

DeepCeres: red neuronal artificial para la segmentación de lóbulos del cerebelo a partir de imágenes multimodales de RM de alta resolución

Sergio Morell Ortega ¹, Marina Ruiz ¹, Pierrick Coupé ² y J.V Manjón ¹

¹ Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, España

² CNRS, Univ. Bordeaux, Bordeaux INP, LABRI, UMR5800, PICTURA, F-33400 Talence, Francia
Email: sermoor1@teleco.upv.es

Resumen

En el presente trabajo se ha diseñado e implementado una red neuronal artificial (RNA) volumétrica para la segmentación de lóbulos cerebelares a partir de imágenes multimodales (T1 y T2) de resonancia magnética nuclear de alta resolución (0.5 mm^3). La herramienta permite mejorar la precisión en la estimación de volúmenes de subestructuras cerebelares (lóbulos) a partir de imágenes de RM de alta resolución, aumentando la capacidad de herramientas actuales (CERES, RASCAL, etc) que funcionan a resolución estándar (1 mm^3) y no pueden medir con precisión la intrincada estructura cerebelar (con grososres corticales submilimétricos en algunas áreas).

1. Introducción

El cerebelo es la parte del encéfalo situada en la la región posterior de la cabeza, entre el cerebro y el tronco encefálico. Aunque sólo representa entorno al 10 % de la masa total del encéfalo, contiene aproximadamente el 50 % de las neuronas del cerebro, gracias al elevado grado de empaquetamiento neuronales en la superficie de los lóbulos, las subdivisiones anatómicas del mismo [1].

La era moderna en el estudio del cerebelo comenzó hace 120 años, cuando Santiago Ramón y Cajal publicó su primer trabajo. Este trabajo ha sido continuado en los últimos años, y tanto la comprensión de la función normal del cerebelo como de sus patologías, se ha consolidado como un foco activo en el campo de la neurociencia [2]. Durante décadas, se ha tenido la noción de que el cerebelo estaba exclusivamente involucrado en el control motor, pero en los últimos años la neurociencia entorno a esta estructura ha sufrido un cambio de paradigma. Las observaciones neuroanatómicas de imágenes cerebrales han recontextualizado el cerebelo como un nodo crítico en los circuitos neuronales distribuidos que sirven al comportamiento. La idea de que el cerebelo modula el pensamiento y la emoción de la misma manera que modula el control motor, avanza en la comprensión de los mecanismos de la cognición, y abre nuevas oportunidades terapéuticas en neurología conductual y neuropsiquiatría [3]. Por lo tanto, existe la necesidad de disponer de nuevos métodos que permitan el estudio objetivo y cuantitativo del cerebelo, dada la complejidad del problema derivada de la baja resolución de la imágenes en la clínica o la falta de herramientas automáticas precisas, entre otros factores.

El presente trabajo presenta una solución automática, robusta y eficiente para el problema de la segmentación del cerebelo, facilitando así el progreso en la comprensión

en profundidad de esta estructura. Para ello se hace uso de algoritmos basados en redes neuronales convolucionales, que en los últimos años han demostrado una gran aplicabilidad en el campo del análisis de imagen médica, complementando e incluso mejorando los resultados de algoritmos basados en enfoques clásicos [4].

2. Materiales

En este trabajo se han usado 75 casos de la base de datos del *Human Connectome Project* con imágenes de resonancia magnética (RM) multimodales T1 y T2 de una resolución de $0,7 \text{ mm}^3$.

Para poder generar las segmentaciones de los 75 casos se ha empleado el *software* CERES [5], teniendo en cuenta que este método era el mejor según la comparativa del estado del arte [6]. Como resultado se generó mapas de segmentación con 28 etiquetas, siguiendo el protocolo de segmentación de Schmahmann [7] recogido en [6].

2.1. Generación de los datos de entrenamiento

Para generar las segmentaciones de los casos se ha llevado a cabo el siguiente preprocesado: eliminación de ruido, corrección de inhomogeneidad en el espacio nativo, registro afín al espacio MNI152 (a una resolución de $0,5 \text{ mm}^3$), corrección de inhomogeneidad en el espacio MNI, recorte local del área del cerebelo y normalización de intensidad. Teniendo en cuenta que el *software* de segmentación CERES funciona a una resolución de 1 mm^3 se utilizó un decimado basado en *strides* para la descomposición y posterior reconstrucción de las segmentaciones, tal y como se observa en la Figura 1.

2.2. Corrección manual de la materia blanca

Tras procesar las imágenes se observó un subsegmentación de la sustancia blanca dentro de los lóbulos del cerebelo. En consecuencia se realizó un proceso de mejora de la segmentación de dicha estructura por parte de un anotador experto.

El resultado se puede observar en la Figura 2, donde se representa la sustancia blanca para ambos hemisferios, también conocida como árbol de vida por los anatomistas. Para aumentar el tamaño muestral se realiza el volteado horizontal de los casos, generando 150 casos. De los cuales 130 se usaron en la partición de entrenamiento, 10 en la de validación y 10 en test.

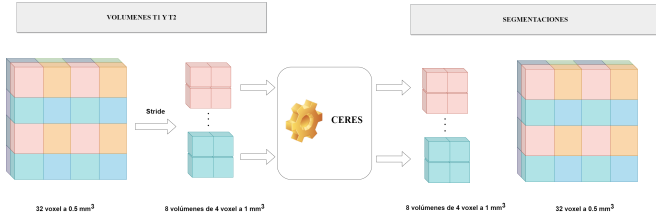


Figura 1: Método para generar las segmentaciones de la librería con CERES.

3. Métodos

3.1. Diseño de las arquitecturas de *Deep Learning*

Teniendo en cuenta las limitaciones computacionales propias al tratar con datos volumétricos de alta dimensionalidad (tensores de entrada de 252x156x162 voxels), con dos canales (T1 y T2) y con una alta densidad de etiquetas (27), se enfocó el diseño en una arquitectura en cascada. Tal y como se puede ver en la Figura 3 primero se realiza una segmentación del cerebelo en hemisferio derecho, izquierdo y fondo. Esta segmentación sirve como máscara cerebelo-fondo para los datos de entrada de la segunda red, facilitándole la atención intra-cerebelar.

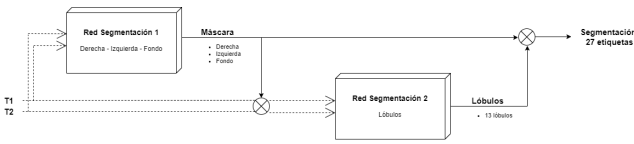


Figura 3: Arquitectura de segmentación.

En cuanto a las redes empleadas en las dos etapas, se ha experimentado con dos topologías:

- La clásica U-net [8], ampliamente validada por la comunidad científica en tareas de segmentación.
- Una adaptación de la anterior realizada en el grupo de investigación donde se sustituye la parte del *encoder* por sucesivas capas de *average pooling*. La arquitectura que hemos denominado DPN (*Deep Piramidal Network*) libera aproximadamente la mitad de memoria que una U-net completa, y funciona usando un enfoque *coarse-to-fine*, empezando con una imagen de una resolución 1/8 de la original y va refinando el resultado a medida que avanza el *decoder*.

Ambas dos topologías han sido entrenadas en las dos tareas (red de derecha-izquierda y lóbulos) para minimizar la siguiente función de pérdida

$$\mathcal{L} = \log \left(\frac{(1 - \sum_{c=1}^C Dice(y_c, \hat{y}_c)) + BCE(y, \hat{y})}{2} \right) \quad (1)$$

Donde C son las clases (etiquetas), \hat{y}_c es el tensor de salida para la clase c e \hat{y}_c es el tensor de salida para esa misma

clase. Dice es el coeficiente empleado para medir la similitud entre las segmentaciones (1 cuando son idénticas y 0 al contrario), y BCE es la entropía binaria cruzada. Adicionalmente, el operador logaritmo ha permitido aliviar los problemas de desvanecimiento del gradiente y acelerar la convergencia durante el entrenamiento.

3.2. Introducción de atlas como canal adicional

Hasta el advenimiento de las técnicas de aprendizaje profundo la mayoría de los métodos (como CERES) estaban basados en uso de un atlas. Estos métodos se caracterizaban por su robustez e interpretabilidad, a costa de una precisión limitada y con un coste computacional elevado. Actualmente, los métodos basados en aprendizaje profundo permiten obtener precisiones más elevadas con un mayor rendimiento temporal y computacional, pero en ocasiones carecen de la robustez de los métodos pasados.

Debido a esto se propone incorporar una presegmentación auxiliar como una información *a priori* generada mediante métodos basados en atlas. De esta forma el modelo cuenta con una información previa que le facilita la comprensión semántica de las estructuras del cerebelo.

Con el objetivo de aumentar la robustez en la generación de la segmentación auxiliar, se emplea un enfoque multi-atlas, donde se realiza un registro no lineal de cada imagen de la librería de entrenamiento (casos de HCP con sus T1 y correspondientes segmentaciones) a la imagen a segmentar. Posteriormente se aplican los registros a las segmentaciones, creándose así una librería específica de segmentaciones. Finalmente, se fusionan siguiendo el algoritmo de *Weighted Majority Voting*, donde para cada píxel se asigna la etiqueta resultado de pesar las etiquetas de cada caso de la librería por la similitud del nivel de intensidad de la T1 con respecto a la T1 del casos a segmentar.

Cabe destacar que el registro no lineal con técnicas habituales tiene un elevado coste temporal, así como la fusión de las etiquetas. Para la tarea de registro se ha empleado una red basada en aprendizaje profundo (una adaptación de *voxelmorph*[9]), y para la fusión de las etiquetas se ha implementado un versión multiproceso en *OpenMP* integrable en Python mediante la API *Numpy-C*. La construcción de la segmentación auxiliar *a priori* se podría resumir en la Figura 4.

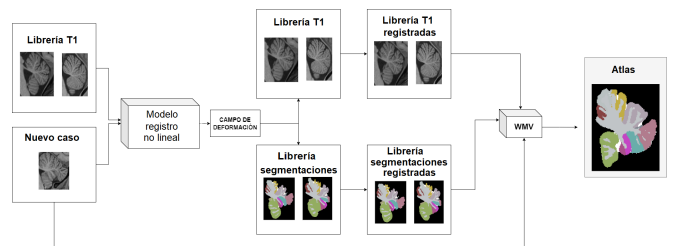


Figura 4: Método para la construcción de la segmentación auxiliar (multi-atlas)

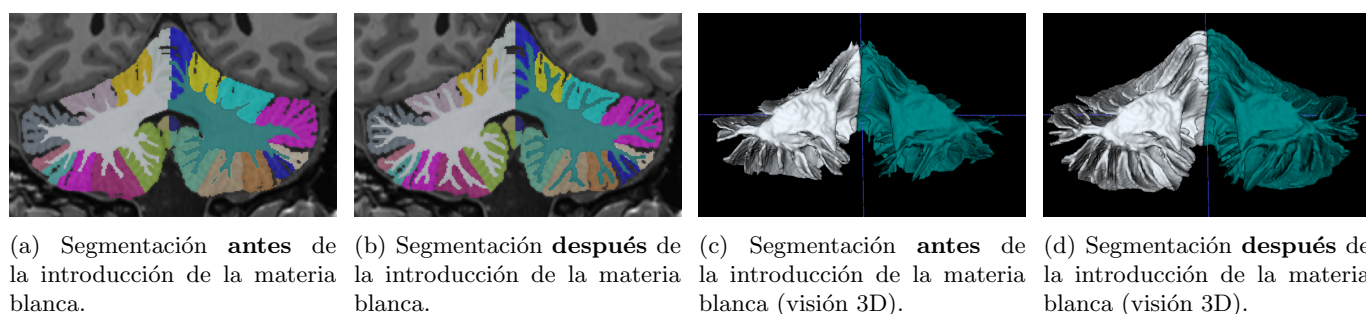


Figura 2: Vista coronal de las segmentaciones antes de la corrección manual y después.

4. Resultados y discusión

4.1. Resultados generales

Todos los experimentos se realizaron en una plataforma DGX1 de Nvidia con 8 GPUs V100 con 32 GB de memoria cada una en un entorno Linux (Ubuntu 19). Se utilizó la versión 1.15 de Tensorflow y la versión 2.2.4 de keras. Adicionalmente se realizaron los tests en equipos con GPUs con 16 GB de memoria para validar su aplicabilidad en equipos con menos prestaciones computacionales.

Se analizó el impacto general de las distintas arquitecturas UNET (3 433 473 parámetros) vs DPN (696 177 parámetros) o el uso de información *a priori* (con o sin atlas). En la Tabla 1 se recogen los valores del índice DICE de las variantes estudiadas y su comparación con el método de referencia CERES.

Nombre	Dice medio	Dice Cerebelo
CERES	0.7729 ± 0.0427	0.9377 ± 0.0090
2 DPN	0.9248 ± 0.0367	0.9887 ± 0.0028
2 Unet	0.9295 ± 0.0438	0.9900 ± 0.0022
2 DPN multi-atlas	0.9342 ± 0.0342*	0.9917 ± 0.0015
2 Unet multi-atlas	0.9347 ± 0.0339*	0.9903 ± 0.0020

Tabla 1: Análisis de las arquitecturas propuestas. (Resaltado el mejor de los resultados en cada caso en negrita). * Diferencias significativas para un test Wilcoxon $p < 0,05$)

En primer lugar, es posible observar la mejora sustancial con respecto al índice Dice del método CERES. No obstante, esta comparativa no es completa, pues los datos de empleados para el *test* fueron distintos. Sin embargo, da una idea orientativa de la ganancia obtenida tras la mejora de la librería, la introducción de la potenciación en T2 y la disponibilidad de imágenes de alta resolución.

Es posible ver en la Tabla 1 como la introducción del atlas ha tenido un impacto importante a nivel de lóbulos y de cerebelo completo, tanto para los modelos basados en la arquitectura DPN como en la Unet. En ambos ca-

sos, conceptualmente la red se especializa en la extracción semántica de los detalles finos de los lóbulos, sin necesidad de extraer características desde cero, además de acelerar la convergencia de los entrenamientos.

4.2. Impacto de la multimodalidad y alta resolución

Tras la selección de los modelos con atlas como aquellos con mejor desempeño, se ha comprobado el impacto que tiene el uso de imágenes de alta resolución ($0,5 \text{ mm}^3$ de resolución espacial) y la introducción de la imagen potenciada en T2.

Para ello se han entrenado los modelos de nuevo siguiendo la misma estrategia que los originales, pero en este caso eliminando el canal de la imagen potenciada en T2 o diezmando los volúmenes de entrada para reducir la resolución. Los resultados se pueden ver en la Tabla 2.

Resolución, Multimodalidad	Tipo de red	Dice medio
BR, T1	2 DPN	0.9117 ± 0.0473
BR, T1	2 UNET	0.9079 ± 0.0520
BR, T1+T2	2 DPN	0.9138 ± 0.0468
BR, T1+T2	2 UNET	0.9088 ± 0.0468
AR, T1	2 DPN	0.9239 ± 0.0353
AR, T1	2 UNET	0.9276 ± 0.0390
AR, T1 + T2	2 DPN	0.9332 ± 0.0439
AR, T1 + T2	2 UNET	0.9345 ± 0.0493

Tabla 2: Impacto de la resolución y la multimodalidad para los dos mejores modelos con atlas específico (para las 2 redes). AR: Alta Resolución (0.5 mm^3) y BR: Baja Resolución (1 mm^3). (* Mejores resultados resaltados en negrita).

Es posible observar como el aumento de resolución tiene un gran impacto sobre el índice Dice a nivel de lóbulos (0.9345 vs 0.9088 en UNET y 0.9332 vs 0.9138 en DPN). Cabe recordar la intrincada estructura del cerebelo, donde existe un elevado grado de empaquetamiento de tejidos y una superficie con un alto grado de textura, la folia, con grosores corticales submilimétricos. Esto provoca problemas de volumen parcial en imágenes de resolución estándar (1 mm^3) donde distintos tejidos aportan información sobre un mismo voxel, impidiendo su clasificación independiente. En alta resolución los problemas de volumen parcial se reducen. Por otra parte, pese a

que la ganancia de introducir la T2 es menor (0.9345 vs 0.9276 en UNET y 0.9332 vs 0.9239 en DPN), se observó un impacto positivo ya que permite reducir algunos falsos positivos y negativos al dar información de otra modalidad para una misma estructura anatómica, como podría ser la sustancia blanca y el líquido (vasos sanguíneos) que brillan de forma similar en T1 y distinta en T2.

Finalmente, con respecto a la disponibilidad en la práctica clínica de datos multimodales y de alta resolución, cabe destacar que datos de esta calidad son escasos. No obstante, y pese a que los modelos propuestos presentan rendimientos superiores a otros métodos incluso a una resolución estándar (1 mm³) se está trabajando en el diseño de métodos basados en aprendizaje profundo para sintetizar volúmenes T2 y aumentar de resolución de los volúmenes T1 y T2 usando técnicas de superresolución.

4.3. Ensemble de modelos

Analizando los resultados de las dos arquitecturas usadas, vimos que algunas estructuras son segmentadas mejor por un modelo en comparación con el otro, por lo que se propone combinación de ambas arquitecturas para mejorar resultado en términos de índice Dice. En consecuencia se estudió la generación de un *ensemble* de modelos para la segmentación del cerebelo. **Como resultado, el modelo dual combinado presenta un índice dice a nivel de lóbulo de 0.9403 ± 0.0296 , en comparación con el mejor modelo individual (Unet con atlas) con un Dice de 0.9347 ± 0.00339 (* Diferencias en las medias significativas para un test Wilconxon $p < 0,05$).**

4.4. Comparativa con los métodos del estado del arte

A modo de conclusión es posible realizar una comparativa semicuantitativa del método desarrollado y el estado del arte. Esta comparativa no puede ser totalmente cuantitativa puesto que los datos sobre los que se ha realizado son distintos, pero permite tener una noción del lugar que ocuparía la herramienta descrita en el panorama actual.

Como se puede apreciar en la Tabla 3 el método propuesto es muy superior (0.94) al segundo de la comparativa (0.87) lo que pone de manifiesto lo competitivo de la solución propuesta.

	Acapulco	Cerebnet	Ceres	Método propuesto
Dice Medio	0.77	0.87	0.7729	0.9403

Tabla 3: Comparativa con el resto de métodos del estado del arte.

5. Conclusiones

La fusión de los enfoques clásicos basados en atlas y los de aprendizaje profundo ha demostrado ser efectiva

para mejorar los resultados en ambas arquitecturas. Esta combinación ha dado lugar a modelos más precisos, robustos y eficientes, incluso superando a modelos más potentes pero sin la información previa. Además, se han explorado alternativas a la red UNET clásica, como la DPN, que ha mostrado buenos resultados con muchos menos parámetros (entorno a 1/5 parte), lo que reduce el riesgo de sobreajuste y mejora la generalización. La combinación de varios modelos, como UNET y DPN, ha proporcionado los mejores resultados, reduciendo la varianza del error de clasificación en la segmentación del cerebelo. Finalmente, la comparación con otros métodos del estado del arte, tanto clásicos como basados en *Deep Learning*, muestra que el enfoque propuesto es altamente competitivo, con índices Dice promedio considerablemente altos (0.87 vs 0.94), a pesar de las limitaciones de los datos y protocolos utilizados.

Agradecimientos

Este trabajo ha sido desarrollado gracias al proyecto PID2020-118608RB-I00 del Ministerio de Ciencia e Innovación de España.

Referencias

- [1] Martin I Sereno et al. The human cerebellum has almost 80 % of the surface area of the neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, 117(32):19538–19543, aug 2020.
- [2] Handbook of the Cerebellum and Cerebellar Disorders. *Handbook of the Cerebellum and Cerebellar Disorders*, 2022.
- [3] Jeremy D. Schmahmann et al. The Theory and Neuroscience of Cerebellar Cognition. <https://doi.org/10.1146/annurev-neuro-070918-050258>, 42:337–364, jul 2019.
- [4] Ana Barragán-Montero et al. Artificial intelligence and machine learning for medical imaging: A technology review. *Physica Medica*, 83:242–256, mar 2021.
- [5] Jose E. Romero et al. CERES: A new cerebellum lobule segmentation method. *NeuroImage*, 147:916–924, feb 2017.
- [6] Aaron Carass et al. Comparing fully automated state-of-the-art cerebellum parcellation from magnetic resonance images. *NeuroImage*, 183:150–172, dec 2018.
- [7] JD Schmahmann et al. *MRI atlas of the human cerebellum*. 2000.
- [8] Olaf Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, 9351:234–241, 2015.
- [9] Guha Balakrishnan et al. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, sep 2018.