

Reconstrucción 3D Automatizada de la Columna Vertebral

Deep Learning aplicado a Imagen Médica

Rafael Carrillo Daniel Chico Miguel L. Rodríguez Robin J. Padilla

Visión por Computador
Universidad de Granada
ETSIIT

Curso 2025-2026

- 1 Motivación
- 2 Dataset y Preprocesamiento
- 3 Arquitecturas de Red
- 4 Resultados Experimentales
- 5 Conclusiones y Trabajo Futuro

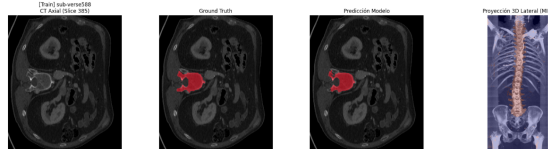
¿Por qué automatizar la segmentación vertebral?

Problema Clínico:

- Segmentación manual: 2-4 horas/paciente
- Alta variabilidad inter-observador
- Propenso a errores por fatiga

Aplicaciones:

- Planificación quirúrgica robótica
- Diseño de prótesis 3D personalizadas
- Diagnóstico de patologías (escoliosis, fracturas)



Ejemplo se segmentación de una muestra

1. Heterogeneidad de Datos

Múltiples escáneres, protocolos, resoluciones anisótropicas

2. Complejidad Anatómica

- 26 vértebras individuales (C1-L5, sacro)
- Morfología variable inter-paciente
- "Efecto tubo": vértebras adyacentes se fusionan

3. Limitaciones Computacionales

Volúmenes 3D masivos ($512 \times 512 \times 600$ vóxeles)

→ Requiere estrategias de procesamiento eficientes

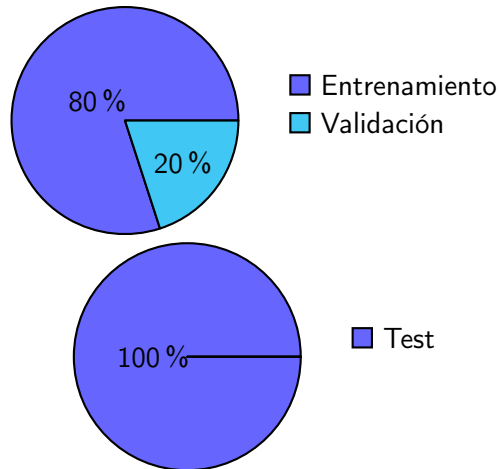
Dataset VerSe (Vertebrae Segmentation Challenge)

Características:

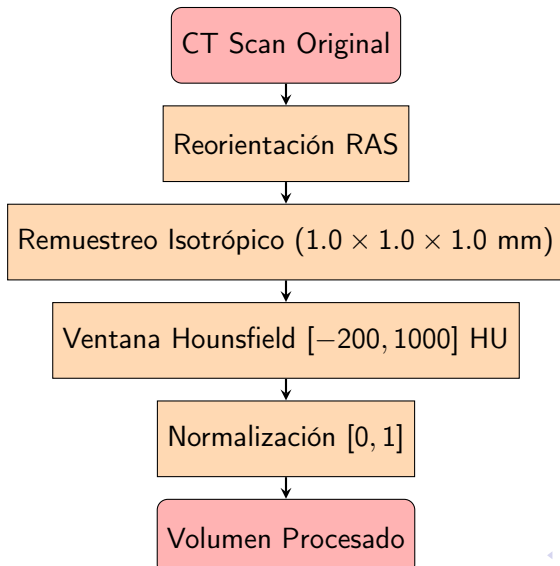
- **Multicéntrico:** 140 escaneos CT
- **Heterogéneo:** Múltiples escáneres y protocolos
- **Casos complejos:**
 - Patologías degenerativas
 - Implantes metálicos
 - Deformidades severas (escoliosis)

División:

- Train: 80 % (112 pacientes)
- Validación: 20 % (28 pacientes)
- Test: 73 pacientes



Pipeline de Preprocesamiento



Escala Radiométrica:

Tejido	HU
Aire	-1000
Grasa	-100 a -50
Agua	0
Músculo	+40 a +80
Hueso	+200 a +1000
Metal	> +3000

Ventana Ósea:

$$I_{norm} = \frac{\text{clip}(I, -200, 1000)}{1200}$$

Elimina:

- Artefactos metálicos
- Ruido de aire/grasa

→ Centra atención en estructuras óseas

Características:

- Encoder-Decoder simétrico
- **Bloques Residuales**: $\mathcal{F}(x) + x$
- Skip connections multi-escala
- Convoluciones 3D nativas

Configuración:

- Canales: [16, 32, 64, 128]
- 2 bloques residuales/nivel
- Entrada: 160^3 vóxeles
- Salida: 2 clases (fondo/hueso)

Innovación:

- **Encoder**: Swin Transformer 3D
- Auto-atención jerárquica
- **Decoder**: Convolutacional

Configuración:

- Feature size: 48
- Patches: $4 \times 4 \times 4$
- SW-MSA (Shifted Window)
- Complejidad: $O(n)$

Patching Volumétrico

- Parches de $160 \times 160 \times 160$ vóxeles
- **Foreground Oversampling**: 50 % parches con hueso
- Evita aprender solo del fondo vacío

Función de Pérdida Híbrida: DiceCELoss

$$\mathcal{L}_{total} = \mathcal{L}_{Dice} + \lambda \mathcal{L}_{CE}$$

- **Dice**: Maximiza solapamiento volumétrico (robusto a desbalanceo)
- **Cross-Entropy**: Penaliza errores a nivel de vóxel

Optimización

AdamW | LR: 10^{-4} | Batch: 2-8 | Early Stopping (paciencia: 5)

- Ventana: 160^3 con solapamiento 25 %
- Promedio ponderado Gaussiano en superposiciones
- Reduce artefactos de borde
- Reconstruye volúmenes completos sin límite de tamaño

1. Coeficiente Dice (DSC):

$$DSC = \frac{2|P \cap G|}{|P| + |G|}$$

- Mide solapamiento volumétrico
- Rango: $[0, 1]$ (1 = perfecto)
- Métrica principal en MICCAI

2. Distancia Hausdorff (HD95):

- Evalúa precisión de contornos
- Percentil 95 (elimina outliers)
- Unidades: milímetros
- Crítico para forma geométrica

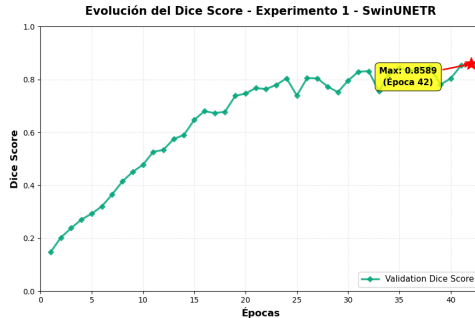
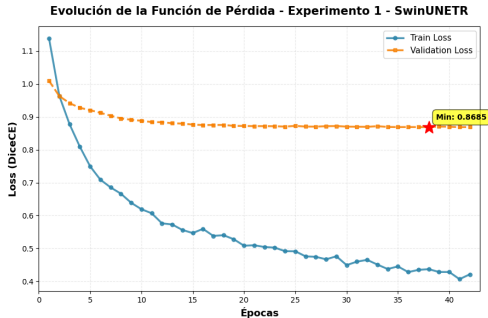
DSC ↑ HD95 ↓

Modelo	Dice Score \uparrow	HD95 (mm) \downarrow	Época Mejor
Baseline U-Net 3D	0.852	4.52	-
ResUNet 3D	0.8274	50.08	64/100
Swin UNETR	0.8689	-	49/50

Observaciones:

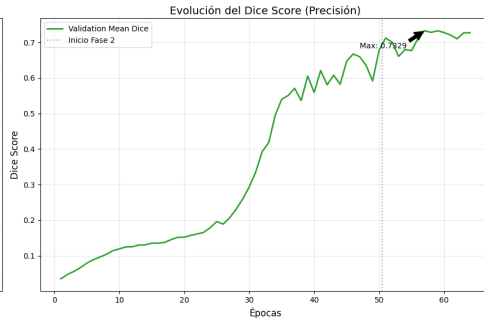
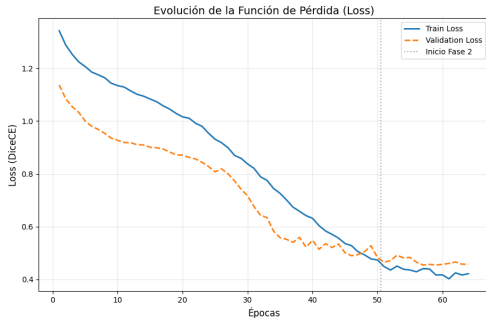
- ResUNet: **+1.7 %** Dice vs baseline
- ResUNet: HD95 elevado sugiere refinamiento de bordes necesario

Curvas de Aprendizaje: Swin UNETR



- Convergencia rápida (época 20-30)
- Dice máximo: 0.8689 (época 44)
- El gap nos hace pensar en la existencia de overfitting

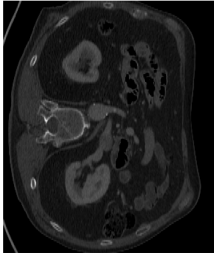
Curvas de Aprendizaje: ResUNet 3D



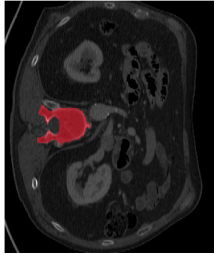
- Entrenamiento en 2 fases (50+50 épocas)
- Mejor validación loss: época 58
- Margen de mejora con ajuste fino

Resultados Cualitativos

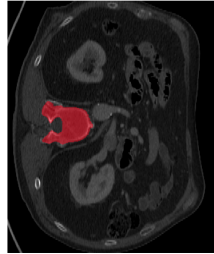
[Train] sub-verse588
CT Axial (Slice 385)



Ground Truth



Predicción Modelo



Proyección 3D Lateral (MIP)



Sujeto: sub-verse588

- Separación correcta de vértebras adyacentes
- Ausencia de "efecto tubo"
- Preservación de morfología vertebral

Resultados Cualitativos SwingUNetR

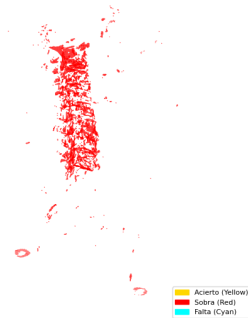
1. GT Real (Verde)



[Test] Test_sub-verse712_img.pt
2. Predicción (Gris)

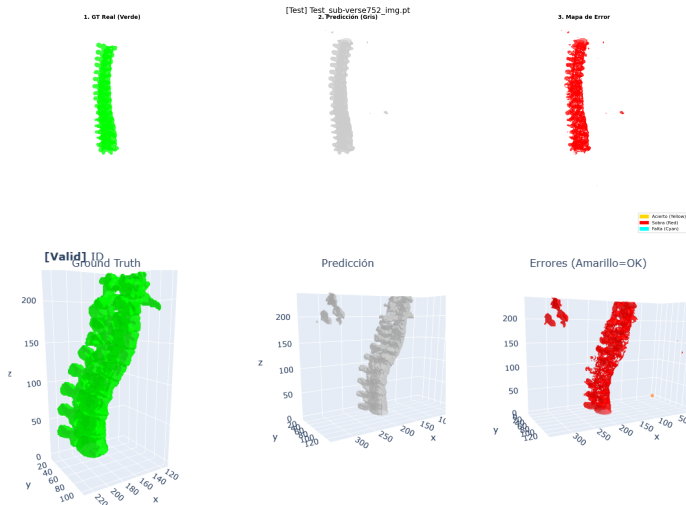


3. Mapa de Error



Sujeto: sub-verse712

Resultados Cualitativos ResUNet3D



Sujetos: sub-verse752 y sub-verse643

Conclusiones Principales

1. Preprocesamiento es Crítico

Reorientación RAS + remuestreo isotrópico → **Esencial** para convergencia

2. Swin UNETR Supera CNN Puras

Auto-atención global captura mejor contexto anatómico (+1.7 % Dice)

3. Sliding Window Viable

Reconstrucción de volúmenes completos en hardware limitado

Limitaciones:

- Requiere GPUs de alta gama (A100 40GB)
- Segmentación binaria (no identifica vértebras individuales)

① Segmentación Multi-clase:

- Transfer learning para identificar C1-L5 individualmente
- Dataset con etiquetas por vértebra

② Validación Clínica:

- Evaluación por radiólogos certificados
- Integración con PACS hospitalarios

¡Gracias por su atención!

¿Preguntas?

`{rafacarrillo, dachival, miluroma, robinjr}@correo.ugr.es`

-  Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. MICCAI 2015.
-  Milletari, F., Navab, N., & Ahmadi, S. A. (2016). *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. 3DV 2016.
-  Sekuboyina, A., et al. (2021). *VerSe: A Vertebrae Labelling and Segmentation Benchmark for Multi-detector CT Images*. Medical Image Analysis.
-  Hatamizadeh, A., et al. (2022). *Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images*. BrainLes Workshop @ MICCAI 2021.