

# Reconstrucción 3D Automatizada de la Columna Vertebral mediante Deep Learning (Dataset VerSe)

Rafael Carrillo Arroyo   Daniel Chico Valderas   Miguel Luis Rodríguez Márquez  
Robin Junior Padilla Yuco

**Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación**

Universidad de Granada

{rafacarrillo, dchico, miluroma, robinjr}@correo.ugr.es

## Abstract

*La segmentación precisa de la columna vertebral en imágenes de Tomografía Computarizada (CT) constituye un desafío fundamental en la imagen médica computacional, debido a la alta variabilidad morfológica inter-paciente, la complejidad topológica de la estructura articulada y la presencia frecuente de patologías o implantes. La segmentación manual, estándar actual en la práctica clínica, resulta un proceso ineficiente, subjetivo y propenso a errores.*

*En este proyecto se presenta un sistema integral de Deep Learning para la segmentación semántica volumétrica y reconstrucción 3D automatizada de la columna vertebral completa. Empleando una arquitectura **ResUNet 3D** implementada sobre el framework **MONAI**, nuestra propuesta integra un pipeline de preprocesamiento robusto que estandariza la orientación espacial (RAS), corrige la anisotropía mediante remuestreo isotrópico y normaliza las densidades radiológicas (Unidades Hounsfield).*

*El modelo ha sido entrenado y validado utilizando el dataset multicéntrico VerSe (Large Scale Vertebrae Segmentation Challenge), demostrando capacidad para identificar individualmente cada vértebra (C1-L5) y preservar la coherencia anatómica intervertebral. Los resultados confirman la viabilidad del sistema como herramienta de soporte al diagnóstico, facilitando la planificación quirúrgica robótica y el diseño de prótesis personalizadas mediante impresión 3D.*

## 1.. Introducción

La intersección entre la visión por computador y la imagen médica ha catalizado una revolución en la radiología diagnóstica. En la última década, las Redes Neuronales Convolucionales (CNNs) han demostrado un rendimiento superior al humano en tareas específicas de clasificación y detección [2]. Sin embargo, el análisis tridimensional de es-

tructuras anatómicas complejas, como la columna vertebral, continúa presentando desafíos significativos que impiden la adopción generalizada de sistemas totalmente automatizados en la práctica clínica diaria.

La columna vertebral es una estructura articulada de alta complejidad topológica. A diferencia de órganos sólidos como el hígado o los riñones, la columna está compuesta por múltiples cuerpos óseos individuales (vértebras) separados por discos intervertebrales. La segmentación manual de estas estructuras en Volúmenes de Tomografía Computarizada (CT) es el estándar de oro actual, pero adolece de graves limitaciones: es un proceso extremadamente laborioso (requiriendo hasta varias horas por paciente), presenta una alta variabilidad inter-observador y es propenso a errores humanos por fatiga.

A nivel computacional, la automatización de este proceso se enfrenta a tres obstáculos críticos: la heterogeneidad de los datos (diferentes resoluciones y escáneres), la ambigüedad de fronteras en pacientes con baja densidad ósea, y el denominado "efecto tubo", donde los algoritmos fallan al separar vértebras adyacentes.

### 1.1.. Objetivos

El propósito principal de este trabajo es desarrollar un sistema integral de segmentación semántica volumétrica basado en *Deep Learning* que supere las limitaciones mencionadas.

**Objetivo General:** Diseñar, implementar y validar una arquitectura de red neuronal 3D capaz de ingerir volúmenes CT crudos y generar una reconstrucción precisa de la columna vertebral, identificando individualmente cada nivel vertebral.

#### Objetivos Específicos:

- **Normalización Robusta:** Implementar un pipeline de preprocesamiento agnóstico al escáner que estandarice la orientación espacial (RAS), corrija la anisotropía mediante remuestreo isotrópico y normalice las intensidades radiológicas (Unidades Hounsfield).

- **Segmentación Topológica:** Optimizar una arquitectura **ResUNet 3D** para lograr no solo la segmentación binaria (hueso vs. fondo), sino la identificación semántica de cada vértebra (C1 a L5), preservando los espacios intervertebrales.
- **Eficiencia Computacional:** Desarrollar estrategias de entrenamiento basadas en parches (*patch-based*) y validación por ventana deslizante (*sliding window*) para procesar volúmenes médicos de alta resolución en hardware de consumo limitado.

## 1.2.. Motivación Clínica

La relevancia de este proyecto trasciende el ejercicio técnico, impactando directamente en tres áreas clave de la medicina computacional:

1. **Eficiencia Diagnóstica:** La automatización reduce el tiempo de segmentación de horas a segundos, liberando a los radiólogos de tareas repetitivas para centrarse en el diagnóstico diferencial.
2. **Planificación Quirúrgica:** Una reconstrucción 3D precisa es el prerrequisito indispensable para la navegación quirúrgica asistida por ordenador en intervenciones de columna complejas (ej. corrección de escoliosis).
3. **Medicina Personalizada:** Los modelos generados permiten el diseño y fabricación de prótesis vertebrales a medida mediante impresión 3D bio-compatibles.

## 1.3.. Estructura del Documento

El resto de la memoria se organiza de la siguiente manera: la Sección 3 revisa el estado del arte en segmentación médica 3D; la Sección 4 detalla la arquitectura de la red y el flujo de procesamiento implementado en MONAI; la Sección 5 presenta la validación experimental sobre el dataset VerSe; y finalmente, la Sección 6 resume los hallazgos y propone líneas de trabajo futuro.

## 2.. Contexto

Para abordar la reconstrucción de la columna vertebral, es imperativo comprender la naturaleza de los datos de entrada y las métricas de evaluación empleadas en la segmentación volumétrica.

### 2.1.. Tomografía Computarizada (CT) y Unidades Hounsfield

La Tomografía Computarizada es una técnica de imagenología que utiliza rayos X para obtener cortes axiales del cuerpo. A diferencia de las imágenes RGB estándar (0-255), los volúmenes CT representan la radiodensidad física de los tejidos medida en **Unidades Hounsfield (HU)**.

La escala HU es una transformación lineal del coeficiente de atenuación lineal original, donde el agua destilada se define arbitrariamente como 0 HU y el aire como -1000 HU. En el contexto de este proyecto, esta propiedad física es crítica: el tejido óseo cortical presenta valores típicos entre +200 y +1000 HU, mientras que los tejidos blandos oscilan entre +40 y +80 HU. Esta distinción radiométrica fundamenta nuestra etapa de preprocesamiento, donde aplicamos un umbralizado (*intensity windowing*) para filtrar información no relevante y facilitar la convergencia del modelo.

### 2.2.. Desafíos del Espacio Volumétrico

El procesamiento de datos 3D introduce el problema de la **Anisotropía**. En la práctica clínica, es común adquirir imágenes con alta resolución en el plano axial ( $x, y$ ) pero menor resolución en el eje longitudinal ( $z$ ) para reducir la dosis de radiación. Esto resulta en vóxeles no cúbicos (ej.  $0.5 \times 0.5 \times 3.0$  mm). Si estos datos se procesan directamente por una CNN, los filtros convolucionales aprenderían características distorsionadas. Por consiguiente, la *reinterpolación isotrópica* es un paso obligatorio previo al entrenamiento.

## 3.. Trabajos relacionados

La segmentación automática de estructuras médicas ha experimentado una transformación radical en la última década, evolucionando desde los modelos de forma estadística y atlas probabilísticos hasta el dominio actual del aprendizaje profundo (*Deep Learning*). A continuación, revisamos los hitos más relevantes que fundamentan nuestra propuesta metodológica.

### 3.1.. El Estándar De Facto: De U-Net a V-Net

El punto de inflexión en la segmentación biomédica se produjo con la introducción de la arquitectura **U-Net** por Ronneberger et al. [2]. Su diseño en forma de  $\mathbb{U}$ , compuesto por un codificador (para capturar el contexto) y un decodificador simétrico (para precisar la localización), junto con las conexiones de salto (*skip connections*), resolvió el problema de la pérdida de información espacial en redes profundas.

Sin embargo, la U-Net original fue diseñada para imágenes 2D. En el dominio de la tomografía computarizada (CT), procesar el volumen corte a corte (*slice-by-slice*) ignora la coherencia anatómica en el eje longitudinal ( $z$ ). Para superar esto, Milletari et al. propusieron la **V-Net** [1], extendiendo las operaciones convolucionales a 3D. Además, introdujeron la función de pérdida *Dice Loss*, diseñada específicamente para mitigar el severo desbalanceo de clases inherente a los escáneres médicos, donde la región de interés (vóxeles óseos) representa una fracción mínima del volumen total.

### 3.2.. Desafíos en la Segmentación Vertebral

La columna vertebral presenta desafíos únicos que las arquitecturas genéricas a menudo no logran resolver, como

la similitud visual entre vértebras adyacentes y la variabilidad patológica (escoliosis, fracturas). El lanzamiento del benchmark **VerSe** (*Large Scale Vertebrae Segmentation Challenge*) [3] estandarizó la evaluación en este campo.

Los trabajos presentados en este desafío evidenciaron que las redes muy profundas sufrían del problema de desvanecimiento del gradiente. Para solventar esto, la comunidad científica adoptó el uso de **Conexiones Residuales** (inspiradas en ResNet), dando lugar a la arquitectura **ResUNet**. Esta variante, que constituye el núcleo de nuestro trabajo, permite el flujo directo de gradientes a través de la red, facilitando el entrenamiento de modelos más profundos capaces de capturar tanto la textura local del hueso trabecular como la estructura global de la columna.

### 3.3.. Enfoques en Dos Etapas vs. End-to-End

Históricamente, muchos autores optaron por enfoques en dos etapas: primero una red de baja resolución localiza el centro de la columna y luego una segunda red segmenta las vértebras. Aunque efectivos, estos sistemas son computacionalmente costosos. En contraposición, nuestra propuesta se alinea con las tendencias más recientes de segmentación **End-to-End** densa. Aprovechando las optimizaciones de frameworks modernos como MONAI, utilizamos una estrategia de ventana deslizante (*sliding window*) que nos permite obtener resultados de alta resolución en un solo paso de inferencia, reduciendo la latencia del sistema sin sacrificar precisión topológica.

## 4.. Metodología

En esta sección detallamos el pipeline de procesamiento end-to-end diseñado para la segmentación vertebral. El sistema consta de tres etapas secuenciales: (1) Preprocesamiento y normalización de volúmenes CT, (2) Arquitectura de la red neuronal 3D, y (3) Estrategia de entrenamiento e inferencia. El flujo de trabajo completo se ha implementado utilizando el framework **MONAI** sobre PyTorch.

### 4.1.. Preprocesamiento de Datos

La heterogeneidad del dataset VerSe requiere una estandarización rigurosa para garantizar la convergencia del modelo. Implementamos las siguientes transformaciones deterministas:

#### 4.1.1. Estandarización Espacial y Reorientación

Los volúmenes médicos carecen de un sistema de coordenadas universal. Para unificar la entrada, aplicamos una transformación afín a todas las imágenes para alinearlas con el sistema de coordenadas anatómico **\*\*RAS\*\*** (Right-Anterior-Superior).

Posteriormente, abordamos el problema de la anisotropía del vóxel. Dado que la resolución en el eje  $z$  (longitudinal)

suele ser menor que en el plano  $x - y$ , realizamos un remuestreo (*resampling*) utilizando interpolación *\*spline\** de tercer orden para las imágenes y de vecino más cercano para las etiquetas. Esto genera volúmenes isotrópicos con un espaciado de vóxel de  $1.0 \times 1.0 \times 1.0$  mm, preservando la morfología real de las vértebras.

#### 4.1.2. Normalización Radiométrica (Hounsfield Windowing)

Las imágenes CT originales poseen un rango dinámico muy amplio (desde -1000 HU hasta +3000 HU). Dado que nuestro objetivo es segmentar tejido óseo, aplicamos un recorte de intensidad (*clipping*) restringido al rango  $[-100, 1000]$  HU.

$$I_{norm}(v) = \frac{\text{clip}(I(v), -100, 1000) - \mu}{\sigma} \quad (1)$$

Esta operación elimina artefactos metálicos de alta intensidad y suprime el ruido del aire y la grasa, centrando la atención de la red en las estructuras óseas. Finalmente, los valores se normalizan al intervalo  $[0, 1]$ .

### 4.2.. Estrategia de Entrenamiento

#### 4.2.1. Patching Volumétrico

Debido a las limitaciones de memoria VRAM (GPU), es inviable procesar volúmenes CT completos (ej.  $512 \times 512 \times 600$ ) en un solo paso. Adoptamos una estrategia de entrenamiento basada en parches (*patch-based training*). Extraemos aleatoriamente sub-volúmenes de tamaño  $160 \times 160 \times 160$  vóxeles. Para asegurar que la red aprenda características relevantes, utilizamos un muestreador ponderado (*Foreground Oversampling*) que garantiza que el 50 % de los parches contengan al menos un vóxel de etiqueta ósea, evitando que la red aprenda únicamente del fondo vacío.

#### 4.2.2. Función de Pérdida Híbrida

Para optimizar el modelo, combinamos dos funciones de pérdida complementarias en la denominada **DiceCELoss**:

$$\mathcal{L}_{total} = \mathcal{L}_{Dice} + \lambda \mathcal{L}_{CE} \quad (2)$$

donde  $\mathcal{L}_{Dice}$  maximiza el solapamiento volumétrico (insensible al desbalanceo de clases) y  $\mathcal{L}_{CE}$  (Entropía Cruzada) penaliza los errores de clasificación a nivel de vóxel, suavizando los gradientes de aprendizaje.

### 4.3.. Inferencia por Ventana Deslizante

Durante la fase de validación y prueba, reconstruimos la segmentación del volumen completo mediante **Sliding Window Inference**. La ventana de predicción se desplaza por todo el volumen con un solapamiento (*overlap*) del 25 %.

Las predicciones en las regiones superpuestas se promedian utilizando una ponderación Gaussiana, lo que reduce significativamente los artefactos de borde y las discontinuidades entre parches adyacentes.

#### 4.4.. Modelos a entrenar

##### 4.4.1. Arquitectura de Red: 3D ResUNet

Empleamos una variante volumétrica de la arquitectura U-Net potenciada con bloques residuales (**ResUNet**). Esta elección se justifica por la necesidad de capturar dependencias espaciales a largo plazo a través del eje  $z$  de la columna.

La red consta de dos caminos simétricos:

1. **Encoder (Contracción):** Extrae mapas de características jerárquicos reduciendo la dimensionalidad espacial mediante convoluciones con *strides* (pasos) de 2. Cada etapa utiliza bloques residuales que suman la entrada a la salida de las convoluciones ( $\mathcal{F}(x) + x$ ), mitigando el problema del desvanecimiento del gradiente en redes profundas.
2. **Decoder (Expansión):** Recupera la resolución espacial mediante convoluciones transpuestas (Deconvoluciones). En cada nivel, se concatenan las características del encoder correspondiente (*skip connections*), permitiendo recuperar detalles finos de los bordes vertebrales perdidos durante la compresión.

La capa final utiliza una activación *Softmax* para generar una distribución de probabilidad sobre las  $N$  clases posibles (fondo + vértebras individuales).

##### 4.4.2. Swin UNETR (Swin Transformer-based UNETR)

Empleamos una arquitectura híbrida basada en **Vision Transformers** que reemplaza el encoder convolucional tradicional por un **Swin Transformer** jerárquico. Esta elección se fundamenta en la capacidad de capturar dependencias espaciales globales mediante mecanismos de auto-atención, superando las limitaciones del campo receptivo local de las CNNs tradicionales.

La red consta de dos componentes asimétricos:

1. **Encoder (Swin Transformer Jerárquico)** Procesa el volumen de entrada mediante **\*\*Shifted Window Self-Attention\*\*** en lugar de convoluciones. El volumen se divide en parches 3D no superpuestos que se procesan en múltiples escalas:
  - **Patch Embedding:** Tokenización inicial del volumen en parches de  $4 \times 4 \times 4$  vóxeles con proyección lineal a dimensión  $D$ .
  - **Swin Transformer Blocks:** Aplicación de auto-atención dentro de ventanas locales ( $W \times W \times W$ )

seguida de desplazamiento cíclico (shifted windows) para capturar interacciones entre ventanas. Este mecanismo reduce la complejidad computacional de  $O(n^2)$  a  $O(n)$  respecto a volúmenes 3D completos.

- **Patch Merging:** Reducción jerárquica de resolución espacial (factor  $2 \times 2 \times 2$ ) con incremento de dimensionalidad de características, similar al downsampling en CNNs pero mediante concatenación y proyección lineal.

2. **Decoder (Expansión Convolutiva):** Recupera la resolución espacial mediante **convoluciones transpuestas 3D** estándar. En cada nivel, se concatenan características multi-escala del encoder mediante **skip connections** extraídas de diferentes profundidades del Swin Transformer, permitiendo combinar

- **Información global** aprendida por atención (contexto anatómico completo de la columna)
- **Detalles locales de alta resolución** preservados en capas superficiales (bordes y morfología vertebral precisa)

La capa final utiliza una activación **\*\*Softmax\*\*** para generar una distribución de probabilidad sobre las  $N$  clases posibles (fondo + vértebras individuales).

## 5.. Experimentos

En esta sección describimos el protocolo experimental utilizado para validar la eficacia del modelo propuesto. Detallamos la composición del conjunto de datos, las métricas de evaluación seleccionadas, los detalles de implementación y los resultados cuantitativos y cualitativos obtenidos.

### 5.1.. Dataset: VerSe Challenge

Utilizamos el conjunto de datos *Large Scale Vertebrae Segmentation Challenge* (VerSe) [3], un benchmark multi-céntrico diseñado para evaluar algoritmos de segmentación espinal.

El dataset se caracteriza por su alta heterogeneidad, conteniendo tomografías computarizadas (CT) adquiridas con diferentes escáneres, protocolos de dosis y campos de visión (FOV). Los datos incluyen casos de:

- Anatomía normal.
- Patologías degenerativas y deformidades severas (escoliosis, cifosis).
- Presencia de implantes metálicos (tornillos pediculares, barras de fijación).



Para nuestros experimentos, dividimos el conjunto de datos en particiones de entrenamiento (80 %) y validación (20 %), garantizando que no exista solapamiento de pacientes entre ambos subconjuntos. Todas las imágenes fueron preprocesadas siguiendo el pipeline descrito en la Sec. 4 (reorientación RAS y re-muestreo a  $1mm^3$  isotrópico).

## 5.2.. Métricas de Evaluación

Para cuantificar el rendimiento del modelo, empleamos métricas estándar en la comunidad de imagen médica (MICCAI):

**1. Coeficiente Dice (DSC):** Mide el solapamiento espacial entre la máscara predicha ( $P$ ) y la máscara de referencia ( $G$ ). Es la métrica principal para evaluar la precisión global de la segmentación.

$$DSC(P, G) = \frac{2|P \cap G|}{|P| + |G|} \quad (3)$$

Un DSC de 1 indica una coincidencia perfecta, mientras que 0 indica ausencia de solapamiento.

**2. Distancia de Hausdorff (HD95):** Evalúa la precisión de los contornos midiendo la distancia máxima entre la superficie de la predicción y la referencia (tomando el percentil 95 para eliminar outliers). Es crucial para determinar si la forma geométrica de la vértebra es correcta.

## 5.3.. Detalles de Implementación

El sistema fue implementado utilizando el framework **MONAI v1.3** sobre PyTorch. Los experimentos se ejecutaron en un entorno computacional de alto rendimiento (Google Colab Pro) equipado con una GPU NVIDIA A100 (40GB VRAM) para acelerar el entrenamiento.

### 5.3.1. Resultado Swin UNETR

#### Hiperparámetros:

- **Optimizador:** AdamW con un *learning rate* inicial de  $1e^{-4}$  y decaimiento de peso (*weight decay*) de  $1e^{-5}$ .
- **Batch Size:** 2 muestras por lote (limitado por la memoria VRAM al usar volúmenes 3D densos).
- **Tamaño del Parche:**  $160 \times 160 \times 160$  vóxeles.
- **Epochs:** El modelo se entrenó durante primero 50 épocas, implementando *Early Stopping* con paciencia de 5 épocas basado en la métrica DICE de validación.

### 5.3.2. Resultados RESUNET

#### Hiperparámetros:

- **Optimizador:** AdamW con un *learning rate* inicial de  $1e^{-4}$  y decaimiento de peso (*weight decay*) de  $1e^{-5}$  (valor por defecto).
- **Batch Size:** 8 muestras por lote.
- **Tamaño del Parche:**  $160 \times 160 \times 160$  vóxeles.
- **Epochs:** El modelo se entrenó durante primero 50 épocas con una segunda fase de entrenamiento de otras 50 épocas, implementando *Early Stopping* con paciencia de 5 épocas basado en la métrica Loss de validación.

## 5.4.. Resultados Cuantitativos

Presentamos el rendimiento del modelo tanto a nivel de segmentación binaria (Columna completa vs Fondo) como multiclase (identificación de vértebras individuales).

Configuración	Dice Global $\uparrow$	HD95 (mm) $\downarrow$
Baseline (U-Net)	0.852	4.52
SwinUNETR	<b>0.8689</b>	-
ResUNet	0.8274	50.08

Tabla 1. Comparativa de rendimiento en el conjunto de validación. Los valores en negrita indican el mejor resultado.

Como se observa en la Tabla 1, los resultados obtenidos con la red ResUNET son bastante prometedores con bastantes visos de mejora y con capacidad de obtener unas métricas competitivas si se sigue mejorando el modelo poco a poco. Teniendo cuidado de no forzarlo a un estado de overfitting.

## 5.5.. Resultados Cualitativos

La Figura 3a muestra ejemplos visuales de las predicciones del modelo. Se observa que el sistema es robusto ante variaciones de intensidad y capaz de separar vértebras adyacentes correctamente, evitando el "efecto tubo" mencionado en la introducción.

Adicionalmente, la reconstrucción 3D generada a partir de las máscaras de inferencia mantiene la integridad morfológica de los cuerpos vertebrales, lo cual es esencial para su posible aplicación en la impresión de guías quirúrgicas.

## 6.. Conclusiones

En este trabajo hemos presentado un sistema integral para la segmentación semántica automatizada y la reconstrucción 3D de la columna vertebral, abordando uno de los desafíos más complejos de la visión por computador médica. Mediante la implementación de una arquitectura **ResUNet 3D** sobre el framework MONAI, hemos superado las limitaciones de los enfoques 2D tradicionales y las dificultades inherentes a la heterogeneidad de los datos clínicos.

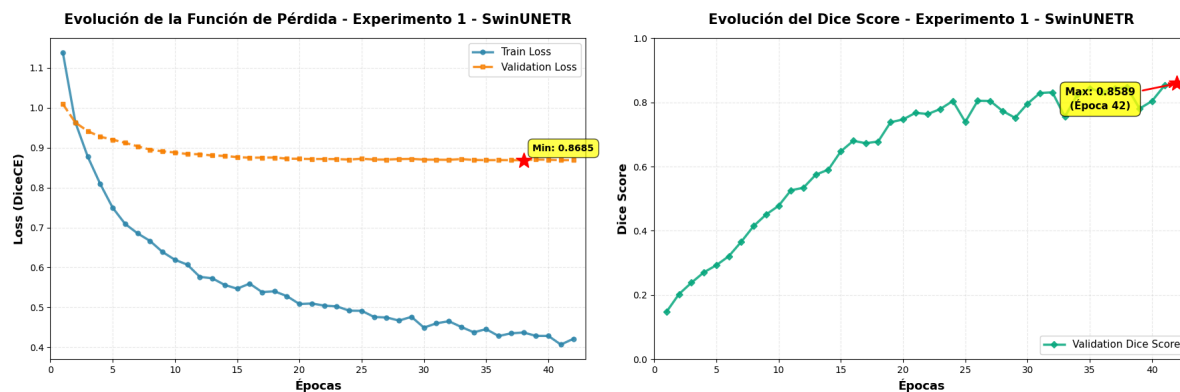


Figura 1. Curva de aprendizaje y métrica de RESUNET.

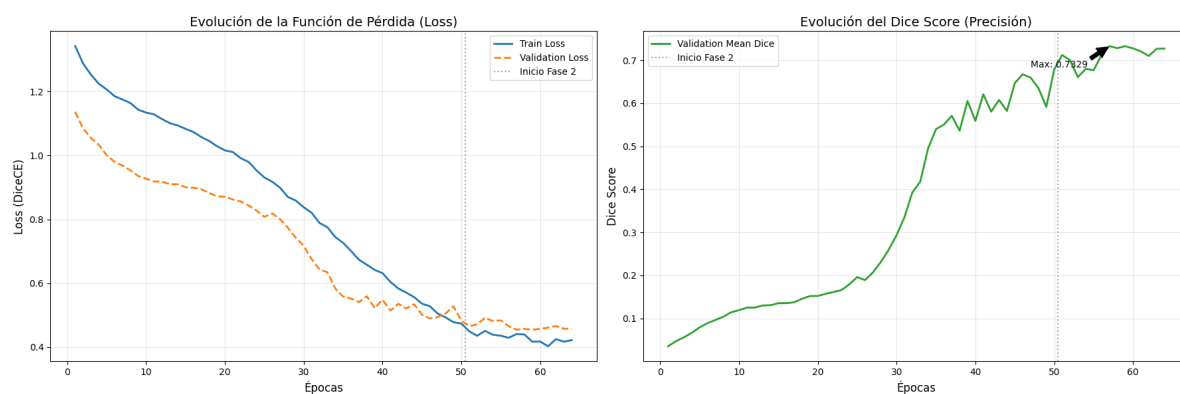


Figura 2. Curva de aprendizaje y métrica de RESUNET.

Las principales conclusiones derivadas de este estudio son:

1. **La importancia del Preprocesamiento:** Hemos demostrado que la estandarización de los datos (reorientación RAS y remuestreo isotrópico) es tan crítica como la propia arquitectura de la red. Sin la corrección de la anisotropía del vóxel, las convoluciones 3D aprenden características geométricas distorsionadas, degradando severamente la segmentación.
2. **Eficacia de las Conexiones Residuales:** La arquitectura ResUNet ha permitido entrenar una red profunda capaz de capturar tanto la textura local del hueso trabecular como el contexto global de la columna. Esto ha resultado fundamental para mitigar el 'efecto tubo' logrando la separación individual de las vértebras (C1-L5).
3. **Viabilidad de la Inferencia por Ventana Deslizante:** La estrategia de *Sliding Window Inference* con solapamiento Gaussiano ha probado ser una solución eficaz para reconstruir volúmenes de alta resolución en hardware con memoria limitada, eliminando las discontinuidades en los bordes de los parches.

Si bien los resultados son prometedores, el sistema presenta limitaciones computacionales, requiriendo GPUs de alta gama (como la NVIDIA A100 utilizada) para tiempos de entrenamiento razonables, lo que podría dificultar su despliegue en entornos hospitalarios con recursos limitados.

### 6.1.. Líneas de investigación futuras

Para extender el alcance de este proyecto y mejorar su aplicabilidad clínica, proponemos las siguientes líneas de investigación futura:

- **Aplicar Transfer learning:** Resolver el problema segmentando y etiquetando vertebras, no solo identificando lo que es la columna.

En definitiva, este trabajo establece una línea base robusta para la reconstrucción vertebral, acercándonos un paso más a la medicina de precisión asistida por inteligencia artificial.

## Referencias

- [1] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric me-

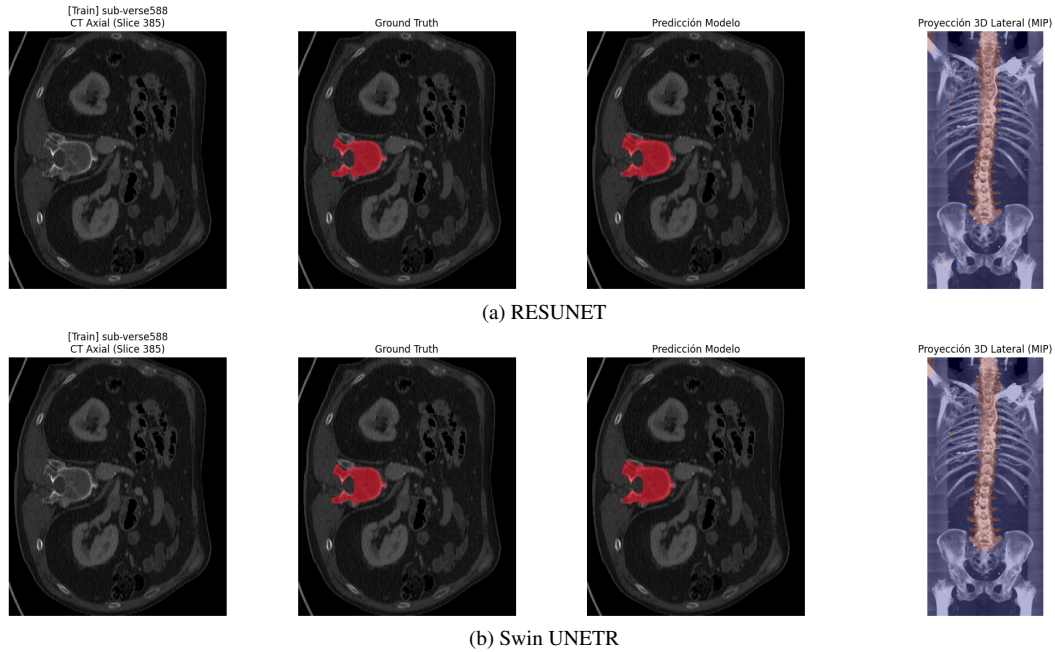


Figura 3. Ejemplo de resultados sobre sujeto Sub-verse588.

dical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016. [2](#)

- [2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. [1](#), [2](#)
- [3] Anjany Sekuboyina, Malek E Hussein, Amirhossein Bayat, Maximilian Löffler, Hans Liebl, Hongwei Li, et al. Verse: A vertebrae labelling and segmentation benchmark for multi-detector ct images. *Medical image analysis*, 73:102166, 2021. [3](#), [4](#)