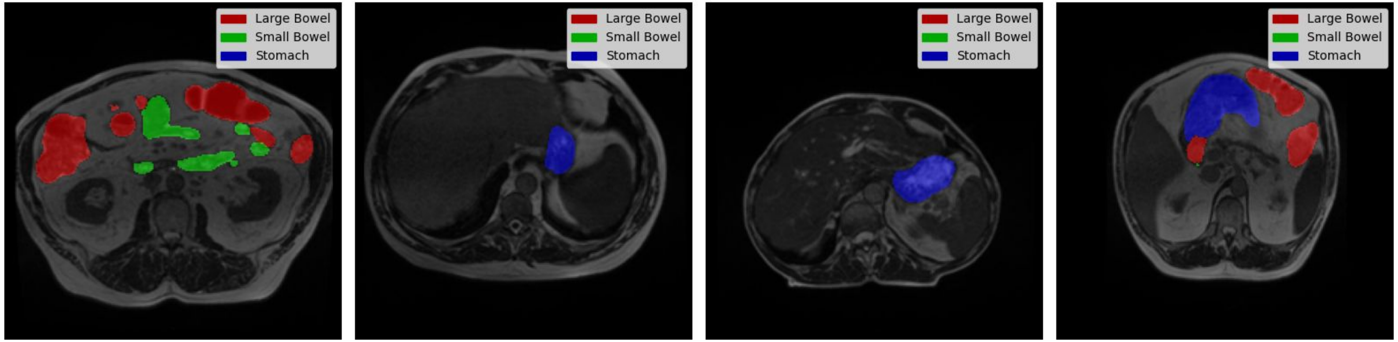
U-Net

Resultados

Imágenes + máscaras



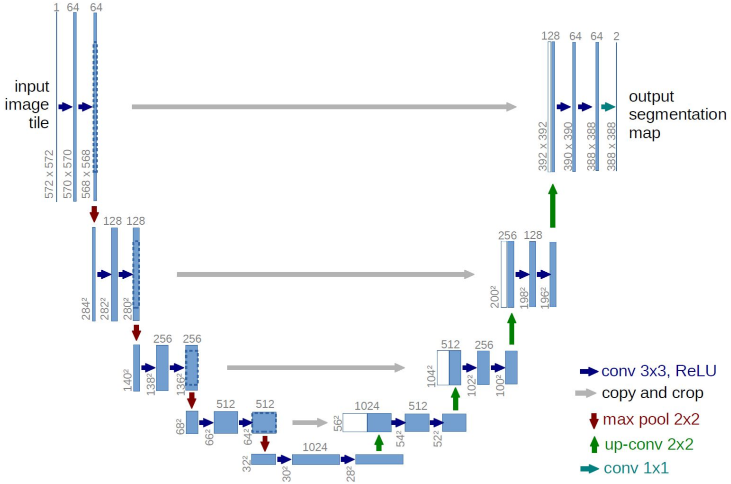
Predicciones

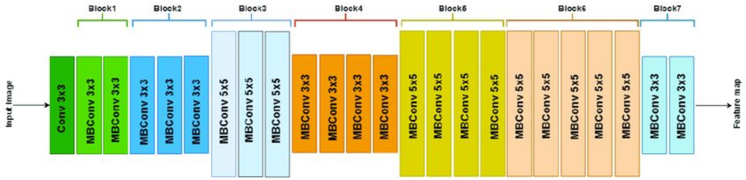


U-Net

Modelo 2

- Backbone: EfficientNet-B1
    - Pre entrenamiento en ImageNet



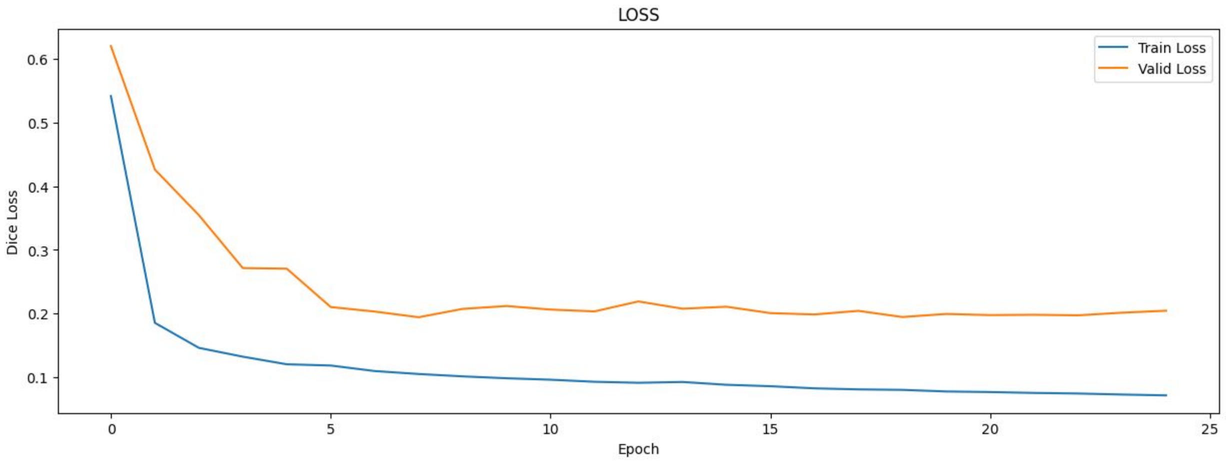


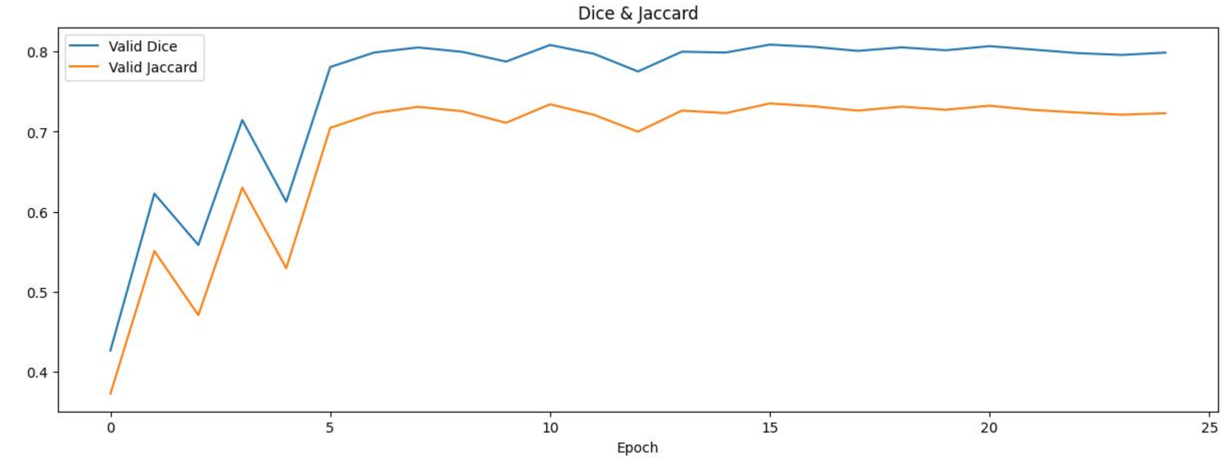
EffUnet-SpaGen: An Efficient and Spatial Generative Approach to Glaucoma Detection

U-Net

Entrenamiento

- Valid Dice: 0.8087
- Valid IoU: 0.7352

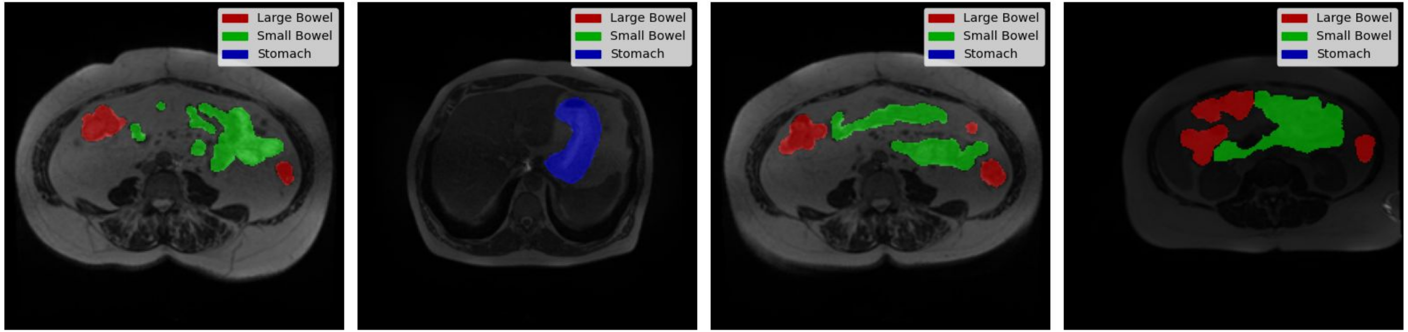




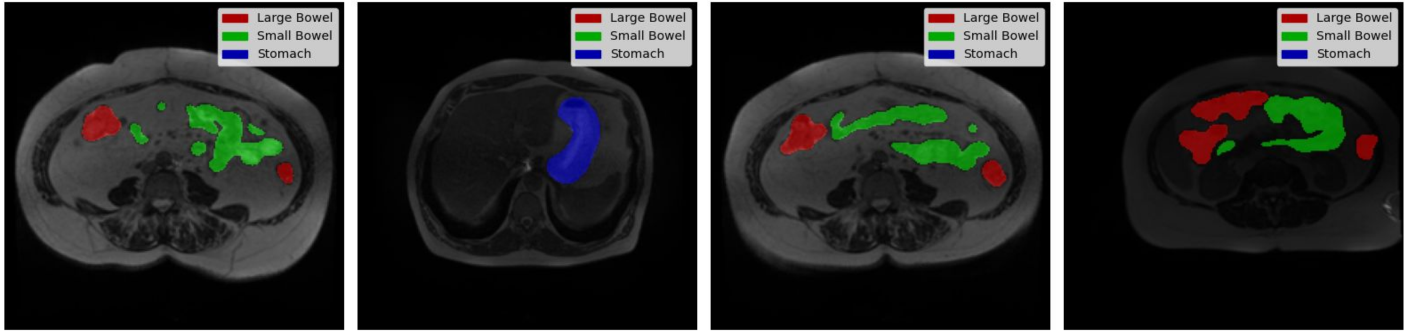
U-Net

Resultados

Imágenes + máscaras



Predicciones



Conclusiones

- Se logró realizar la tarea de segmentación con diferentes grados de éxito.
- MASK R-CNN
  - El transfer learning realizado ha logrado disminuir en gran medida el tiempo de entrenamiento necesario.
  - El aumento de datos permite extender el entrenamiento por más épocas antes de llegar a un sobre entrenamiento.
- U-Net
  - El transfer learning realizado no aportó mejoras significativas en el tiempo de entrenamiento ni las métricas obtenidas sin él.
- Ambos modelos pueden mejorar si se realizan entrenamientos durante un tiempo más prolongado.