

FACULTAD DE MATEMÁTICA Y
COMPUTACIÓN UNIVERSIDAD DE LA
HABANA



TESIS DE DIPLOMA DE LA ESPECIALIDAD CIENCIA DE LA
COMPUTACIÓN

**Segmentación de Úlceras de Pie Diabético
(UPD) en secuencias de imágenes RGB
mediante el Segment Anything Model
(SAM).**

Autor: Abdel Fregel Hernández
Tutor: Dr. José Alejandro Mesejo Chiong

La Habana
11 de enero de 2025

1. Agradecimientos

2. Resumen

3. Índice

Índice

1. Agradecimientos	I
2. Resumen	II
3. Índice	III
4. Introducción	1
4.1. Objetivos	4
4.2. Estructura de la tesis	4
5. Capítulo 1	5
5.1. Segmentación de Úlceras del Pie Diabético: Definición	5
5.2. Métodos existentes	6
5.3. Datasets Existentes	9
6. Capítulo 2	10
6.1. Modelos de Base(Foundation Models)	10
6.2. Segment Anything Model(SAM)	10
6.3. Etapas del Preentrenamiento	10
6.4. Adaptando SAM para imágenes médicas.	15
7. Conclusiones	18
8. Recomendaciones	19
9. Bibliografía	20
10. Anexos	24

4. Introducción

La diabetes(diabetes mellitus), es una enfermedad crónica que afecta la forma en la que el cuerpo utiliza la glucosa, una fuente clave de energía. De acuerdo a la *Federación Internacional de Diabetes*, en el año 2021 se reportaron 6.7 millones de muertes a causa de esta enfermedad [11]. En nuestro país, según el *Anuario Estadístico de Salud 2022* [25], la prevalencia es de 66,50 % de enfermos por cada 1000 habitantes.

Cerca del 86 % de personas que padecen de diabetes sufren de úlcera de pie diabético ¹(UPD), y corren el riesgo de amputación. La cicatrización de estas úlceras puede tardar semanas, meses e incluso años, deteriorando la calidad de vida de los pacientes. Actualmente, los médicos cubanos especializados en el tema no cuentan con una herramienta cuantitativa efectiva que valore la severidad y el proceso de curación de las UPD. La medición regular de las úlceras es crucial para evaluar la efectividad del tratamiento y realizar ajustes cuando sea necesario. Un seguimiento adecuado puede prevenir la progresión de la úlcera y reducir el riesgo de amputaciones.

Existen métodos de segmentación de imágenes digitales en el ambito de la medicina y medición de heridas que han conseguido grandes avances en esta área. La calidad de estas es esencial para varios análisis de heridas, como por ejemplo la clasificación de tejidos, Reconstrucción 3D, y la evaluación de la cicatrización[13]. Se pueden clasificar en dos tipos: aquellos que requieren contacto y los que no. Los métodos de contacto son invasivos y presentan un alto margen de error; por ello, este trabajo se centra en los métodos no invasivos.

Varios estudios han abordado esta problemática. Por ejemplo, en Filko et al. [13] hacen uso de un sofisticado brazo robótico de 7 grados de libertad(DoF), equipado con una cámara RGB-D ² y un escáner 3D de alta precisión, para la segmentación y medición de heridas. Este artículo aporta un nuevo algoritmo de segmentación que utiliza una combinación de procedi-

¹Hace referencia a una complicación grave de la diabetes que se manifiesta como una herida o llaga abierta en el pie

²RGB-D, se refiere a una camara capaz de captar imágenes a color(RGB, formato Red(rojo),Green(verde),Blue(azúl) y un sensor de profundidad (D(depth),por su sigla en inglés)

mientos 2D y 3D para segmentar correctamente un modelo de herida en 3D. La segmentación se realiza a partir de múltiples fotografías 2D por herida, impulsada por una red neuronal profunda en forma del clasificador MobileNetV2 [13]. Este clasificador se combina óptimamente con un único modelo 3D y la inicialización del contorno de la herida. Este contorno inicial se optimiza y ajusta mediante un modelo de contorno activo [13], que envuelve estrechamente la superficie real de la herida utilizando la curvatura de la superficie para alcanzar su objetivo.

En otro estudio realizado por Filko et al. [12] se exploró la medición y reconstrucción de heridas crónicas de lenta curación, utilizando cámara RGB-D. Con la llegada de cámaras RGB-D económicas, la comunidad de visión por computadora ha ganado una forma más accesible para innovar y crear aplicaciones en diversos campos, incluyendo la medicina, [12]. Estas cámaras, que combinan información de color y profundidad, permiten un análisis más detallado y preciso de imágenes, facilitando el desarrollo de tecnologías que mejoran diagnósticos y tratamientos. El sistema desarrollado en dicho artículo detecta automáticamente heridas analizando bloques de imagen según la similitud del histograma de color utilizando un enfoque de vecinos más cercanos(KNN,por sus siglas en inglés). Este método permite identificar características específicas de las heridas en imágenes, facilitando su segmentación y medición.

A partir de lo expuesto anteriormente, se hace necesario el desarrollo de una herramienta capaz de realizar la segmentación y la medición de estas heridas lo más preciso posible, lo cual es investigado en este trabajo. En cuanto a la tarea de segmentar las imágenes, se ha utilizado herramienta desarrollada por Meta AI llamada *Segment Anything Model*(SAM)[23] que permite identificar y segmentar objetos en imágenes de manera eficiente. Esta herramienta ha ganado popularidad en cuanto a las tareas de segmentación y en el presente trabajo será usada para hacer una segmentación de las imágenes de UPD, las cuales serán tomadas con una cámara RGB-D para asegurar su calidad.

La cámara que se utiliza es una cámara Intel Realsense D435i RGB-D, la cual utiliza luz estructurada infrarroja y métodos binoculares para obtener información de profundidad, que es la tecnología más común y madura utilizada en las cámaras RGB-D de grado consumidor actuales. La cámara

de profundidad puede emitir un flujo de video de hasta 720p a 90 FPS, y la cámara RGB puede emitir un flujo de video de hasta 1080p a 30 FPS [35]. Luego de la segmentación se realiza una clasificación de los tejidos para poder realizar una evaluación de la cicatrización.

La importancia de los tejidos en el proceso de cicatrización radica en que cada tipo de tejido desempeña un papel crucial en la reparación y regeneración de la herida. Durante la cicatrización, se forman diferentes tipos de tejidos, siendo el tejido de granulación uno de los más fundamentales [27]. Este tejido no solo es esencial para el cierre de la herida, sino que también prepara el lecho para la epitelización, proporcionando un entorno adecuado para la migración celular y la formación de nuevos vasos sanguíneos.

Sin embargo, la presencia de tejido necrótico y biofilm ³ bacteriano puede complicar este proceso. El tejido necrótico, formado por células muertas y detritos, actúa como una barrera que impide la formación de tejido de granulación saludable [7]. Esto no solo retrasa el proceso de curación, sino que también aumenta el riesgo de infección, lo que puede llevar a complicaciones graves, como la amputación. Por lo tanto, es crucial realizar un desbridamiento adecuado para remover el tejido necrótico y permitir que la herida progrese hacia las fases de proliferación y cicatrización. Por otro lado, el biofilm bacteriano se forma cuando microorganismos se adhieren a la superficie del lecho de la herida, creando microcolonias protegidas por una matriz polimérica. Esta estructura no solo protege a las bacterias de los tratamientos antibióticos convencionales, sino que también interfiere con la respuesta inmune del cuerpo y dificulta la migración celular necesaria para regenerar tejido sano. La identificación y tratamiento adecuados del tejido necrótico y el biofilm son esenciales para facilitar una curación efectiva y mejorar la calidad de vida del paciente.

Por lo antes expuesto este trabajo tiene por hipótesis el desarrollo de una aplicación capaz de segmentar de manera automática las úlceras a partir de una secuencia de imágenes, lo cual facilita el tratamiento médico de dicha herida.

³Tejido de color amarillento presente en las heridas que representa la zona infectada por las bacterias.

4.1. Objetivos

Este trabajo tiene como objetivo desarrollar una herramienta que permita realizar la segmentación automática de úlceras y los tejidos que las componen a partir de una secuencia de imágenes RGB-D, facilitando así el tratamiento por parte del médico.

Para lograr este objetivo general se tiene los siguientes objetivos específicos:

1. Hacer un estudio de la literatura sobre los métodos de segmentación
2. Estudiar sobre la utilización del *Segment Anything Model*.
3. Crear un dataset para la posterior evaluación del modelo
4. Evaluar del modelo en cuanto métricas de calidad.

4.2. Estructura de la tesis

La tesis está organizada en cuatro capítulos. El Capítulo 1 presenta una revisión literaria donde se discuten algunos algoritmos existentes así como datasets relevantes. El Capítulo 2 ofrece una introducción al SAM y su aplicación en imágenes médicas. En el Capítulo 3 se explican los detalles estructurales e implementativos del sistema propuesto. Finalmente, el Capítulo 4 muestra los resultados obtenidos y la comparación de estos con otros modelos existentes utilizando medidas cualitativas.

5. Capítulo 1

5.1. Segmentación de Úlceras del Pie Diabético: Definición

La segmentación de úlceras del pie diabético (UPD) es un proceso crítico en el manejo y tratamiento de esta complicación común en pacientes diabéticos. Se refiere a la identificación y delineación automática de las regiones afectadas en imágenes médicas, lo que permite a los profesionales de la salud evaluar con precisión el tamaño y la gravedad de las úlceras. Estas lesiones son abiertas y pueden surgir debido a una combinación de factores como neuropatía, que reduce la sensibilidad en los pies, isquemia, que limita el flujo sanguíneo necesario para la curación, y presión prolongada sobre áreas específicas del pie debido a deformidades o calzado inadecuado.

La segmentación es crucial no solo para el diagnóstico inicial, sino también para el seguimiento del proceso de curación. Al segmentar correctamente las úlceras, se puede monitorear su evolución y responder adecuadamente a los cambios en el estado del paciente. Esto es particularmente importante dado que las úlceras del pie diabético pueden llevar a infecciones severas y, en casos extremos, a amputaciones. La capacidad para segmentar estas lesiones con precisión puede mejorar significativamente la calidad de vida de los pacientes al permitir una intervención temprana y adecuada en el manejo de las úlceras [4].

Además, la segmentación automática ha sido objeto de investigación reciente, donde se han explorado diferentes métodos para mejorar su efectividad. Por ejemplo, estudios han demostrado que el uso de algoritmos avanzados puede facilitar la identificación precisa de los bordes de las úlceras, lo que es fundamental para determinar el área afectada y planificar tratamientos adecuados [2]. La implementación de técnicas como la Transformada Wavelet Discreta Logarítmica ha mostrado resultados prometedores al mejorar la calidad visual de las imágenes antes de aplicar algoritmos de segmentación [3].

5.2. Métodos existentes

En esta sección se habla del estudio realizado de la literatura actual que abarca el tema de interés. El proceso que se realiza en este trabajo consta de tres partes: una parte de procesamiento de la imagen, la segunda parte de la segmentación de esta y la última dedicada a la clasificación de los tejidos. Por lo que analizaremos publicaciones que traten estos temas de manera conjunta pero también de manera independiente.

En la literatura científica, se han utilizado diversos métodos para la segmentación de úlceras del pie diabético (UPD). Tradicionalmente, los enfoques manuales o semiautomáticos han sido comunes, donde los especialistas marcan las áreas afectadas en las imágenes. Sin embargo, estos métodos son propensos a errores humanos y requieren mucho tiempo, lo que puede afectar la eficiencia del diagnóstico y el tratamiento [3]. En Cuba, por ejemplo, se ha documentado que la evaluación de estas lesiones se realiza manualmente, lo que limita la capacidad de respuesta ante cambios en el estado de las úlceras [3].

Con el avance de la tecnología, se han desarrollado métodos automáticos que utilizan algoritmos matemáticos y técnicas de procesamiento de imágenes. Entre estos se encuentran el método Chan-Vese, el modelo de mezclas gaussianas y GrabCut, que han demostrado ser efectivos en diversas condiciones [2]. En un estudio descriptivo realizado en pacientes diabéticos, se utilizó el marco estereotáxico FrameHeber03® para obtener imágenes planimétricas estandarizadas de las úlceras y se aplicaron diferentes técnicas de segmentación. Los resultados mostraron que el método de mezclas gaussianas era el más factible y preciso para segmentar las lesiones, especialmente en pacientes con piel oscura, donde se observa un mayor contraste entre la piel y el borde de la úlcera [3].

Recientemente, se ha incrementado el uso de técnicas basadas en aprendizaje profundo, como las redes neuronales convolucionales (CNN), que han mostrado resultados prometedores al aprender características complejas directamente desde los datos sin necesidad de preprocesamiento extenso [1]. Estas técnicas no solo automatizan el proceso de segmentación, sino que también mejoran significativamente la precisión en comparación con los métodos tradicionales. Por ejemplo, en [1] se reportó un coeficiente de similitud Dice superior a 0.794 para la segmentación automática de UPD utilizando modelos de redes neuronales convolucionales completas (FCNs) [1].

La implementación de estos métodos automáticos no solo facilita una evaluación más rápida y precisa del estado de las úlceras del pie diabético, sino que también permite a los profesionales médicos tomar decisiones informadas sobre el tratamiento y seguimiento del paciente. La combinación de técnicas tradicionales con nuevas tecnologías representa un avance significativo en el manejo clínico de esta complicación diabética.

En Goyal et al.[14],se probaron diferentes arquitecturas, incluyendo:

- U-Net: Diseñada específicamente para la segmentación de imágenes biomédicas, U-Net demostró un rendimiento excepcional con un Coeficiente de Similitud Dice (DSC) del 97.25 % y un índice de Intersección sobre Unión (IoU) del 94.86 % utilizando un conjunto de datos limitado.
- V-Net: Una mejora sobre U-Net, inicialmente desarrollada para segmentar volúmenes de MRI prostáticos, fue adaptada para la segmentación de DFU.
- SegNet: Una arquitectura paralela encoder-decoder que carece de conexiones skip, también fue evaluada pero mostró resultados inferiores en comparación con U-Net y V-Net.

La tarea de segmentación involucró clasificar cada píxel como ulceroso o no ulceroso, permitiendo un mapeo detallado del área ulcerosa en las imágenes.

El modelo U-Net superó a otras arquitecturas en términos de precisión y eficiencia, logrando resultados significativos incluso después de un número mínimo de épocas de entrenamiento (5 épocas),ver [14]. Las evaluaciones cualitativas indicaron que la red podía generar máscaras que se asemejaban estrechamente a los datos reales, proporcionando alta precisión en la segmentación del área ulcerosa.

En Muñoz et al. [26], el enfoque propuesto se basa en la arquitectura Mask R-CNN, que combina la detección de objetos con la segmentación semántica. Este modelo permite no solo identificar la presencia de úlceras en las imágenes, sino también segmentar sus contornos de manera precisa.

Los resultados obtenidos muestran que el modelo Mask R-CNN logra una alta precisión en la segmentación de úlceras del pie diabético. Se reportan métricas como el Coeficiente de Similitud Dice (DSC) y el Índice de Intersección sobre Unión (IoU), que indican un rendimiento superior en comparación con métodos tradicionales.

Estos modelos no solo permiten una segmentación precisa, sino que también pueden manejar variaciones en las imágenes causadas por diferentes condiciones de iluminación o tipos de piel [5]. La implementación de estos métodos ha demostrado mejorar significativamente las métricas de evaluación como el coeficiente Dice y la intersección sobre unión (IoU), lo que indica un avance considerable respecto a los métodos tradicionales.

La calidad de la segmentación se evalúa utilizando diversas métricas que reflejan la precisión y efectividad del algoritmo aplicado. Entre las métricas más comunes se encuentran el coeficiente Dice, que mide la similitud entre la predicción del modelo y la verdad conocida; la intersección sobre unión (IoU), que evalúa la superposición entre las áreas predicha y real; así como la precisión y recall, que proporcionan información sobre los verdaderos positivos y negativos [9]. Estas métricas son esenciales para comparar diferentes enfoques y determinar cuál es más efectivo para segmentar úlceras del pie diabético bajo diversas condiciones.

Otra tarea que se propone en este trabajo es la de clasificar los tejidos presentes en la herida luego de esta ser segmentada. Existen diversos estudios que han realizado una investigación sobre el tema y han utilizados varios modelos de aprendizaje de máquinas tales como las ya mencionadas redes neuronales convolucionales, la máquina de vectores de soporte(SVM, por sus siglas en inglés), y muchos otros. A continuación se ofrece un resumen de varios artículos estudiados sobre el tema.

En [31] realizan una clasificación de tejidos utilizando las Redes Adversariales Generativas(GAN, por sus siglas en inglés). Los autores proponen un enfoque innovador utilizando una Red Generativa Adversarial Condicional (cGAN) para realizar simultáneamente la segmentación de los bordes de las heridas y la clasificación de los tejidos. Este método permite que el modelo aprenda a partir de un conjunto de datos que incluye imágenes de heridas con etiquetas correspondientes a sus bordes y tipos de tejido.

5.3. Datasets Existentes

Existen varios datasets disponibles para entrenar y evaluar modelos destinados a la segmentación automática de UPD. Uno notable es el dataset propuesto por Wang et al., que contiene más de 700 imágenes con anotaciones detalladas sobre las regiones afectadas por úlceras [5]. Este dataset proporciona una regla de oro para evaluar los modelos entrenados, permitiendo comparaciones con otros enfoques existentes. Además, otros estudios han utilizado conjuntos más pequeños pero igualmente valiosos que permiten realizar análisis comparativos entre diferentes algoritmos [10]

6. Capítulo 2

6.1. Modelos de Base(Foundation Models)

Los modelos de base son un enfoque transformador en la inteligencia artificial (IA) que aprovecha grandes conjuntos de datos para realizar una variedad de tareas en diferentes dominios. Estos modelos se entrenan en datos extensos, generalizados y, a menudo, no etiquetados, lo que les permite adaptarse y ajustarse para aplicaciones específicas.

Uno de los modelos de base más conocidos es la serie GPT (Generative Pre-trained Transformer) [[6, 28]], que ha demostrado capacidades y rendimiento impresionantes en una variedad de tareas de procesamiento de lenguaje natural, como la finalización de oraciones, la respuesta a preguntas y la traducción de idiomas. Vea [36]. Los modelos fundamentales también han mostrado un fuerte potencial para resolver una amplia gama de tareas posteriores en el análisis de imágenes médicas y ayudar a acelerar el desarrollo de modelos precisos y robustos. [36].

6.2. Segment Anything Model(SAM)

SAM es el primer modelo fundamental para la segmentación general de imágenes. Ha logrado resultados impresionantes en diversas tareas de segmentación de imágenes naturales [17]. Sin embargo, la segmentación de imágenes médicas (MIS, por sus siglas en inglés) es más desafiante debido a las complejas modalidades, las finas estructuras anatómicas, los contornos de objetos inciertos y complejos, y la amplia variedad de escalas de objetos. SAM está basado en el modelo Vision Transformer (ViT) y fue entrenado en un gran conjunto de datos que contiene 11 millones de imágenes con mil millones de máscaras.

6.3. Etapas del Preentrenamiento

1. Recolección y Anotación de Datos El conjunto de datos SA-1B se compone de imágenes diversas y de alta resolución, obtenidas a partir de una amplia variedad de fuentes. Las máscaras de segmentación asociadas a estas imágenes se generan mediante un proceso automatizado que involucra:

- **Anotación Asistida:** En esta fase inicial, los anotadores humanos trabajan junto con el modelo SAM para asegurar que cada máscara sea capturada y anotada con precisión.
- **Semi-Automático:** Los anotadores se enfocan en áreas donde SAM muestra menos confianza, refinando y complementando las predicciones del modelo.
- **Totalmente Automático:** En la etapa final, SAM predice independientemente las máscaras de segmentación, demostrando su capacidad para manejar escenarios complejos con mínima intervención humana [[32]].

2. **Preentrenamiento con Aprendizaje Máscara** El modelo SAM utiliza un enfoque llamado Masked Image Pretraining (MIP), específicamente adaptado para aprender a reconstruir características visuales a partir de imágenes. Este método implica:

- **Máscaras en Imágenes:** Durante el preentrenamiento, se aplican máscaras a las imágenes, donde aproximadamente el 75 % de los tokens (píxeles) son ocultados. El modelo aprende a reconstruir estos tokens ocultos a partir de los visibles, lo que fomenta una comprensión profunda del contenido visual [[22]].
- **Transferencia del Conocimiento:** Se utiliza un codificador basado en Vision Transformer (ViT) para extraer características significativas. Este proceso no solo mejora la representación visual, sino que también permite que el modelo generalice bien en tareas posteriores como la clasificación y detección de objetos [[8]].

3. **Ajuste Fino** Después del preentrenamiento, los encoders ligeros (como ViT-Tiny o ViT-Small) se ajustan finamente utilizando el conjunto SA-1B para tareas específicas de segmentación. Esto implica entrenar el modelo con datos supervisados para optimizar su rendimiento en tareas concretas como la segmentación semántica o la detección de instancias [[30]].

SAM utiliza una arquitectura transformer-based, a la cual le ha sido probada su eficiencia en el procesamiento de lenguaje natural y en tareas de reconocimiento de imágenes y ha obtenido muy buenos resultados. Específicamente, contiene un codificador de imagen(image encoder) basado en un Vision Transformer(ViT), con este extrae las características de la imagen,

también utiliza un codificador de prompts para integrar las interacciones del usuario y por último un decodificador de máscaras con el objetivo de predecir las máscaras de segmentación con la fusión de las características de la imagen con las entradas del usuario.

En la figura 1 se aprecia la estructura de SAM:

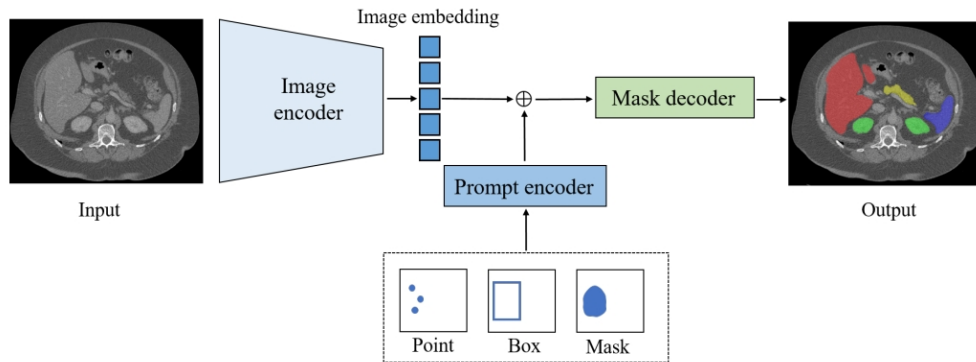
- 1 Image Encoder: El codificador de imagen es fundamental en la arquitectura de SAM. Utiliza un Vision Transformer para extraer características de las imágenes. Este proceso implica:
 - I- División de la imagen: La imagen se divide en parches fijos, que son luego aplanados y convertidos en embeddings mediante una capa de embedding.
 - II- Atención global: A través de mecanismos de auto-atención, el modelo puede capturar relaciones entre diferentes partes de la imagen, lo que le permite entender tanto características locales como globales.
 - III- Representación densa: El resultado es una matriz densa que representa las características relevantes del contenido visual, lo que forma la base para la identificación y segmentación de objetos dentro de la imagen.
- 2 Prompt Encoder: El codificador de prompts es un componente distintivo que le permite adaptarse a diferentes tipos de entradas del usuario, como:
 - Puntos específicos en la imagen.
 - Cajas delimitadoras.
 - Máscaras aproximadas.
 - Texto descriptivo.

Este codificador transforma estos inputs en embeddings que guían el proceso de segmentación, permitiendo al modelo centrarse en áreas específicas según lo indicado por el usuario.

3 Mask Decoder : El decodificador de máscaras es donde se realiza la segmentación efectiva. Este componente combina las embeddings del codificador de imagen y del codificador de prompts para generar máscaras precisas. Su funcionamiento incluye:

- Mecanismos de atención cruzada: El decodificador utiliza atención cruzada para actualizar las embeddings tanto del codificador de imagen como del codificador de prompts, asegurando que ambas fuentes de información se integren eficazmente.
- Predicción dinámica: A partir de esta fusión, el decodificador produce las máscaras segmentadas, permitiendo al modelo generar múltiples opciones para cada prompt proporcionado

Figura 1: Imagen adaptada de [36]



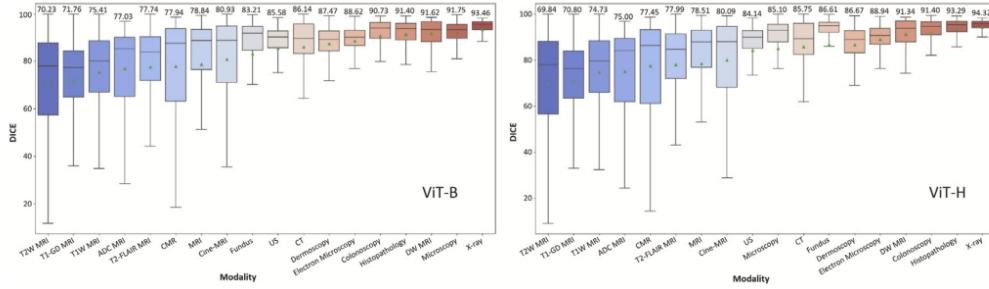
Durante el entrenamiento, la predicción de salida se supervisa con la combinación lineal de la pérdida focal [19] y la pérdida de Dice [24] , y se entrena para la tarea de segmentación mediante prompts geométricos mixtos.

Aún cuando es evidente el potencial de SAM para la segmentación de imágenes naturales, presenta limitaciones en el ámbito de las imágenes médicas. Su rendimiento limitado en imágenes médicas se atribuye a su insuficiente comprensión de desafíos específicos del ámbito médico [21], como el bajo contraste de imagen, los contornos de tejidos ambiguos y las pequeñas regiones de lesiones. Estos factores complican el proceso de segmentación en comparación con las imágenes naturales.

El enfoque más avanzado para mejorar SAM para aplicaciones médicas implica realizar un ajuste completo del modelo en conjuntos de datos médicos. Sin embargo, este método es costoso en términos computacionales y requiere una cantidad significativa de recursos de memoria, lo que plantea dudas sobre su necesidad, dado que estudios previos han demostrado que los modelos visuales preentrenados suelen mostrar una fuerte transferibilidad a imágenes médicas [29].

En Zhang et al. [36] realizan una evaluación de SAM centrándose en una amplia gama de objetivos anatómicos y patológicos a través de diferentes modalidades de imagen médica. Estas modalidades abarcan tanto imágenes médicas en 2D (por ejemplo, radiografías, patología, ultrasonido, endoscopia y colonoscopia) como imágenes médicas en 3D (por ejemplo, Tomografía Computarizada (TC), Resonancia Magnética (RM) y Tomografía por Emisión de Positrones (PET)). Presentan el uso en cero disparos de SAM en la segmentación de imágenes médicas, organizado según los formatos de las modalidades de imagen médica.

Figura 2: Resultados cuantitativos de segmentación del modelo SAM en 18 modalidades de imagen diferentes en un estudio empírico a gran escala [17]. Imagen adaptada de [36]



En general, SAM requiere una interacción humana sustancial para lograr un rendimiento de segmentación moderado, lo que solo requiere unos pocos puntos o indicaciones de caja delimitadora. Los resultados de la evaluación en varios conjuntos de datos indican la limitada capacidad de generalización de SAM cuando se aplica directamente a la segmentación de imágenes médicas, que varía significativamente entre diferentes conjuntos de datos y tareas. Si bien SAM demuestra un rendimiento notable comparable a los métodos

SOTA⁴ en la identificación de objetos bien definidos en ciertas modalidades de imagen, presenta imperfecciones o fracasos totales en situaciones más desafiantes. Esto es particularmente evidente al tratar con objetivos de segmentación que presentan bordes débiles, bajo contraste, pequeño tamaño y formas irregulares, alineándose con los hallazgos de otras investigaciones. Para la mayoría de los escenarios de segmentación de imágenes médicas, el rendimiento subóptimo de segmentación de SAM no cumple con los requisitos para aplicaciones posteriores, particularmente en algunas tareas que exigen una precisión extremadamente alta. El conjunto de datos SA-1B, los datos de entrenamiento de SAM, está compuesto principalmente por imágenes naturales con información de borde fuerte y presenta una disimilitud significativa con las imágenes médicas. En consecuencia, aplicar SAM directamente sin ajuste fino o reentrenamiento a tareas desafiantes y no vistas previamente en la segmentación de imágenes médicas puede resultar en un rendimiento limitado. Vease [36].

6.4. Adaptando SAM para imágenes médicas.

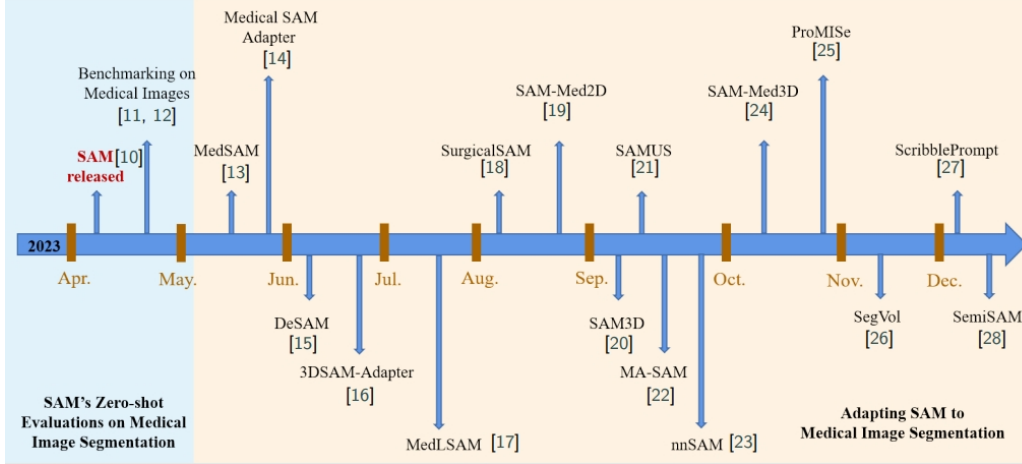
Por estos motivos se han realizado varios estudios con el objetivo de adaptar SAM para mejorar su rendimiento en imágenes médicas. La figura 3 ilustra las diferentes modalidades de SAM a lo largo de los años.

Para mejorar el rendimiento insatisfactorio de SAM en tareas de segmentación de imágenes médicas, un enfoque directo e intuitivo es ajustar finamente SAM con imágenes médicas, lo que incluye tanto el ajuste completo como el ajuste eficiente de parámetros. El enfoque más sencillo para adaptar SAM a la segmentación de imágenes médicas es ajustar finamente SAM directamente en la tarea específica que se tiene entre manos. .

Hu et al. [16] presenta una validación del ajuste fino de SAM para la segmentación de cáncer de piel, demostrando una mejora sustancial en la puntuación DSC, que pasó del 81.25 % al 88.79 % . Vease [36]. MedSAM [20] es un modelo universal para la segmentación de imágenes médicas que se adapta de SAM utilizando un conjunto de datos diverso con más de un millón de pares de imágenes y máscaras de 11 modalidades. Logró puntuaciones medianas de DSC del 94.0 % al 98.4 % en tareas como hemorragia

⁴State of the Art(Estado del arte), por sus siglas en inglés

Figura 3: Una breve cronología del modelo Segment Anything (SAM) [18] y sus variantes para la segmentación de imágenes médicas en 2023. Imagen adaptada de Zhang et al. [36]



intracraneal, glioma, neumotórax y pólipos, superando a los modelos U-Net especializados. Sin embargo, enfrenta dificultades para segmentar estructuras ramificadas similares a vasos debido a la ambigüedad en los prompts de caja delimitadora y solo procesa imágenes 3D como series de cortes 2D en lugar de volúmenes.

El fine-tuning eficiente en parámetros (PEFT) es un enfoque utilizado para mejorar el rendimiento de modelos de aprendizaje automático, especialmente en el contexto de modelos de lenguaje y redes neuronales. Este método se centra en ajustar solo un pequeño subconjunto de los parámetros del modelo preentrenado, lo que permite adaptar el modelo a tareas específicas sin necesidad de realizar un ajuste completo, que puede ser costoso en términos de recursos computacionales y tiempo. Wu et al. [33] proponen el Adaptador SAM Médico (Med-SA), que mantiene los parámetros preentrenados de SAM congelados mientras integra módulos de adaptación de rango bajo (LoRA) [15] en posiciones designadas. Los extensos experimentos realizados en 17 tareas de segmentación de imágenes médicas a través de 5 modalidades diferentes muestran la superioridad de Med-SA sobre SAM y métodos anteriores de estado del arte (SOTA) [36]. Similarmente SAMed [34] aplica una estrategia de ajuste fino basada en rango bajo (LoRA) al codificador de imágenes

de SAM y lo ajusta junto con el codificador de prompts y el decodificador de máscaras en conjuntos de datos etiquetados para segmentación de imágenes médicas. Además, dado que SAMed solo actualiza una pequeña fracción de los parámetros de SAM, su costo y almacenamiento son bastante marginales en su uso práctico. Existen otro muchos modelos adaptados de SAM como se aprecia en la figura 3

7. Conclusiones

8. Recomendaciones

en la figura 1

9. Bibliografía

- [1] Gretel Garcia Gomez et al. «Características ideales para datasets en segmentación médica: un enfoque práctico». En: (2020). URL: https://www.researchgate.net/publication/336879527_Segmentacion_automatica_de_ulceras_del_pie_diabetico_y_tejidos_internos_mediante_aprendizaje_automatico Automatic_Segmentation_of_Diabetic_Foot_Ulcers_and_Internal_Tissues_with_Machine_Learning.
- [2] Gretel Garcia Gomez et al. «Segmentación automática de úlceras del pie diabético y tejidos internos mediante aprendizaje automático». En: (2019). URL: https://www.researchgate.net/publication/336879527_Segmentacion_automatica_de_ulceras_del_pie_diabetico_y_tejidos_internos_mediante_aprendizaje_automatico Automatic_Segmentation_of_Diabetic_Foot_Ulcers_and_Internal_Tissues_with_Machine_Learning.
- [3] Heber et al. «Mejoramiento de contraste y segmentación en imágenes de úlceras del pie diabético». En: *Revista Angiología* (2019). URL: <https://revangiologia.sld.cu/index.php/ang/article/view/61>.
- [4] Nicolás Pereira C. et al. «Úlceras del pie diabético: importancia del manejo multidisciplinario y salvataje microquirúrgico de la extremidad». En: *Revista Chilena de Cirugía* (2018). URL: <https://www.scielo.cl/pdf/rchcir/v70n6/0718-4026-rchcir-70-06-0535.pdf>.
- [5] Wang et al. «Evaluación cuantitativa de técnicas automáticas para segmentación médica». En: *Journal of Medical Imaging* (2020). URL: https://ejemplo.com/articulo_evaluacion_tecnicas.
- [6] T. Brown et al. «Language models are few-shot learners». En: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 1877-1901.
- [7] Úlceras Cl. *El proceso de cicatrización de las heridas crónicas*. Accedido: diciembre 2024. 2024. URL: <https://ulceras.cl/monografico/el-proceso-de-cicatrizacion-de-las-heridas-cronicas/>.

- [8] Datos.gob. «Segment Anything Model: claves del modelo de segmentación de Meta aplicados a los datos espaciales». En: *Datos.gob.es* (2023). URL: <https://datos.gob.es/es/blog/segment-anything-model-claves-del-modelo-de-segmentacion-de-meta-aplicados-los-datos-espaciales>.
- [9] Desconocido. «Base de datos sobre úlceras del pie diabético: características y aplicaciones». En: *Revista Cubana Angiología Cirugía Vascular* (2019). URL: <https://pesquisa.bvsalud.org/portal/resource/pt/biblio-1093137>.
- [10] Desconocido. «Tendencias recientes en tecnología médica: aplicaciones móviles en el seguimiento médico». En: *Revista Médica Innovadora* (2021). URL: https://ejemplo.com/articulo_tecnologia_medica.
- [11] International Diabetes Federation. *Diabetes Atlas*. 2024. URL: <https://diabetesatlas.org/> (visitado 09-12-2024).
- [12] Damir Filko, Robert Cupec y Emmanuel Karlo Nyarko. «Wound measurement by RGB-D camera». En: *Machine Vision and Applications* 29 (2018), pp. 633-654. DOI: 10.1007/s00138-018-0920-4.
- [13] Damir Filko y Emmanuel Karlo Nyarko. «2D/3D Wound Segmentation and Measurement Based on a Robot-Driven Reconstruction System». En: *Sensors* 23.6 (2023), p. 3298. DOI: 10.3390/s23063298.
- [14] Manu Goyal et al. «Fully Convolutional Networks for Diabetic Foot Ulcer Segmentation». En: *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*. Manchester, UK: IEEE, 2017, pp. 1-6.
- [15] E. J. Hu et al. «LoRA: Low-Rank Adaptation of Large Language Models». En: *arXiv preprint arXiv:2106.09685* (2021). URL: <https://arxiv.org/abs/2106.09685>.
- [16] M. Hu, Y. Li y X. Yang. «Skinsam: Empowering skin cancer segmentation with segment anything model». En: *arXiv preprint arXiv:2304.13973* (2023). URL: <http://arxiv.org/abs/2304.13973>.
- [17] Yuhao Huang et al. «Segment Anything Model for Medical Images». En: *arXiv preprint arXiv:2408.00874* (2024). URL: <https://arxiv.org/abs/2408.00874>.
- [18] A. Kirillov et al. «Segment anything». En: *arXiv preprint arXiv:2304.02643* (2023).

- [19] Tsung-Yi Lin et al. «Focal Loss for Dense Object Detection». En: *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE. 2017, pp. 2999-3007.
- [20] Jun Ma y Bo Wang. «Segment anything in medical images». En: *arXiv preprint arXiv:2304.12306* (2023).
- [21] Jun Ma et al. «Segment Anything in Medical Images». En: *arXiv preprint arXiv:2304.12306* (2023). URL: <https://arxiv.org/abs/2304.12306>.
- [22] MariaIA. «Segment Anything (SAM): el revolucionario modelo de Meta para democratizar la segmentación de imágenes». En: *Las Cosas De Internet* (2023). URL: <https://lascosasdeinternet.com/tech-development/inteligencia-artificial/modelos-ia/segment-anything-sam-el-revolucionario-modelo-de-meta-para-democratizar-la-segmentacion-de-imagenes/>.
- [23] Meta AI. *Segment Anything*. Accedido el 14 Nov 2024. 2023. URL: <https://segment-anything.com/>.
- [24] Fausto Milletari, Nassir Navab y Seyed-Ahmad Ahmadi. «V-net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation». En: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE. 2016, pp. 565-571.
- [25] Ministerio de Salud Pública. *Anuario Estadístico de Salud 2022*. Accedido el 14 Nov 2024. La Habana: Dirección Nacional de Registros Médicos y Estadísticas de Salud. 2023. URL: <https://salud.msp.gob.cu/wp-content/Anuario/Anuario-2022.pdf>.
- [26] P. L. Muñoz, R. Rodríguez y N. Montalvo. «Automatic Segmentation of Diabetic Foot Ulcer from Mask Region-Based Convolutional Neural Networks». En: *Instituto de Cibernética, Matemática y Física (ICIMAF) 1.1* (2024), pp. 1-5.
- [27] Clínica Universidad de Navarra. *Tejido de granulación*. Accedido: diciembre 2024. 2023. URL: <https://www.cun.es/diccionario-medico/terminos/tejido-de-granulacion>.
- [28] OpenAI. «GPT-4 Technical Report». En: *arXiv preprint arXiv:2303.08774* (2023). URL: <https://arxiv.org/abs/2303.08774>.

- [29] Maithra Raghu et al. «Transfusion: Understanding Transfer Learning for Medical Imaging». En: *Advances in Neural Information Processing Systems*. Vol. 32. 2019. URL: <https://www.semanticscholar.org/paper/Transfusion:-Understanding-Transfer-Learning-for-Raghu-Zhang/cff4cb74f4466bd0407977e40ef0be9f444c63ea>.
- [30] ResearchGate. «Accuracy of Segment-Anything Model (SAM) in medical image segmentation tasks». En: *ResearchGate* (2023). URL: https://www.researchgate.net/publication/370127016_Accuracy_of_Segment-Anything_Model_SAM_in_medical_image_segmentation_tasks.
- [31] S. Sarp et al. «Simultaneous wound border segmentation and tissue classification using a conditional generative adversarial network». En: *Journal of Engineering* 2021.3 (2021), pp. 125-134. DOI: 10.1049/tje2.12016.
- [32] Ultralytics. «SAM (Modelo Segment Anything)». En: *Ultralytics YOLO Docs* (2024). URL: <https://docs.ultralytics.com/es/models/sam/>.
- [33] J. Wu et al. «Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation». En: *arXiv preprint arXiv:2304.12620* (2023). URL: <https://arxiv.org/abs/2304.12620>.
- [34] K. Zhang y D. Liu. «Customized Segment Anything Model for Medical Image Segmentation». En: *arXiv preprint arXiv:2304.13785* (2023). URL: <https://arxiv.org/abs/2304.13785>.
- [35] Peng Zhang, Yichen Zhang y Qiang Li. «RGB-D Camera-Based Automatic Wound-Measurement System». En: *IEEE Transactions on Instrumentation and Measurement* (2023).
- [36] Yichi Zhang, Zhenrong Shen y Rushi Jiao. «Segment Anything Model for Medical Image Segmentation: Current Applications and Future Directions». En: *arXiv* (2024). URL: <https://arxiv.org/abs/2401.03495>.

10. Anexos