

Santiago de Cali, 10 de marzo de 2024.

Doctora

Dr. Gloria Inés Álvarez Vargas

Directora Maestría en Ciencia de Datos.

Facultad de Ingeniería y Ciencias

Pontificia Universidad Javeriana de Cali

Con el fin de cumplir con los requisitos exigidos por la Universidad para optar por el título de Magíster en Ciencia de Datos, nos permitimos presentar a su consideración el proyecto denominado **“Análisis comparativo de Modelos de Segmentación en imágenes de tomografía computarizada (CT) del área abdominal”**, el cual fue realizado por los estudiantes Jan Polanco Velasco con código 2336323 , Stefania Astudillo Bello con código 8976067 pertenecientes a la Maestría en Ciencia de Datos, bajo la dirección del profesor Dr. Julián Gil González .

El suscrito director del Proyecto Aplicado autoriza para que se proceda a hacer la evaluación de este proyecto, toda vez que ha revisado cuidadosamente el documento y avala que ya se encuentra listo para ser presentado y sustentado oficialmente.

Atentamente,

Jan Polanco Velasco
C.C 1062'290.362 de Santander de Quilichao

Stefania Astudillo Bello
C.C 1143'849.764 de Santiago de Cali

Dr. Julián Gil González
C.C 1088'286.439 de Pereira

Santiago de Cali, 10 de marzo de 2024.

Señores

Pontificia Universidad Javeriana Cali.

Dr. Gloria Inés Álvarez Vargas

Directora de Carrera de Maestría en Ciencia de Datos.

Facultad de Ingeniería y Ciencias

Con el fin de cumplir con los requisitos exigidos por la Universidad para llevar a cabo el Proyecto de Grado y posteriormente optar por el título de Magíster en Ciencia de Datos, me permito presentar como Director al Dr. Julián Gil González identificado con C.C. 1088'286.439 de Pereira del Proyecto de Grado denominado **“Análisis comparativo de Modelos de Segmentación en imágenes de tomografía computarizada (CT) del área abdominal”**, el cual será realizado por los estudiantes, Jan Polanco Velasco con código 2336323 , Stefania Astudillo Bello con código 8976067 .

Atentamente,

Jan Polanco Velasco
C.C 1062'290.362 de Santander de Quilichao

Stefania Astudillo Bello
C.C 1143'849.764 de Santiago de Cali

Dr. Julián Gil González
C.C 1088'286.439 de Pereira

Pontificia Universidad Javeriana Cali
Facultad de Ingeniería y Ciencias.
Maestría en Ciencia de Datos.
Anteproyecto de Grado.

Análisis comparativo de Modelos de Segmentación en imágenes de tomografía computarizada (CT) del área abdominal

Jan Polanco Velasco
Stefania Astudillo Bello

Director: Dr. Julián Gil González

10 de marzo de 2024



Ficha Resumen

TÍTULO: “Análisis comparativo de Modelos de Segmentación en imágenes de tomografía computarizada (CT) del área abdominal”

1. **ÁREA DE TRABAJO:** Ciencia de Datos aplicada al campo de la medicina, específicamente en el análisis de imágenes de tomografía computarizada (CT) del área abdominal.
2. **TIPO DE PROYECTO:** Investigación aplicada
3. **ESTUDIANTES:** Jan Polanco Velasco y Stefania Astudillo Bello .
4. **CORREO ELECTRÓNICO:** janpolanco@javerianacali.edu.co, stefania19@javerianacali.edu.co.
5. **DIRECCIÓN:** Carrera 100 No 34-96, Apto 402 Torre 8. Santiago de Cali - Valle del Cauca. Carrea 22 No 5-44
6. **TELÉFONOS:** +57 321 4497869 y +57 300 6306997
7. **DIRECTOR:** Dr. Julián Gil González
8. **VINCULACIÓN DEL DIRECTOR:** Profesor asistente
9. **CORREO ELECTRÓNICO DEL DIRECTOR:** julian.gil@javerianacali.edu.co
10. **PALABRAS CLAVE:** segmentación de imágenes, tomografía computarizada, área abdominal, Deep Learning, Machine Learning, precisión, recall, F1-score, coeficiente Dice.
11. **FECHA DE INICIO:** 18 de Septiembre de 2023
12. **FECHA DE FINALIZACIÓN:** 5 de Octubre de 2024
13. **RESUMEN:** El presente proyecto de investigación tiene como objetivo desarrollar y comparar modelos de segmentación de Deep Learning para la segmentación de imágenes de tomografía computarizada (CT) del área abdominal. El enfoque tradicional de segmentación manual realizado por radiólogos expertos es laborioso, propenso a errores y consume mucho tiempo. Por lo tanto, se busca utilizar técnicas de Machine Learning, como el *Deep Learning*, para mejorar la precisión y eficiencia de este proceso.
La segmentación de imágenes de CT abdominal consiste en identificar y delimitar estructuras anatómicas o patológicas en las imágenes, lo que permite un diagnóstico más preciso y un tratamiento más efectivo de enfermedades abdominales. En este proyecto, se implementarán y entrenarán modelos ampliamente utilizados en tareas de segmentación, como U-Net, SegNet y Mask R-CNN. Posteriormente, se evaluará el rendimiento de cada modelo utilizando métricas de precisión, recall, F1-score, coeficiente Dice, entre otras.
La importancia de este proyecto radica en la necesidad de una segmentación precisa de las

imágenes de CT abdominal para facilitar el diagnóstico y tratamiento de enfermedades. El uso de técnicas de Deep Learning tiene el potencial de mejorar la precisión y eficiencia de este proceso, lo que resulta en una identificación más precisa y rápida de las regiones de interés. Además, la automatización de este proceso mediante técnicas de Machine Learning puede ahorrar tiempo y recursos, además de reducir la influencia de errores humanos, obteniendo así resultados más consistentes y confiables.

El proyecto seguirá una serie de pasos para alcanzar sus objetivos. En primer lugar, se recopilarán y procesarán los datos del CT abdominal. Luego, se implementarán y entrenarán diversos modelos de segmentación de Deep Learning, ajustando los parámetros adecuados para el problema específico. A continuación, se evaluará el rendimiento de cada modelo utilizando métricas relevantes, como precisión y recall, y se realizará un análisis comparativo de los resultados obtenidos. La investigación y comparación de diferentes modelos de Deep Learning permitirá determinar cuál es el más efectivo en la segmentación de imágenes de CT en órganos abdominales.

Índice general

1. Definición Del Problema	3
1.1. Planteamiento Del Problema	3
1.2. Formulación del Problema	4
2. Objetivos Del Proyecto	7
2.1. Objetivo General	7
2.2. Objetivos Específicos	7
3. Marco Teórico y Antecedentes	9
3.1. Marco Teórico	9
3.2. Antecedentes	11
4. Justificación	19
5. Objetivo 1	23
5.1. Preprocesamiento de la base de datos	23
5.1.1. Descripción de la base de datos	24
5.2. Propuestas de pre-procesamiento para las imágenes.	25
5.2.1. Preprocesamiento y Eliminación de Ruido en Imágenes de TC	25
5.2.2. Ampliación de Datos (<i>Data Augmentation</i>)	26
6. Objetivo 2	31
6.1. Implementación de modelos de Deep Learning	31
6.1.1. Modelo U-Net	31
6.1.2. Modelo SegNet	33
6.1.3. Modelo Mask R-CNN	33
6.1.4. División de datos y validación cruzada	33
6.1.5. Entrenamiento de modelos	33
7. Objetivo 3	35
7.1. Validación del modelo obtenido	35
7.1.1. Evaluación de rendimiento	35
7.1.2. Iteración y ajuste	35
7.1.3. Interpretación de métricas	35
7.1.4. Análisis de resultados y comparación	35
7.1.5. Generación de conclusiones y recomendaciones	36

8. Conclusiones y trabajos futuros	37
8.1. Conclusiones	37
8.2. Trabajos futuros	37
9. Anexos	39
Bibliografía	41

Índice de figuras

5.1. Imágen y mascara original del paciente cero	24
5.2. Imágen y mascara sin camilla del paciente cero	27
5.3. Imágen del paciente cero aumentada	29

Índice de cuadros

Introducción

Definición Del Problema

1.1. Planteamiento Del Problema

La tomografía computarizada (CT) es una técnica de diagnóstico por imágenes que permite obtener cortes axiales del cuerpo humano y reconstruirlos en diferentes planos. Esta técnica es útil para detectar y evaluar diversas patologías, como lesiones en órganos abdominales ocasionadas por diferentes enfermedades. Los problemas con la tomografía radican en la similitud de las tonalidades de ciertos órganos y en las características propias de las imágenes de CT, como el ruido, la baja resolución y el contraste limitado [16].

Los especialistas médicos son los encargados de la cuantificación de la forma y el tamaño de los órganos, el diagnóstico de enfermedades, la evaluación de la respuesta al tratamiento y la planificación y navegación quirúrgica. En [16], este proceso requiere una segmentación manual de las regiones, pero es un proceso lento, tedioso y propenso a errores humanos. Por ello, se requiere de métodos automáticos o semiautomáticos que faciliten esta tarea y que permitan obtener resultados precisos, rápidos y reproducibles.

Otros problemas o desafíos asociados a los métodos de segmentación utilizados están relacionados con la reserva que tienen los centros médicos, la disponibilidad limitada de datos y códigos, así como la baja eficiencia computacional. Estas limitaciones dificultan la aplicación generalizada de la segmentación automática de imágenes CT abdominales en entornos clínicos[16]

Por lo tanto, el DL es una técnica de *Machine Learning* (ML) que ha demostrado su eficacia en diversos problemas de visión artificial, como la clasificación, detección y segmentación de imágenes. Estas redes son capaces de aprender características jerárquicas y abstractas a partir de los datos de entrada, sin necesidad de una extracción manual de rasgos. Además, pueden adaptarse a diferentes dominios y tipos de imágenes, lo que las hace adecuadas para el análisis de imágenes médicas. A pesar de los avances en la segmentación de imágenes CT utilizando redes neuronales profundas, aún existen desafíos importantes por abordar.

Los desafíos que tiene el DL en la segmentación de imágenes CT están relacionados con la precisión y robustez, es decir, las redes neuronales profundas deben ser capaces de segmentar con precisión las regiones de interés en las imágenes CT, incluso en presencia de ruido, artefactos y variaciones en la forma y tamaño de las estructuras anatómicas. Otro problema está relacionado con la generalización, es decir, las redes neuronales deben poder generalizar el conocimiento aprendido

durante el entrenamiento para segmentar diferentes tipos de imágenes CT, incluyendo diferentes modalidades, adquisiciones y pacientes. Esto implica superar la variabilidad intrínseca de las imágenes médicas.

Y por último otro gran desafío es la eficiencia computacional, es decir, dado que las imágenes CT suelen tener una gran resolución espacial y se generan en grandes volúmenes, es necesario desarrollar técnicas que permitan segmentar las imágenes de manera rápida y eficiente, sin comprometer la calidad de los resultados [25].

Por lo tanto, en este proyecto se plantea la necesidad de abordar la problemática con la segmentación de órganos abdominales en imágenes de tomografía computarizada (CT) a partir de la comparación de distintas redes neuronales profundas ya establecidas, con el objetivo de desarrollar un modelo de comparación robusto y eficiente con métricas que sean precisas y que permitan establecer cual modelo es el adecuado para el problema de segmentación. Este enfoque se espera que tenga un alto grado de generalización y sea capaz de procesar grandes volúmenes de datos CT en tiempo real, lo que podría mejorar significativamente el diagnóstico y tratamiento de enfermedades basadas en imágenes médicas.

1.2. Formulación del Problema

La articulación precisa del problema constituye una piedra angular en cualquier investigación científica, pues establece el fundamento sobre el cual se orienta un estudio riguroso, canalizando los esfuerzos hacia la consecución de hallazgos significativos.

En el marco del análisis comparativo de modelos de segmentación aplicados a imágenes de tomografía computarizada (CT) del área abdominal, la formulación del problema facilita la identificación del objetivo principal. Esta formulación orienta la investigación al definir las cuestiones específicas que requieren atención, tales como las técnicas y algoritmos idóneos para el preprocesamiento de imágenes, los modelos de *Deep Learning* más eficaces para la segmentación, y los métodos de validación apropiados para evaluar el rendimiento del modelo resultante.

Pregunta Problema: ¿Cuál modelo de segmentación se mostró mas eficiente para lograr una segmentación precisa y confiable en imágenes de tomografía computarizada (CT) del área abdominal, al exhibir los mejores resultados en términos de precisión, *recall*, y otras métricas pertinentes, a través de un análisis comparativo de diferentes modelos?

Las cuestiones de sistematización planteadas fueron las siguientes:

- ¿Qué técnicas y algoritmos se evidenciaron como los más adecuados para llevar a cabo un preprocesamiento exhaustivo de las imágenes de tomografía computarizada (CT) del área abdominal, con la finalidad de eliminar el ruido y optimizar la calidad de las imágenes?

-
- ¿Cuáles fueron los tres modelos de *Deep Learning* más eficaces y apropiados para la segmentación de las imágenes de tomografía computarizada (CT) del área abdominal, tomando en consideración la precisión y confiabilidad en los resultados de segmentación?
 - ¿Cómo se efectuó la validación del modelo de segmentación obtenido, evaluando su desempeño en términos de precisión, *recall* y otras métricas relevantes, garantizando su robustez y generalización en diferentes conjuntos de datos?

Objetivos Del Proyecto

2.1. Objetivo General

Realizar un análisis comparativo de modelos de segmentación en imágenes de tomografía computarizada (CT) del área abdominal, con el fin de determinar el modelo más efectivo para obtener una segmentación precisa y confiable, que logre los mejores resultados en términos de precisión, *recall* y otras métricas relevantes.

2.2. Objetivos Específicos

- Realizar un procesamiento previo y exhaustivo de las imágenes de tomografía computarizada (CT) del área abdominal, con el objetivo de eliminar el ruido y mejorar la calidad de las imágenes, facilitando así el proceso de segmentación.
- Implementar y entrenar tres modelos definidos de *Deep Learning* para la segmentación de las imágenes de tomografía computarizada (CT) del área abdominal, utilizando modelos adecuados y técnicas avanzadas de *Machine Learning*, con el propósito de lograr una segmentación precisa y confiable.
- Realizar la validación del modelo de segmentación obtenido, evaluando su desempeño en términos de precisión, *recall* y otras métricas relevantes.

Marco Teórico y Antecedentes

3.1. Marco Teórico

La segmentación de imágenes médicas es una tarea fundamental en el análisis de imágenes médicas, que consiste en identificar y delimitar regiones de interés en las imágenes, como órganos, lesiones y tejidos. La segmentación tiene muchas aplicaciones clínicas, como el diagnóstico, el tratamiento, la planificación quirúrgica y la evaluación de resultados. Sin embargo, la segmentación de imágenes médicas también presenta muchos desafíos, como la variabilidad anatómica, la calidad de imagen, el ruido, la oclusión y la falta de datos etiquetados [7].

Entre las diferentes modalidades de imagen médica, las imágenes de tomografía computarizada (CT) son ampliamente utilizadas para la segmentación de órganos y lesiones debido a su alta resolución espacial y su capacidad para capturar estructuras internas. Las imágenes de CT se obtienen mediante un equipo de rayos X y tecnología computarizada que permite obtener imágenes de cortes axiales del cuerpo. Estas imágenes representan la atenuación de los rayos X al atravesar los tejidos y se expresan mediante números CT o coeficientes de atenuación [10].

Para segmentar las imágenes de CT, se han desarrollado diversos métodos y algoritmos basados en técnicas clásicas o en aprendizaje automático. Entre ellos, los métodos basados en redes neuronales convolucionales (CNN) han demostrado un gran potencial y rendimiento debido a su capacidad para extraer características complejas y proporcionar resultados precisos y eficientes. Las CNN son modelos de aprendizaje profundo que utilizan operaciones de convolución para extraer características de las imágenes. [13].

Sin embargo, las CNN también tienen algunas limitaciones, como la necesidad de una gran cantidad de datos etiquetados para entrenarlas, la dificultad para capturar el contexto global y local de las imágenes y la falta de generalización a nuevos objetos o dominios. Para superar estas limitaciones, se han propuesto algunos métodos innovadores que combinan las CNN con otras técnicas o paradigmas.

Uno de ellos es el *Segment Anything Model* (SAM), que es un modelo de segmentación fundacional que puede segmentar cualquier objeto en una imagen en función de indicaciones proporcionadas por el usuario, como puntos, cajas o textos. SAM está basado en redes neuronales transformadoras (TNN) o simplemente como “Transformers” y se entrena con el conjunto de datos de segmentación más grande disponible. SAM ha logrado resultados impresionantes en la segmentación de imágenes

naturales, pero su rendimiento en la segmentación de imágenes médicas requiere una mayor validación [12].

Para evaluar la calidad o el rendimiento de los métodos o algoritmos de segmentación, se utilizan diferentes medidas o métricas que pueden ser objetivas o subjetivas, dependiendo de si se basan en criterios matemáticos o en la percepción humana. Para esto se pueden considerar: La función de pérdidas (*Loss Function*), Intersección sobre Unión (*Intersection over Union IOU*), Coeficiente de Dice, precisión y exhaustividad (*recall*), para esto se detallarán de forma general las distintas métricas.

La función de pérdida es una métrica utilizada en *Machine Learning* para evaluar qué tan bien se están realizando las predicciones de un modelo en comparación con los valores reales. Esta función cuantifica la discrepancia o error entre las predicciones del modelo y las etiquetas reales de los datos. La Entropía Cruzada (*Cross Entropy*) es una función de pérdida comúnmente utilizada en problemas de clasificación (véase Ecuación 3.1). Esta función compara la distribución de probabilidades predicha por el modelo con la distribución de probabilidades reales de las clases de los datos.

$$CE(y, \hat{y}) = - \sum_{i=1}^n y_i \log(\hat{y}_i). \quad (3.1)$$

Donde CE es la función de pérdida de Entropía Cruzada, y es el vector de probabilidades reales de las clases, \hat{y} es el vector de probabilidades predicho por el modelo y n es el número de clases. La fórmula realiza una sumatoria sobre todas las clases y compara la probabilidad real y_i con la probabilidad predicha \hat{y}_i , calculando el logaritmo de esta última. El signo negativo se utiliza para invertir la dirección de la pérdida, ya que se busca minimizar la función de pérdida.

La función de Intersección sobre Unión (IoU) es una métrica comúnmente utilizada para evaluar la precisión de la detección o segmentación de objetos en tareas de visión por computador (véase Ecuación 3.2). Mide la proporción de píxeles comunes entre la imagen segmentada y la imagen de referencia (*ground truth*).

$$IoU = \frac{A \cap B}{A \cup B} = \frac{\text{Área de Intersección}}{\text{Área de Unión}}. \quad (3.2)$$

La IoU proporciona un valor entre 0 y 1, donde 0 indica una falta de superposición entre las regiones y 1 indica una coincidencia perfecta entre ellas. Un valor más alto de IoU indica una mayor precisión en la detección o segmentación de los objetos.

El coeficiente de Dice es una medida que evalúa la superposición espacial entre dos conjuntos, como imágenes segmentadas y una imagen de referencia (véase Ecuación 3.3). A diferencia de

otras métricas, el Coeficiente de Dice penaliza tanto los falsos positivos como los falsos negativos. Se calcula mediante el doble del área de intersección entre las máscaras predichas y verdaderas, dividido por el área total de ambas máscaras. Un valor más alto indica una mejor coincidencia entre las máscaras [19], [22], [6].

$$\text{Dice} = \frac{2TP}{2TP + FP + FN}. \quad (3.3)$$

La precisión es una medida que evalúa la exactitud de un algoritmo al clasificar píxeles como parte de la región de interés (*véase Ecuación 3.4*). Se calcula como la proporción de píxeles correctamente clasificados como pertenecientes al objeto, respecto a todos los píxeles clasificados como tal [3].

$$\text{Precisión} = \frac{TP}{TP + FP}. \quad (3.4)$$

El *Recall* es una medida que evalúa la exhaustividad de un algoritmo al clasificar píxeles como parte de la región de interés (*véase Ecuación 3.5*). Se calcula como la proporción de píxeles correctamente clasificados como pertenecientes al objeto, respecto a todos los píxeles que realmente pertenecen al objeto [3].

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (3.5)$$

La segmentación puede influir en la eficiencia y la precisión de los modelos, así como en la capacidad para manejar grandes volúmenes de datos y complejidad anatómica o de pequeños elementos [1], [14], [16].

3.2. Antecedentes

Para comenzar una investigación sólida, es fundamental familiarizarse con las investigaciones previas realizadas a nivel mundial y publicadas en revistas indexadas. Estas investigaciones previas proporcionan una base sólida para construir los antecedentes más relevantes.

Se utilizó *Google Scholar*. Se estableció un rango temporal entre 2021 y 2023 para seleccionar los artículos más pertinentes y actualizados en el área de modelos de segmentación de imágenes de tomografía computacional. Estos artículos seleccionados en Google Scholar actúan como punto de partida para la construcción de los antecedentes más relevantes de la investigación actual.

La segmentación de imágenes de CT es una tarea fundamental en el análisis de imágenes médicas, que consiste en identificar y delimitar regiones de interés (*Region Of Interest ROI*) en diversas

imágenes médicas, como órganos, lesiones y tejidos. La segmentación precisa es esencial para muchas aplicaciones clínicas, como el diagnóstico de enfermedades, la planificación del tratamiento y el seguimiento de la progresión de la enfermedad. Los modelos basados en DL han mostrado una gran promesa en la segmentación de imágenes de CT debido a su capacidad para aprender características complejas de las imágenes y proporcionar resultados de segmentación precisos y eficientes.

Sin embargo, los modelos actuales suelen estar adaptados a modalidades y objetivos específicos de imagen, y su capacidad de generalización es limitada. Por lo tanto, desarrollar modelos fundamentales que puedan adaptarse a diversas modalidades y objetivos de imagen es de suma importancia para avanzar en el análisis de imágenes médicas. En este contexto, se han realizado diversas investigaciones para proponer y evaluar diferentes métodos y modelos para la segmentación de imágenes de CT, especialmente en el ámbito del tórax y el abdomen. A continuación, se presentan algunos ejemplos recientes de estos trabajos:

En el artículo **“An automatic method for segmentation of liver lesions in computed tomography images using deep neural networks”** realizado por Araújo et al. (2021) se presenta un método computacional para segmentar lesiones hepáticas en imágenes de tomografía computarizada (CT). El método propuesto utiliza dos modelos de redes neuronales convolucionales profundas y técnicas de procesamiento de imágenes. El método se evaluó usando un conjunto de 131 imágenes de CT del conjunto de datos LiTS (Liver Tumor Segmentation Challenge) [1].

El mejor resultado fue un coeficiente de correlación de Matthews (MCC) del 83,62 %, una sensibilidad del 83,86 %, una especificidad del 99,96 %, un coeficiente de Dice del 82,99 %, un error de superposición volumétrica (VOE) del 27,89 % y una diferencia relativa de volumen (RVD) del 1,69 %. Los autores muestran que el problema de la segmentación de lesiones hepáticas en imágenes de CT se puede resolver eficientemente mediante el uso de redes neuronales convolucionales profundas para definir el alcance del problema y segmentar las lesiones con precisión [1].

El proyecto actual se inspira en el enfoque exitoso de Araújo et al. (2021), pero se diferencia al ampliar el alcance a toda el área abdominal, ajustar las técnicas de procesamiento previo y considerar la evaluación con métricas específicas para la anatomía abdominal.

Por su parte, Radiuk (2020) elabora una investigación llamada **“Applying 3D U-Net Architecture to the Task of Multi-Organ Segmentation in Computed Tomography”** donde propone un modelo modificado de redes neuronales convolucionales totalmente (FCNs) dedicadas al procesamiento de imágenes volumétricas de tomografía computarizada (CT) en las tareas de segmentación semántica automática. El método propuesto se entrenó y probó en un conjunto de datos compilado manualmente de escaneos de CT. La modelo mejorada de 3D U-Net logró un puntaje promedio del coeficiente de similitud de Sørensen-Dice (SDSC) del 84,8 % en el subconjunto de prueba entre varios órganos abdominales. El autor también comparó su modelo con los resultados reconocidos del estado del arte y demostró que los modelos basados en 3D U-Net podrían lograr

un rendimiento y una eficiencia competitivos en la tarea de segmentación multiorgánica [20].

El trabajo de Radiuk (2020), ofrece una perspectiva valiosa al proponer una arquitectura específica para la segmentación multiorgánica con un énfasis en el uso de modelos 3D U-Net. El proyecto actual puede beneficiarse de estas ideas al adaptar la arquitectura, considerar métricas específicas y explorar la eficiencia en la segmentación de órganos en el área abdominal.

En su trabajo llamado **“When SAM Meets Medical Images: An Investigation of Segment Anything Model (SAM) on Multi-phase Liver Tumor Segmentation”**, Hu y Li (2023) exploran la capacidad de SAM para el análisis de imágenes médicas, especialmente para la segmentación de tumores hepáticos multifase (MPLiTS), en términos de indicaciones, resolución de datos y fases. SAM con tres variantes (ViT-B, ViT-L, ViT-H) se valida en un conjunto de datos interno que recopila 388 pacientes con 1552 volúmenes de tomografía computarizada con contraste mejorado multifase (CECT). Los resultados muestran que SAM podría no alcanzar el rendimiento esperado para MPLiTS. Sin embargo, los resultados cualitativos muestran que SAM es una herramienta poderosa para mejorar la eficiencia de la anotación significativamente [11].

Aunque los resultados muestran que SAM puede no alcanzar el rendimiento esperado para MPLiTS, se destaca que SAM es una herramienta poderosa para mejorar la eficiencia de la anotación. Esto puede proporcionar información valiosa para el proyecto actual al considerar no solo el rendimiento cuantitativo sino también aspectos prácticos como la eficiencia en el proceso de anotación. El trabajo de Hu y Li aporta al proyecto actual al explorar la aplicación de SAM en la segmentación de tumores hepáticos multifase. El proyecto puede beneficiarse de esta investigación al adaptar las variantes de modelos y considerar la eficiencia práctica en la anotación de imágenes en el área abdominal.

En el proyecto **“Lung computed tomography image segmentation based on U-Net network fused with dilated convolution”** Chen et al. (2021) proponen una red U-Net (DC-U-Net) fusionada con convolución dilatada y comparan los resultados de la segmentación de imágenes de tomografía computarizada (CT) de pulmón con DC-U-Net, Otsu y el crecimiento regional. Utilizan la intersección sobre la unión (IOU), el coeficiente de Dice y la precisión para evaluar el rendimiento de los tres algoritmos. Los resultados muestran que la imagen segmentada por DC-U-Net es más cercana a la verdad de fondo que las otras dos. El IOU de DC-U-Net es 0.9627 y el coeficiente de Dice es 0.9743, que es cercano a 1 y mucho más alto que los otros dos algoritmos. Los autores concluyen que este método acelera la segmentación, simplifica los pasos de la segmentación de imágenes médicas y proporciona una mejor segmentación para los tejidos posteriores de vasos sanguíneos pulmonares, tráquea y otros [3].

El trabajo de Chen et al. (2021) ofrece un enfoque específico para la segmentación de imágenes de tomografía computarizada de pulmón utilizando la arquitectura DC-U-Net. Mientras que el proyecto actual se centra en el área abdominal, puede tomar ideas sobre arquitecturas de redes y

métricas de evaluación para adaptarlas a sus necesidades específicas.

En el artículo **“Lung Segmentation on High-Resolution Computerized Tomography Images Using Deep Learning: A Preliminary Step for Radiomics Studies”**, Comelli et al. (2020) tienen como objetivo identificar un enfoque de segmentación de aprendizaje profundo automático, preciso y rápido, aplicado al parénquima, utilizando un conjunto de datos muy pequeño de imágenes de tomografía computarizada de alta resolución de pacientes con fibrosis pulmonar idiopática. De esta manera, pretenden mejorar la metodología realizada por los operadores sanitarios en los estudios radiómicos, donde se deben utilizar métodos de segmentación independientes del operador para identificar correctamente el objetivo y, en consecuencia, el modelo de predicción basado en la textura [4].

Los autores investigan dos modelos de aprendizaje profundo: El primero es U-Net, ya utilizado en muchas tareas de segmentación de imágenes biomédicas, y el segundo es E-Net, utilizado para tareas de segmentación de imágenes en coches autónomos, donde la disponibilidad de hardware es limitada y la segmentación precisa es crítica para la seguridad del usuario. Su pequeño conjunto de datos de imágenes está compuesto por 42 estudios de pacientes con fibrosis pulmonar idiopática, de los cuales solo 32 se utilizaron para la fase de entrenamiento. Comparan el rendimiento de los dos modelos en términos de la similitud de su resultado de segmentación con el estándar de oro y en términos de sus requisitos de recursos.

Los resultados muestran que E-Net se puede utilizar para obtener una segmentación precisa (coeficiente de similitud de dice = 95.90 %), rápida (20.32 s) y clínicamente aceptable de la región pulmonar. Los autores concluyen que los modelos de aprendizaje profundo se pueden aplicar eficientemente para segmentar y cuantificar rápidamente el parénquima de los pacientes con fibrosis pulmonar, sin supervisión alguna del radiólogo, con el fin de producir resultados independientes del usuario [4].

El trabajo de Comelli et al. (2020) proporciona una perspectiva valiosa al aplicar modelos de aprendizaje profundo para la segmentación del parénquima pulmonar. Aunque el contexto es diferente, el proyecto actual puede beneficiarse de la elección de modelos y la evaluación de rendimiento presentadas en este trabajo.

En el artículo **“Kidney segmentation from computed tomography images using deep neural network”** Batista da Cruz et al. (2020) proponen un método automático para delimitar los riñones en imágenes de tomografía computarizada (CT) usando técnicas de procesamiento de imágenes y redes neuronales convolucionales profundas (CNNs) para minimizar los falsos positivos. El método propuesto tiene cuatro pasos principales: (1) adquisición del conjunto de datos KiTS19, (2) reducción del alcance usando AlexNet, (3) segmentación inicial usando U-Net 2D y (4) reducción de falsos positivos usando procesamiento de imágenes para mantener los elementos más grandes (riñones) [5].

El método propuesto se evaluó en 210 CTs del conjunto de datos KiTS19 y obtuvo el mejor resultado con un coeficiente de Dice promedio de 96.33 %, un índice de Jaccard promedio de 93.02 %, una sensibilidad promedio de 97.42 %, una especificidad promedio de 99.94 % y una precisión promedio de 99.92 %. En el desafío KiTS19, presentó un coeficiente de Dice promedio de 93.03 %. Los autores concluyen que el problema de segmentación de riñones en CT se puede resolver eficientemente usando redes neuronales profundas para definir el alcance del problema y segmentar los riñones con alta precisión y con el uso de técnicas de procesamiento de imágenes para reducir los falsos positivos [5].

El trabajo de Batista da Cruz et al. (2020) proporciona un enfoque eficiente para la segmentación de riñones en imágenes de tomografía computarizada, destacando el uso de redes neuronales profundas y técnicas de procesamiento de imágenes. Aunque el contexto es diferente, el proyecto actual puede beneficiarse de la metodología estructurada y la atención específica a la reducción de falsos positivos presentes en este trabajo.

En el artículo **“Deep Learning for Image Segmentation: A Focus on Medical Imaging”** Khalifa y Badr (2023) presentan una revisión de los modelos y mecanismos de redes neuronales convolucionales profundas (CNN) más avanzados utilizados en la segmentación de imágenes. Los autores clasifican los modelos de segmentación según su modelo y su principio de funcionamiento principal. Luego, describen las categorías de CNN y discuten varios modelos dentro de cada categoría [13].

A diferencia de otras revisiones existentes, se analizan varias aplicaciones con múltiples adaptaciones de los modelos dentro de cada categoría. Un resumen comparativo se incluye para dar al lector una visión general de los modelos utilizados en diferentes aplicaciones y conjuntos de datos. Este estudio se centra en las aplicaciones de segmentación de imágenes médicas, donde se ilustran los modelos más utilizados y se sugieren otros modelos prometedores que han demostrado su éxito en diferentes dominios. Finalmente, el trabajo actual discute las limitaciones y soluciones actuales junto con las tendencias futuras en el campo [13].

La revisión de Khalifa y Badr (2023) proporciona un marco integral que abarca modelos, adaptaciones, aplicaciones y tendencias en el ámbito de la segmentación de imágenes mediante CNN, con un énfasis específico en la imagen médica. El proyecto actual puede beneficiarse de esta revisión al obtener una visión global de las opciones disponibles y comprender las tendencias actuales en el campo de la segmentación de imágenes médicas.

En el artículo **“Deep learning approaches to biomedical image segmentation”** Haque y Neubert (2020) presentan una revisión de los métodos y mecanismos de redes neuronales convolucionales profundas (CNN) más avanzados utilizados en la segmentación de imágenes. Los autores revisan los conceptos básicos de los métodos de aprendizaje profundo y ofrecen una visión general

de las implementaciones exitosas que involucran la segmentación de imágenes para diferentes aplicaciones médicas [7].

Además, se incluye un resumen comparativo para dar al lector una idea de los modelos utilizados en diferentes aplicaciones y conjuntos de datos. Este estudio se centra en las aplicaciones de segmentación de imágenes biomédicas, donde se ilustran los modelos más utilizados y se sugieren otros modelos prometedores que han demostrado su éxito en diferentes dominios. Finalmente, el trabajo actual discute las limitaciones y soluciones actuales junto con las tendencias futuras en el campo [7].

La revisión de Haque y Neubert (2020) ofrece una perspectiva integral sobre los enfoques de segmentación de imágenes biomédicas mediante CNN, con un enfoque particular en conceptos básicos, implementaciones exitosas y tendencias futuras. El proyecto actual puede beneficiarse de esta revisión al obtener información detallada sobre modelos exitosos y considerar las tendencias actuales en el campo de la segmentación abdominal.

En el artículo **“Accuracy of Segment-Anything Model (SAM) in Medical Image Segmentation Tasks”** He et al. (2023) evalúan la generalización de cero disparos de SAM en 12 conjuntos de datos de imágenes médicas diferentes. Los autores comparan la precisión de SAM con la de cinco algoritmos de segmentación basados en aprendizaje profundo diseñados específicamente para imágenes médicas. Los autores utilizan diferentes indicadores para medir la dificultad de segmentación y el contraste entre el objetivo y el fondo. Los resultados muestran que SAM tiene una precisión significativamente menor que los otros algoritmos en todos los conjuntos de datos, especialmente en imágenes 3D, regiones objetivo que son pequeñas y de bajo contraste, y casos más complejos. Los autores sugieren algunas posibles mejoras para aumentar la precisión de SAM en imágenes médicas [8].

El trabajo de He et al. (2023) ofrece una evaluación crítica de la precisión de SAM en tareas de segmentación de imágenes médicas, identificando sus limitaciones en ciertos escenarios. Aunque la evaluación se centra en el ámbito médico, el proyecto actual puede aprender de estas limitaciones y sugerencias de mejoras para ajustar su enfoque de segmentación abdominal según las necesidades específicas del conjunto de datos y la tarea a realizar.

En el artículo **“Generalist Vision Foundation Models for Medical Imaging: A Case Study of Segment Anything Model on Zero-Shot Medical Segmentation”**, Shi et al. (2023) examinan el reciente modelo de segmentación de cualquier cosa (SAM) en imágenes médicas y reportan resultados cuantitativos y cualitativos de segmentación de cero disparos en nueve conjuntos de datos de segmentación de imágenes médicas, que abarcan diversas modalidades de imagen, como tomografía de coherencia óptica (OCT), resonancia magnética (MRI) y tomografía computarizada (CT), así como diferentes aplicaciones que incluyen dermatología, oftalmología y radiología. Los autores comparan el rendimiento de SAM con el de otros modelos de visión generalista y específicos para imágenes médicas [21].

Los autores encuentran que mientras que SAM presenta un rendimiento de segmentación notable en imágenes del dominio general, su capacidad de segmentación de cero disparos sigue siendo limitada para imágenes fuera de distribución, por ejemplo, imágenes médicas. Además, SAM muestra un rendimiento inconsistente de segmentación de cero disparos en diferentes dominios médicos no vistos [21].

Para ciertos objetivos estructurados, por ejemplo, vasos sanguíneos, la segmentación de cero disparos de SAM falló completamente. En contraste, un simple ajuste fino del mismo con una pequeña cantidad de datos podría conducir a una mejora notable de la calidad de la segmentación, mostrando el gran potencial y la viabilidad de utilizar SAM ajustado para lograr una segmentación precisa de imágenes médicas para un diagnóstico de precisión. El estudio indica la versatilidad de los modelos de visión generalista en la imagen médica y su gran potencial para lograr el rendimiento deseado mediante el ajuste fino y finalmente abordar los desafíos asociados al acceso a conjuntos de datos médicos grandes y diversos en apoyo del diagnóstico clínico [21].

El trabajo de Shi et al. (2023) ofrece una evaluación crítica de SAM en el contexto de imágenes médicas, destacando sus limitaciones en la segmentación de cero disparos y su capacidad de ajuste fino para mejorar la precisión en objetivos médicos específicos. Este estudio puede proporcionar valiosas lecciones sobre la aplicabilidad de SAM en el proyecto actual y resalta la importancia del ajuste fino para adaptar modelos generalistas a tareas médicas específicas.

En el artículo **“Segment Anything in Medical Images”**, Ma y Wang (2023) presentan MedSAM, el primer intento de extender el éxito del modelo de segmentación de cualquier cosa (SAM) a las imágenes médicas, con el objetivo de crear una herramienta universal para la segmentación de diversos objetivos médicos. Específicamente, los autores curan un conjunto de datos de imágenes médicas a gran escala, que abarca más de 200.000 máscaras en 11 modalidades diferentes. Luego, desarrollan un método simple de ajuste fino para adaptar SAM a la segmentación general de imágenes médicas. Experimentos exhaustivos en 21 tareas de segmentación 3D y 9 tareas de segmentación 2D demuestran que MedSAM supera al modelo SAM predeterminado con un coeficiente de similitud de dice (DSC) promedio de 22.5% y 17.6% en las tareas de segmentación 3D y 2D, respectivamente. El código y el modelo entrenado están disponibles públicamente en <https://github.com/bowang-lab/MedSAM> [15].

El trabajo de Ma y Wang (2023) presenta MedSAM como una extensión exitosa de SAM para la segmentación de imágenes médicas, destacando la importancia de un conjunto de datos extenso y diversificado, así como un método de ajuste fino para mejorar el rendimiento. Este artículo puede inspirar al proyecto actual al proporcionar un ejemplo de cómo adaptar modelos generales para tareas médicas específicas.

En el artículo **“Segment Anything Model for Medical Images?”**, Huang et al. (2023)

analizan el rendimiento del modelo de segmentación de cualquier cosa (SAM) en imágenes médicas, y reportan resultados de segmentación de cero disparos en 52 conjuntos de datos de segmentación de imágenes médicas, que abarcan 16 modalidades, 68 objetos y 553K rebanadas. Los autores comparan diferentes estrategias de prueba de SAM en el conjunto de datos COSMOS 553K y evalúan la influencia de diferentes factores (como la complejidad del borde basada en Fourier y el tamaño de los objetos segmentados) en el rendimiento de la segmentación [12].

Los resultados muestran que SAM tiene un mejor rendimiento con pistas manuales como puntos y cajas para la percepción de objetos en imágenes médicas, lo que lleva a un mejor rendimiento en el modo de indicación que en el modo automático. Sin embargo, SAM muestra un rendimiento variable en algunos objetos y modalidades específicos, y es imperfecto o incluso falla totalmente en otras situaciones. Los autores concluyen que la capacidad de segmentación de cero disparos de SAM no es suficiente para garantizar su aplicación directa a la segmentación de imágenes médicas [12].

El trabajo de Huang et al. (2023) ofrece una evaluación crítica del rendimiento de SAM en imágenes médicas, identificando tanto fortalezas como limitaciones. Este análisis puede ser fundamental para el proyecto actual al proporcionar información sobre la viabilidad de SAM en el contexto de segmentación abdominal y resaltar áreas donde podrían ser necesarios ajustes o estrategias adicionales.

En el artículo “**Fast and Low-GPU-memory abdomen CT organ segmentation: The FLARE challenge**”, Ma et al. (2022) presentan el primer desafío de segmentación de órganos abdominales que evalúa tanto la precisión como la eficiencia de los métodos de segmentación. Los autores organizan un conjunto de datos grande y diverso, que incluye 511 casos de 11 centros médicos, y abarca 13 órganos abdominales [16].

Los autores proponen dos medidas de clasificación para evaluar el rendimiento de los métodos participantes: una basada en el coeficiente de similitud de dice (DSC) y otra basada en el tiempo de inferencia y el consumo de memoria GPU. Los resultados muestran que los métodos basados en redes neuronales convolucionales (CNN) superan a los basados en redes neuronales transformadoras (TNN) en términos de precisión, pero tienen una mayor variabilidad en términos de eficiencia. Los autores concluyen que el desafío FLARE proporciona un punto de referencia integral y desafiante para la segmentación de órganos abdominales en CT y fomenta el desarrollo de métodos más precisos y eficientes para la práctica clínica [16].

El trabajo de Ma et al. (2022) presenta el desafío FLARE como una iniciativa para evaluar métodos de segmentación de órganos abdominales, abordando tanto la precisión como la eficiencia. Este enfoque integral puede inspirar al proyecto actual a considerar medidas de evaluación más amplias y a buscar un equilibrio entre la precisión y la eficiencia en la implementación de métodos de segmentación abdominal.

Justificación

El desarrollo del proyecto de investigación propuesto, titulado “Análisis comparativo de Modelos de Segmentación en imágenes de tomografía computarizada (CT) del área abdominal”, se presenta como una iniciativa altamente significativa y pertinente en el contexto de la radiología y la medicina moderna. La segmentación de imágenes de tomografía computarizada del área abdominal es fundamental para identificar y delimitar estructuras anatómicas y patológicas en imágenes médicas, desempeñando un papel crucial en el diagnóstico y tratamiento de diversas enfermedades. En este sentido, la integración de técnicas de *Machine Learning*, particularmente el aprendizaje profundo, se vislumbra como un enfoque vanguardista que puede transformar radicalmente este proceso esencial.

El enfoque tradicional de la segmentación manual, llevado a cabo por radiólogos expertos, presenta limitaciones significativas en términos de tiempo, precisión y consistencia. La automatización de este proceso mediante técnicas de *Machine Learning* no solo ahorra tiempo y recursos, sino que también minimiza los errores humanos inherentes al proceso manual. La fatiga, el sesgo y la variabilidad interobservador, factores recurrentes en la segmentación manual, se eliminan con la implementación de modelos de *Machine Learning*, lo que resulta en resultados más precisos y consistentes.

Cabe destacar que los equipos de TC modernos han experimentado mejoras significativas en la resolución de imagen, esto significa que las imágenes de TC actuales pueden capturar detalles anatómicos mucho más pequeños y sutiles que en el pasado, la segmentación automatizada se vuelve esencial para analizar estas imágenes de alta resolución de manera eficiente y precisa, ya que sería extremadamente laborioso realizar una segmentación manual detallada en estructuras tan minuciosas.

Uno de los pilares fundamentales que respalda la viabilidad de este proyecto es la disponibilidad de un conjunto de datos extenso y diverso. La calidad y diversidad de los datos son esenciales para garantizar que el modelo de *Machine Learning* pueda generalizar y segmentar una amplia variedad de tejidos, órganos y estructuras anatómicas presentes en las imágenes de tomografía computarizada. La representatividad del conjunto de datos es crucial para evitar sesgos y garantizar la robustez y la precisión del modelo. La capacidad de obtener resultados precisos en diferentes tipos de casos clínicos es esencial para avanzar hacia una medicina de precisión, permitiendo diagnósticos más precisos y tratamientos más efectivos para las enfermedades abdominales.

El impacto potencial de este proyecto es significativo y multifacético. En el ámbito de la investigación médica, el proyecto contribuirá al avance general de la ciencia médica al mejorar los

métodos de segmentación y análisis de imágenes. Este avance no solo beneficiará a los investigadores, sino que también proporcionará herramientas más precisas y confiables para los profesionales de la salud, mejorando así la calidad de la atención médica en general.

En el ámbito clínico, la implementación de técnicas de segmentación mejoradas tendrá un impacto directo en el diagnóstico y tratamiento de enfermedades abdominales. La obtención de resultados más precisos y confiables permitirá a los médicos tomar decisiones fundamentadas sobre los tratamientos, lo que podría conducir a una mejora significativa en la calidad de vida de los pacientes. Además, la eficiencia en la atención médica se verá notablemente incrementada, ya que los profesionales de la salud podrán acceder a información detallada de manera más rápida y precisa.

Desde una perspectiva académica, el desarrollo de este proyecto no solo demuestra la excelencia académica de los estudiantes involucrados, sino que también demuestra su capacidad para implementar tecnologías de vanguardia en un contexto médico. Las habilidades adquiridas en este proyecto serán invaluable para la formación profesional de los estudiantes, equipándolos con conocimientos especializados en ciencia de datos y medicina, preparándolos para abordar desafíos futuros en estos campos interdisciplinarios.

El proceso de realizar un análisis comparativo de arquitecturas de deep learning proporciona una oportunidad valorable para la formación y capacitación de estudiantes e investigadores, a través de este ejercicio, se adquiere una experiencia práctica en la implementación y evaluación de modelos de aprendizaje profundo, así como una comprensión más profunda de los desafíos y consideraciones específicas en el contexto médico, dicho análisis comparativo impulsa el avance tanto en la investigación médica como en el desarrollo tecnológico, al tiempo que proporciona una valiosa experiencia de aprendizaje para los involucrados en el proyecto.

Una de las razones para seleccionar la arquitectura U-Net en el desarrollo del proyecto es que este ha sido ampliamente adoptado en la comunidad científica y médica debido a su efectividad en la segmentación de imágenes médicas, al implementar U-Net en el proyecto, se estaría utilizando una arquitectura establecida y validada, lo que facilitaría la comparación con otros enfoques y contribuiría al avance general en el campo de la segmentación de imágenes médicas, contribuyendo así a la calidad y precisión de los resultados, así como al avance continuo en el campo de la medicina de imágenes.

La segmentación de imágenes de tomografía computarizada (CT) del área abdominal es una tarea importante y desafiante, ya que permite obtener información anatómica y funcional de los órganos internos, que puede ser útil para el diagnóstico, el tratamiento y el seguimiento de diversas enfermedades [9]. Sin embargo, la segmentación de estas imágenes presenta varios problemas, como la baja resolución, el ruido, el contraste, la variabilidad y la complejidad de las estructuras, y la presencia de artefactos y patologías [24].

Para resolver estos problemas, se han propuesto diferentes modelos de segmentación basados en *Deep Learning*, que son capaces de aprender de forma automática las características relevantes de las imágenes, y de generar segmentaciones precisas y confiables [18]. Entre estos modelos, se pueden destacar los siguientes: U-Net, es un modelo de segmentación basado en una arquitectura de codificador-decodificador, que utiliza conexiones de salto entre las capas para conservar la información espacial y mejorar la precisión de la segmentación. U-Net ha demostrado ser muy efectivo para la segmentación de imágenes médicas, especialmente para el hígado, el páncreas y los riñones [24]

SegNet, es otro modelo de segmentación basado en una arquitectura de codificador-decodificador, que se diferencia de U-Net en que utiliza un método de agrupación máxima con índices, que reduce el número de parámetros y mejora la eficiencia del modelo. SegNet ha sido aplicado con éxito para la segmentación de imágenes de CT del abdomen, logrando resultados comparables o superiores a otros métodos [23].

Mask R-CNN, es un modelo de segmentación que combina dos tareas: la detección de objetos y la segmentación de instancias. Utiliza una red residual con una capa de pirámide de características, y dos ramas paralelas, una para la detección de objetos y otra para la segmentación de instancias. Mask R-CNN ha mostrado un alto rendimiento para la segmentación de imágenes de CT del abdomen, especialmente para el hígado, el páncreas y los riñones [26];[2].

Estos modelos se han seleccionado por ser algunos de los más populares y efectivos para la segmentación de imágenes de CT del abdomen, según la literatura científica disponible. Estos modelos han demostrado tener un alto rendimiento y una buena precisión para segmentar diferentes estructuras y órganos abdominales, como el hígado, el páncreas y los riñones [26]; [2]. Además, estos modelos tienen la ventaja de ser flexibles y adaptables a diferentes condiciones y escenarios, como la variabilidad anatómica, la calidad de las imágenes, la presencia de artefactos o patologías [24];[2].

Objetivo 1

5.1. Preprocesamiento de la base de datos

El preprocesamiento de datos constituye una etapa crucial en el proceso de análisis y modelado de datos, especialmente en el ámbito de la segmentación de imágenes médicas. En el contexto de las imágenes de tomografía computarizada (CT) del área abdominal, una adecuada preparación de los datos es indispensable para garantizar la precisión y la eficacia de los modelos de segmentación empleados. Esta fase inicial involucra una serie de procedimientos meticulosos que buscan mejorar la calidad y la estructura de los datos, facilitando así su posterior análisis. Entre estos procedimientos se incluyen la corrección de artefactos, la normalización, la estandarización, la segmentación de regiones de interés y la ampliación de datos.

La corrección de artefactos es fundamental para eliminar o minimizar los errores inherentes o las anomalías presentes en las imágenes, como el ruido o las distorsiones, que pueden obstaculizar la interpretación correcta de la información contenida. Por otro lado, la normalización y la estandarización son procedimientos esenciales que permiten homogenizar las escalas de intensidad y las unidades de medida, proporcionando una base consistente para el análisis y la comparación entre diferentes conjuntos de datos o imágenes.

La segmentación de las regiones de interés, por su parte, se enfoca en aislar y destacar las áreas específicas dentro de las imágenes que son relevantes para el estudio, permitiendo una concentración efectiva de los recursos computacionales y analíticos en los elementos cruciales. Además, la ampliación de datos mediante técnicas como la rotación, el escalamiento o la transformación, contribuye a expandir y diversificar el conjunto de datos disponible, enriqueciendo así el entrenamiento y la validación de los modelos, y promoviendo una mejor generalización de estos.

A través del preprocesamiento, se busca no solo optimizar la calidad de los datos, sino también estructurarlos de manera que faciliten la extracción de características y la aplicación de técnicas avanzadas de aprendizaje automático y *deep learning*. La correcta ejecución de esta etapa preparatoria es vital para maximizar el rendimiento y la precisión de los modelos de segmentación en la tarea de identificación y delimitación de órganos y estructuras en las imágenes de tomografía computarizada del área abdominal, lo cual es esencial para un análisis diagnóstico confiable y una interpretación clínica precisa. En resumen, el preprocesamiento de datos se presenta como un pilar fundamental en la ruta hacia la obtención de insights valiosos y aplicaciones prácticas efectivas en el campo de la imagenología médica y la inteligencia artificial.

5.1.1. Descripción de la base de datos

El conjunto de datos utilizado en este trabajo de grado se adquirió a través de la plataforma *Grand Challenge*, una iniciativa dedicada al fomento del desarrollo integral de soluciones de aprendizaje automático aplicadas a imágenes biomédicas. Específicamente, los datos provienen del desafío *MICCAI 2021 FLARE Challenge: Fast and Low GPU memory Abdominal oRgan sEgmentation*, organizado en el año 2021. La documentación proporcionada en la plataforma detalla que el conjunto de datos es amplio y diversificado, compuesto por imágenes proporcionadas por 11 centros médicos distintos [17].

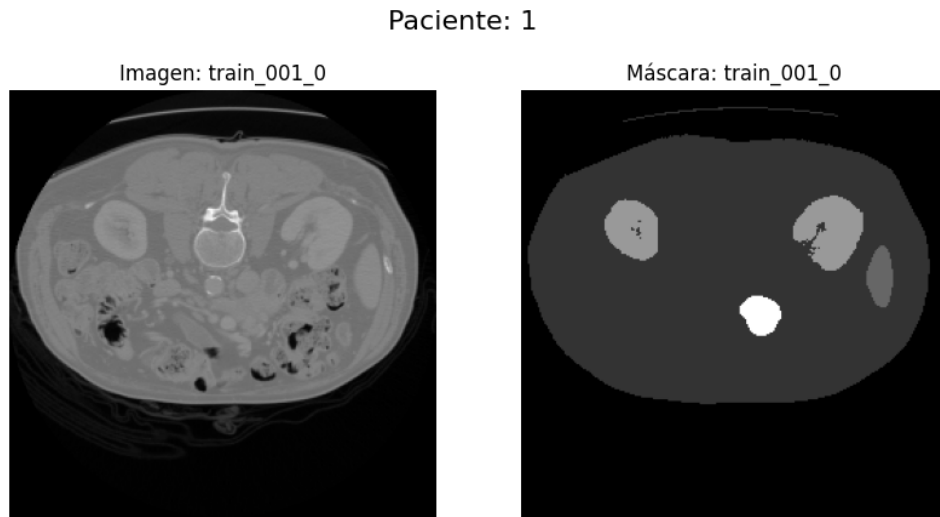


Figura 5.1: Imágen y mascara original del paciente cero

La figura adjunta ilustra la imagen y la mascara correspondiente al paciente No 1 del conjunto de entrenamiento (*train*) y su corte (*slice*) inicial. Para tener una comprensión más profunda acerca del conjunto de datos, se analizó detenidamente la información contenida en sus dos subconjuntos, denominados *train* y *test*. El conjunto *train* se organiza en dos directorios: imágenes y máscaras, albergando cada uno 2725 archivos en formato numpy (*.*np*y). Este conjunto incorpora datos de 48 pacientes, cada uno con un número variado de cortes (*slices*). En promedio, se registraron alrededor de 57 *slices* por paciente, con una desviación estándar de 15.32.

Por otro lado, el conjunto *test* también se estructura en dos directorios: imágenes y máscaras, conteniendo cada uno 513 archivos en formato *.*np*y. Este subconjunto comprende datos de 10 pacientes, cada uno con un número distintivo de *slices*. En promedio, se observaron alrededor de 51 *slices* por paciente, con una desviación estándar de 17.12.

En términos generales como información y metadata, cada imagen posee una resolución de 256 (ancho) x 256 (alto) píxeles. Cada archivo **.npy* se halla en formato Float32, mientras que el tipo de dato de cada archivo es *uint8*, con un orden en memoria descrito como fila principal orden Fortran: false. El conjunto de datos, en su totalidad, tiene 3,238 imágenes, de las cuales 2,725 corresponden al conjunto de entrenamiento, representando el 84 % del total, y 513 al conjunto de prueba, equivalente al 16 % del total. A nivel de distribución de pacientes, el 82.76 % se asignó al conjunto de entrenamiento y el 17.24 % al conjunto de prueba.

5.2. Propuestas de pre-procesamiento para las imágenes.

La tomografía computarizada (TC) es una técnica de imagen médica fundamental que proporciona vistas detalladas de estructuras internas del cuerpo humano. Estas imágenes son cruciales para el diagnóstico y seguimiento de diversas condiciones médicas. Sin embargo, para aprovechar al máximo la información que estas imágenes proporcionan, es esencial realizar un preprocesamiento adecuado y una eliminación eficiente del ruido.

En este trabajo de grado, se aplican varias técnicas de preprocesamiento de imágenes para preparar los datos de tomografía computarizada (CT) del área abdominal antes de utilizarlos en los modelos de segmentación. Estas técnicas son esenciales para garantizar la consistencia y calidad de los datos de entrada para los algoritmos de segmentación. A continuación, se describen detalladamente las técnicas de preprocesamiento utilizadas:

5.2.1. Preprocesamiento y Eliminación de Ruido en Imágenes de TC

El propósito del preprocesamiento en imágenes de TC es múltiple. En primer lugar, busca mejorar la calidad de la imagen, que puede verse comprometida por diversos factores como el ruido inherente al proceso de adquisición de imágenes o las interferencias causadas por elementos externos, como la camilla en la que se acuesta el paciente. Además, el preprocesamiento tiene como objetivo preparar las imágenes para su posterior análisis y procesamiento, en particular para las técnicas de segmentación avanzadas utilizando modelos de aprendizaje profundo por ejemplo U-Net, SegNet y Mask R-CNN. Estos modelos requieren imágenes de alta calidad y uniformidad para funcionar con precisión, lo que hace que el preprocesamiento sea un paso indispensable en el flujo de trabajo de análisis de imágenes médicas.

Dentro de esta subsección, abordaremos específicamente la eliminación de la camilla en las imágenes de TC y residuos adicionales como cobijas, mantas, etc. La presencia de estos elementos puede introducir artefactos que distorsionan los resultados de la segmentación, llevando a interpretaciones erróneas o inexactas. Para resolver este problema se planteó un código en *Python* para la modificación total del *dataset* agregando una mascara que analiza el 50 % de las filas de la matriz (parte superior de todas las mascaras).

La Función `clean_noise_from_mask()`: Esta función tiene como objetivo limpiar ruido de una matriz de máscara. La matriz de la máscara es una matriz bidimensional que donde la camilla tiene el valor 1 y el fondo (*background*) tiene el valor de 0, esta función recibe a `mask_data` que es la matriz de la máscara representada como un array de NumPy y el porcentaje de filas iniciales a considera con `m_percentage` este es un valor flotante que indica el porcentaje de la máscara que se procesará para la eliminación de ruido.

La función itera sobre las filas de la matriz de máscara desde un índice calculado en base al porcentaje proporcionado hacia atrás. Luego, llama a la función `process_columns()` para procesar las columnas de cada fila en esa sección de la máscara. Esta función examina las sumas de columnas y reemplaza los valores en columnas específicas si se cumplen ciertas condiciones (columnas con sumas mayores a 0 y menores o iguales a 3, considerando valores iguales a 1).

La función `verify_noise_removal()`, es una duplicación `clean_noise_from_mask()`. Ambas funciones tienen la misma estructura y utilizan la misma lógica para procesar la matriz de máscara. Sin embargo, este proceso realiza una segunda validación para evitar algún ruido adicional.

La función `process_columns()` se encarga de procesar las columnas de la matriz de máscara hasta una fila específica (`row_limit`). Para cada columna, calcula la suma de los elementos que son iguales a 1 en la sección de la máscara hasta la fila `row_limit`. Luego, identifica las columnas que cumplen ciertas condiciones (suma mayor a 0 y menor o igual a 3) y reemplaza los valores en esas columnas con 0 si originalmente eran 1.

Una vez resuelto el ruido del 50 % superior de la mascara usamos la función `apply_mask_to_image()` que utiliza todo el fondo de la mascara igual a cero y lo aplica sobre la imagen con un valor de -1042, de esta forma ya no se tiene ningún ruido en la imagen relacionado con camillas, cobijas, mantas, etc. (véase Imagen 5.2). La elección entre guardar imágenes o máscaras se determina por el parámetro ‘`select_item`’. El código de eliminación de la camilla se encuentra disponible en el repositorio de Github con el nombre 02 - `denoise.ipynb` en la carpeta *Notebooks*. A continuación se visualiza la imagen del paciente cero sin camilla.

5.2.2. Ampliación de Datos (*Data Augmentation*)

En el ámbito del aprendizaje profundo (*Deep Learning*), la técnica conocida como *data augmentation* se convierte en un componente esencial para enriquecer tanto la cantidad como la variabilidad de los conjuntos de datos utilizados en el entrenamiento de modelos. Este aspecto es particularmente crítico en campos donde la adquisición de datos se presenta como una tarea costosa o restringida, como sucede comúnmente en el ámbito de las imágenes médicas. En nuestra investigación, la implementación de *Data Augmentation* se lleva a cabo haciendo uso de la librería *imgaug* de Python, la cual proporciona una amplia gama de técnicas para la mejora de imágenes.

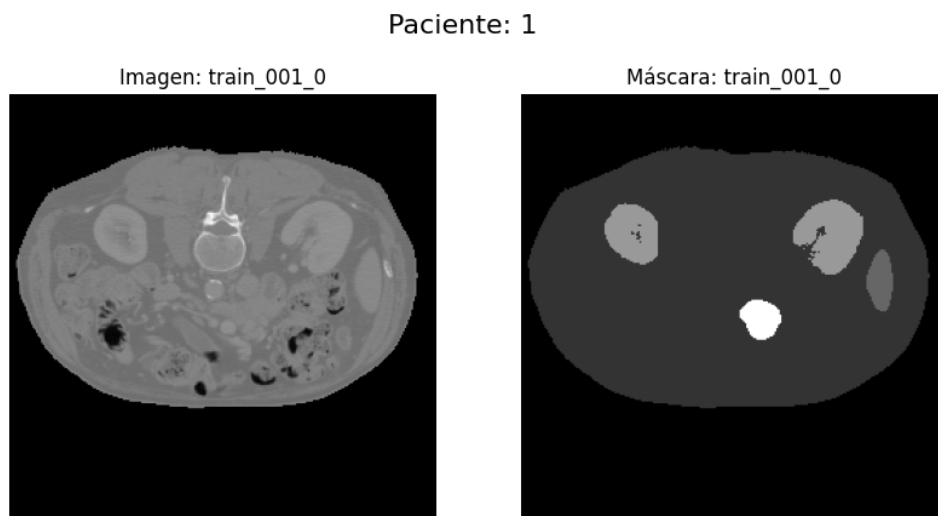


Figura 5.2: Imagen y mascara sin camilla del paciente cero

Dentro del contexto específico de nuestro proyecto, nos centramos principalmente en las transformaciones geométricas proporcionadas por *imgaug*, tales como rotaciones, traslaciones, escalado y volteo, con el propósito de generar variaciones en las imágenes de tomografía computacional (CT).

La ampliación de datos se conceptualiza como un proceso dinámico que busca crear nuevas instancias de entrenamiento a través de la aplicación de transformaciones específicas a las imágenes existentes. Estas transformaciones, aunque no alteran la información de la clase, introducen variabilidad, desempeñando un papel crucial al mejorar la capacidad del modelo para generalizar. Este enriquecimiento artificial de la diversidad de datos contribuye significativamente a la mejora de la generalización de los modelos resultantes.

Es importante destacar que el proceso de *data augmentation* debe ser ejecutado con cautela. Es necesario encontrar un equilibrio óptimo en términos de cantidad y tipo de augmentación para evitar el sobreajuste del modelo. Este cuidado garantiza que las imágenes modificadas mantengan su relevancia y realismo en relación con la problemática médica que se pretende abordar.

El código implementado para la *data augmentation* se encuentra disponible para su revisión en el repositorio de GitHub, cuyo enlace se proporciona en el siguiente *link*: [GitHub](#). Para proporcionar una visión tangible de los resultados obtenidos a través de esta técnica, se presenta a continuación la visualización de la imagen correspondiente al paciente inicial en nuestro conjunto de datos.

Las técnicas de *Data Augmentation* que se pueden utilizar para la segmentación de imágenes de tomografía computacional son:

1. Rotación:

- Las imágenes de tomografía pueden rotarse en diferentes ángulos para simular variaciones en la posición del paciente durante la exploración.
- La rotación puede ser en 2D o 3D, dependiendo de la naturaleza del volumen de datos.

2. Zoom y Escala:

- La aplicación de zoom y escala simula variaciones en la resolución espacial de las imágenes.
- Puede ayudar al modelo a adaptarse a diferentes niveles de detalle presentes en diferentes estudios de tomografía.

3. Deformación Elástica:

- Introduce deformaciones locales en la imagen para simular posibles distorsiones anatómicas o artefactos en las imágenes de tomografía.

4. Flip Horizontal/Vertical:

- Reflejar horizontal o verticalmente las imágenes ayuda a enseñar al modelo sobre simetrías y mejora su capacidad para reconocer estructuras desde diferentes perspectivas.

5. Ajustes de Contraste y Brillo:

- Variaciones en el contraste y el brillo replican las diferencias en la iluminación durante la adquisición de imágenes.

6. Corte Aleatorio:

- Realizar cortes aleatorios en las imágenes en 3D para imitar variaciones en la orientación y posición del objeto escaneado.

7. Adición de Ruido:

- Introducir ruido artificial en las imágenes ayuda al modelo a ser más robusto frente a posibles imperfecciones en los datos reales.

Como se observa en la imagen, en la parte izquierda tenemos la imagen original sin ruido y en la parte derecha tenemos la imagen desplazada y escalada. Este código en Python, basado en la librería *imgaug*, ofrece una solución efectiva para aplicar transformaciones geométricas a las imágenes de tomografía, generando así variaciones artificiales que mejoran la capacidad de generalización del modelo.

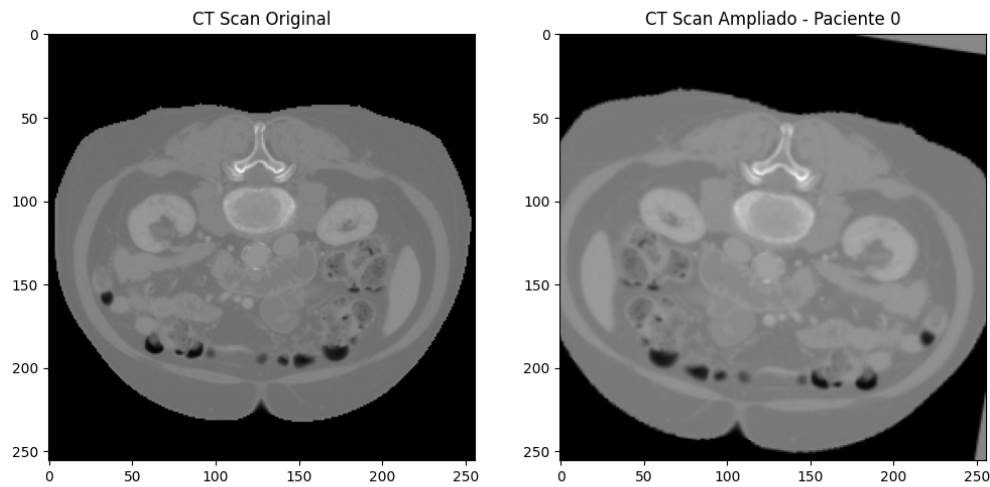


Figura 5.3: Imagen del paciente cero aumentada

La función `apply_data_augmentation()` toma como entrada una lista de rutas de archivos que contienen imágenes de tomografía en formato numpy (`.npy`). A continuación, se proporciona una explicación detallada del código:

Listing 5.1: Implementación de *Data Augmentation* con *imgaug*

```

1 def apply_data_augmentation(images_list, output_folder="augmented_images"):
2     os.makedirs(output_folder, exist_ok=True)
3
4     for i, image_path in enumerate(images_list):
5         ct_scan = np.load(image_path)
6
7         augmenter = iaa.Sequential([
8             iaa.Affine(rotate=(-10, 10)), # Rotación de la imagen
9             iaa.Fliplr(0.5),             # Volteos horizontales
10            iaa.Crop(percent=(0, 0.1))    # Recortar imágenes
11        ])
12
13        # Aplicar la ampliación de datos a la imagen de tomografía actual
14        augmented_ct_scan = augmenter(images=ct_scan)
15
16        output_path = os.path.join(output_folder, ...
17                                   f'train_{str(i).zfill(3)}_augmented.npy')
18        np.save(output_path, augmented_ct_scan)
19
20        # Mostrar la imagen original y la imagen ampliada

```

```
20     fig, ax = plt.subplots(1, 2, figsize=(12, 6))
21     ax[0].imshow(ct_scan, cmap='gray')
22     ax[0].set_title(f'CT Scan Original - Paciente {i}')
23     ax[1].imshow(augmented_ct_scan, cmap='gray')
24     ax[1].set_title(f'CT Scan Ampliado - Paciente {i}')
25     plt.savefig(f'train_{str(i).zfill(3)}_augmented.png')
26     plt.show()
```

La función comienza por crear un directorio de salida, denominado `augmented_images`, donde se almacenarán las imágenes ampliadas. Luego, itera sobre la lista de imágenes originales de tomografía, aplicando una secuencia de transformaciones definida por el objeto `augmenter`. Estas transformaciones incluyen rotaciones en el rango de -10 a 10 grados, volteos horizontales con una probabilidad del 50 %, y recortes de hasta el 10 % de la imagen.

Posteriormente, la imagen resultante de la ampliación se guarda en formato `.npy` en el directorio de salida con un nombre de archivo específico. Además, se genera una visualización comparativa entre la imagen original y la imagen ampliada, mostrándolas lado a lado. Estas visualizaciones se guardan como archivos de imagen en formato `.png`.

Esta implementación facilita la generación eficiente y automatizada de un conjunto de datos ampliado, enriqueciendo así la diversidad de muestras disponibles para el entrenamiento de modelos de segmentación en tomografía computacional. Este enfoque se integra de manera efectiva en la sección de *Data Augmentation*, contribuyendo a mejorar la robustez y generalización del modelo resultante.

Objetivo 2

6.1. Implementación de modelos de Deep Learning

la implementación de modelos de Deep Learning para la segmentación de imágenes de TC abdominal aporta precisión, eficiencia y consistencia a la práctica médica, y tiene un impacto significativo en el diagnóstico, el tratamiento y la investigación médica en el campo de la radiología y la medicina moderna, por tal motivo se aplicaran una serie de pasos para la implementación de cada modelo de aprendizaje profundo

6.1.1. Modelo U-Net

Como primera instancia importamos las bibliotecas necesarias de PyTorch para trabajar con redes neuronales y transformaciones de imágenes.

Posteriormente se implementó una Clase DoubleConv el cual define un bloque básico utilizado en la arquitectura U-Net, este bloque consiste en dos capas convolucionales, cada una seguida de una capa de normalización por lotes (Batch Normalization) y una función de activación ReLU. Las capas convolucionales tienen un tamaño de kernel de 3x3 y un relleno de 1 para mantener el tamaño de la entrada, no se usa un sesgo (bias) en estas capas convolucionales. Esto se traduce de la siguiente manera: Imagina que quieres construir una máquina que pueda entender imágenes y decirte qué hay en ellas, esta máquina necesita aprender cosas complicadas sobre las imágenes, como bordes, formas y colores. La máquina toma la imagen y la pasa a través de dos capas especiales llamadas convoluciones, estas capas ayudan a la máquina a comprender mejor los detalles de la imagen. Después de cada convolución, la máquina ajusta la imagen usando una fórmula matemática llamada "normalización por lotes", esto ayuda a que la máquina aprenda de manera más eficiente. Finalmente, la máquina aplica una especie de filtro llamado ReLU para resaltar características importantes en la imagen.

La Clase UNET init construye la arquitectura U-Net con dos partes principales: la parte descendente (down) y la parte ascendente (up), la parte descendente consiste en una serie de bloques DoubleConv seguidos de max-pooling, el número de características (features) se duplica a medida que se desciende en la red, mientras que la parte ascendente consiste en convoluciones transpuestas seguidas de bloques DoubleConv, estas convoluciones transpuestas permiten aumentar el tamaño de las características para reconstruir la imagen de salida. Se utiliza una lista de características

(features) que especifica el número de características en cada capa. Es decir que la máquina que construimos anteriormente es bastante buena para entender las imágenes, pero necesita ser más inteligente para hacer cosas más complejas. Así que creamos una máquina aún más grande que se parece a un U invertido (de ahí el nombre UNET), esta máquina puede hacer dos cosas importantes: mirar detalles pequeños y juntar toda la información para entender la imagen completa. La parte superior de la máquina (la parte "down") mira los detalles pequeños de la imagen y los entiende muy bien usando los bloques DoubleConv que hicimos antes. La parte inferior de la máquina (la parte "up") usa esa información para juntar todos los detalles y reconstruir la imagen completa. Usa un truco llamado "conexiones de salto" para recordar detalles importantes de la imagen original. Al final, la máquina nos da una versión mejorada de la imagen que contiene toda la información que aprendió.

la clase UNET forward define el paso hacia adelante de la red U-Net. En el paso hacia adelante, las características se propagan hacia abajo a través de los bloques DoubleConv y el muestreo máximo, luego pasan por un cuello de botella (bottleneck) y finalmente se propagan hacia arriba a través de las convoluciones transpuestas y las conexiones de salto (skip connections). La función forward de la clase UNET es como una receta para que nuestra máquina haga su trabajo. Primero, la máquina mira los detalles importantes en la imagen, esto significa que busca cosas como bordes y formas que puedan ser útiles para entender la imagen, utiliza los bloques DoubleConv, que son como herramientas especiales para ver estos detalles. Después de mirar los detalles, la máquina toma una foto de ellos y los guarda para más tarde, Esta foto se llama "conexiones de salto" es como una nota recordatoria. Ahora, la máquina quiere ver la imagen de manera más clara, para hacer esto, reduce la imagen, como si estuviera viéndola desde más arriba. esto se hace mediante el muestreo máximo, que es como un filtro que ayuda a la máquina a ver mejor los detalles pequeños. Después de ver los detalles, la máquina quiere entender la imagen completa, para hacer esto, pasa la imagen por un lugar llamado "cuello de botella" (bottleneck), esto ayuda a que la máquina capture la esencia de la imagen en su conjunto. Ahora que la máquina ha visto los detalles y comprende la imagen en su conjunto, quiere mejorarla, así que amplía la imagen, como si estuviera viendo una versión más grande de ella, utiliza convoluciones transpuestas, que son como una lupa que hace que la imagen sea más grande. Luego, la máquina toma las notas que tomó antes (las conexiones de salto) y las pone junto con la imagen ampliada, esto ayuda a que la imagen mejorada tenga tanto los detalles finos como la visión general de la imagen. Después de hacer todos estos pasos, la máquina nos da la imagen mejorada como resultado final, es como si hubiera tomado una foto borrosa y la hubiera hecho nítida y clara.

La función de prueba (test) crea una entrada de ejemplo, instancia el modelo U-Net y realiza una predicción con la entrada, luego, verifica que las dimensiones de la salida coincidan con las dimensiones de la entrada. Para asegurarnos de que nuestra máquina funcione bien, le damos una imagen simple para entender. La dejamos trabajar y luego comparamos lo que nos da con lo que le dimos. Si lo que nos devuelve se parece mucho a lo que le dimos, significa que nuestra máquina funciona bien.

6.1.2. Modelo SegNet

Explicar el código de implementación

6.1.3. Modelo Mask R-CNN

Explicar el código de implementación

6.1.4. División de datos y validación cruzada

Dividir el conjunto de datos en conjuntos de entrenamiento, validación y prueba, aplicar técnicas de validación cruzada para evaluar la robustez de los modelos y su capacidad de generalización en diferentes escenarios clínicos.

6.1.5. Entrenamiento de modelos

Realizar el entrenamiento de cada modelo utilizando las imágenes de entrenamiento, ajustando los pesos y parámetros a través de técnicas de aprendizaje profundo, como la retropropagación, evaluar el progreso del entrenamiento y establecer criterios de parada para evitar el sobreajuste.

Objetivo 3

7.1. Validación del modelo obtenido

Esta validación tiene como objetivo determinar cuán eficaz y confiable es el modelo en la tarea de segmentación, y se llevará a cabo de la siguiente manera:

7.1.1. Evaluación de rendimiento

Evaluar el rendimiento de los modelos en términos de precisión, sensibilidad, especificidad, coeficiente de Dice y otras métricas relevantes. Comparar los resultados obtenidos por los tres modelos para determinar su capacidad para segmentar estructuras anatómicas y patologías en las imágenes de TC abdominal.

7.1.2. Iteración y ajuste

Si la validación inicial revela deficiencias en el rendimiento del modelo, se pueden realizar ajustes en los hiperparámetros o en la arquitectura del modelo y se puede realizar una nueva validación. Este proceso de iteración se realizará hasta lograr un modelo con un rendimiento aceptable en términos de precisión y *recall*.

7.1.3. Interpretación de métricas

La interpretación de las métricas de evaluación, en particular la precisión y el *recall*, permitirá comprender el equilibrio entre la capacidad del modelo para identificar las regiones de interés (precisión) y su capacidad para no pasar por alto regiones de interés importantes (*recall*). Un equilibrio óptimo es esencial para la toma de decisiones clínicas confiables.

7.1.4. Análisis de resultados y comparación

Analizar en detalle los resultados de la segmentación obtenidos por cada modelo y comparar su eficiencia en términos de precisión y tiempo de procesamiento, identificar las fortalezas y debilidades de cada enfoque en la segmentación de imágenes de TC abdominal.

7.1.5. Generación de conclusiones y recomendaciones

Concluir sobre cuál de los modelos (U-Net, SegNet, Mask R-CNN) es más adecuado para la segmentación precisa y confiable de imágenes de TC abdominal en función de los resultados obtenidos y proporcionar recomendaciones para la implementación en contextos clínicos y futuras investigaciones.

Conclusiones y trabajos futuros

8.1. Conclusiones

8.2. Trabajos futuros

CAPÍTULO 9

Anexos

Bibliografía

- [1] José Denes Lima Araújo, Luana Batista da Cruz, Jonnison Lima Ferreira, Otilio Paulo da Silva Neto, Aristófanés Corrêa Silva, Anselmo Cardoso de Paiva, and Marcelo Gattass. An automatic method for segmentation of liver lesions in computed tomography images using deep neural networks. *Expert Systems with Applications*, 180:115064, 2021.
- [2] Jinzheng Cai, Le Lu, Zizhao Zhang, Fuyong Xing, Lin Yang, and Qian Yin. Pancreas segmentation in mri using graph-based decision fusion on convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 442–450. Springer, 2016.
- [3] Kuan-bing Chen, Ying Xuan, Ai-jun Lin, and Shao-hua Guo. Lung computed tomography image segmentation based on u-net network fused with dilated convolution. *Computer Methods and Programs in Biomedicine*, 207:106170, 2021.
- [4] Albert Comelli, Claudia Coronello, Navdeep Dahiya, Viviana Benfante, Stefano Palmucci, Antonio Basile, Carlo Vancheri, Giorgio Russo, Anthony Yezzi, and Alessandro Stefano. Lung segmentation on high-resolution computerized tomography images using deep learning: a preliminary step for radiomics studies. *Journal of Imaging*, 6(11):125, 2020.
- [5] Luana Batista da Cruz, Jose Denes Lima Araujo, Jonnison Lima Ferreira, Joao Otavio Bandeira Diniz, Aristofanes Correa Silva, João Dallyson Sousa de Almeida, Anselmo Cardoso de Paiva, and Marcelo Gattass. Kidney segmentation from computed tomography images using deep neural network. *Computers in Biology and Medicine*, 123:103906, 2020.
- [6] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [7] Intisar Rizwan I Haque and Jeremiah Neubert. Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, 18:100297, 2020.
- [8] Sheng He, Rina Bao, Jingpeng Li, P Ellen Grant, and Yangming Ou. Accuracy of segment-anything model (sam) in medical image segmentation tasks. *arXiv preprint arXiv:2304.09324*, 2023.
- [9] Hector Henriquez, Alejandro Freundlich, Dra, and Glasinovic Andrea. Liver segmentation model on computed tomography images using fully convolutional neural network. 10 2020.
- [10] Godfrey N Hounsfield. Computerized transverse axial scanning (tomography): Part 1. description of system. *The British journal of radiology*, 46(552):1016–1022, 1973.

- [11] Chuanfei Hu and Xinde Li. When sam meets medical images: An investigation of segment anything model (sam) on multi-phase liver tumor segmentation. *arXiv preprint arXiv:2304.08506*, 2023.
- [12] Yuhao Huang, Xin Yang, Lian Liu, Han Zhou, Ao Chang, Xinrui Zhou, Rusi Chen, Junxuan Yu, Jiongquan Chen, Chaoyu Chen, et al. Segment anything model for medical images? *arXiv preprint arXiv:2304.14660*, 2023.
- [13] Ali F Khalifa and Eman Badr. Deep learning for image segmentation: A focus on medical imaging. *CMC-COMPUTERS MATERIALS & CONTINUA*, 75(1):1995–2024, 2023.
- [14] Chae Eun Lee, Minyoung Chung, and Yeong-Gil Shin. Voxel-level siamese representation learning for abdominal multi-organ segmentation. *Computer methods and programs in biomedicine*, 213:106547, 2022.
- [15] Jun Ma and Bo Wang. Segment anything in medical images. *arXiv preprint arXiv:2304.12306*, 2023.
- [16] Jun Ma, Yao Zhang, Song Gu, Xingle An, Zhihe Wang, Cheng Ge, Congcong Wang, Fan Zhang, Yu Wang, Yinan Xu, et al. Fast and low-gpu-memory abdomen ct organ segmentation: the flare challenge. *Medical Image Analysis*, 82:102616, 2022.
- [17] Jun Ma, Yao Zhang, Song Gu, Xingle An, Zhihe Wang, Cheng Ge, Congcong Wang, Fan Zhang, Yu Wang, Yinan Xu, Shuiping Gou, Franz Thaler, Christian Payer, Darko tern, Edward G.A. Henderson, Dónal M. McSweeney, Andrew Green, Price Jackson, Lachlan McIntosh, Quoc-Cuong Nguyen, Abdul Qayyum, Pierre-Henri Conze, Ziyang Huang, Ziqi Zhou, Deng-Ping Fan, Huan Xiong, Guoqiang Dong, Qiongjie Zhu, Jian He, and Xiaoping Yang. Fast and low-gpu-memory abdomen ct organ segmentation: The flare challenge. *Medical Image Analysis*, 82:102616, 2022.
- [18] Arturo Orellana García and Luis Manuel García Portal. Técnicas de segmentación y procesamiento para la detección de carcinomas renales en imágenes de tomografía abdominal. *Revista Cubana de Informática Médica*, 12(2), 2020.
- [19] David MW Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*, 2020.
- [20] Pavlo Radiuk. Applying 3d u-net architecture to the task of multi-organ segmentation in computed tomography. *Applied Computer Systems*, 25(1):43–50, 2020.
- [21] Peilun Shi, Jianing Qiu, Sai Mu Dalike Abaxi, Hao Wei, Frank P-W Lo, and Wu Yuan. Generalist vision foundation models for medical imaging: A case study of segment anything model on zero-shot medical segmentation. *Diagnostics*, 13(11):1947, 2023.
- [22] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15(1):1–28, 2015.

-
- [23] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and R Summers. Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *IEEE CVPR*, volume 7, page 46. sn, 2017.
 - [24] Melanie Yusta Gómez, Marlen Pérez Díaz, Rubén Orozco Morales, and Xiomara Plasencia Hernández. Segmentación del hígado en imágenes de tomografía computarizada. *MediSur*, 20(2):257–271, 2022.
 - [25] Fan Zhang, Yu Wang, and Hua Yang. Efficient context-aware network for abdominal multi-organ segmentation. *arXiv preprint arXiv:2109.10601*, 2021.
 - [26] Zhuotun Zhu, Yingda Xia, Wei Shen, Elliot Fishman, and Alan Yuille. A 3d coarse-to-fine framework for volumetric medical image segmentation. In *2018 International conference on 3D vision (3DV)*, pages 682–690. IEEE, 2018.