

ESTRATEGIA DE APRENDIZAJE PROFUNDO PARA  
LA SEGMENTACIÓN NO SUPERVISADA EN  
PATOLOGÍA DIGITAL

JUAN SEBASTIÁN DÍAZ GUTIÉRREZ

JULIÁN CAMILO CAMACHO TORRES

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
PROGRAMA DE PREGRADO EN INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA

2024

**ESTRATEGIA DE APRENDIZAJE PROFUNDO PARA  
LA SEGMENTACIÓN NO SUPERVISADA EN  
PATOLOGÍA DIGITAL**

**Trabajo de grado para optar al título de Ingenieros de Sistemas**

**JUAN SEBASTIÁN DÍAZ GUTIÉRREZ**

**JULIÁN CAMILO CAMACHO TORRES**

**Director:**

David Romo Bucheli, Doctor en Ingeniería

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
PROGRAMA DE PREGRADO EN INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA

2024

## DEDICATORIA

A mi madre Ana y a mi padre Nelson, que amo con todo mi corazón y me han apoyado incondicionalmente con cariño y consejos para la vida.

A mis demás familiares que han creído en mí y en mi futuro.

A mi novia, quien ha sido mi acompañante en mi carrera y nos hemos brindado apoyo mutuo.

A mis amigos y compañeros, los cuales hemos compartido y nos han ayudado mucho en la carrera, especialmente a Juan, quien es un gran amigo y que nos hemos apoyado durante la carrera.

**Julián Camilo Camacho Torres**

## DEDICATORIA

A mi padre Rafael y mi madre Ana, a quienes amo incondicionalmente y me han brindado absolutamente todo su apoyo y amor en esta travesía llamada vida; a mi hermana Gabriela por ser mi cómplice y más grande admiradora en todo lo que me he propuesto desde que tengo memoria. A todos mis amigos y compañeros que han sido de gran apoyo durante este proyecto, especialmente a Julián quien no solo es un gran amigo, sino mi compañero en esta investigación.

A mi hermanos más que mis amigos, Oscar Blanco, Luis Torres, Angel Gonzalez, Camilo Jerez y Cristian Chiquillo, quienes no solo me apoyaron sino que me elevaron en esta magnífica etapa de mi vida.

A mi hermosa guardiana, mi abuela Farina, quien me motivó a emprender esta aventura en la UIS.

A Eli, a quien tengo tanto para agradecerle.

Por mí y por tí Richard Gerard Díaz Cujia, espero estés orgulloso donde sea que estés.

**Juan Sebastián Díaz Gutiérrez**

## AGRADECIMIENTOS

Los autores expresan su agradecimiento:

Nos gustaría extender nuestro más sincero agradecimiento a todas las personas que contribuyeron de manera invaluable a la realización de esta tesis. En primer lugar, queremos expresar nuestro profundo reconocimiento a nuestro asesor académico, David Romo, cuya orientación experta y dedicación guiaron cada paso de este arduo proceso de investigación. Sus consejos perspicaces y constante apoyo nos brindaron una visión clara y enriquecedora para abordar los desafíos que enfrentamos. Además, deseamos agradecer a nuestros familiares y amigos, cuyo aliento y comprensión nos impulsaron a perseverar en este viaje académico. Su paciencia y motivación inquebrantables fueron el motor que nos permitió alcanzar este logro tan significativo.

En este punto culminante de nuestra trayectoria académica, queremos expresar nuestro más profundo agradecimiento a todas las personas que contribuyeron de manera desinteresada a esta tesis.

## GLOSARIO

**Clúster** se refiere a un grupo de diferentes píxeles o regiones con características similares.

**Etiqueta** es una etiqueta o categoría asignada a cada ejemplo o instancia en un conjunto de datos de entrenamiento. En tareas de aprendizaje supervisado, las etiquetas son las respuestas o resultados conocidos que se utilizan para entrenar un modelo. Estas etiquetas son esenciales para que el modelo aprenda a hacer predicciones precisas.

**Histología** la histología es el estudio de la anatomía microscópica de los tejidos. Es la contrapartida microscópica de la anatomía macroscópica, que estudia estructuras más grandes visibles sin microscopio. La histología es una rama de la biología que se utiliza en muchos campos diferentes, como la medicina, la odontología, la veterinaria y la biología.

**Inteligencia artificial** es una rama de la informática y la ciencia de la computación que se enfoca en la creación de sistemas y programas capaces de realizar tareas que, normalmente, requieren de la inteligencia humana. Estos sistemas utilizan algoritmos y modelos matemáticos avanzados para procesar datos, aprender de la experiencia y tomar decisiones con base en patrones y reglas previamente establecidas.

**Modelo** es una representación matemática o computacional que captura ciertas características o patrones de un conjunto de datos.

**Redes Convolucionales** son un tipo de arquitectura de redes neuronales profundas diseñadas específicamente para el procesamiento de datos que tienen una estructura de cuadrícula, como imágenes, señales de audio o datos en formato de serie temporal. Estas redes se han vuelto ampliamente populares y efectivas en la tarea de reconocimiento de patrones en imágenes, y

han demostrado su eficacia en otras tareas relacionadas con la visión por computadora y el procesamiento de señales.

**Redes Neuronales** son modelos computacionales inspirados en la estructura y funcionamiento del cerebro humano. Estas redes están diseñadas para aprender patrones y relaciones complejas a partir de datos de entrada, mediante la utilización de múltiples capas de unidades de procesamiento interconectadas, conocidas como neuronas artificiales. Cada neurona toma las entradas, las procesa mediante una función de activación y produce una salida.

**Segmentación** es una técnica que consiste en dividir una imagen en regiones o segmentos significativos basados en ciertas características o propiedades comunes. El objetivo de la segmentación es identificar y separar objetos, áreas o patrones específicos dentro de una imagen para facilitar su análisis y comprensión.

# ÍNDICE GENERAL

	Pág
INTRODUCCIÓN . . . . .	14
<b>1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA . . . . .</b>	<b>17</b>
<b>2. OBJETIVOS . . . . .</b>	<b>20</b>
2.1. OBJETIVO GENERAL . . . . .	20
2.2. OBJETIVOS ESPECÍFICOS . . . . .	20
<b>3. MARCO TEÓRICO Y TRABAJOS PREVIOS . . . . .</b>	<b>21</b>
3.1. HISTOPATOLOGÍA . . . . .	21
3.2. PATOLOGÍA DIGITAL . . . . .	22
3.3. APRENDIZAJE DE MÁQUINA . . . . .	23
3.3.1. Aprendizaje supervisado . . . . .	24
3.3.2. Segmentación supervisada . . . . .	25
3.3.3. Aprendizaje no supervisado . . . . .	26
3.3.4. Segmentación no supervisada . . . . .	28
<b>4. MATERIALES . . . . .</b>	<b>30</b>
4.1. GlaS CHALLENGE . . . . .	30
4.2. PARTICIÓN DEL DATASET . . . . .	30
<b>5. MÉTODOS . . . . .</b>	<b>32</b>
5.1. ARQUITECTURA ReDO: UNSUPERVISED OBJECT SEGMENTATION BY RE-DRAWING . . . . .	32
5.2. SGSCN: SPATIAL GUIDED SELF-SUPERVISED CLUSTERING NETWORK . . . . .	34

5.3. MEDIAN POOLING . . . . .	36
5.4. MUMFORD-SHAH LOSS . . . . .	36
5.5. APRENDIZAJE SUPERVISADO: ARQUITECTURA U-NET . . . . .	38
5.6. MÉTRICAS DE EVALUACIÓN . . . . .	40
<b>6. CONFIGURACIÓN EXPERIMENTAL . . . . .</b>	<b>43</b>
6.1. PREPROCESAMIENTO DE LAS IMÁGENES . . . . .	43
6.2. DETALLES EXPERIMENTALES . . . . .	44
6.3. ARQUITECTURA U-NET . . . . .	49
<b>7. EVALUACIÓN Y RESULTADOS . . . . .</b>	<b>52</b>
7.1. ReDO . . . . .	52
7.2. SGSCN . . . . .	53
7.3. U-NET . . . . .	55
7.4. RESULTADOS CUANTITATIVOS . . . . .	56
<b>8. CONCLUSIONES Y PERSPECTIVAS . . . . .</b>	<b>61</b>
<b>BIBLIOGRAFÍA . . . . .</b>	<b>62</b>
<b>ANEXOS . . . . .</b>	<b>66</b>

# ÍNDICE DE FIGURAS

	Pág
Figura 1. Aprendizaje de máquina . . . . .	24
Figura 2. Arquitectura ReDO . . . . .	32
Figura 3. Arquitectura SGSCN . . . . .	35
Figura 4. Ejemplo de pooling layer con un tamaño de agrupación de 2x2. . . . .	36
Figura 5. Representación de la función MS Loss para la segmentación no supervisada de imágenes, tomada de <sup>1</sup> . . . . .	37
Figura 6. Arquitectura U-Net . . . . .	39
Figura 7. Esquema del sistema de medición de la similitud estructural (SSIM) <sup>2</sup> . . . . .	41
Figura 8. Hiperparámetros U-Net . . . . .	49
Figura 9. Resultados cualitativos ReDO . . . . .	52
Figura 10. Resultados cualitativos SGSCN . . . . .	53
Figura 11. Grilla de Resultados cualitativos . . . . .	54
Figura 12. Resultados cualitativos Unet . . . . .	55
Figura 13. Diagrama de caja para los resultados de la métrica SSIM. . . . .	58
Figura 14. Diagrama de caja para los resultados de la métrica MIOU. . . . .	59
Figura 15. Resultados encontrados sobre el dataset All Gland, en la cual el color amarillo hace referencia a la segmentación de una glándula y el azul al fondo. . . . .	59

---

<sup>1</sup> Boah Kim y Jong Chul Ye. “Mumford–Shah loss functional for image segmentation with deep learning”. En: *IEEE Transactions on Image Processing* 29 (2019), págs. 1856-1866.

<sup>2</sup> Zhou Wang et al. “Image quality assessment: from error visibility to structural similarity”. En: *IEEE transactions on image processing* 13.4 (2004), págs. 600-612.

## ÍNDICE DE CUADROS

	<b>Pág</b>
Tabla 1. Hiperparámetros utilizados con el método ReDO . . . . .	45
Tabla 2. Hiperparámetros utilizados en el método SGSCN . . . . .	46
Tabla 3. Hiperparámetros utilizados con Median Pooling . . . . .	48
Tabla 4. Resultados cuantitativos para los modelos de segmentación . . . . .	57

## RESUMEN

**TÍTULO:** ESTRATEGIA DE APRENDIZAJE PROFUNDO PARA LA SEGMENTACIÓN NO SUPERVISADA EN PATOLOGÍA DIGITAL \*

**AUTORES:** JUAN SEBASTIÁN DÍAZ GUTIÉRREZ, JULIÁN CAMILO CAMACHO TORRES \*\*

**PALABRAS CLAVE:** Aprendizaje profundo, patología digital, segmentación no supervisada.

**DESCRIPCIÓN:** La histopatología es área fundamental de la medicina que se ocupa del estudio de las enfermedades a través de la observación microscópica de tejidos y células. Utilizando técnicas de tinción y análisis microscópico, los histopatólogos examinan muestras de tejido obtenidas mediante biopsias, autopsias o cirugías para identificar y diagnosticar enfermedades.

Por otro lado, la inteligencia artificial (IA) ha demostrado ser una herramienta valiosa en el campo de la histopatología al mejorar la precisión y eficiencia en el análisis de muestras de tejido. Mediante algoritmos de aprendizaje automático, la IA puede ayudar en la detección y clasificación de células y estructuras patológicas, permitiendo una identificación más rápida y precisa de enfermedades.

Con este trabajo, esperamos explorar la capacidad de generalización y robustez de una representación obtenida a través de modelos de segmentación no supervisada en el área de la histopatología. Esta técnica puede mejorar el tiempo en el diagnóstico de los pacientes debido a que los métodos no supervisados no requieren de etiquetas para generar una representación que tiene potencialmente una mejor generalización.

---

\* Trabajo de investigación

\*\* Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: David Romo Bucheli, PhD

## ABSTRACT

**TITLE:** Deep learning strategies for unsupervised segmentation in digital pathology. \*

**AUTHOR:** JUAN SEBASTIÁN DÍAZ GUTIÉRREZ, JULIÁN CAMILO CAMACHO TORRES \*\*

**KEYWORDS:** Deep Learning, digital pathology, unsupervised segmentation

**DESCRIPTION:** Histopathology is a fundamental area of medicine that deals with the study of disease through microscopic observation of tissues and cells. Using microscopic staining and analysis techniques, histopathologists examine tissue samples obtained by biopsy, autopsy or surgery to identify and diagnose disease.

On the other hand, artificial intelligence (AI) has proven to be a valuable tool in the field of histopathology by improving accuracy and efficiency in the analysis of tissue samples. Through machine learning algorithms, AI can aid in the detection and classification of pathological cells and structures, enabling faster and more accurate identification of diseases.

With this work, we hope to explore the generalizability and robustness of a Representation obtained through unsupervised segmentation models in the area of histopathology. This technique can improve efficiency and objectivity in its diagnosis of patients because unsupervised methods do not require labels to generate a representation that potentially has better generalization.

---

\* Research work

\*\* Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: David Romo Bucheli

## INTRODUCCIÓN

En el campo de la medicina, el diagnóstico preciso y temprano de enfermedades es de vital importancia para mejorar los resultados clínicos y la calidad de vida de los pacientes. En el ámbito de la histopatología, donde los patólogos analizan muestras de tejido para identificar y clasificar enfermedades, se han realizado grandes avances en las últimas décadas. Sin embargo, el creciente volumen de datos digitales generados por la patología digital ha planteado desafíos significativos en términos de tiempo y claridad en el momento de analizar de estas muestras. <sup>1</sup>

La patología digital ha revolucionado la forma en que se diagnostican las enfermedades al permitir la digitalización y el análisis computarizado de imágenes de tejido. Mediante el uso de escáneres de alta resolución, las muestras de tejido se convierten en imágenes digitales que pueden ser almacenadas, visualizadas y compartidas electrónicamente <sup>23</sup>. Esto ha facilitado el acceso remoto a expertos en patología, la colaboración entre profesionales de la salud y la creación de grandes repositorios de datos patológicos <sup>4</sup>.

Sin embargo, el análisis manual de estas imágenes digitales puede ser laborioso y propenso a

---

<sup>1</sup> N. Gurcan Metin et al. “Histopathological image analysis: a review”. En: *IEEE Rev Biomed Eng* (2009), págs. 147-71147-71. DOI: 10.1109/RBME.2009.2034865.

<sup>2</sup> Min Chen et al. “Deep feature learning for medical image analysis with convolutional autoencoder neural network”. En: *IEEE Transactions on Big Data* 7.4 (2017), págs. 750-758.

<sup>3</sup> Laura Barisoni y B. Hodgins Jeffrey. “Digital pathology in nephrology clinical trials, research, and pathology practice”. En: *Current opinion in nephrology and hypertension* (2017), págs. 450-459. DOI: 10.1097/MNH.0000000000000360.

<sup>4</sup> L. Pantanowitz, JH. Sinard y WH. Henricks. “Validating whole slide imaging for diagnostic purposes in pathology”. En: *guideline from the College of American Pathologists Pathology and Laboratory Quality Center. Arch Pathol Lab Med* (2013), págs. 1710-22. DOI: 10.5858/arpa.2013-0093-CP.

errores humanos <sup>5</sup>. Aquí es donde entra en juego el aprendizaje de máquina y, más específicamente, el aprendizaje no supervisado. El aprendizaje no supervisado es una rama del aprendizaje automático que permite a los algoritmos identificar patrones y estructuras en los datos sin la necesidad de etiquetas o supervisión previa <sup>6</sup>.

En el área de la patología digital, el aprendizaje no supervisado ha mostrado un gran potencial para mejorar la eficiencia en el análisis de imágenes de tejido, permitiendo presentar resultados sin la presencia de vastos datos etiquetados <sup>7</sup>. Estos algoritmos pueden identificar características relevantes en las imágenes, agrupar muestras similares y descubrir patrones que pueden ser indicativos de ciertas enfermedades o pronósticos. Al combinar el conocimiento experto de los patólogos con las capacidades del aprendizaje no supervisado, es posible mejorar la velocidad y precisión del diagnóstico, así como descubrir nuevas relaciones y conocimientos en el campo de la patología <sup>8</sup>, con la ayuda de las técnicas supervisadas se prevé una reducción significativa en el muy influyente error humano.

El propósito de este proyecto es desarrollar un modelo de segmentación no supervisada para patología digital. El modelo utilizará imágenes histológicas no etiquetadas como datos de entrenamiento. Esto permitirá desarrollar una representación clara que facilite la identificación de patrones y anomalías en la anatomía microscópica de los tejidos. Una representación robusta de

---

<sup>5</sup> Camilo Andrés Pérez-Montiel. “Nueva clasificación molecular del carcinoma de endometrio: impacto en el diagnóstico histopatológico, tratamiento y pronóstico”. En: *MedUNAB* 24 (2022), 365–374.

<sup>6</sup> Y. Bengio, A. Courville y P. Vincent. “Representation Learning: A Review and New Perspectives”. En: *IEEE Trans Pattern Anal Mach Intell* (2013), págs. 1798-1828.

<sup>7</sup> D. Komura y S. Ishikawa. “Machine Learning Methods for Histopathological Image Analysis”. En: *Computational and structural biotechnology* (2018). DOI: 10.1016/j.csbj.2018.01.001.

<sup>8</sup> F. Steiner David et al. “Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes”. En: *Am J Surg Pathol (The American journal of surgical pathology)* (2018), págs. 1636-1646. DOI: 10.1097/PAS.0000000000001151.

estas imágenes facilitará la identificación de patrones visuales asociados a patologías y a tejidos sanos, lo cual proporcionará una base sólida para la investigación y el diagnóstico médico en este tipo de imágenes.

## 1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

La patología digital consiste en un conjunto de tecnologías computacionales para el intercambio, gestión, adquisición e interpretación de imágenes de histopatología. En el desarrollo de la patología digital, se han planteado diversas estrategias de aprendizaje automático para el análisis computarizado de imágenes de histopatología, y se ha demostrado que es posible identificar y evaluar automáticamente características visuales patológicas. Estas herramientas computacionales podrían dar soporte al proceso de diagnóstico médico, de tal manera que los patólogos realicen diagnósticos más precisos.

Dentro de los esquemas de entrenamiento para modelos de aprendizaje máquina utilizados en patología digital, se destaca el entrenamiento supervisado. En este tipo de entrenamiento, se le proveen al modelo imágenes junto con sus correspondientes anotaciones o etiquetas, realizadas por expertos. De esta manera, el algoritmo aprende una representación interna que permite distinguir entre las diferentes categorías asociadas a las imágenes. Este proceso se logra mediante la minimización del error de clasificación durante el entrenamiento, a través de algoritmos de optimización como retropropagación. Además, en tareas de clasificación que involucran imágenes, es común emplear modelos de segmentación. Estos modelos se caracterizan por dividir las imágenes en segmentos o clústeres de píxeles, que contienen patrones similares. Normalmente, el modelo entrena con un conjunto de imágenes y anotaciones realizadas a nivel de píxel. Estas anotaciones son realizadas por patólogos, en el caso de la patología digital. Este proceso de anotación requiere la delimitación de las estructuras anatómicas microscópicas, lo cual requiere bastante tiempo y es tedioso, esto debido a que manualmente los expertos deben anotar cada una de las secciones histológicas. Adicionalmente, la capacidad de generalización de la segmentación supervisada ha demostrado ser limitada si no se dispone de una gran cantidad de imágenes y etiquetas. Un ejemplo muy popular en los modelos de aprendizaje de segmentación es la red neuronal convolucional U-Net, que se desarrolló para la segmentación de imágenes biomédicas.

Por otro lado, el agrupamiento o análisis de clústeres es una técnica de aprendizaje no supervisado en la que se agrupan objetos o datos similares en conjuntos o grupos basándose en sus características comunes. El objetivo del análisis de clústeres es descubrir patrones o estructuras ocultas en los datos sin la necesidad de etiquetas o categorías predefinidas. Existen varios métodos de análisis de clústeres, cada uno con sus propias características, por ejemplo: K-means, en el cual la división de los datos se realiza en un número predefinido de clústeres de acuerdo a un centroide (punto central) más cercano en el espacio multidimensional. Por otro lado, en el análisis de clústeres jerárquico se construye una jerarquía de clústeres de forma divisiva o aglomerativa. Otro algoritmo es DBSCAN (Density-Based Spatial Clustering of Applications with Noise), que se basa en la densidad de los puntos en el espacio multidimensional de los datos. Se identifican áreas densas de puntos y considera puntos cercanos dentro de estas áreas como parte del mismo clúster. Es importante señalar que, el análisis de clústeres se centra en la agrupación de datos en múltiples clústeres, pero es pertinente resaltar que, al aplicar estos métodos en imágenes, no se tiene en cuenta la disposición o cercanía espacial. Para incorporar información del contexto espacial, se han desarrollado técnicas de segmentación no supervisada que se enfocan en dividir una imagen o conjunto de datos en segmentos o regiones significativas, incluyendo así información espacial contextual de la imagen. Esta información adicional da como resultado una representación potencialmente más robusta, ya que los algoritmos cuentan con información más pertinente al momento de generar sus representaciones. Entre los métodos de segmentación supervisada observados en la literatura, podemos encontrar métodos basados en superpíxeles, que son regiones compactas y coherentes que agrupan píxeles con características similares de color, textura y proximidad espacial. También existen enfoques basados en la segmentación de imágenes usando grafos, en el cual los píxeles se consideran nodos y se crean enlaces entre los nodos vecinos. Esta representación se puede fusionar gradualmente en función de un criterio de peso de borde para crear segmentaciones plausibles. En cuanto a las técnicas de segmentación no supervisada, basada en aprendizaje profundo, algunos enfoques como ReDO (ReDrawing

of Objects) que aprovecha la potencia de las redes generativas adversariales (GAN) para lograr resultados de segmentación de objetos precisos y fiables.

La aplicación de técnicas de segmentación no supervisada en el campo de la histopatología digital esta siendo explorado en trabajos recientes. Estos mecanismos tienen un potencial interesante para generar representaciones robustas en el área de la patología digital. Es por esta razón que el presente trabajo se enfocará en explorar la utilización de este tipo de métodos en la histología digital con el objetivo de evaluar su capacidad de generalización al representar imágenes histológicas.

## **2. OBJETIVOS**

### **2.1. OBJETIVO GENERAL**

Desarrollar un modelo de aprendizaje profundo para obtener una representación que capture la variabilidad de los patrones visuales en el conjunto de datos, incluso en presencia de ruido y distorsión en histología mediante la segmentación no supervisada de imágenes de histopatología.

### **2.2. OBJETIVOS ESPECÍFICOS**

1. Seleccionar al menos una base de datos abierta para problemas de segmentación en imágenes de histopatología.
2. Seleccionar algoritmos de segmentación supervisados y no supervisados disponibles en la literatura para evaluar su desempeño en la base de datos previamente seleccionada.
3. Implementar el modelo de aprendizaje de segmentación no supervisado para obtener una representación que capture la variabilidad de los patrones visuales en el conjunto de datos, incluso en presencia de ruido y distorsión en histopatología.
4. Evaluar la representación obtenida por el modelo de aprendizaje de segmentación no supervisado propuesto.

### 3. MARCO TEÓRICO Y TRABAJOS PREVIOS

#### 3.1. HISTOPATOLOGÍA

La histopatología es una rama de la patología que se encarga del estudio de las enfermedades a través del análisis microscópico de los tejidos biológicos. Combina la histología, que es el estudio de los tejidos en su estructura normal, con el análisis de las alteraciones que ocurren en estos tejidos debido a enfermedades.

En la histopatología, se obtienen muestras de tejidos, como biopsias o tejidos extirpados durante cirugías, y se procesan mediante técnicas de fijación, inclusión, corte y tinción. Posteriormente, se examinan estas muestras al microscopio para evaluar la estructura, la organización y las posibles alteraciones en los tejidos <sup>9</sup>.

El análisis histopatológico permite identificar cambios celulares y tisulares asociados con enfermedades, como inflamación, infección, tumores y otros trastornos. Los histopatólogos, médicos especializados en esta disciplina, interpretan los hallazgos microscópicos y proporcionan diagnósticos precisos para guiar el tratamiento y la gestión de las enfermedades <sup>10</sup>.

La histopatología es fundamental en la investigación médica, la clasificación de enfermedades, la evaluación de la efectividad de tratamientos y el pronóstico de los pacientes. También se utiliza en el seguimiento de enfermedades crónicas y en el estudio de los mecanismos de acción de fármacos y terapias.

---

<sup>9</sup> Guy Orchard y Brian Nation. *Histopathology*. Oxford University Press, USA, 2012.

<sup>10</sup> D.M. Fletcher Christopher, R. Goldblum John y D.R. Thompson Lester. *What is Histopathology?* Saunders Ltd, 2012.

### 3.2. PATOLOGÍA DIGITAL

La patología digital, también conocida como patología virtual o patología digitalizada, es una disciplina que se refiere al uso de tecnologías digitales en la práctica de la patología. Consiste en la captura, almacenamiento, análisis y visualización de imágenes digitales de muestras histológicas o citológicas, en lugar del análisis convencional basado en el examen microscópico de muestras de tejido en un portaobjetos <sup>11</sup>.

En la patología digital, las muestras histológicas se digitalizan mediante escáneres de alta resolución, generando imágenes digitales de alta calidad. Estas imágenes se almacenan en sistemas de información de laboratorio y se pueden visualizar y analizar en computadoras, lo que permite a los patólogos acceder a ellas de forma remota <sup>12</sup>.

La patología digital ofrece varias ventajas en comparación con los métodos tradicionales de patología. Permite el almacenamiento y acceso rápido a grandes volúmenes de imágenes, facilitando la búsqueda y recuperación de casos. Además, la digitalización de las muestras permite el análisis y procesamiento de imágenes utilizando herramientas de software especializadas, lo que puede ayudar a mejorar la precisión y eficiencia del diagnóstico patológico.

Otra ventaja importante de la patología digital es la posibilidad de realizar la telepatología, que implica la transmisión y visualización en tiempo real de imágenes digitales de muestras patológicas a través de internet. Esto permite la consulta y colaboración entre patólogos ubicados en diferentes lugares geográficos, lo que puede ser especialmente útil en áreas remotas o en casos que requieren la opinión de expertos.

---

<sup>11</sup> M. B. Marilyn et al. “Digital and Computational Pathology: Bring the Future into Focus”. En: *Pathology Informatics* (), pág. 10. DOI: 10.4103/2153-3539.255259.

<sup>12</sup> Martin Lindvall. “Designing with Machine Learning in Digital Pathology”. En: (2021).

### 3.3. APRENDIZAJE DE MÁQUINA

El aprendizaje de máquina, también conocido como machine learning en inglés, es un campo de estudio de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos que permiten a las máquinas aprender de forma automática a partir de datos sin necesidad de ser programadas explícitamente.

En lugar de seguir instrucciones específicas, los sistemas de aprendizaje de máquina son capaces de analizar grandes conjuntos de datos, identificar patrones y tendencias, y generar predicciones o tomar decisiones basadas en esos patrones aprendidos. Este enfoque se basa en la idea de que las máquinas pueden aprender de la experiencia y mejorar su rendimiento a medida que se les proporciona más información. El aprendizaje de máquina se puede dividir en diferentes tipos, entre los que se incluyen:

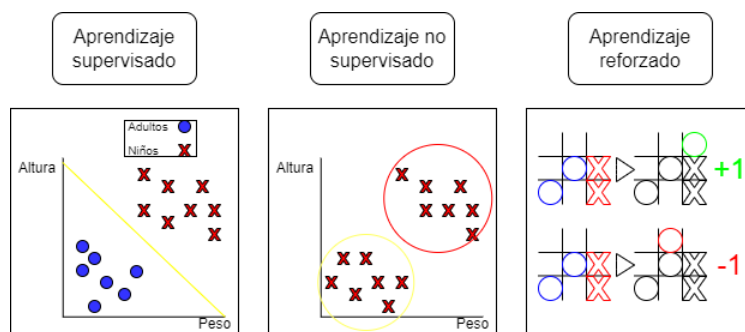
1. Aprendizaje supervisado: Se proporciona algoritmos con datos etiquetados, es decir, se les informa sobre la respuesta correcta para un conjunto de ejemplos. El algoritmo aprende a partir de estos ejemplos y luego puede predecir la respuesta correcta para nuevos datos no etiquetados <sup>13</sup>.
2. Aprendizaje no supervisado: No se proporcionan etiquetas a los algoritmos. En cambio, se les permite descubrir patrones y estructuras ocultas en los datos por sí mismos. Este tipo de aprendizaje se utiliza para agrupar datos, encontrar relaciones o detectar anomalías <sup>14</sup>.
3. Aprendizaje por refuerzo: El algoritmo aprende a través de la interacción con un entorno

---

<sup>13</sup> Trevor Hastie et al. “Overview of supervised learning”. En: *The elements of statistical learning: Data mining, inference, and prediction* (2009), págs. 9-41.

<sup>14</sup> Trevor Hastie et al. “Unsupervised learning”. En: *The elements of statistical learning: Data mining, inference, and prediction* (2009), págs. 485-585.

y recibe recompensas o castigos según las acciones que toma. A medida que interactúa con el entorno, el algoritmo aprende a tomar decisiones que maximicen las recompensas a largo plazo.<sup>15</sup>



**Figura 1.** El aprendizaje supervisado implica entrenar modelos utilizando ejemplos de entrada con sus etiquetas para hacer predicciones precisas, el aprendizaje no supervisado se realiza sin etiquetas y busca descubrir patrones ocultos en los datos y el aprendizaje reforzado se basa en la interacción con el entorno, donde los modelos reciben recompensas o penalizaciones para mejorar su rendimiento en una tarea específica.

El aprendizaje de máquina tiene aplicaciones en una amplia variedad de campos, como la medicina, la industria, el comercio electrónico, la seguridad, la conducción autónoma, la recomendación de contenido, entre otros. Es una herramienta poderosa que permite extraer conocimiento y realizar tareas complejas basadas en datos de manera más eficiente y precisa.

**3.3.1. Aprendizaje supervisado** El aprendizaje supervisado es un enfoque fundamental del aprendizaje automático en el que un algoritmo aprende a partir de un conjunto de datos que contiene pares de entrada-salida. El objetivo es enseñar al algoritmo a asignar entradas a salidas correctas, lo que le permite hacer predicciones precisas sobre datos nuevos y desconocidos. El proceso consiste en entrenar el algoritmo con ejemplos etiquetados, ajustando sus parámetros internos para minimizar la diferencia entre los resultados previstos y las etiquetas reales. Para

<sup>15</sup> M. I. Jordan y T. M. Mitchell. "Machine learning: Trends, perspectives, and prospects." En: *Science (New York, N.Y.)* (2015), págs. 255-60. DOI: 10.1126/science.aaa8415.

ello se utiliza un modelo elegido y una función de pérdida asociada que mide la precisión de la predicción. El algoritmo optimiza iterativamente su modelo utilizando técnicas como el descenso de gradiente. Tras el entrenamiento, el rendimiento del modelo se evalúa en datos de validación o prueba independientes para garantizar que puede generalizarse bien a nuevas situaciones. Este método se utiliza ampliamente en diversos campos para tareas como la clasificación de imágenes, la traducción de idiomas y el diagnóstico de enfermedades, empleando modelos como las redes neuronales y los árboles de decisión para aprender eficazmente relaciones complejas entre entradas y salidas <sup>1617</sup>.

**3.3.2. Segmentación supervisada** La segmentación supervisada es un enfoque de procesamiento de imágenes o análisis de datos en el que se utilizan etiquetas o información previa para entrenar un modelo que pueda asignar automáticamente píxeles, regiones o elementos a diferentes clases o categorías específicas. A diferencia de la segmentación no supervisada, donde el algoritmo busca patrones sin información previa, la segmentación supervisada requiere ejemplos previamente etiquetados para guiar el proceso de segmentación.

En la segmentación supervisada, se entrena un modelo utilizando un conjunto de datos de entrenamiento en el que cada elemento está etiquetado con la clase o categoría a la que pertenece. El modelo aprende a partir de estos ejemplos y busca patrones y características que distingan las diferentes clases. Luego, se utiliza el modelo entrenado para asignar etiquetas a nuevos datos no etiquetados y realizar la segmentación.

La segmentación supervisada tiene la ventaja de proporcionar resultados precisos y controlables,

---

<sup>16</sup> Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas et al. “Supervised machine learning: A review of classification techniques”. En: *Emerging artificial intelligence applications in computer engineering* 160.1 (2007), págs. 3-24.

<sup>17</sup> Trevor Hastie et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009.

ya que se basa en información previa. Sin embargo, requiere un conjunto de datos de entrenamiento adecuado y la creación de etiquetas precisas, lo que puede ser un proceso laborioso y costoso. Además, es importante tener en cuenta que la segmentación supervisada está limitada a las clases o categorías para las cuales se dispone de datos de entrenamiento etiquetados.

**3.3.3. Aprendizaje no supervisado** El aprendizaje no supervisado es un enfoque del aprendizaje automático (machine learning) que se utiliza para extraer patrones y estructuras ocultas en conjuntos de datos no etiquetados. A diferencia del aprendizaje supervisado, donde los datos de entrenamiento están previamente etiquetados con respuestas conocidas, en el aprendizaje no supervisado no se proporcionan etiquetas o información explícita sobre las categorías o clases a las que pertenecen los datos.

Algunos algoritmos de aprendizaje no supervisados son:

1. K-Means: Es un algoritmo de agrupación basado en particiones ampliamente utilizado. Su objetivo es dividir los datos en K clústers, donde cada clúster está representado por su centroide (la media de los puntos de datos en ese clúster). El algoritmo funciona de forma iterativa asignando puntos de datos al centroide más cercano y actualizando después los centroides en función de los clústers recién formados. Este proceso continúa hasta la convergencia. K-Means es conocido por su sencillez y eficacia, pero puede ser sensible a los centros iniciales de los conglomerados <sup>18</sup>.
2. DBSCAN (Density-Based Spatial Clustering of Applications with Noise): DBSCAN es un algoritmo de agrupación basado en la densidad. Agrupa los puntos de datos en función de su densidad dentro de un vecindario definido. Requiere dos parámetros: “epsilon” especifica la distancia máxima entre dos puntos que deben considerarse en el mismo vecindario, y

---

<sup>18</sup> S. Lloyd. “Least squares quantization in PCM”. En: *IEEE Transactions on Information Theory* 28.2 (1982), págs. 129-137. DOI: 10.1109/TIT.1982.1056489.

“minPts” especifica el número mínimo de puntos necesarios para formar una región densa. DBSCAN puede encontrar automáticamente clústers de diferentes formas y manejar datos ruidosos con eficacia <sup>19</sup>.

3. Hierarchical Clustering: El Clustering Jerárquico crea una jerarquía de clústers en forma de árbol, conocida como dendrograma. Para ello, fusiona o divide iterativamente los conglomerados en función de medidas de similitud. Existen dos tipos principales: Agrupamiento jerárquico aglomerativo (ascendente) y divisivo (descendente). El aglomerativo comienza con cada punto de datos como un conglomerado separado y fusiona progresivamente los conglomerados más similares, mientras que el divisivo comienza con todos los puntos de datos de un conglomerado y lo divide recursivamente en conglomerados más pequeños <sup>20</sup>.

En el aprendizaje no supervisado, los algoritmos analizan los datos en busca de similitudes, patrones y estructuras intrínsecas. Su objetivo es agrupar los datos en categorías o clústers basados en la similitud de sus características. Estas agrupaciones pueden ayudar a descubrir relaciones ocultas, identificar anomalías o reducir la dimensionalidad de los datos <sup>21</sup>.

El aprendizaje no supervisado es utilizado en una variedad de aplicaciones, como el análisis de datos, la segmentación de clientes, la detección de anomalías, la recomendación de contenido, la clasificación automática de documentos y la exploración de datos en general. Es una herramienta valiosa para descubrir conocimiento oculto en los datos y generar nuevas ideas o insights.

---

<sup>19</sup> Martin Ester et al. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. En: *Knowledge Discovery and Data Mining*. 1996.

<sup>20</sup> Leonard Kaufman y Peter Rousseeuw. *Finding Groups in Data: An Introduction To Cluster Analysis*. Ene. de 1990. DOI: 10.2307/2532178.

<sup>21</sup> M Emre Celebi y Kemal Aydin. *Unsupervised learning algorithms*. Vol. 9. Springer, 2016.

**3.3.4. Segmentación no supervisada** Por segmentación no supervisada se entiende el proceso de dividir una imagen o un conjunto de datos en regiones o segmentos significativos sin ningún conocimiento previo ni datos de entrenamiento etiquetados. Se trata de una ardua tarea de visión por ordenador cuyo objetivo es descubrir automáticamente patrones, límites o estructuras en los datos<sup>22</sup>.

En la segmentación no supervisada, el algoritmo analiza las características de los datos y busca patrones y similitudes entre ellos. Luego, asigna los datos a grupos o segmentos basados en la similitud de sus características. La idea es que los datos dentro de un grupo sean más similares entre sí que con los datos de otros grupos<sup>23</sup>.

La segmentación no supervisada es útil en diversas aplicaciones en las que se desea explorar y descubrir patrones ocultos o desconocidos en los datos, sin información previa sobre las categorías o grupos a los que pueden pertenecer. Algunas aplicaciones y casos de uso comunes de la segmentación no supervisada incluyen:

1. Segmentación de marketing: Las técnicas de segmentación no supervisada pueden ayudar a identificar distintos segmentos de clientes en función de su comportamiento de compra, preferencias o información demográfica. Esta información puede utilizarse para campañas de marketing dirigidas y experiencias de cliente personalizadas.
2. Análisis de imágenes y vídeos: La segmentación no supervisada se utiliza para identificar regiones de interés, objetos o límites en imágenes y vídeos. Ayuda en tareas como el reconocimiento de objetos, la recuperación de imágenes, el análisis de movimiento y la

---

<sup>22</sup> Jianbo Shi y J. Malik. “Normalized cuts and image segmentation”. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.8 (2000), págs. 888-905. DOI: 10.1109/34.868688.

<sup>23</sup> Iman Aganj et al. “Unsupervised Medical Image Segmentation Based on the Local Center of Mass”. En: *Scientific Reports* (29 August 2018).

comprensión de escenas.

3. Detección de anomalías: Al identificar patrones normales dentro de un conjunto de datos, la segmentación no supervisada puede ayudar a detectar anomalías o valores atípicos que se desvían del comportamiento esperado. Esto resulta útil en diversos ámbitos, como la detección de fraudes, la detección de intrusiones en la red y el control de calidad.

## 4. MATERIALES

### 4.1. GlaS CHALLENGE

El dataset GlaS (Gland Segmentation in Colon Histology Images) es una base de datos construida para entrenar modelos de segmentación de glándulas de colon <sup>24</sup>. Incluye una gran colección de imágenes histológicas de alta resolución de muestras de tejido colon.

El dataset GlaS sirve como base ideal para investigaciones detalladas sobre las técnicas de segmentación de imágenes con una aplicación específica en el campo médico en este estudio. Para explorar métodos, algoritmos y modelos novedosos de segmentación automatizada de glándulas, proporciona un conjunto de datos completo y real. En última instancia, esto ayudará a avanzar en el diagnóstico médico, el tratamiento y la comprensión de las enfermedades relacionadas con el colon. Este tipo de investigación es una contribución al campo más amplio del análisis de imágenes médicas y tiene el potencial de mejorar los resultados y servicio de atención médica brindado al paciente.

### 4.2. PARTICIÓN DEL DATASET

La necesidad de organizar adecuadamente los conjuntos de datos para su posterior procesamiento es común en el proceso de administrar y trabajar con ellos. La asignación de un fold a cada fila del conjunto de datos es una estrategia efectiva, lo que implica categorizar cada instancia en una de varias particiones o conjuntos predefinidos. En este caso, se han establecido tres categorías principales para nuestra base de datos: “train”, “testA” y “testB”.

---

<sup>24</sup> Korsuk Sirinukunwattana et al. “Gland segmentation in colon histology images: The glas challenge contest”. En: *Medical image analysis* 35 (2017), págs. 489-502.

El conjunto de datos cuenta con 16 pacientes en total, entre todos los pacientes hay 74 muestras benignas y 91 malignas.

El conjunto de datos para el entrenamiento contiene un 80 %, el conjunto de datos de validación contiene un 10 % y finalmente decidimos hacer una distribución homogénea (entre muestras benignas y malignas) en nuestro conjunto de test, para así poder evaluar los resultados generados de manera justa. Para lograr esto, se hizo un muestreo estratificado de nuestro conjunto de datos a nivel de cada paciente, de eso sacamos 3 pacientes cuyas muestras contenían la misma proporción de datos tanto benignos como malignos, estos representaban el 10 % de las muestras totales.

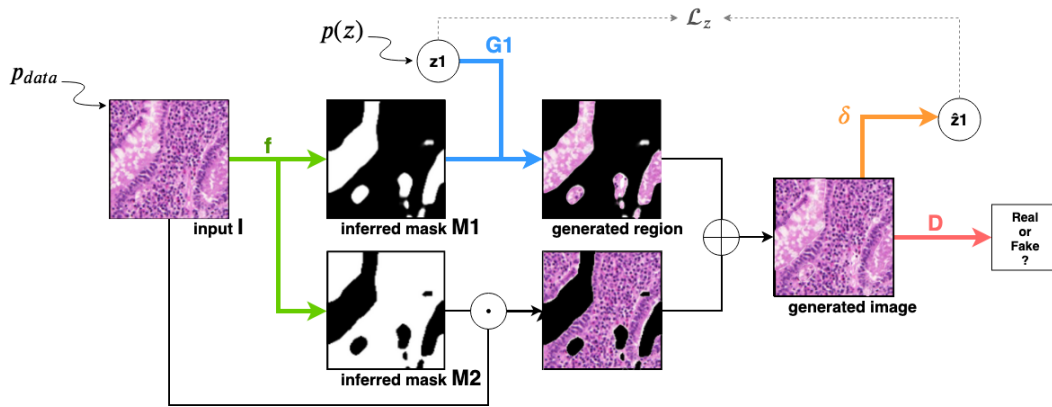
Asignar una categoría a cada fila del conjunto de datos se hace para optimizar la carga de datos durante el entrenamiento y la evaluación de modelos. Por ejemplo, cuando estamos desarrollando y ajustando modelos, podemos inicialmente cargar solo los datos de entrenamiento al separar los datos de entrenamiento de las test A y B. Esto acelera el proceso de desarrollo y ahorra recursos computacionales.

Podemos cargar solo los datos correspondientes a esas categorías cuando llega el momento de evaluar el rendimiento del modelo en los conjuntos de test A o B, lo que facilita una carga de datos rápida y eficiente. Esto es especialmente útil cuando hay grandes conjuntos de datos y no es necesario cargar todo el conjunto de datos en memoria en todo momento.

## 5. MÉTODOS

### 5.1. ARQUITECTURA ReDO: UNSUPERVISED OBJECT SEGMENTATION BY REDRAWING

ReDO es una técnica innovadora de segmentación de objetos no supervisada cuyo objetivo es separar los objetos de interés del fondo en imágenes sin anotaciones previas ni etiquetado manual. Utiliza algoritmos de aprendizaje profundo y aprovecha la potencia de las redes generativas adversariales (GAN) para lograr resultados de segmentación de objetos precisos y fiables. Al explorar la estructura y los patrones inherentes a las imágenes <sup>25</sup>, ReDO proporciona una solución eficiente y efectiva para la difícil tarea de la segmentación de objetos, permitiendo segmentar imágenes sin la presencia de vastos datos etiquetados.



**Figura 2.** ReDO comienza con una segmentación inicial y un modelo generativo que aprende la distribución de las segmentaciones. Luego, se realiza un proceso iterativo en donde las segmentaciones son redibujadas basándose en las imágenes originales y las segmentaciones generadas por el modelo <sup>26</sup>. Este proceso se repite para mejorar la calidad de la segmentación sin la necesidad de etiquetas.

La función de costo general de ReDO es una combinación de la pérdida adversaria, la pérdida

<sup>25</sup> Mickaël Chen, Thierry Artières y Ludovic Denoyer. “Unsupervised Object Segmentation by Redrawing”. En: *Advances in neural information processing systems* 32 (27 May 2019).

de reconstrucción y la pérdida de máscara.

$$L = L_{adv} + L_{rec} + \lambda L_{mask}$$

Donde:

- $L_{adv}$  es la pérdida adversaria.
- $L_{rec}$  es la pérdida de reconstrucción.
- $L_{mask}$  es la pérdida de máscara.
- $\lambda$  es un hiperparámetro que controla el peso de la pérdida de máscara.

La pérdida adversarial es una acción entre el generador y el discriminador. El generador intenta crear imágenes que tengan más probabilidades de engañar al discriminador, mientras que el discriminador intenta distinguir entre imágenes reales y falsas. La pérdida adversarial se define de la siguiente manera:

$$L_{adv} = -\log(D(G(I)))$$

Donde:

- $G(I)$  es la imagen redibujada creada por el generador a partir de la imagen de entrada  $I$ .
- $D(I)$  es la salida del discriminador para la imagen  $I$ .

La pérdida de reconstrucción se utiliza para entrenar al generador para que reconstruya la imagen original a partir de la imagen redibujada. Cuando mencionamos la imagen redibujada, hacemos referencia a la imagen que es generada reemplazando las partes correspondientes a las máscaras inferidas. Esto se hace mediante el uso de una red de reconstrucción entrenada para minimizar la diferencia entre la imagen original y la imagen reconstruida. La pérdida de reconstrucción se define de la siguiente manera:

$$L_{rec} = ||I - R(G(I))||^2$$

Donde:

- R(I) es la imagen reconstruida creada por la red de reconstrucción a partir de la imagen redibujada G(I)

La pérdida de máscara se utiliza para alentar al generador a crear imágenes redibujadas que tengan una gran superposición con la máscara de segmentación de verdad sobre el terreno. La pérdida de máscara se define de la siguiente manera:

$$L_{Mask} = ||M - M'||^2$$

Donde:

- M es la máscara de segmentación de verdad fundamental
- M' es la máscara de segmentación prevista

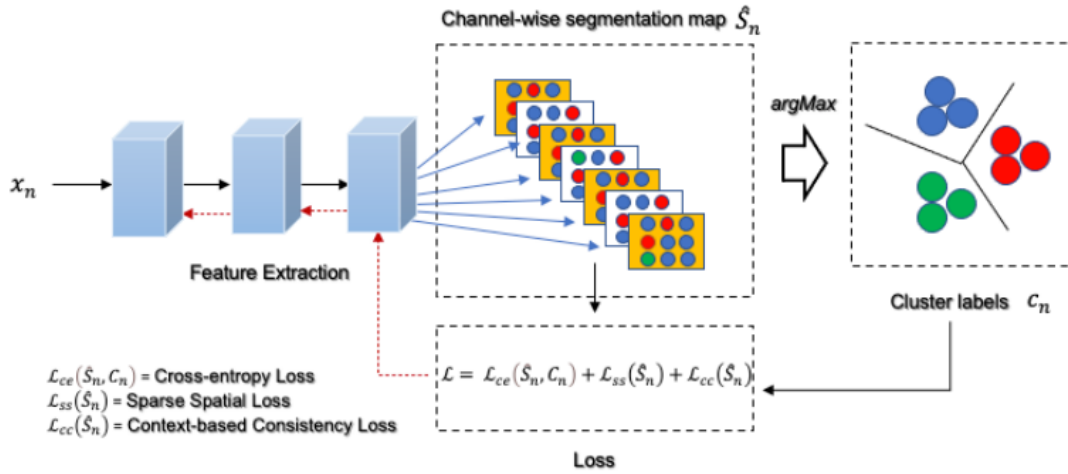
## 5.2. SGSCN: SPATIAL GUIDED SELF-SUPERVISED CLUSTERING NETWORK

El Spatial Guided Self-supervised Clustering Network (SGSCN) es un método propuesto para la segmentación de imágenes médicas. Fue desarrollado por Euijoon Ahn, Dagan Feng y Jinman Kim. <sup>27</sup>.

El SGSCN aborda el reto de los métodos de aprendizaje profundo supervisado para la segmentación de imágenes médicas, que a menudo requieren grandes cantidades de datos de entrenamiento etiquetados. Dado que la cantidad de datos anotados no ha aumentado proporcionalmente a

---

<sup>27</sup> Euijoon Ahn, Dagan Feng y Jinman Kim. “A spatial guided self-supervised clustering network for medical image segmentation”. En: (2021), págs. 379-388.



**Figura 3.** En primer lugar, la red recibe datos de entrada no etiquetados, como imágenes o vectores de características. A continuación, emplea mecanismos de orientación espacial, como la atención espacial o las matrices de afinidad. Por último, la red refina los resultados de la agrupación de forma iterativa, aprovechando la orientación espacial para mejorar aún más la calidad de la agrupación.

la expansión de los repositorios de datos de imágenes médicas, el SGSCN fue diseñado para superar esta limitación <sup>27</sup>.

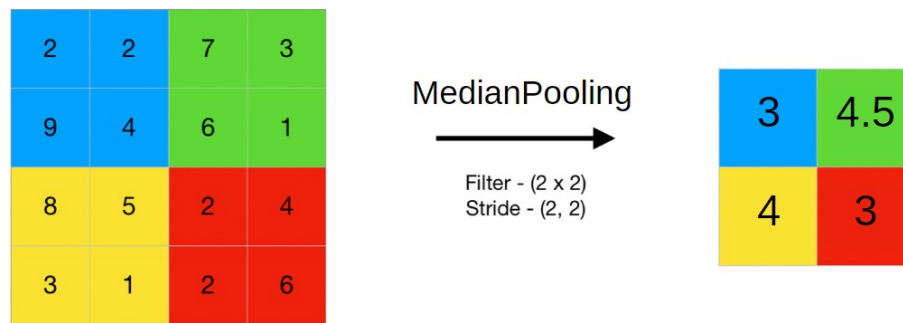
El SGSCN introduce múltiples funciones de pérdida para ayudar a agrupar los píxeles de la imagen que están conectados espacialmente y tienen representaciones de características similares. Aprende las representaciones de las características y la asignación de agrupación de cada píxel de extremo a extremo a partir de una única imagen.

Además, el SGSCN propone una pérdida de consistencia basada en el contexto que delinea mejor la forma y los límites de las regiones de la imagen. Esta función de pérdida obliga a que todos los píxeles pertenecientes a un conglomerado estén espacialmente cerca del centro del conglomerado <sup>27</sup>.

El método se evaluó en 2 conjuntos de datos públicos de imágenes médicas y se comparó con los

métodos de agrupación convencionales y autosupervisados existentes. Los resultados experimentales mostraron que el SGSCN era el más preciso para la segmentación de imágenes médicas <sup>27</sup>.

### 5.3. MEDIAN POOLING

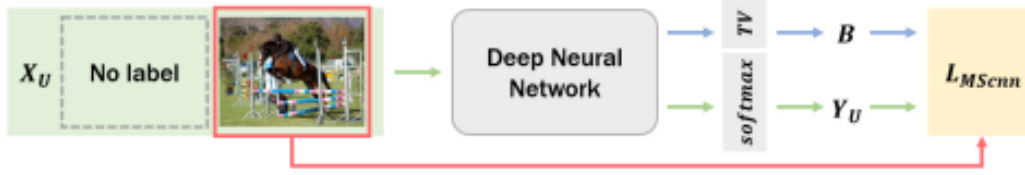


**Figura 4.** Ejemplo de pooling layer con un tamaño de agrupación de 2x2.

Median pooling en la segmentación de imágenes consiste en deslizar una ventana por una imagen o mapa de características. En cada posición, la mediana del valor del píxel dentro de la vecindad de la ventana sustituye al valor del píxel original. Este proceso ayuda a reducir las dimensiones espaciales al tiempo que preserva detalles importantes como los bordes y las características finas, lo que lo hace ventajoso para tareas como la segmentación de objetos en imágenes. A diferencia de la agrupación de promedios, la agrupación de medianas es menos sensible a los valores atípicos, lo que contribuye a su eficacia a la hora de mantener información crucial para una segmentación precisa. Sin embargo, puede introducir una ligera sobrecarga computacional debido al cálculo de la mediana.

### 5.4. MUMFORD-SHAH LOSS

Mumford-Shah Loss (MS) integra la fidelidad de los datos con respecto a la imagen de entrada, y los términos de regularización para guiar la segmentación de imágenes en el aprendizaje



**Figura 5.** Representación de la función MS Loss para la segmentación no supervisada de imágenes, tomada de <sup>28</sup>.

profundo.

Al formular la segmentación como un problema de optimización, equilibra la fidelidad a los datos de la imagen con la suavidad de los límites de segmentación. Durante el entrenamiento, las redes neuronales convolucionales (CNN) se entrenan para minimizar la pérdida de MS, fomentando la producción de segmentaciones que representan con precisión el contenido de la imagen al tiempo que mantienen límites coherentes. Este enfoque fomenta la preservación de los límites, la suavidad y la robustez, lo que lo hace especialmente eficaz para tareas en las que es crucial una delineación precisa de los objetos <sup>29</sup>. Al incorporar el conocimiento previo de la estructura de la imagen en el proceso de entrenamiento, la función de pérdida MS mejora el rendimiento de los modelos de aprendizaje profundo para la segmentación de imágenes, lo que conduce a mejores resultados de segmentación.

La función de pérdida Mumford–Shah está definida por la siguiente ecuación:

$$L_{MS_{cnn}(\Theta, x)} = \sum_{n=1}^N \int_{\Omega} |x(r) - c_n|^2 y_n(r) dr + \lambda \sum_{n=1}^N \int_{\Omega} |\nabla y_n(r)| dr$$

<sup>29</sup>.

---

<sup>29</sup> Boah Kim y Jong Chul Ye. “Mumford–Shah loss functional for image segmentation with deep learning”. En: *IEEE Transactions on Image Processing* 29 (2019), págs. 1856-1866.

Donde:

- $x(r)$  es la entrada.
- $y_n(r) := y_n(r; \Theta)$  es la salida de la capa softmax.
- $c_n := c_n(\Theta) = \frac{\int_{\Omega} x(r)y_n(r;\Theta)dr}{\int_{\Omega} y_n(r;\Theta)dr}$

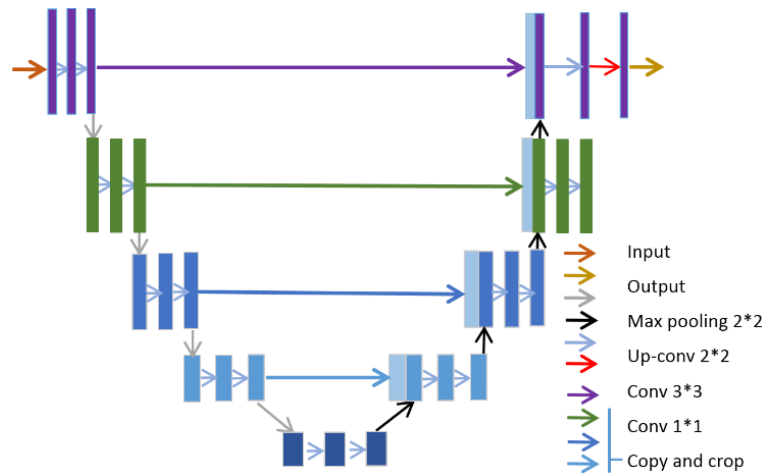
## 5.5. APRENDIZAJE SUPERVISADO: ARQUITECTURA U-NET

U-Net es una arquitectura de red neuronal convolucional (CNN) ampliamente utilizada para tareas de segmentación semántica en visión por ordenador. Fue presentada por primera vez por Olaf Ronneberger, Philipp Fischer y Thomas Brox en su artículo de 2015 titulado “U-Net: Convolutional Networks for Biomedical Image Segmentation”<sup>30</sup>.

La arquitectura U-Net aborda el desafío de segmentar con precisión imágenes en diferentes clases o regiones de interés, por lo que es particularmente popular en el dominio biomédico. Se ha aplicado con éxito a tareas como la segmentación celular, la segmentación de órganos y el análisis de imágenes médicas. Además, U-Net resulta útil en otros campos, como el análisis de imágenes de satélite y la conducción autónoma de automóviles.

---

<sup>30</sup> Olaf Ronneberger, Philipp Fischer y Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer, 18 November 2015.



**Figura 6.** Arquitectura U-Net para la segmentación semántica. Se compone de un codificador, que captura características de la imagen en diferentes niveles de abstracción usando capas convolucionales y max-pooling, y de un decodificador, compuesto por capas de convolución transpuesta y de concatenación que generan una máscara con las dimensiones de la imagen original <sup>31</sup>.

U-Net consta de dos componentes principales: la ruta de contracción (codificador) y la ruta de expansión (decodificador). La ruta de contracción de U-Net se compone de capas convolucionales seguidas de operaciones de max pooling. Este método capta las características de alta resolución y bajo nivel reduciendo gradualmente las dimensiones espaciales de la imagen de entrada. Por otro lado, la ruta de expansión aplica convoluciones transpuestas, también conocidas como deconvoluciones o capas de sobremuestreo, se utilizan para sobremuestrear los mapas de características de la ruta de codificación en la ruta de expansión U-Net <sup>32</sup>.

---

<sup>32</sup> O. Ronneberger, P.Fischer y T. Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. En: LNCS 9351 (2015). (available on arXiv:1505.04597 [cs.CV]), págs. 234-241.

## 5.6. MÉTRICAS DE EVALUACIÓN

En esta sección se discutirá a profundidad las métricas de validación que fueron utilizadas para estimar la reconstrucción de los vectores embebidos, para esto se tuvo en cuenta los siguientes términos:

- **Intersection over Union (IoU):** Intersección sobre unión, también conocida como IoU, es una métrica ampliamente utilizada en tareas de visión por computadora y aprendizaje automático, especialmente en el ámbito de la detección de objetos y la segmentación de imágenes. Se utiliza para evaluar la precisión de la localización o segmentación de objetos midiendo el solapamiento entre las cajas o regiones predichas y las regiones verdaderas.

El cálculo de la métrica IoU se basa en la relación entre el área de solapamiento entre las regiones predichas y las regiones de la verdad en el terreno, así como el área de su unión. Se puede expresar matemáticamente como:

$$IoU = \frac{M \cap R}{M \cup R}$$

Dónde:

M = Máscara predicha

R = Máscara real

- **Mean Intersection over Union (MIOU):** La intersección media sobre la unión (MIOU) es una métrica utilizada habitualmente para evaluar el rendimiento de los modelos de segmentación semántica en tareas de visión por ordenador. Mide el solapamiento entre las máscaras de segmentación predichas y las máscaras reales para cada clase en un conjunto de datos de imágenes.

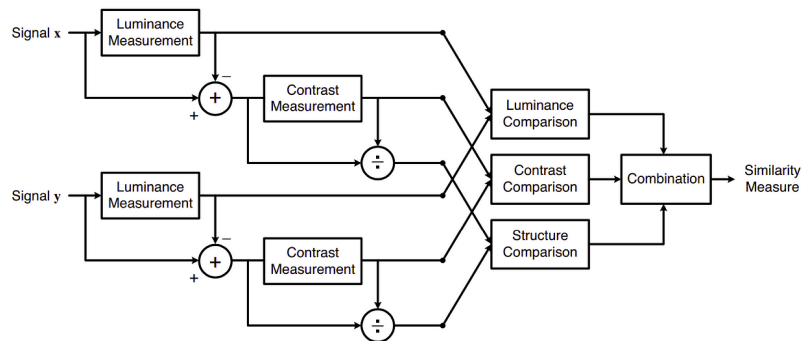
Para calcular el MIOU, primero se calcula la intersección sobre la unión (IoU), también conocida como índice de Jaccard, para cada clase. El IoU se define como la relación entre

la intersección de las máscaras de predicción y de la verdad sobre el terreno y su unión. A continuación, se calcula la media de IoU de todas las clases para obtener la puntuación MIoU:

$$MIoU = \frac{\sum_{i=1}^n IoU_i}{N}$$

Dónde N es el número de total de clases.

- **Structural Similarity Index Measure(SSIM):**



**Figura 7.** Esquema del sistema de medición de la similitud estructural (SSIM)<sup>33</sup>.

SSIM son las siglas de Structural Similarity Index Measure (medida del índice de similitud estructural). Es una métrica utilizada para medir la similitud entre dos imágenes. Tiene en cuenta tres componentes de una imagen: luminancia, contraste y estructura. El SSIM compara patrones locales de intensidades de píxeles que han sido normalizados para la luminancia y el contraste. Suele utilizarse en tareas de procesamiento de imágenes y visión por ordenador para evaluar la calidad de imágenes reconstruidas o comprimidas, así como para evaluar el rendimiento de algoritmos de restauración de imágenes<sup>34</sup>.

---

<sup>34</sup> Zhou Wang et al. “Image quality assessment: from error visibility to structural similarity”. En: *IEEE transactions on image processing* 13.4 (2004), págs. 600-612.

El SSIM suele calcularse utilizando una ventana deslizante para comparar las partes correspondientes de los píxeles de dos imágenes. El resultado es un valor entre -1 y 1, donde 1 indica una similitud perfecta entre las dos imágenes. Los valores SSIM más altos implican una mayor similitud.

La fórmula para SSIM entre dos imágenes  $x$  y  $y$  suele representarse como:

$$SSIM(x, y) = \left( \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \right)$$

Donde:

- $\mu_x$  y  $\mu_y$  son los valores medios de los píxeles de las imágenes  $x$  e  $y$  respectivamente.
- $\sigma_x^2$  y  $\sigma_y^2$  son las varianzas  $x$  e  $y$  respectivamente.
- $\sigma_{xy}$  es la covarianza de  $x$  e  $y$ .
- $c_1$  y  $c_2$  son constantes añadidas para evitar errores de división por cero y estabilizar la fórmula.

Para nuestro proceso de evaluación, los valores de las constantes  $c_1$  y  $c_2$  fueron 0.01 y 0.03 respectivamente.

## 6. CONFIGURACIÓN EXPERIMENTAL

Una parte importante de la investigación de análisis de imágenes médicas son las representaciones claras de imágenes histológicas. Una variedad de aplicaciones médicas, como la detección temprana de enfermedades, la evaluación de la eficacia del tratamiento y la investigación científica, requieren la capacidad de extraer características relevantes y significativas de estas imágenes. Este capítulo examinará el aprendizaje supervisado como punto de partida para establecer un punto de referencia sólido para el análisis de imágenes histológicas e indagará en la configuración de las alternativas no supervisadas.

El enfoque presentado en este trabajo trata del reconocimiento de características visuales, implementando y evaluando métodos de segmentación no supervisada.

### 6.1. PREPROCESAMIENTO DE LAS IMÁGENES

Se descubrió durante una revisión exhaustiva de las imágenes que presentaban una variabilidad significativa en términos de dimensiones. Se encontraron imágenes con diferentes anchos (de 567 a 775 píxeles) y alturas (de 530 a 522 píxeles). Sin embargo, se encontró que las dimensiones estandarizadas de la mayoría de las imágenes eran de 775 píxeles de ancho por 522 píxeles de alto.

Esta discrepancia en las dimensiones de las imágenes puede deberse a una variedad de razones, incluidas fuentes de datos diferentes, cámaras con configuraciones diferentes o procesos de redimensionamiento previos. La identificación de esta variabilidad es un paso crucial en la gestión de datos porque puede afectar el proceso de procesamiento y análisis de imágenes.

En proyectos de visión por computadora, es común realizar un preprocesamiento para estandarizar las dimensiones de las imágenes, lo que implica ajustarlas a un tamaño común, por

esta razón las imágenes fueron reescaladas a una dimensión de 512 x 512, para garantizar un análisis coherente y preciso. Dado que todas las imágenes tienen la misma resolución y relación de aspecto, este método facilita la comparación y el procesamiento posterior de las imágenes.

En resumen, identificar la variabilidad de las dimensiones de las imágenes es un paso crucial en la preparación de datos para proyectos de visión por computadora, y la estandarización de las dimensiones puede ser una estrategia efectiva para garantizar un análisis coherente y resultados más precisos.

## **6.2. DETALLES EXPERIMENTALES**

Una vez se ha completado el proceso de entrenamiento del modelo ReDO, el siguiente paso en nuestra metodología es realizar inferencias sobre las imágenes del dataset de evaluación, y evaluar el desempeño. Los pasos que serán aquí descritos son meramente necesarios para el modelo ReDO, debido a que este está condicionado a realizar segmentaciones meramente con imágenes de  $128 \times 128$ , entonces para adaptar las imágenes de entrada al formato requerido por el modelo de segmentación, se implementa un proceso de preprocesamiento adicional en este caso, que consiste en un método de recorte determinístico aleatorio de  $128 \times 128$  píxeles. Este método elimina los píxeles excesivos de la imagen original y solo mantiene la región central importante. Posteriormente, se utiliza el modelo entrenado para crear una máscara predictiva que represente las áreas segmentadas de la imagen. Los resultados finales de segmentación se obtienen al fusionar estas máscaras predictivas. Este método de preprocesamiento proporciona una metodología controlada y sistemática para preparar las imágenes de entrada durante la fase de inferencia. Esto garantiza la coherencia en el tamaño y la calidad de los datos procesados, lo que contribuye a la precisión y confiabilidad de los resultados obtenidos durante el proceso de segmentación de imágenes.

Los parámetros utilizados en primera instancia para el modelo de segmentación ReDO fueron:

Parámetros	Valores
NX	3
NZ	64
NF	64
NRESM	3
NMASKS	2
initOrthoGain	0.8
wdcay	1e-4
wrecZ	5
lrG,lrD, lrZ	1e-4
lrM	1e-5
nIteration	20000
ITERSTEPS	500
NTEST	5
BATCHSIZE	20

**Tabla 1.** Hiperparámetros utilizados con el método ReDO

**SIZEX = 128:** El tamaño de redimensionamiento se estableció en 128. Este valor se eligió para equilibrar la eficiencia computacional y el nivel de detalle en las imágenes.

**NX = 3:** El número de canales se estableció en 3, correspondiente al espacio de color RGB.

**NZ = 64:** Utilizamos una dimensionalidad del espacio latente de 64.

**NF = 64:** El número de filtros se estableció en 64, proporcionando un buen equilibrio entre la complejidad del modelo y la eficiencia computacional.

**NRESM = 3:** Utilizamos 3 bloques residuales en nuestra arquitectura del modelo.

**NMASKS = 2:** El número de máscaras se estableció en 2, permitiendo una mejor segmentación de objetos.

**initOrthoGain = 0.8:** La ganancia ortogonal se estableció en 0.8, lo que ayudó a estabilizar el proceso de entrenamiento.

**wdecay = 1e-4**: Se utilizó una decadencia de peso de 1e-4 para prevenir el sobreajuste.

**wrecZ = 5**: El peso para la pérdida de reconstrucción en el espacio latente se estableció en 5.

**lrG = lrD = lrZ = 1e-4** y **lrM = 1e-5**: Estas tasas de aprendizaje se eligieron después de varias rondas de experimentación, con el objetivo de garantizar un aprendizaje estable y eficiente.

**nIteration = 20000** y **ITERSTEPS = 500**: Entrenamos nuestro modelo durante 20,000 iteraciones con 500 pasos por iteración.

**NTEST = 5**: Realizamos pruebas en cinco conjuntos de datos diferentes para validar nuestros resultados.

**BATCHSIZE = 20**: Se utilizó un tamaño de lote de 20 para el entrenamiento.

Así mismo, se probó otro enfoque con el modelo de segmentación Spatial Guided Self-supervised Clustering Network (SGSCN), esto con el objetivo de tener distintas referencias para comparar nuestros resultados cualitativos y cuantitativos.

Los hiperparámetros que estuvieron involucrados en nuestros experimentos fueron los siguientes:

Parámetros	Valores
nChannel	100
maxIter	50
minLabels	3
LR	0.1
nConv	2
STEPSIZECE	1
STEPSIZES	5
CENTER	False

**Tabla 2.** Hiperparámetros utilizados en el método SGSCN

**nCHANNEL = 100:** Representa el número de canales. Controla el número de canales en las capas convolucionales de la red de segmentación.

**maxIter = 50:** Especifica el número de iteraciones máximas. Determina el número máximo de iteraciones de entrenamiento durante el proceso de optimización.

**minLabels = 3:** Establece el número mínimo de etiquetas. Especifica el número mínimo de etiquetas necesarias en el mapa de segmentación.

**LR = 0.1:** Significa tasa de aprendizaje. Determina la velocidad a la que se actualizan los parámetros del modelo durante la optimización (por ejemplo, utilizando el descenso por gradiente estocástico).

**nCONV = 2:** Representa el número de capas convolucionales. Controla el número de capas convolucionales en la red de segmentación.

**STEPSIZECE = 1:** Establece el tamaño del paso para la pérdida de entropía cruzada. Determina el peso de la pérdida de entropía cruzada durante el entrenamiento.

**STEPSIZESS = 5:** Establece el tamaño del paso para la pérdida espacial dispersa. Determina el peso de la pérdida espacial dispersa, que fomenta la suavidad espacial en las predicciones de segmentación.

**CENTER = FALSE:** Es un indicador booleano. Si se especifica al ejecutar el script, activa el uso de la pérdida de coherencia basada en el contexto durante el entrenamiento.

Durante los experimentos aplicando una capa de Median Pooling, los parámetros utilizados en ambos modelos de segmentación fueron:

**KERNELSIZE = 5:** Tamaño del núcleo de agrupación.

**STRIDE = 1:** Distancia de agrupamiento.

**PADDING = 0:** Pool padding.

**SAME = True:** Anula el relleno y aplica el mismo relleno.

Parámetros	Valores
KERNELSIZE	5
STRIDE	1
PADDING	0
SAME	True

**Tabla 3.** Hiperparámetros utilizados con Median Pooling

Así mismo, para la implementación de la Mumford–Shah Loss, se utilizaron los siguientes parámetros:

**OUTPUT:** Utilizamos la máscara generada para este tensor que representa la salida del conjunto de niveles predicho por la red neuronal.

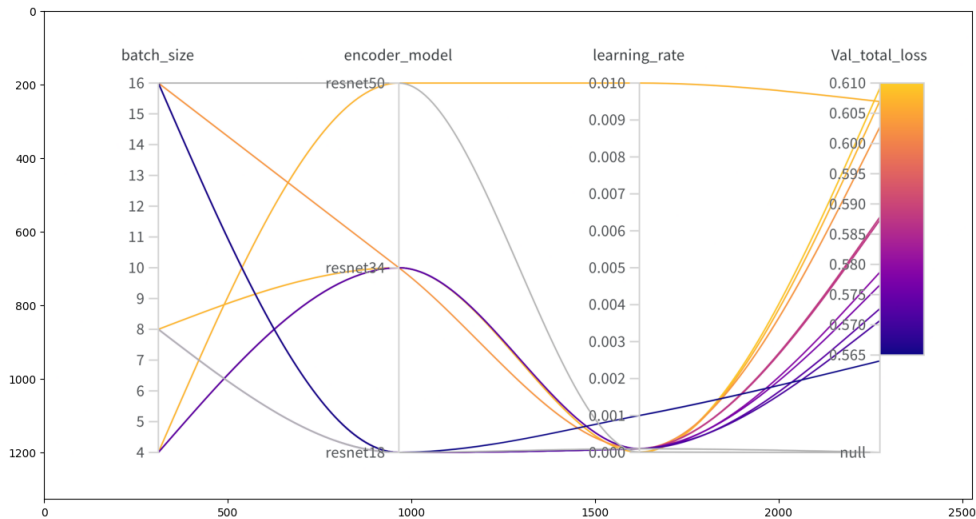
**TARGET:** Utilizamos la máscara verdadera este tensor que representa el conjunto de niveles verdaderos.

**INPUT:** Utilizamos la máscara generada para este tensor que representa la entrada a la función de pérdida de gradiente.

### 6.3. ARQUITECTURA U-NET

En cuanto nuestro aprendizaje supervisado, encontramos un conjunto de hiperparámetros que demostraron ser extremadamente efectivos en nuestro entorno de entrenamiento para optimizar el rendimiento de nuestro modelo. Nos decidimos por un valor de 16 para el tamaño de lote, lo que nos permitió equilibrar la velocidad de entrenamiento con la utilización eficiente de recursos computacionales.

Optamos por el modelo de codificador ResNet-18, que ha sido comprobado y efectivo. Su capacidad para mantener un número relativamente bajo de parámetros mientras extrae características importantes de las imágenes de entrada lo llevó a esta elección. La implementación de ResNet-18 en nuestro modelo de referencia generó alrededor de 14 millones de parámetros, específicamente 14,328,209, lo que garantiza una representación de datos sólida sin aumentar la complejidad.



**Figura 8.** Representación de la pérdida de validación con respecto a los hiperparámetros de `batch_size`, `encoder_model` y `learning_rate`. De las diferentes combinaciones se puede concluir que la mejor combinación, es decir la que representa el la menor pérdida de validación, es utilizar un `batch_size` de 16, con el `encoder_model` `resnet18` y una tasa de aprendizaje de 0.001.

Establecimos un valor de 0.001 para la tasa de aprendizaje, que demostró ser adecuada para el

proceso de entrenamiento de nuestro modelo. Esta tasa de aprendizaje permitió un equilibrio entre la convergencia rápida en las primeras etapas del entrenamiento y la estabilidad en las etapas posteriores.

La arquitectura U-Net sirvió como base para el modelo de referencia. Esta arquitectura es ampliamente utilizada en aplicaciones de procesamiento de imágenes médicas y más porque es conocida por su capacidad para tareas de segmentación de imágenes. El modelo U-Net configurado tenía aproximadamente 14 millones de parámetros, lo que lo hacía apropiado para nuestra tarea.

**ReDO base** Experimento utilizando ReDO base, tal como se describió en la Subsección 7.1

**SGSCN base** Experimento utilizando SGSCN base, tal como se describió en la Subsección 7.2.

**ReDO\_mp** Experimento con la arquitectura ReDO con capa de MedianPooling que permite reducir la resolución de las características extraídas manteniendo el valor mediano en una ventana deslizable.

**SGSCN\_mp** Experimento con la arquitectura SGSCN con capa de MedianPooling que permite reducir la resolución de las características extraídas manteniendo el valor mediano en una ventana deslizable.

**ReDO\_ms** Experimento con la arquitectura ReDO que incluye la función de costo Mumford-Shah.

**SGSCN\_ms** Experimento con la arquitectura SGSCN que incluye la función de costo Mumford-Shah.

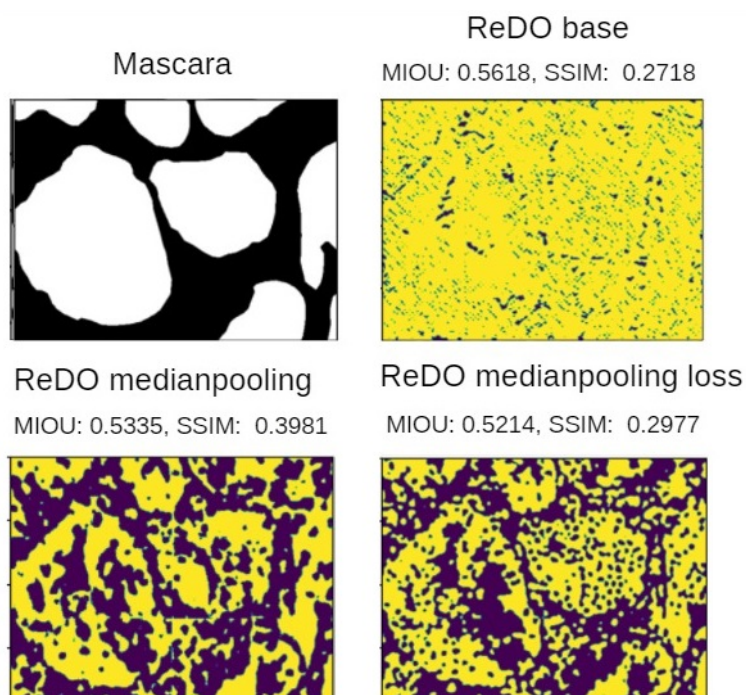
**All\_gland** Experimento con modelo el cual asume que las todas imágenes introducidas constan totalmente de una glándula, es decir, una matriz que contiene 1s en su totalidad. A través de este método se establece una cota inferior para el desempeño de los modelos.

**U-Net** Experimento utilizando U-Net base, tal como se describio en la Subsección 7.3.

## 7. EVALUACIÓN Y RESULTADOS

### 7.1. ReDO

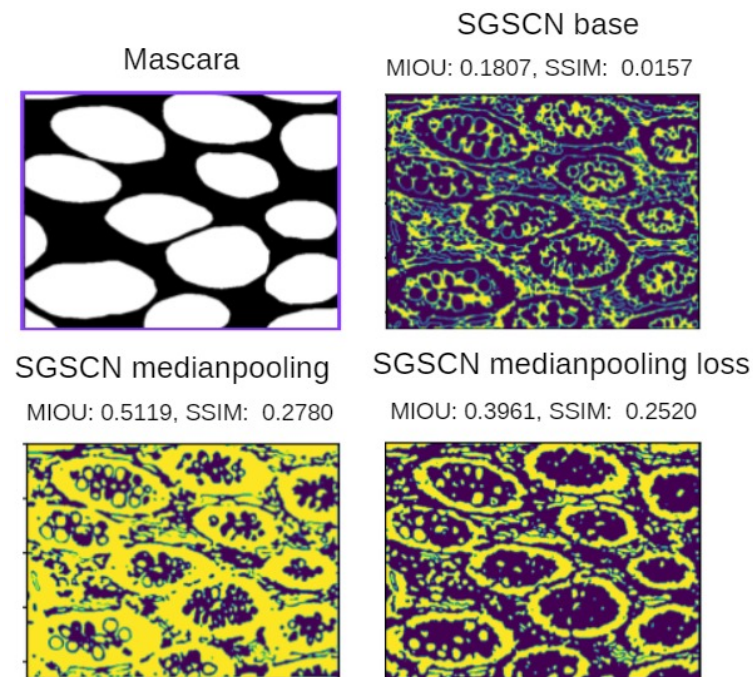
Después de completar el proceso de entrenamiento del modelo ReDO, el siguiente paso importante en nuestro flujo de trabajo es hacer conclusiones sobre nuestras imágenes. Utilizamos el modelo entrenado para crear una máscara predictiva que represente las áreas segmentadas de la imagen. Los resultados finales de segmentación se obtienen al fusionar estas máscaras predictivas. Resultados cualitativos de esta inferencia se pueden observar en la Figura 9.



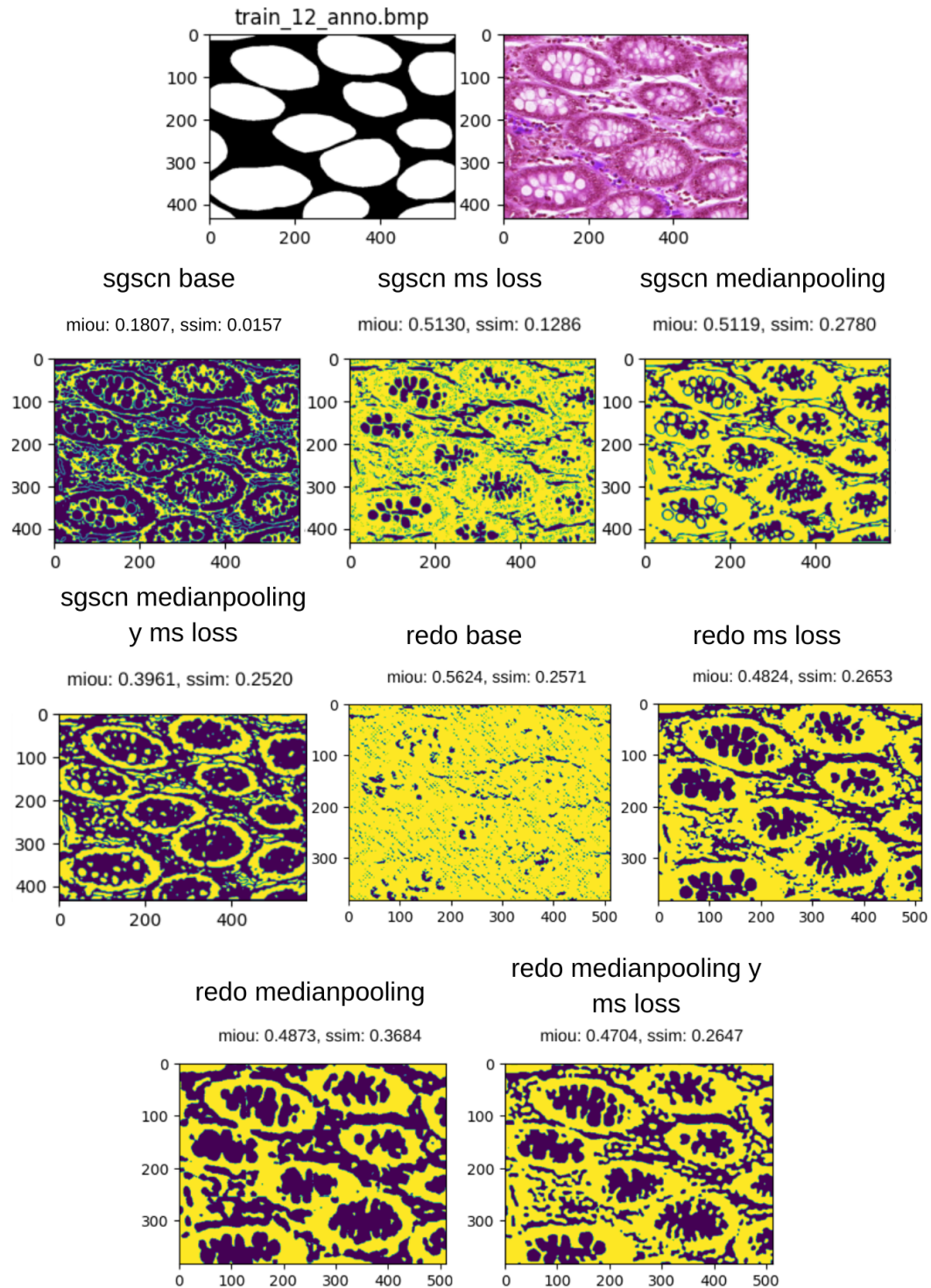
**Figura 9.** Segmentación no supervisada con el modelo ReDO, en la cual el color amarillo hace referencia a la segmentación de una glándula y el azul al fondo.

## 7.2. SGSCN

Por otro lado, la implementación del modelo SGSCN no requirió de ningún tipo de inferencia por separado para luego hacer la unión y así poder generar resultados. Este modelo nos permitió generar los resultados cualitativos observados en la Figura 10 sin necesidad de realizar algún paso intermedio.



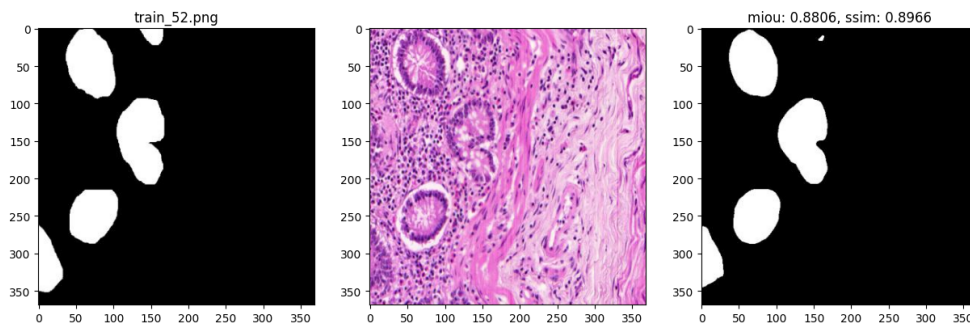
**Figura 10.** Segmentación no supervisada con el modelo SGSCN, en la cual el color amarillo hace referencia a la segmentación de una glándula y el azul al fondo.



**Figura 11.** Grilla con múltiples representaciones de la misma máscara con los diferentes métodos y modelos, en la cual el color amarillo hace referencia a la segmentación de una glándula y el azul al fondo.

### 7.3. U-NET

El propósito inicial de nuestro trabajo es explorar alternativas no supervisadas para así poder evaluar su desempeño en comparación a las ya conocidas estrategias supervisadas. Por este motivo, generamos resultados cualitativos de la estrategia supervisada en cuestión (Arquitectura U-Net), para así evaluar su desempeño y tener un margen de referencia, los resultados cualitativos se encuentran ilustrados en la Figura 12.



**Figura 12.** Segmentación supervisada con la arquitectura U-Net.

## 7.4. RESULTADOS CUANTITATIVOS

A lo largo de nuestros experimentos utilizamos las métricas:

- SSIM, o Structural Similarity Index Measure, es una métrica utilizada para cuantificar la similitud entre dos imágenes. Tiene en cuenta factores como la luminancia, el contraste y la estructura, y proporciona un valor numérico entre -1 y 1, donde los valores más altos indican una mayor similitud.
- La Mean Intersection Over Union (MIOU) es una métrica utilizada habitualmente en tareas de visión por ordenador como la segmentación semántica. Mide la precisión de los modelos de segmentación calculando el solapamiento entre las máscaras de segmentación previstas y las reales.

Esto para poder hacer un análisis con más profundidad y certeza sobre el desempeño de los modelos al momento de generar las imágenes.

Los resultados cuantitativos se encuentran ilustrados en la Tabla 4.

Método	MIOU	SSIM
SGSCN Base	0.304	0.247
SGSCN con medianpooling	0.308	<b>0.323</b>
SGSCN con MS loss	0.290	0.304
SGSCN con medianpooling y MS loss	0.366	0.317
ReDO Base	<b>0.446</b>	0.235
ReDO con medianpooling	0.362	0.306
ReDO con MS loss	0.392	0.263
ReDO con medianpooling y MS loss	0.380	0.262
U-Net	0.781	0.709
All gland	0.467	0.003

**Tabla 4.** Resultados cuantitativos para los modelos de segmentación

Con respecto a la tabla, los métodos base son SGSCN Base y ReDO Base. Los resultados de MIOU y SSIM para estos métodos son 0,304046 y 0,247216 para SGSCN Base, y 0,446511 y 0,235906 para ReDO Base respectivamente.

Cuando se utiliza median pooling y MS loss, los resultados son los siguientes:

- Para SGSCN con medianpooling y MS loss, los valores de MIOU y SSIM son 0,308437 y 0,323208 respectivamente.
- Para ReDO con medianpooling y MS loss, los valores de MIOU y SSIM son 0,3623835 y 0,30607 respectivamente.

Comparando estos resultados con los métodos base, podemos ver que:

- Para SGSCN, el uso de medianpooling y MS loss mejora el valor de MIOU (de 0,304046 a 0,366518), y mejora el valor de SSIM (de 0,263529 a 0,337833).
- Para ReDO, el uso de medianpooling y MS loss disminuye el valor de MIOU (de 0,446511 a 0,380918), pero mejora ligeramente el valor de SSIM (de 0,235906 a 0,26261).

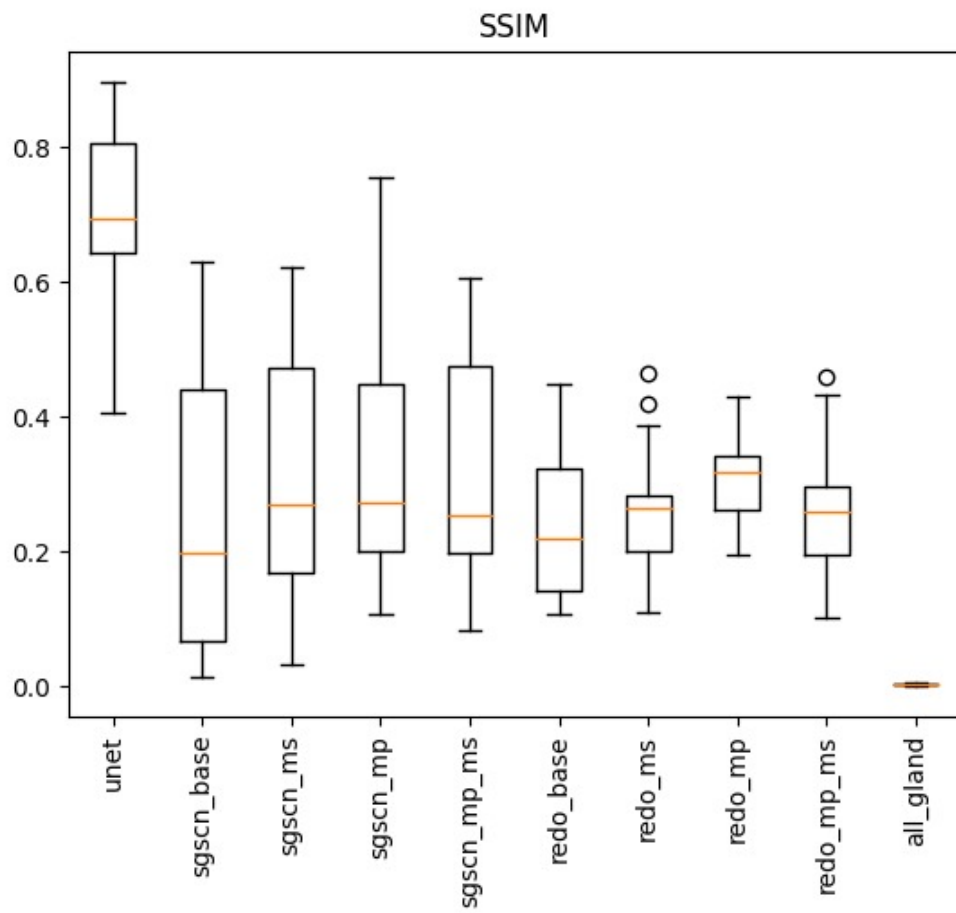
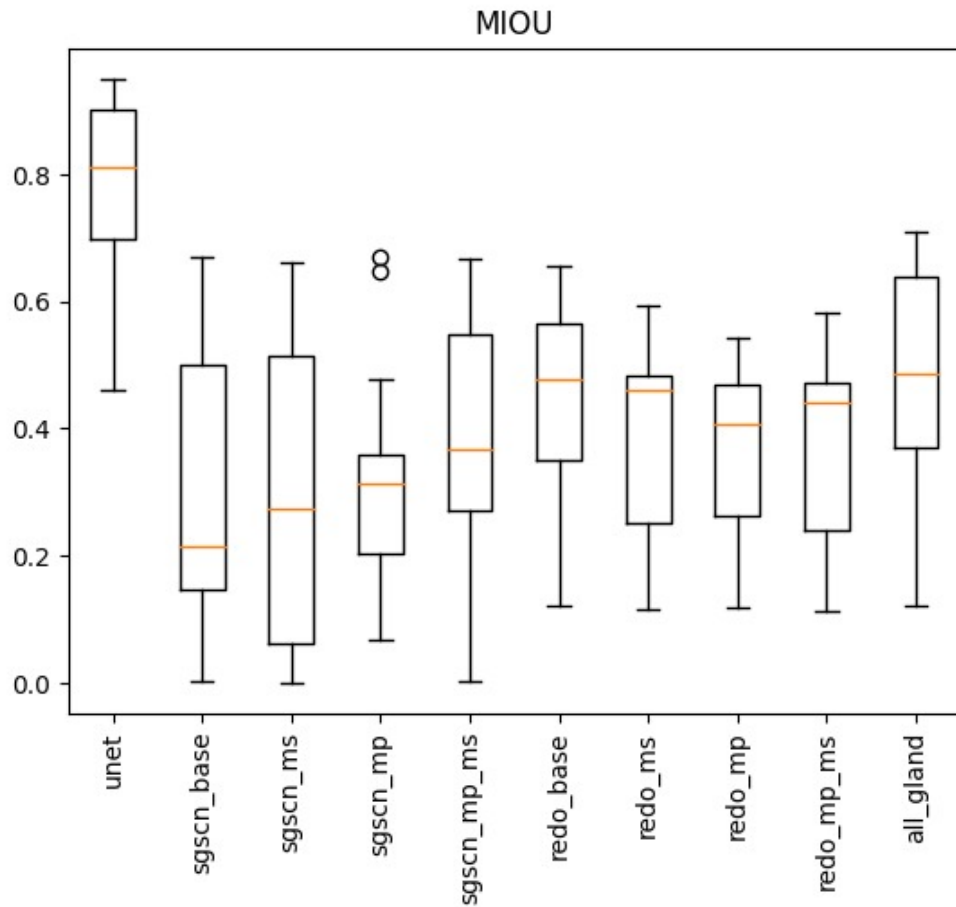
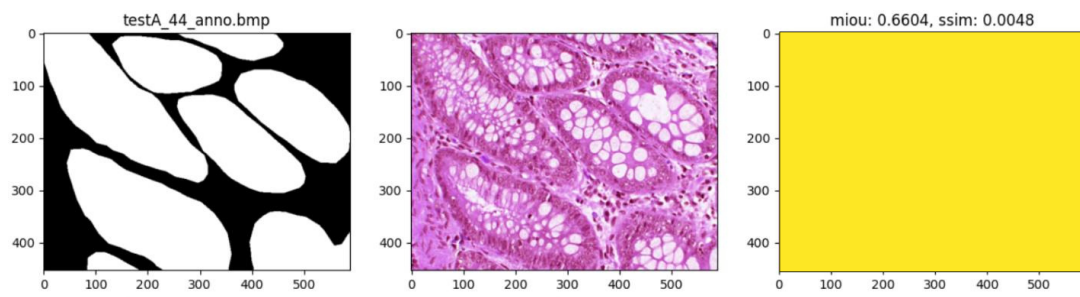


Figura 13. Diagrama de caja para los resultados de la métrica SSIM.



**Figura 14.** Diagrama de caja para los resultados de la métrica MIOU.



**Figura 15.** Resultados encontrados sobre el dataset All Gland, en la cual el color amarillo hace referencia a la segmentación de una glándula y el azul al fondo.

MIOU es la métrica más utilizada en el campo de la segmentación supervisada para medir el rendimiento de un modelo de aprendizaje automático en diferentes tareas. Lamentablemente

para el caso de la segmentación no supervisada puede ser problemática. Por ejemplo, en este caso las máscaras generadas por el método All Gland tienen un rendimiento alto, pues gran parte de las imágenes contenidas en los campos visuales corresponden a glándulas. Sin embargo este comportamiento en la métrica no es deseable, pues la estructura de las máscaras no tienen ningún parecido estructural con las máscaras delineadas por los expertos. Por esa razón SSIM es una mejor métrica para describir el rendimiento en nuestros experimentos, tal como se muestra en la Figura 15.

## 8. CONCLUSIONES Y PERSPECTIVAS

Podemos concluir basándonos en los hallazgos de este estudio que las estrategias de aprendizaje profundo para la segmentación no supervisada de la patología digital tienen un gran potencial para cambiar el campo de la medicina diagnóstica. Los modelos de segmentación no supervisados demostraron ser capaces de identificar patrones junto con anomalías en los tejidos del colon con gran eficiencia y precisión.

Se concluyó que técnicas y métodos como Median Pooling, SGSCN, Mumford-Shah Loss y ReDO pueden mejorar la calidad de la segmentación y contribuir al momento de generar representaciones más robustas. Estos hallazgos resaltan la importancia de investigar y experimentar con una variedad de métodos para maximizar la eficacia de nuestros modelos.

Tomando como referencia el aumento más significativo que en nuestro caso fue con el modelo SGSCN, donde en su instancia base obtenemos un resultado de 0,247216, observamos que aplicando la capa de Median Pooling nos otorga un mejor resultado 0,323208, en lugar de si aplicamos la función de pérdida Mumford-Shah 0,304634 o si aplicamos la función de pérdida y la capa de Median Pooling 0,317799.

## BIBLIOGRAFÍA

- Aganj, Iman et al. “Unsupervised Medical Image Segmentation Based on the Local Center of Mass”. En: *Scientific Reports* (29 August 2018) (vid. pág. 28).
- Ahn, Euijoon, Dagan Feng y Jinman Kim. “A spatial guided self-supervised clustering network for medical image segmentation”. En: (2021), págs. 379-388 (vid. págs. 34-36).
- Barisoni, Laura y B. Hodgins Jeffrey. “Digital pathology in nephrology clinical trials, research, and pathology practice”. En: *Current opinion in nephrology and hypertension* (2017), págs. 450-459. DOI: 10.1097/MNH.0000000000000360 (vid. pág. 14).
- Bengio, Y., A. Courville y P. Vincent. “Representation Learning: A Review and New Perspectives”. En: *IEEE Trans Pattern Anal Mach Intell* (2013), págs. 1798-1828 (vid. pág. 15).
- Celebi, M Emre y Kemal Aydin. *Unsupervised learning algorithms*. Vol. 9. Springer, 2016 (vid. pág. 27).
- Chen, Mickaël, Thierry Artières y Ludovic Denoyer. “Unsupervised Object Segmentation by Redrawing”. En: *Advances in neural information processing systems* 32 (27 May 2019) (vid. pág. 32).
- Chen, Min et al. “Deep feature learning for medical image analysis with convolutional autoencoder neural network”. En: *IEEE Transactions on Big Data* 7.4 (2017), págs. 750-758 (vid. pág. 14).
- Christopher, D.M. Fletcher, R. Goldblum John y D.R. Thompson Lester. *What is Histopathology?* Saunders Ltd, 2012 (vid. pág. 21).

- David, F. Steiner et al. “Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes”. En: *Am J Surg Pathol (The American journal of surgical pathology)* (2018), págs. 1636-1646. DOI: 10.1097/PAS.0000000000001151 (vid. pág. 15).
- Ding, Yi, Fujuan Chen y Yang Zhao. “A Stacked Multi-Connection Simple Reducing Net for Brain Tumor Segmentation”. En: *University of Electronic Science and Technology of China* (2019). DOI: 10.1109/ACCESS.2019.2926448 (vid. pág. 39).
- Ester, Martin et al. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. En: *Knowledge Discovery and Data Mining*. 1996 (vid. pág. 27).
- Hastie, Trevor et al. “Overview of supervised learning”. En: *The elements of statistical learning: Data mining, inference, and prediction* (2009), págs. 9-41 (vid. pág. 23).
- Hastie, Trevor et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009 (vid. pág. 25).
- Hastie, Trevor et al. “Unsupervised learning”. En: *The elements of statistical learning: Data mining, inference, and prediction* (2009), págs. 485-585 (vid. pág. 23).
- Jordan, M. I. y T. M. Mitchell. “Machine learning: Trends, perspectives, and prospects.” En: *Science (New York, N.Y.)* (2015), págs. 255-60. DOI: 10.1126/science.aaa8415 (vid. pág. 24).
- Kaufman, Leonard y Peter Rousseeuw. *Finding Groups in Data: An Introduction To Cluster Analysis*. Ene. de 1990. DOI: 10.2307/2532178 (vid. pág. 27).
- Kim, Boah y Jong Chul Ye. “Mumford–Shah loss functional for image segmentation with deep learning”. En: *IEEE Transactions on Image Processing* 29 (2019), págs. 1856-1866 (vid. pág. 37).

- Komura, D. y S. Ishikawa. “Machine Learning Methods for Histopathological Image Analysis”. En: *Computational and structural biotechnology* (2018). DOI: 10.1016/j.csbj.2018.01.001 (vid. pág. 15).
- Kotsiantis, Sotiris B, Ioannis Zaharakis, P Pintelas et al. “Supervised machine learning: A review of classification techniques”. En: *Emerging artificial intelligence applications in computer engineering* 160.1 (2007), págs. 3-24 (vid. pág. 25).
- Lindvall, Martin. “Designing with Machine Learning in Digital Pathology”. En: (2021) (vid. pág. 22).
- Lloyd, S. “Least squares quantization in PCM”. En: *IEEE Transactions on Information Theory* 28.2 (1982), págs. 129-137. DOI: 10.1109/TIT.1982.1056489 (vid. pág. 26).
- Marilyn, M. B. et al. “Digital and Computational Pathology: Bring the Future into Focus”. En: *Pathology Informatics* (), pág. 10. DOI: 10.4103/2153-3539.255259 (vid. pág. 22).
- Metin, N. Gurcan et al. “Histopathological image analysis: a review”. En: *IEEE Rev Biomed Eng* (2009), págs. 147-71147-71. DOI: 10.1109/RBME.2009.2034865 (vid. pág. 14).
- Orchard, Guy y Brian Nation. *Histopathology*. Oxford University Press, USA, 2012 (vid. pág. 21).
- Pantanowitz, L., JH. Sinard y WH. Henricks. “Validating whole slide imaging for diagnostic purposes in pathology”. En: *guideline from the College of American Pathologists Pathology and Laboratory Quality Center. Arch Pathol Lab Med* (2013), págs. 1710-22. DOI: 10.5858/arpa.2013-0093-CP (vid. pág. 14).
- Pérez-Montiel, Camilo Andrés. “Nueva clasificación molecular del carcinoma de endometrio: impacto en el diagnóstico histopatológico, tratamiento y pronóstico”. En: *MedUNAB* 24 (2022), 365–374 (vid. pág. 15).

- Ronneberger, O., P.Fischer y T. Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. En: LNCS 9351 (2015). (available on arXiv:1505.04597 [cs.CV]), págs. 234-241 (vid. pág. 39).
- Ronneberger, Olaf, Philipp Fischer y Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer, 18 November 2015 (vid. pág. 38).
- Shi, Jianbo y J. Malik. “Normalized cuts and image segmentation”. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.8 (2000), págs. 888-905. DOI: 10.1109/34.868688 (vid. pág. 28).
- Sirinukunwattana, Korsuk et al. “Gland segmentation in colon histology images: The glas challenge contest”. En: *Medical image analysis* 35 (2017), págs. 489-502 (vid. pág. 30).
- Wang, Zhou et al. “Image quality assessment: from error visibility to structural similarity”. En: *IEEE transactions on image processing* 13.4 (2004), págs. 600-612 (vid. pág. 41).

## ANEXOS

Anexo A. Productos académicos.

- Juan, D., Julián, C & David, R. (2024, Mayo). Estrategia de aprendizaje profundo para la segmentación no supervisada en patología digital. Aceptado para presentar en III Conferencia Colombiana de Matemáticas Aplicadas e Industriales 2024 (MAPI 3).

Anexo B. Función para hacer el cálculo de las métricas.

```
def calculate_precision(y_true, y_pred):  
    """  
    Calcula la precision de dos matrices de segmentacion.  
  
    Args:  
        y_true: La matriz de segmentacion de referencia.  
        y_pred: La matriz de segmentacion predicha.  
  
    Returns:  
        La precision.  
    """  
    intersection = np.sum((y_true & y_pred) == 1)  
    union = np.sum((y_true | y_pred) == 1)  
  
    if union == 0:  
        return 0  
    else:  
        return intersection / union  
  
def calculate_recall(y_true, y_pred):  
    """  
    Calcula el recall de dos matrices de segmentacion.  
  
    Args:  
        y_true: La matriz de segmentacion de referencia.
```

y\_pred: La matriz de segmentacion predicha.

Returns:

El recall.

```
"""
```

```
intersection = np.sum((y_true & y_pred) == 1)
```

```
ground_truth = np.sum(y_true == 1)
```

```
if ground_truth == 0:
```

```
    return 0
```

```
else:
```

```
    return intersection / ground_truth
```

```
def calculate_metrics(y_true, y_pred):
```

```
"""
```

Calcula el miou, SSIM, F1 y DICE de dos matrices de segmentacion.

Args:

y\_true: La matriz de segmentacion de referencia.

y\_pred: La matriz de segmentacion predicha.

Returns:

Una lista con el MIoU, SSIM, F1 y DICE.

```
"""
```

```
intersection = np.sum(((y_true/255>0.5) & (y_pred/255>0.5)) == 1)
```

```
union = np.sum(((y_true/255>0.5) | (y_pred/255>0.5)) == 1)
```

```

if union == 0:
    MIOU = 0
else:
    MIOU = intersection / union

SSIM = metrics.structural_similarity(y_true/255>0.5, y_pred,
data_range=1)

precision = calculate_precision((y_true/255>0.5), (y_pred/255>0.5))
recall = calculate_recall((y_true/255>0.5), (y_pred/255>0.5))

if precision + recall == 0:
    F1 = 0
else:
    F1 = 2 * (precision * recall) / (precision + recall)

intersection = np.sum((y_true/255>0.5) * (y_pred/255>0.5))
union = np.sum((y_true/255>0.5)) + np.sum((y_pred/255>0.5))

if union == 0:
    DICE = 0
else:
    DICE = (2 * intersection) / (union + 1e-12)

return [MIOU, SSIM, F1, DICE]

```