

Proyecto Final: Métodos de DL para análisis de tractografía mediante datos de dMRI

Tópicos Selectos de Ciencia de Datos: Deep Learning
Centro de Investigación en Matemáticas, A.C.

Gabriel Alejandro Aguilar Farrera
David Alejandro Muro Campa

Contenido

1. Objetivos del Proyecto
2. Aspectos generales
3. Introducción
4. Metodología
5. Evaluación
6. Resultados

Objetivos del Proyecto

Aspectos Generales

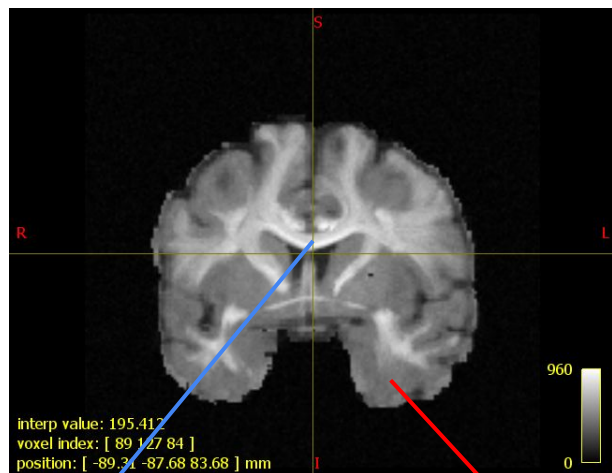


Fig. (1): MRI (Vista Coronal).

Materia Blanca
(Cuerpo Calloso).

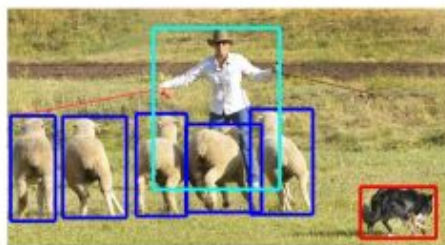
Materia Gris

1. **Tractografía de materia blanca** es un procedimiento que se usa para poner en manifiesto los tractos neuronales, así como determinar conectividad cerebral. Utiliza técnicas especiales de imagen por resonancia magnética (MRI por sus siglas en inglés) y de procesamiento de imágenes. [6]
2. Determinar la conectividad cerebral es uno de los principales objetivos médicos del análisis moderno de imágenes de resonancia magnética dado que este conocimiento se relaciona con el estado de algunas enfermedades fisiológicas del cerebro (como la apoplejía o embolia cerebral) y también puede ser usada para investigar desórdenes neurológicos (por ejemplo problemas del lenguaje) y en general el desarrollo cerebral [1].
3. La **segmentación de los tractos de la materia blanca** permite un análisis detallado de los tractos individuales de la materia blanca. Ayuda a caracterizar el cerebro sano e identificar áreas que contienen morfología anormal en cerebros enfermos. Tales análisis de tractometría dependen intrínsecamente de la calidad de la segmentación de tractos subyacentes, y por lo tanto, requieren una delimitación precisa.

Introducción: ¿Qué es la segmentación semántica?



(a) Image classification



(b) Object localization



(c) Semantic segmentation



(d) This work

Fig. (2): Segmentación Semántica [5].

Our Work

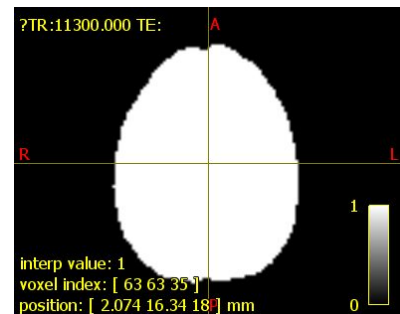


Fig. (3): Máscara del Cerebro completo.

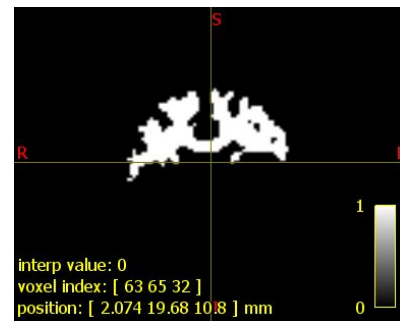


Fig. (4): Máscara del Cuerpo Calloso.

Introducción: Arquitectura U-NET Part I

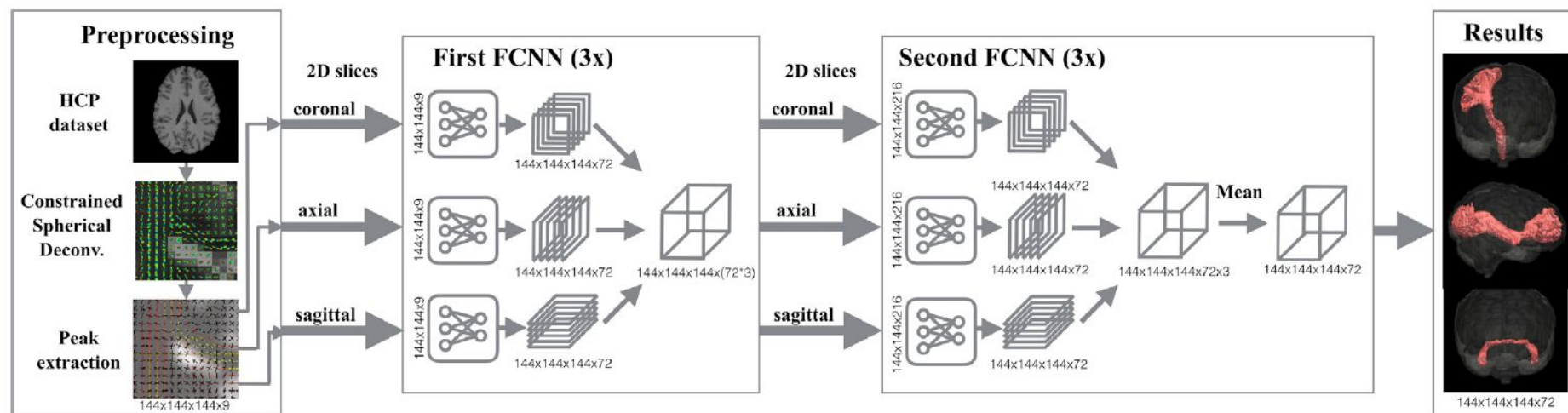


Fig. (5): TracSeg Segmentation Pipeline.

Introducción: Arquitectura U-NET Part II

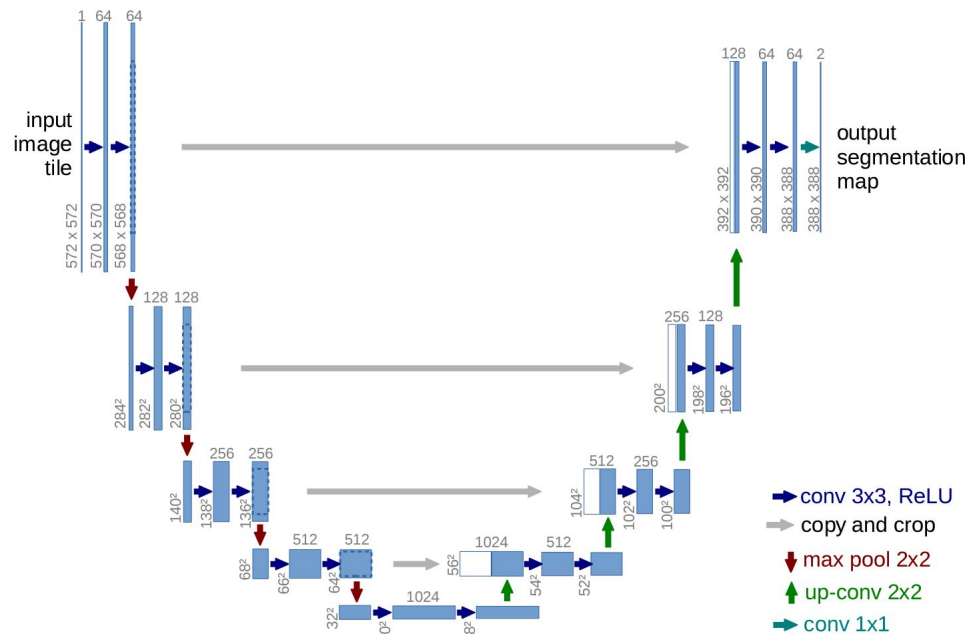
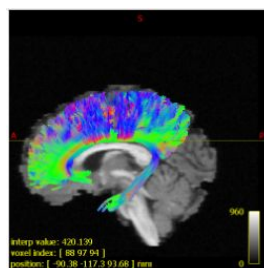


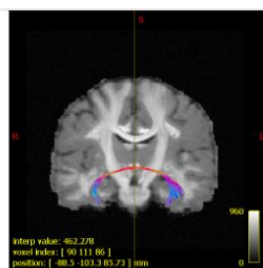
Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Metodología

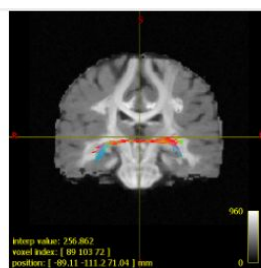
Los datos corresponden a la base de datos Renault 2023, los cuales se obtuvieron del **ISMRM 2015 Tractography Challenge**. Los datos contienen 21 bundles o máscaras (archivos con extensión .nii), tractografía por cada uno de los bundles (archivos con extensión .tck), la imagen pesada por difusión (DWI, por sus siglas en inglés), la tabla de gradientes (archivo con extensión .bvec) y finalmente el archivo de b-valores (archivo con extensión .tck). **Todo esto es corresponde a un solo paciente.**



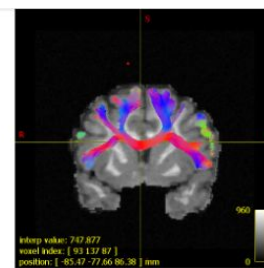
Cíngulo Der-Izq



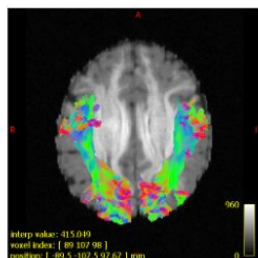
Comisura Anterior



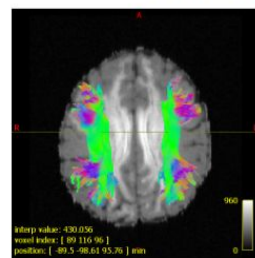
Comisura del Papez



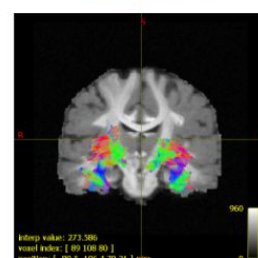
Cuerpo Calloso



Fascículo Longitudinal Inferior Der-Izq



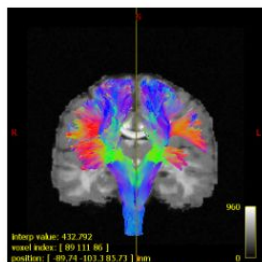
Fascículo Longitudinal Superior Der-Izq



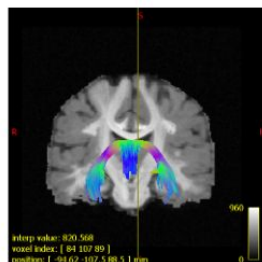
Fascículo Uncinado Der-Izq

Metodología

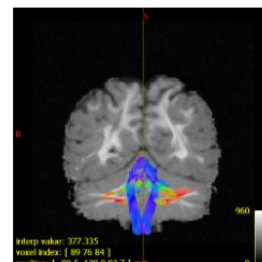
Los datos corresponden a la base de datos Renault 2023, los cuales se obtuvieron del **ISMRM 2015 Tractography Challenge**. Los datos contienen 21 bundles o máscaras (archivos con extensión .nii), tractografía por cada uno de los bundles (archivos con extensión .tck), la imagen pesada por difusión (DWI, por sus siglas en inglés), la tabla de gradientes (archivo con extensión .bvec) y finalmente el archivo de b-valores (archivo con extensión .tck). **Todo esto es corresponde a un solo paciente.**



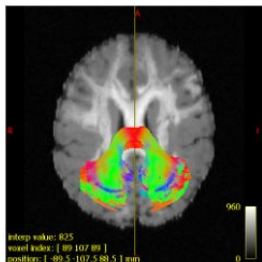
Pedúnculos Cerebrales Der-Izq



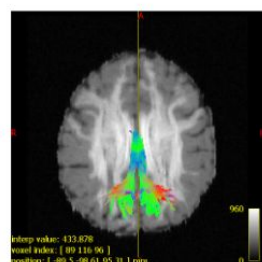
Fórnix



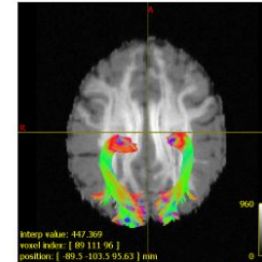
Pedúnculo Cerebeloso Inferior Der-Izq



Pedúnculo Cerebeloso Medio



Pedúnculo Cerebeloso Superior Der-Izq



Radiación Óptica Der-Izq

Metodología

1. Adquisición de Datos de Pre Entrenamiento

- Descargar datos del Human Connectome Project (HCP) que serán los datos que utilizaremos para el pre entrenamiento.
- Utilizar la imagen de difusión de los datos descargados para calcular los picos utilizando la técnica de "Constrained Spherical Deconvolution" (CSD).
- Una vez obtenidos los picos podemos obtener las segmentaciones de los bundles con TractSeg. TractSeg nos da 72 bundles.

The screenshot shows the 'CONNECTOME db' website interface. At the top, there's a search bar and navigation links. The main heading is 'Download Packages: WU-Minn HCP Data - 1200 Subjects'. Below this, there's a section 'Select Packages to Download:' with instructions to click an icon or package title to add it to the download queue. On the right, it shows 'Total Queued: 0 packages: 0 files, 0 B'. Below this, there are buttons for 'Select All', 'Clear Selection', and 'Download Packages'. On the left, there are filters for 'Session Type' (3T MRI), 'Processing Level' (Preprocessed), and 'Modalities' (Structural, Resting State Denoised, Resting State, Task, Diffusion). The 'Diffusion' modality is selected. A package titled 'Diffusion Preprocessed' is listed with a description: 'This package contains Diffusion data preprocessed with the HCP diffusion pipeline (updated with EDDY 5.0.10), including diffusion weighting (bvals), direction (bvecs), time series, brain mask, a file (grad_dev.nii.gz) that can be used to account for gradient nonlinearities during model fitting, and log files of EDDY processing.' It also shows '492 of 526 subjects OK' and '7,380 files, 600.44 GB'. A button '+ queue for download' is next to it. At the bottom, there's a 'Reset Filters' button and another set of 'Select All', 'Clear Selection', and 'Download Packages' buttons, with 'Total Queued: 0 packages: 0 files, 0 B'.

Metodología

2. Identificación de Bundles

- Analizar las segmentaciones obtenidas y clasificar los bundles en dos categorías: "bundles existentes" y "bundles novel".
- Bundles existentes son aquellos bundles de nuestros datos que TractSeg ya segmenta.
- Bundles novel son aquellos bundles de nuestros datos que TractSeg no segmenta.

Bundles Existentes	Bundles Novel
CA	BPS_left
CC	BPS_right
ICP_left	CP
ICP_right	Fornix
ILF_left	SLF_left
ILF_right	SLF_right
MCP	Cingulum_left
OR_left	Cingulum_right
OR_right	-
SCP_left	-
SCP_right	-
UF_left	-
UF_right	-

Cuadro 1: Tabla de Bundles Existentes y Novel.

Metodología

3. Extracción de Datos para el Pre entrenamiento y el Entrenamiento:

- Extraer los picos y las segmentaciones correspondientes a los bundles existentes y novel de los datos de pre entrenamiento.
- Organizar los datos para que cada paciente tenga los picos y los bundles existentes, estableciendo así los conjuntos de pre entrenamiento.
- Organizar los datos de nuestro único paciente con los picos y los bundles novel

```
Pretraining/HCP_preproc/201818/  
    '-> mrtrix_peaks.nii.gz  
    '-> bundle_masks_13.nii.gz  
Pretraining/HCP_preproc/205220/  
    '-> mrtrix_peaks.nii.gz  
    '-> bundle_masks_13.nii.gz  
...
```

```
Training/HCP_preproc/Renauld/  
    '-> mrtrix_peaks.nii.gz  
    '-> bundle_masks_8.nii.gz
```

Metodología

4. Fase de Pre Entrenamiento

- Una vez acomodados los datos podemos realizar el pre entrenamiento de tal forma que obtenemos un modelo vía TractSeg como modelo base que segmenta únicamente en los 13 bundles existentes. El pre entrenamiento se realiza siguiendo la estrategia que se utiliza en [2]. TractSeg no está optimizado para realizar el entrenamiento de tu propio modelo por lo que se deben seguir una serie de pasos que se enlistan a continuación
 1. Realizar el procesamiento con *proprocessing.py* en la que hay que adaptar los paths
 2. Adaptar los paths en el archivo *my_custom_experiment.py*
 3. Crear un archivo *config.txt* con el nombre de la carpeta donde están tus datos
 4. Adaptar la función *get_bundle_names()* y *get_labels_file_name()* del script *dataset_especif_utils.py* donde establezcas los nombres de los bundles existentes y novel.
 5. Adaptar el archivo *subjects.py* con los id de los pacientes que vas a utilizar
 6. Adaptar *ExpRunner.py* con los paths de tus datos
 7. Ejecutar *ExpRunner.py*

Metodología

5. Fase de Entrenamiento (Fine Tuning):

- Aplicar data augmentation mediante Random Cut Out [2] a nuestro único paciente para generar 63 nuevas instancias de pacientes. En la siguiente Figura apreciamos como se debe organizar la carpeta
- Asignar 48 instancias para el conjunto de entrenamiento y 15 para el conjunto de validación.
- Realizar fine-tuning siguiendo la estrategia de [2] para ajustar el modelo a la segmentación de los 18 bundles novel. Como su modelo base es TractSeg igual el fine-tuning no está optimizado por lo que se deben hacer una serie de pasos:
 1. Modificar el el archivo *TractSeg_Fewshot_Pretrain_base.py* indicando la fase de fine-tuning, el path del modelo pre entrenado y de los datos que usaremos.
 2. Adaptar el archivo *TractSeg_Fewshot_Pretrain_ExpRunner.py* con los ID de los pacientes que utilizaremos
 3. Ejecutar *ExpRunner.py*.

```
Training/HCP_preproc/daug_01/  
    '-> mrtrix_peaks.nii.gz  
    '-> bundle_masks_8.nii.gz  
Training/HCP_preproc/daug_02/  
    '-> mrtrix_peaks.nii.gz  
    '-> bundle_masks_8.nii.gz  
...
```

Metodología

6. Evaluación del Modelo:

- Para testear el modelo lo evaluamos en los picos de nuestro paciente de entrenamiento ya que no tenemos otros datos para realizar un test. La justificación es que el modelo se entrenó con transformaciones de éste y no con esos datos en sí.

Métrica de Evaluación: Dice Score (DS)

El Coeficiente de Dice (DICE) (índice de Superposición) es la métrica más utilizada en la validación de segmentaciones de volúmenes médicos.

DICE se calcula de la siguiente manera:

$$\text{DICE} = \frac{2 |S_g^1 \cap S_t^1|}{|S_g^1| + |S_t^1|} = \frac{2TP}{2TP + FP + FN}$$

Donde S_g^1 es la segmentación de referencia (Ground Truth) y S_t^1 es la segmentación de prueba

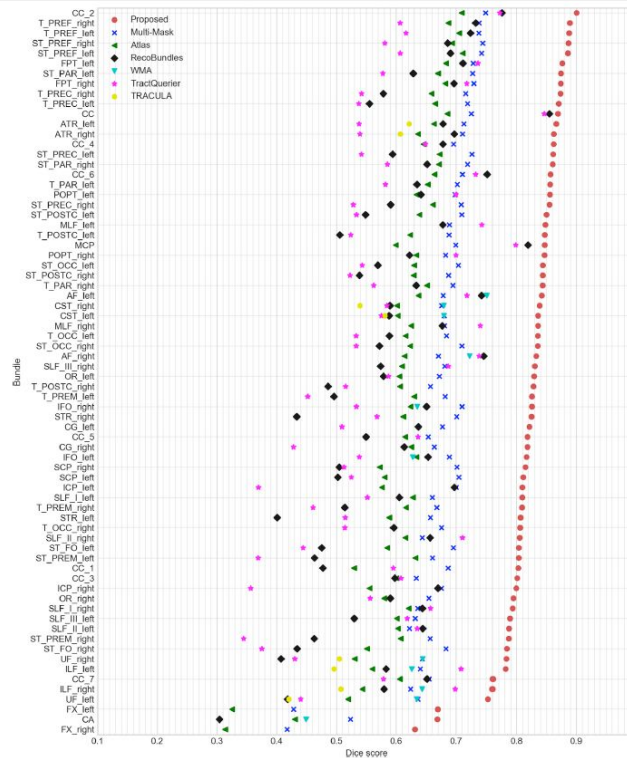
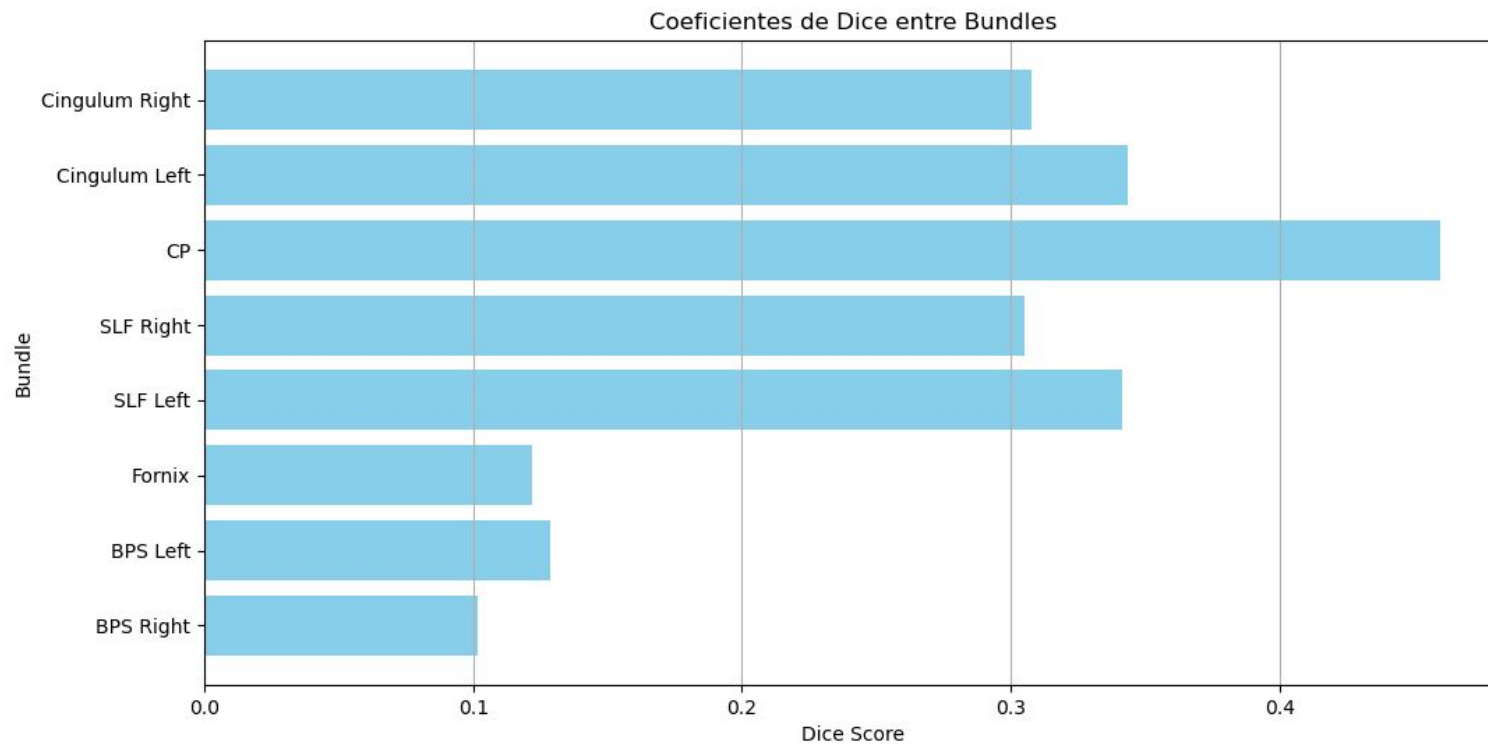
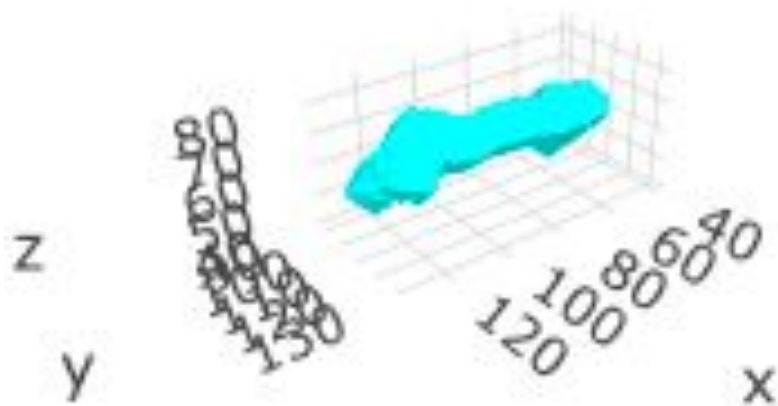


Fig. (6): Dice Score para cada uno de los 72 bundles

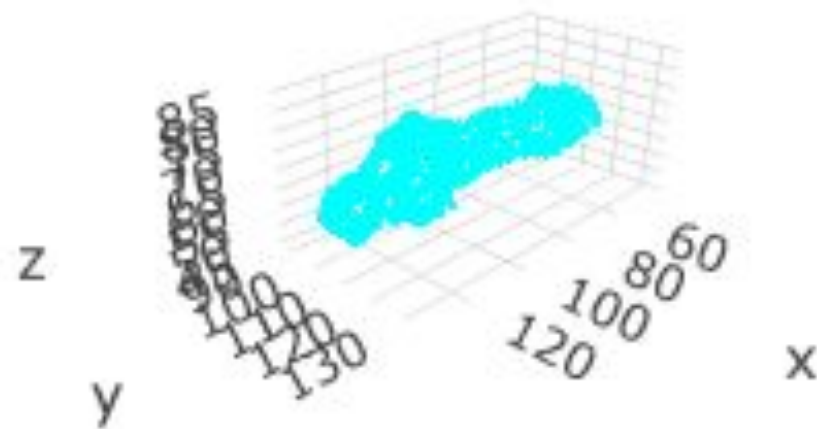
Resultados



Resultados: CP con DS = 0.459

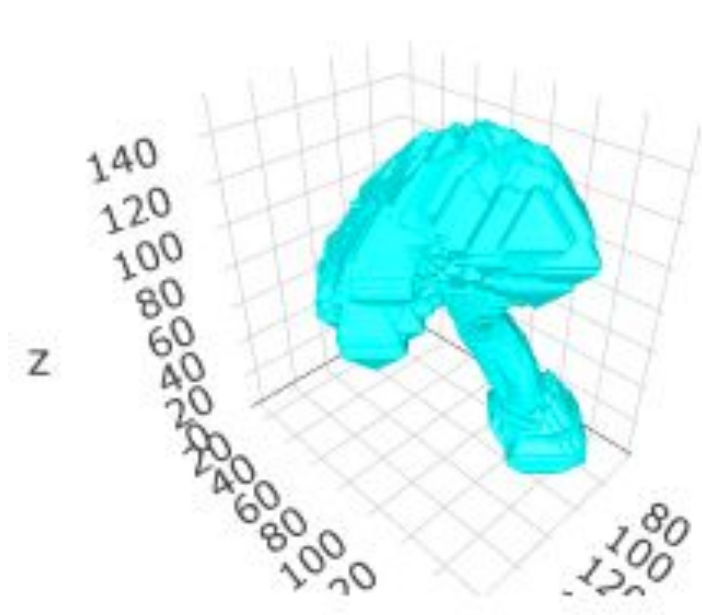


CP original

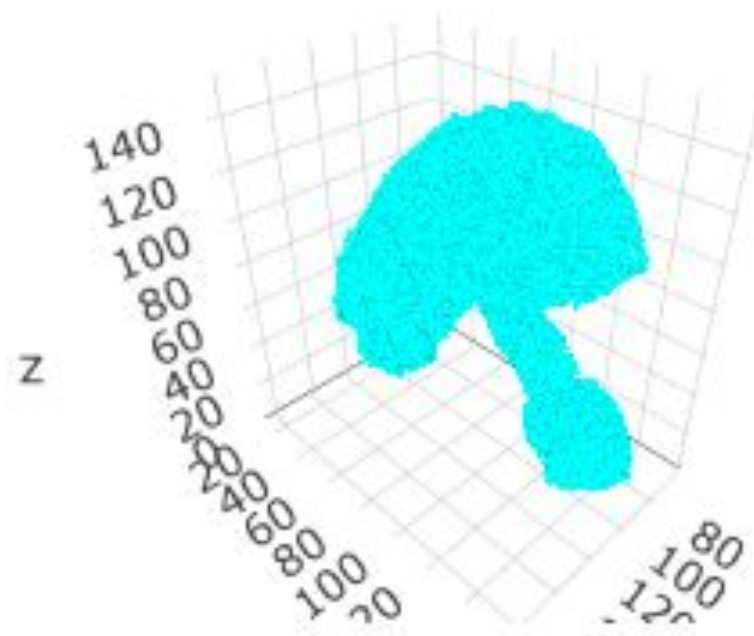


CP predicho

Resultados: BPS left con DS = 0.128



BPS left original



BPS left predicho

Conclusiones

- La metodología, centrada en la identificación y segmentación de "bundles existentes" y "bundles novel", ha demostrado ser una aproximación valiosa, especialmente en situaciones donde la disponibilidad de datos de entrenamiento es limitada, como en el caso de contar con un único paciente para entrenamiento.
- Los resultados obtenidos son aceptables para la complejidad del problema y para la metodología que se implementó.
- La mejora del dice score para datos limitados es aún objeto de investigación pues un problema muy frecuente al que se enfrentan los investigadores y profesionistas en el campo de la neurología y neuroimagen.
- La morfología y tamaño de los bundles influyen mucho en el valor del *dice score*.

Referencias

- [1]: Wasserthal, J., Neher, P., & Maier-Hein, K. H. (2018). TractSeg-Fast and accurate white matter tract segmentation. *NeuroImage*, 183, 239-253.
- [2]: Liu, W., Zhuo, Z., Liu, Y., & Ye, C. (2023). One-shot segmentation of novel white matter tracts via extensive data augmentation and adaptive knowledge transfer. *Medical Image Analysis*, 90, 102968.
- [3]: Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18 (pp. 234-241). Springer International Publishing.
- [4]: Tournier, J. D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., ... & Connelly, A. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage*, 202, 116137.
- [5]: Schilling, K. G., Rheault, F., Petit, L., Hansen, C. B., Nath, V., Yeh, F. C., ... & Descoteaux, M. (2021). Tractography dissection variability: What happens when 42 groups dissect 14 white matter bundles on the same dataset?. *Neuroimage*, 243, 118502.
- [6] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V* 13 (pp. 740-755). Springer International Publishing.
- [7] Aranda Campos, Á. R. (2010). Partículas masivas para tractografía cerebral usando multi-tensores de difusión.
- [8] Alonso Ramírez-Manzanares. Estimación de las trayectorias de manojes de nervios cerebrales, por medio de la restauración y el filtrado de la información inte-voxel en imágenes de tensor de difusión de resonancia magnética. Master's thesis, Centro de Investigación en Matemáticas (CIMAT) A.C., 2003 2003. 1