



Universidad Católica
San Pablo

FACULTAD DE INGENIERÍA Y COMPUTACIÓN

**Escuela Profesional de Ciencia de la
Computación**

Trabajo de Investigación

**Segmentación 3D de Tumores Cerebrales
Eficiente en Memoria**

Presentado por:

Percy Maldonado Quispe

Para Optar el Grado Académico:

Bachiller en Ciencia de la Computación

Asesor: Dr. Juan Carlos Gutiérrez Cáceres

Arequipa, Marzo 2021

A mis padres por siempre estar conmigo, incluso por las distancia, a cada uno de mis profesores, compañeros y amigos que conocí a través de la carrera. A cada uno de ellos que me inspiraron y motivaron en el desarrollo de mi vida profesional.

Agradecimientos

En primer lugar deseo agradecer a Dios por haberme guiado a lo largo de estos cinco años de estudio.

Agradezco a mis padres por el apoyo brindado para forjarme como un profesional.

Agradezco a la universidad *Universidad Católica San Pablo*, por haberme cobijado y brindado la formación que ahora me permitirá ayudar a construir una mejor sociedad.

Agradezco de forma muy especial a mi orientador Prof. Dr. Juan Carlos Gutiérrez Cáceres por haberme guiado en el desarrollo de esta tesis.

Deseo agradecer de forma especial a mis docentes: Prof. Dr Alex J. Cuadros-Vargas, Prof. Dr Rensso Mora, Prof. Dr Jose Eduardo Ochos Luna porque fueron ejemplos que deseo seguir en mi vida profesional.

Resumen

El diagnóstico temprano y la segmentación precisa de los tumores cerebrales son imprescindibles para un tratamiento exitoso. Desafortunadamente, la segmentación manual es lenta, costosa y, a pesar de la amplia experiencia humana, a menudo es inexacta. En este documento, presentamos una arquitectura para la segmentación de tumores basado en imágenes MRI utilizando una red neuronal convolucional 3D regularizada con autoencoder. Entrenamos el modelo con imágenes *Magnetic Resonance Imaging* (MRI) segmentadas manualmente: T1, T1ce, T2 y Flair de 285 pacientes con tumores de gravedad, tamaño y ubicación variables. Luego probamos el modelo utilizando datos independientes de 66 pacientes y segmentamos con éxito los tumores cerebrales en tres subregiones: el núcleo del tumor (TC), el tumor potenciador (ET) y el tumor completo (WT). También se explora pasos de preprocesamiento para mejorar el rendimiento de la segmentación. Es importante destacar que nuestro modelo se implementó en una sola unidad gráfica y, por lo tanto, optimiza la segmentación tumoral para un hardware ampliamente asequible. En resumen, se trata de presentar una solución económica y eficiente en memoria para la segmentación tumoral para respaldar el diagnóstico preciso de los tumores cerebrales.

Palabras claves — Tumores Cerebrales, Imágenes MRI, Aprendizaje Máquina, Redes Neuronales, Codificadores.

Abstract

Early diagnosis and precise segmentation of brain tumors are essential for successful treatment. Unfortunately, manual segmentation is slow, expensive and, despite extensive human experience, it is often inaccurate. In this document, we present an MRI-based tumor segmentation architecture using autoencoder-regularized 3D-convolutional neural network. We trained the model with manually segmented MRI images: T1, T1ce, T2 and Flair of 285 patients with tumors of varying severity, size and location. We then tested the model using independent data from 66 patients and successfully segmented brain tumors into three subregions: the tumor core (TC), enhancer tumor (ET) and whole tumor (WT). Preprocessing steps are also explored to improve segmentation performance. Importantly, our model was implemented in a single graphics unit and, therefore, optimizes tumor segmentation for widely affordable hardware. In summary, we try to present an economical and efficient memory solution for tumor segmentation to support the precise diagnosis of brain tumors.

Keywords — Brain Tumors, MRI Images, Machine Learning, Neural Networks, Encoders.

Abreviaturas

CNN *Convolutional Neural Network*

MRI *Magnetic Resonance Imaging*

BRATS *Brain Tumor Segmentation Challenge*

HGG *High-Grade Gliomas*

LGG *Low-Grade Gliomas*

RM *Resonance Magnetic*

FLAIR *Fluid Attenuation Inversion Recovery*

SVM *Support Vector Machine*

NN *Neural Network*

CSF *Cerebrospinal fluid*

EMMA *Ensembles of Multiple Models and Architectures*

VAE *Variational Autoencoder*

Índice general

1. Introducción	2
1.1. Motivación y Contexto	3
1.2. Planteamiento del Problema	3
1.3. Objetivos	4
1.3.1. Objetivos Específicos	4
1.4. Organización de la Tesis	4
2. Marco Teórico	5
2.1. Introducción a las Redes Neuronales	6
2.2. Capas Convolucionales	6
2.2.1. Imágenes de Entrada	7
2.2.2. Kernels	8
2.2.3. Funciones de Activación	8
2.2.4. Dropout and Pooling Layers	9
2.3. Capas totalmente conectadas	10
2.4. Autoencoder	10
2.5. Variational Autoencoder	11
2.6. Entrenamiento	12
2.7. Secuencias de MR	13
2.8. Metodología	14
3. Estado del Arte	16

4. Propuesta	20
4.1. Arquitectura	20
4.1.1. Optimización	21
4.1.2. Arquitectura	22
4.2. Entrenamiento	23
5. Resultados	24
5.1. Interpretación de los resultados	24
5.1.1. Etiquetas	24
5.1.2. Métricas	24
5.2. Resultados	25
5.3. Segmentación	25
6. Conclusiones	28
Bibliografía	33

Índice de tablas

5.1. Promedio de las métricas Coeficiente Dice y la Distancia Hausdorff (WT, TC, ET)	25
--	----

Índice de figuras

2.1. Ejemplo de Segmentación de Tumores	5
2.2. Neuron	6
2.3. Visualización de las activaciones del kernel almacenadas en el mapa de características, también conocido como mapa de activación	7
2.4. Canales de imágenes	8
2.5. Un ejemplo de un núcleo que reconoce una curva como una característica .	9
2.6. Función de activación ReLu	9
2.7. ReLu vs Leaky ReLu	10
2.8. Un Autoencoder estándar	11
2.9. Un ejemplo de una red neuronal de tipo <i>Variational Autoencoder</i> (VAE) .	12
2.10. De izquierda a derecha: imágenes T1, T1C, T2 y FLAIR.	14
4.1. Arquitectura propuesta	22
5.1. Segmentación en un corte axial, sagital y coronal en el tipo Flair de la imagen MRI.	26
5.2. Segmentación en un corte axial, sagital y coronal en el tipo T1ce de la imagen MRI.	26
5.3. Segmentación en un corte axial, sagital y coronal en el tipo T1 de la imagen MRI.	27

Capítulo 1

Introducción

Según las estadísticas de cáncer de 2019 [Siegel et al., 2019], los tumores cerebrales se consideran la principal causa de mortalidad relacionada con el cáncer y la mortalidad en todo el mundo y se conoce como una de las formas más comunes de cáncer tanto en niños como en adultos. Dentro de los tumores cerebrales mas comunes, tenemos los que comienzan y tienden a permanecer en el cerebro, el cual es el meningioma con un 34 % de presencia, otro tipo de tumor es el llamado glioma, que surgen del tejido circundante en el cerebro, forma parte del 30 % del todos los tumores en el cerebro, sin embargo este glioma representa el 80 % de los tumores malignos, convirtiéndolo así en el tumor más común que causa la muerte.

Los tumores cerebrales se refieren a un grupo mixto de neoplasias que se originan en tejidos intracraneales y las meninges con grados de malignidad que van desde benignos a agresivos. Cada tipo de tumor tiene su propia biología, tratamiento y pronóstico y es probable que cada uno esté causado por diferentes factores de riesgo. Incluso los tumores “benignos” pueden ser letales debido a su sitio en el cerebro, su capacidad para infiltrarse localmente y su propensión a transformarse en malignidad. El tipo más común de tumores cerebrales primarios son los gliomas, que surgen de las células gliales del cerebro. Los gliomas pueden ser de subtipos: *High-Grade Gliomas* (HGG) y *Low-Grade Gliomas* (LGG). Los gliomas de alto grado son un tipo agresivo de tumor cerebral maligno que crece rápidamente, generalmente requieren cirugía y radioterapia y tienen un mal pronóstico de supervivencia.

Tenemos que nuestro problema principal es la detección del tumor, esta detección es conocida como segmentación. La segmentación de tumores cerebrales juega un papel importante en el procesamiento de imágenes médicas, la detección temprana de tumores cerebrales mejorará las posibilidades de vida del paciente. En la actualidad el procesamiento de imágenes esta siendo utilizada en gran cantidad en el campo de la medicina para el análisis de imágenes. El análisis de imágenes comúnmente requiere de segmentación de imágenes en regiones homogéneas para la extracción y análisis de características, pero en el caso de MRI, las imágenes son de tipo altamente heterogéneas. Estas imágenes médicas se adquieren mediante distintas modalidades, como la Tomografía computarizada, Imágenes de resonancia magnética, Rayos X, y variante de estas. Entre las diferentes tecnologías de imágenes médicas disponibles, MRI es más favorecida para la clasificación

de tipo de tumor cerebral debido a su naturaleza inofensiva. Esta investigación busca la detección de tumores cerebrales mediante el análisis de imágenes MRI. La resonancia magnética es una herramienta de diagnóstico clave para el análisis de tumores cerebrales, el monitoreo y la planificación de la cirugía. Por lo general, se adquieren varias modalidades complementarias de *Resonance Magnetic* (RM) en 3D, como T1, T1 con agente de contraste (T1c), T2 y *Fluid Attenuation Inversion Recovery* (FLAIR), para enfatizar diferentes propiedades tisulares y áreas de diseminación del tumor. [Myronenko, 2018]

El trabajo se centra en el procesamiento automático de los volúmenes con gliomas HGG y LGG. Debido al crecimiento de los pacientes afectados con tumores cerebrales, también aumenta el número de datos adquiridos y disponibles. Por ello se ve un claro crecimiento en la necesidad de algoritmos automáticos que sean capaces de procesar los datos automáticamente. Este hecho ha provocado un reciente aumento de interés de autores como [Kaur et al., 2017] [Zhao et al., 2018] [Havaei et al., 2017] [Myronenko, 2018], en el diseño de sistemas de procesamiento de imágenes automatizados de alta precisión para la clasificación de tumores cerebrales.

1.1. Motivación y Contexto

En gran parte de los hospitales en el Perú y de muchos países por igual, la evaluación de las imágenes adquiridas son realizadas manualmente de acuerdo con criterios cuantitativos o medidas de diámetro visible más pronunciado en el corte axial [Watanabe et al., 2009]. Por lo tanto las técnicas altamente precisas para analizar automáticamente las MRI tendrían un gran potencial para el diagnóstico y la planificación de la terapia. Sin embargo, [Menze et al., 2015], demostró que incluso la evaluación manual por expertos tiene variaciones significativas en áreas donde los gradientes de intensidad entre las estructura del tumor y el tejido circundante son suaves, haciendo imposible el uso de algoritmos comunes de reconocimiento de patrones clásicos. Como mencionamos, la clasificación del tipo de tumor por inspección humana es una tarea extremadamente lenta y propensa a errores, que depende en gran medida de la experiencia y las habilidades del médico especialista.

1.2. Planteamiento del Problema

Como mencionamos en la sección 1.1, vemos que la detección de tumores cerebrales, recae en dos tipos de soluciones. Primero la evaluación manual que es en gran parte costosa, lenta y sujeta a la experiencia del radiólogo o experto. Segundo tenemos la automatización del proceso de detección, mediante técnicas de segmentación, haciendo uso de *Neural Network* (NN). Estas razones motivan nuestra exploración de una solución de aprendizaje automático que recae en una propuesta flexible y de alta capacidad a la vez que es extremadamente eficiente.

1.3. Objetivos

Esta sección describe los objetivos y metas de la tesis. El objetivo principal es desarrollar una arquitectura para la detección y segmentación de tumores cerebrales en MRI de una o varias secuencias en 3D.

1.3.1. Objetivos Específicos

- Proponer el uso eficiente de memoria y recursos en nuestra arquitectura.
- Proponer una arquitectura *Convolutional Neural Network* (CNN) nueva, tomando en cuenta las arquitecturas existentes y discutir sobre ellas al combinarlas.
- Acelerar el proceso de segmentación de tumores cerebrales haciendo uso de propuestas por distintas técnicas.

1.4. Organización de la Tesis

Este trabajo está organizado de la siguiente manera. El Capítulo 1, muestra la introducción a la tesis. En el Capítulo 2 presentamos los fundamentos teóricos. El estado del arte se presentan en el Capítulo 3. En el Capítulo 4 se introduce nuestra propuesta para la segmentación de tumores cerebrales. El Capítulo 5 está dedicado a los Resultados. Finalmente la conclusión es dada en el capítulo 6, donde los posibles trabajos futuros son descritos también.

Capítulo 2

Marco Teórico

Antes de profundizar los aspectos específicos en las técnicas que ofrecen una arquitectura neuronal para segmentar los tumores, veremos algunos de los fundamentos de las CNN, como principal rama de las NN en la que se basa esta tesis, y la segmentación en general. Cuando discutimos la segmentación en el campo de las imágenes médica, estamos discutiendo la capacidad de clasificar diferentes partes de una imagen médica. Esta segmentación está bastante definida y puede usarse para describir la clasificación a través de una variedad de granularidades. En una granularidad más gruesa, podríamos tener una radiografía de una pierna donde queremos diferenciar y etiquetar los diferentes huesos de la imagen. En una granularidad más fina, podríamos estar interesados en poder identificar y etiquetar diferentes partes de los tumores en una resonancia magnética de un cerebro como se ve en la figura 2.1.

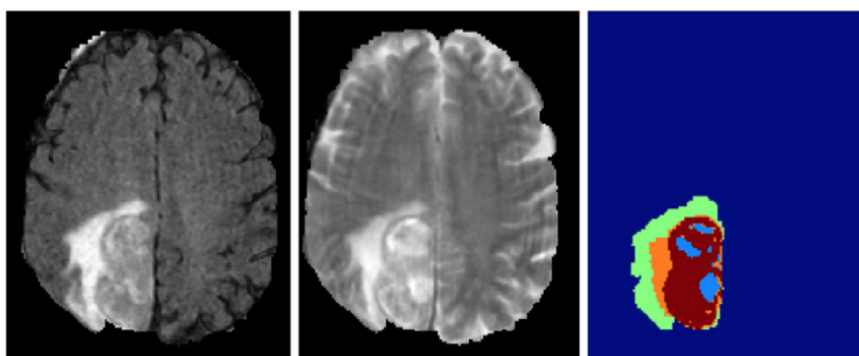


Figura (2.1) Ejemplo de Segmentación de Tumores

Actualmente, la mayor parte de esta segmentación se realiza manualmente por profesionales médicos. Aquí es donde entran las redes neuronales convolucionales. Estas CNN en su mayoría son entrenadas sobre un conjunto de imágenes, las cuales han sido divididas en parches, que han sido debidamente etiquetados por profesionales médicos. Esto le enseña cómo diferenciar las diferentes partes de la imagen por sí mismo. La red luego toma parches de una imagen sin etiquetar como entrada y usa su entrenamiento para intentar etiquetar correctamente la imagen. El objetivo final es que la red genere etiquetas de imágenes correctas. Esto es una breve explicación al proceso que se realiza para segmentar tumores a través de redes neuronales y una descripción del tipo de dato que estamos procesando.

2.1. Introducción a las Redes Neuronales

En su forma más básica, las redes neuronales son reconocedores de patrones modelados en las estructuras neuronales de la corteza cerebral, una parte del cerebro que toma en cuenta los datos sensoriales. Como se ve en la figura 2.2, las redes generalmente están compuestas por capas de nodos que se activan cuando reconocen una determinada entrada.

El resultado de estas activaciones se pasa luego a los nodos vecinos a través de conexiones ponderadas. Después de pasar a través de las capas de nodos y conexiones, los datos resultantes se envían fuera de la red como una forma de salida. Lo que hace que las redes neuronales sean tan poderosas es su capacidad de uso general en una amplia gama de problemas controlados por datos, así como la capacidad de alterar los pesos de las conexiones de los nodos para mejorar la precisión de su salida.

En nuestro caso, estamos interesados en un tipo de red neuronal llamada Convolutional Neural Network. Estas CNN se basan en muchos de los mismos principios que las NN comunes, presenta una capa de entrada, capas ocultas y capas de salida, pero tienen un tipo de capa adicional que se ha encontrado muy útil cuando se trata de aprender cosas acerca de imágenes. En nuestro caso, las entradas de la red serán parches de los píxeles de una resonancia magnética, siendo la salida etiquetas de segmentación.

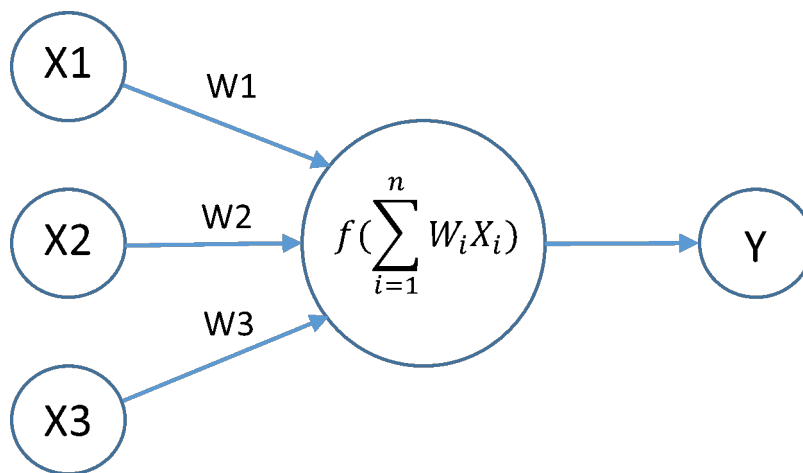


Figura (2.2) Neuron

2.2. Capas Convolucionales

Las capas convolucionales son las que diferencian las CNN de otras redes neuronales, la primera capa de todas las CNN es una convolucional. Estas capas se utilizan para condensar los datos de entrada en patrones de datos reconocidos, reduciendo así el tamaño de los datos y reconociendo cosas de interés. La capa convolucional toma una serie de valores que representan los píxeles o voxels de un parche de la imagen de entrada. Luego, esta capa utiliza lo que se denomina de manera intercambiable filtro, neurona o kernel, que es otra matriz que representa algún tipo de característica, en muchos de los casos que hacen uso de las CNN, hacen uso de kernel aleatorios.

Este kernel se alinea con la esquina superior izquierda de la entrada, el área que cubre se llama campo receptivo. La matriz contenida dentro del campo receptivo se multiplica luego con el kernel mediante la multiplicación de elementos. Las multiplicaciones se suman y almacenan en la misma posición relativa de lo que se llama un mapa de características, como se ve en la figura 2.3

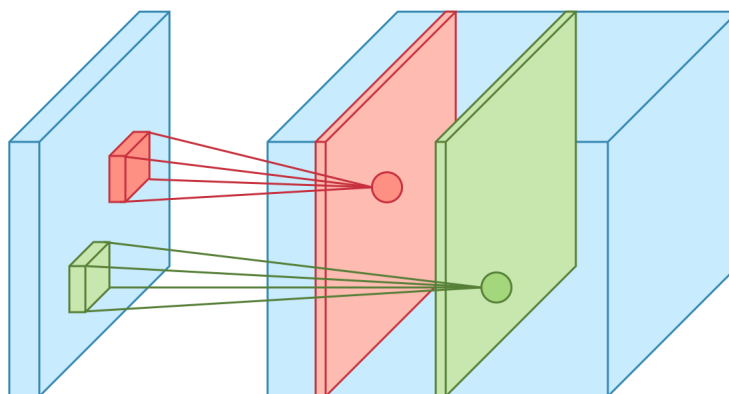


Figura (2.3) Visualización de las activaciones del kernel almacenadas en el mapa de características, también conocido como mapa de activación

El kernel luego se desliza sobre una distancia especificada en la entrada y realiza la misma operación, almacenando el resultado en la siguiente posición de el mapa de características. Lo que terminamos después de todas las posibles convoluciones del kernel y la entrada es un mapa de características completo. El mapa de características es una matriz que contiene todos los resultados de las convoluciones entre el kernel y la entrada. Estos mapas de características se pueden pasar como entrada a las capas futuras.

En otras palabras una convolución se define como una operación en dos funciones. En el análisis de imágenes, una función consiste en valores de entrada en una posición en la imagen, y la segunda función es un filtro (o kernel); Cada uno puede ser representado como una serie de números. Al calcular el producto de puntos entre las dos funciones se obtiene una salida. El filtro se desplaza a la siguiente posición en la imagen según lo definido por la longitud del paso.

El cálculo se repite hasta que se cubre toda la imagen, lo que produce un mapa de características. Este es un mapa donde el kernel está fuertemente activado y es capaz de observar o definir una característica como una línea recta, un punto o un borde curvo. Si una fotografía de una cara se introdujo en una CNN, los filtros descubren inicialmente características de bajo nivel, como líneas y bordes, como vemos en la figura 2.5, detectamos una curva. Estos se acumulan para características progresivamente más altas en capas subsiguientes, a medida que los mapas de características se convierten en entradas para la siguiente capa en la arquitectura CNN.

2.2.1. Imágenes de Entrada

Normalmente las CNN usan imágenes como datos de entrada, pero vemos que se puede procesar datos como musica, vídeo, texto, etc, para así mandarle la entrada a nuestra CNN, y poder entrenar nuestra red neuronal. En nuestro caso, tenemos imágenes

de tipo MRI, lo cual es una imagen en tres dimensiones, pensemos en ello como un cubo. Si nuestra CNN solo trabaja con dos dimensiones, entonces se recorta parches de la imagen en solo dos ejes, y si aplica a una segmentación 3D, se extrae un cubo pequeño como entrada. Por lo general, las imágenes presentan 3 canales de colores, el más común, el RGB: Rojo, Verde y Azul.

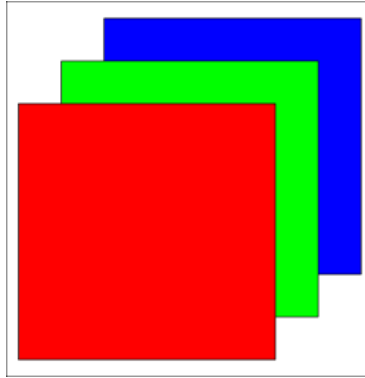


Figura (2.4) Canales de imágenes

2.2.2. Kernels

Los kernels, como se describió anteriormente, son matrices de valores que están destinados a representar características que deben reconocerse. Por ejemplo, un kernel podría contener una característica como una curva, como se muestra en la figura 2.5. Esto podría estar representado por un patrón de números en el kernel. Cuando el kernel se multiplica con la entrada, el resultado será un número mayor si la característica en el kernel es similar a la característica en el campo receptivo. Si la característica descrita por el kernel no está presente en el campo receptivo, entonces el resultado de la multiplicación será menor. Estos reconocimientos de características se almacenan en el mapa de características, donde probablemente se utilizarán como entrada para la siguiente capa. Cuantos más kernels haya, más funciones podrá reconocer la red. Como estos mapas de características se utilizan en capas futuras de operaciones similares, se crea una jerarquía de características con características más complejas que se representan en capas posteriores.

2.2.3. Funciones de Activación

ReLU (Rectified Linear Unit)

ReLU es la función de activación más utilizada en este momento. Es por eso que se utiliza en casi todas las redes neuronales convolucionales o el aprendizaje profundo. Como se puede observar en la figura 2.6, el ReLU está medio rectificado. $f(z)$ es cero cuando z es menor que cero y $f(z)$ es igual a z cuando z es superior o igual a cero.

0	0	0	0	0	0.3	0
0	0	0	0	0.3	0	0
0	0	0	0.3	0	0	0
0	0	0	0.3	0	0	0
0	0	0	0.3	0	0	0
0	0	0	0.3	0	0	0
0	0	0	0	0	0	0

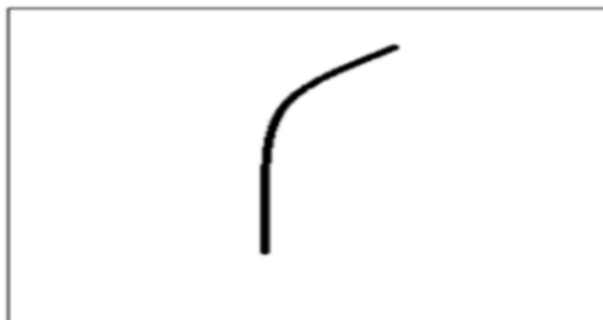


Figura (2.5) Un ejemplo de un núcleo que reconoce una curva como una característica

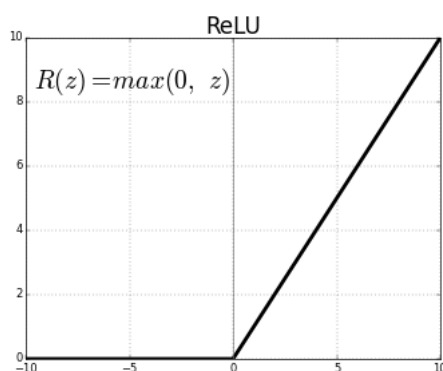


Figura (2.6) Función de activación ReLu

Leaky ReLu

Es un intento de resolver el problema dying de ReLU. Leak ayuda a aumentar el rango de la función ReLU. Por lo general, el valor de a es 0.01 más o menos.

2.2.4. Dropout and Pooling Layers

Dos tipos de capas más utilizadas en redes neuronales convolucionales son las capas de Dropout y Pooling. En primer lugar es la capa de agrupación. Las capas de Pooling tienen un propósito relativamente sencillo; toman los grupos del mapa de entidades de entrada y los reducen a una sola entidad. Un ejemplo de una operación de agrupación es max-pooling. Una capa de agrupación de agrupación máxima dividirá la entrada en grupos y colocará el valor más alto de cada grupo en su lugar correspondiente en el mapa de características de la capa de agrupación. Las capas de pooling se utilizan para reducir

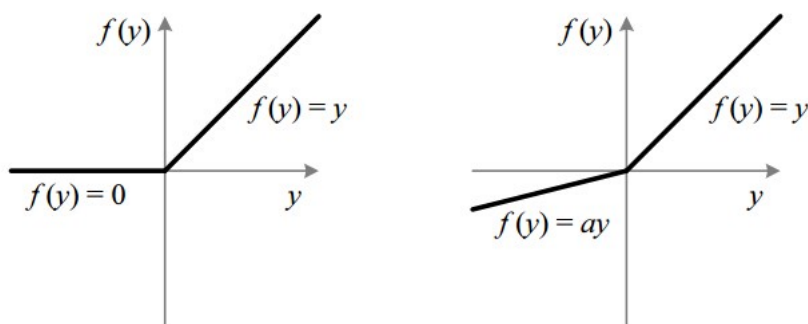


Figura (2.7) ReLu vs Leaky ReLu

drásticamente la cantidad de datos espaciales al eliminar una gran parte de la entrada en un solo paso.

Luego tenemos la capa Dropout, fiel a su nombre, selecciona aleatoriamente las activaciones en el mapa de características de entrada y las ponen a cero. Esto ayuda a garantizar que la red pueda predecir correctamente las clasificaciones, incluso cuando algunas de las activaciones se desactiven. La redundancia proporcionada por las capas de dropout es útil en el proceso de entrenamiento, pero no es aplicable a los datos que no son de entrenamiento. Ambos tipos de capas son herramientas útiles para abordar el problema de sobre ajuste en las CNN. El ajuste excesivo es esencialmente cuando la red se asemeja demasiado al conjunto de datos de entrenamiento y no se generaliza bien a las imágenes que aún no se han visto. Hay muchos otros métodos para abordar el sobre ajuste, pero estos son los dos únicos dentro del alcance de este documento.

2.3. Capas totalmente conectadas

La capa totalmente conectada es a menudo la capa final en la red. Toma como entrada un mapa de características de la capa anterior y devuelve un vector de probabilidades de etiquetas para el píxel central del parche de imagen de entrada. Un CNN podría por ejemplo. estar tratando de segmentar una imagen en lo que es y no es un tumor. La salida es por lo tanto una opción binaria, tumor o no tumoral. La capa totalmente conectada mira las características representadas en el mapa de características y luego proporciona un vector con dos valores, la probabilidad de que la entrada sea un tumor y la probabilidad de que no sea un tumor.

2.4. Autoencoder

Una red de autoencoder es en realidad un par de dos redes conectadas, que son un red convolucional y una deconvolucional, un codificador y un decodificador. Una red de codificador toma una entrada y la convierte en una representación más pequeña y densa, que la red del decodificador puede usar para convertirla nuevamente a la entrada original, como esta representado en la imagen 2.8.

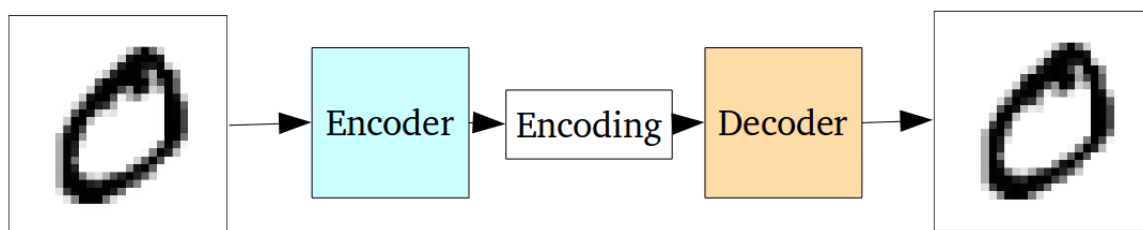


Figura (2.8) Un Autoencoder estándar

Las capas convolucionales de cualquier CNN toman una imagen grande y la convierten en una representación mucho más compacta y densa. Esta representación densa es utilizada luego por la red de clasificadores totalmente conectada para clasificar la imagen.

El codificador es similar, simplemente es una red que toma una entrada y produce una representación mucho más pequeña, que contiene suficiente información para que la siguiente parte de la red la procese en el formato de salida deseado. Por lo general, el codificador se entrena junto con las otras partes de la red, optimizado mediante propagación inversa, para producir codificaciones específicamente útiles para la tarea en cuestión.

Los codificadores automáticos toman esta idea y la cambian ligeramente de cabeza, haciendo que el codificador genere codificaciones específicamente útiles para reconstruir su propia entrada.

Toda la red generalmente se entrena como un todo. La función de pérdida suele ser el error cuadrático medio o la entropía cruzada entre la salida y la entrada, conocida como pérdida de reconstrucción, que penaliza a la red por crear salidas diferentes de la entrada.

Los codificadores automáticos estándar aprenden a generar representaciones compactas y reconstruir bien sus entradas, pero aparte de algunas aplicaciones como la eliminación de ruido de codificadores automáticos, son bastante limitados.

El problema fundamental con los autoencoders, para la generación, es que el espacio latente al que convierten sus entradas y dónde se encuentran sus vectores codificados, puede no ser continuo o permitir una fácil interpolación.

2.5. Variational Autoencoder

Este tipo de CNN, tienen una propiedad fundamentalmente única que los separa de los autoencoders estándares, y es esta propiedad la que los hace tan útiles para el modelado generativo: sus espacios latentes son, por diseño, continuos, permitiendo un muestreo aleatorio e interpolación fáciles. Lo logra haciendo algo que parece bastante sorprendente al principio: hacer que su codificador no genere un vector de codificación de tamaño n , sino que genera dos vectores de tamaño n : un vector de medias μ , y otro vector de desviaciones estándar σ , 2.9.

Esta generación estocástica es utilizada para generar la imagen inicial, significa que,

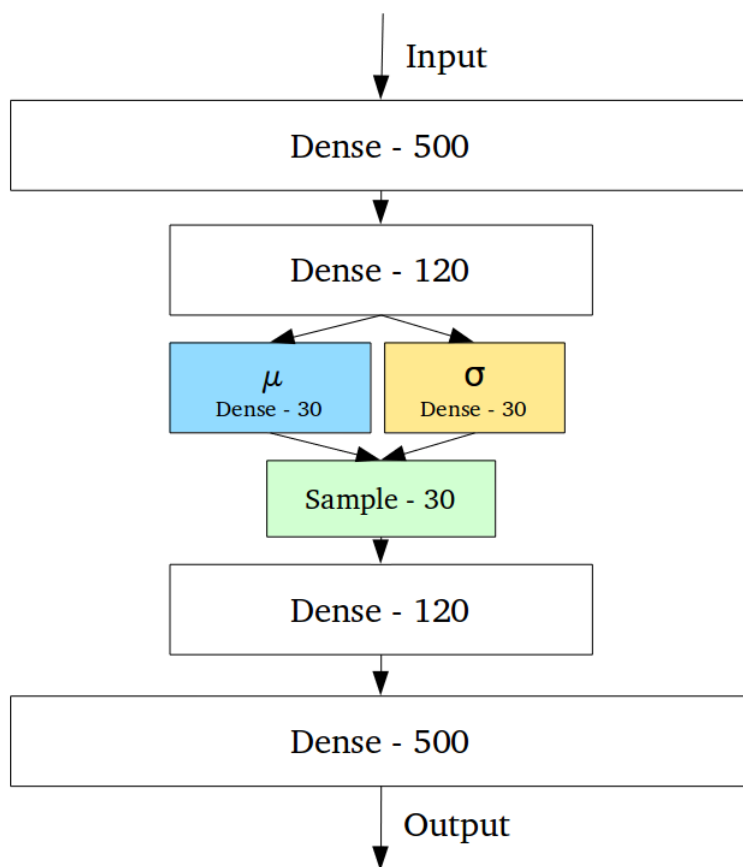


Figura (2.9) Un ejemplo de una red neuronal de tipo VAE

incluso para la misma entrada, mientras que las desviaciones medias y estándar siguen siendo las mismas, la codificación real variará de alguna manera en cada pasada simplemente debido al muestreo.

2.6. Entrenamiento

Ahora que hemos pasado por algunos de los componentes de una CNN, podemos entrar en lo que hace que todo funcione, el proceso de entrenamiento. Antes de pasar a los pasos básicos del proceso de entrenamiento o capacitación, es importante tener en cuenta que se requiere un conjunto de datos de entrenamiento para comenzar el proceso. El conjunto de datos de entrenamiento en nuestro caso serían imágenes médicas que se emparejan con sus etiquetas de verdad de fondo. Estas etiquetas de hechos reales son las etiquetas observadas por un profesional médico en la segmentación manual. De esta manera, tenemos imágenes que la red puede intentar segmentar, comparando sus resultados con las etiquetas de verdad proporcionadas.

Los kernels en la red originalmente comienzan aleatoriamente y, por lo tanto, las probabilidades de salida deberían ser aproximadamente iguales. En el paso hacia adelante, un parche de imagen de los datos de entrenamiento se envía a través de la red y la probabilidad de salida se compara con la probabilidad de etiqueta verdadera propor-

cionada con la imagen de prueba. Esta comparación se realiza a través de una función de pérdida para cuantificar la inexactitud. Al comienzo del proceso de entrenamiento, la pérdida probablemente será muy alta, con el objetivo de minimizar la pérdida tanto como sea posible. La función de pérdida se utiliza en el siguiente paso del proceso denominado paso hacia atrás. En el paso hacia atrás, usted avanza iterativamente a través de la red de salida a entrada, evaluando qué kernels contribuyeron más a la pérdida total y calculando los ajustes de peso que minimizarían dicha pérdida. Una vez completada la pasada hacia atrás, se realiza el paso final, actualización de peso. Este último paso toma la pérdida minimizando los cambios de peso de la fase de paso hacia atrás y los implementa.

La aplicación de este proceso de cuatro pasos para cada imagen en el conjunto de datos de entrenamiento se considera una época; El entrenamiento generalmente requiere muchas épocas. Después de completar el proceso de entrenamiento, la red se puede probar en un conjunto de datos de prueba. El conjunto de datos de prueba, al igual que los datos de entrenamiento, contiene imágenes y sus verdaderas etiquetas. Este conjunto de datos permite la evaluación del rendimiento de las redes antes de utilizarlo en datos sin etiquetar. Los datos en el conjunto de pruebas no pueden contener imágenes del conjunto de entrenamiento debido a la inclinación inherente que la red tiene hacia las imágenes en las que se entrenó. Hay algunas cosas que es importante tener en cuenta sobre el proceso de entrenamiento. En general, cuantas más imágenes en el conjunto de datos de entrenamiento, mejor. Esto puede ser un obstáculo para las aplicaciones de imágenes médicas, debido a la dificultad de recopilar las imágenes adecuadas y el tiempo necesario para etiquetarlas. Existen algunos métodos para aumentar el conjunto de datos para aumentar el tamaño del grupo de imágenes de entrenamiento, como la aplicación de rotaciones, traslaciones y fluctuaciones a las imágenes en el conjunto de entrenamiento. Son preferibles imágenes realmente diferentes [Finzel, 2017] .

2.7. Secuencias de MR

La MRI se usa comúnmente para el análisis y exploración de tumores cerebrales. Existe una variedad de secuencias de MRI, donde cada una de ellas es adecuada para diferentes propósitos de imagen. Hoy en día, es una práctica común en el análisis automático utilizar una combinación de varias secuencias de MRI para lograr resultados más valiosos y precisos. En este trabajo, se utilizan tres secuencias de MRI diferentes, imagen ponderada en T1, imagen ponderada en T2 e imagen FLAIR, que se caracterizarán en breve como se describe en [Armstrong et al., 2004].

- Imagen ponderada en T1. En MRI, T1 se refiere al tiempo que los protones dentro de un tejido necesitan volver al estado de magnetización inicial, que es dado por el campo magnético estático. Las imágenes simples ponderadas en T1 proporcionan mejores detalles anatómicos que las imágenes ponderadas en T2, pero generalmente no aportan información interesante cuando se investiga un tumor cerebral. Sin embargo, se usan en combinación con FLAIR, que se inyecta en el sistema vascular del paciente. El agente de contraste resalta el flujo sanguíneo en las imágenes ponderadas en T1. Esto hace que la parte activa del tumor, así como los vasos, aparezcan hiperintensos y fácilmente distinguibles de los tejidos circundantes. La presencia del

tumor activo a menudo se usa durante la investigación de tumores malignos. Dichas imágenes se denominan imágenes potenciadas en contraste T1z en este trabajo se utilizará la abreviatura T1C.

- Imagen ponderada en T2. T2 se refiere al tiempo que los protones perturbados en una oscilación coherente por pulso de radiofrecuencia requieren perder esta coherencia. Las imágenes ponderadas en T2 son, en comparación con las imágenes T1, más sensibles al contenido de agua y, por lo tanto, a la patología, que, así como elCSF, parece hiperintensiva aquí.
- Imagen FLAIR. Es una secuencia que puede suprimir fluidos y se utiliza para suprimir el *Cerebrospinal fluid* (CSF) en las imágenes del cerebro. Este efecto permite distinguir las lesiones, que siguen siendo hiperintensas como en las imágenes T2, del CSF que se vuelve hipointensivo aquí. Por esa razón, se usa comúnmente en las imágenes de tumores cerebrales.

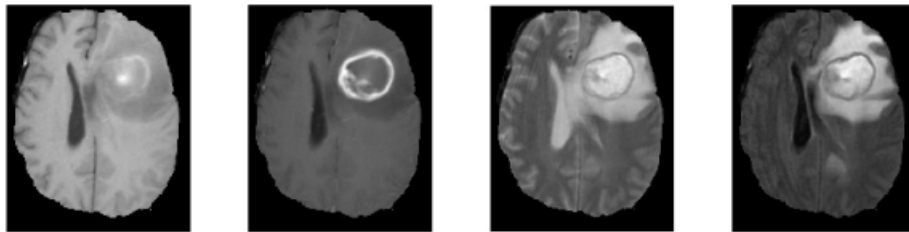


Figura (2.10) De izquierda a derecha: imágenes T1, T1C, T2 y FLAIR.

2.8. Metodología

En el problema de segmentación, existe una metodología para resolver este problema, como se muestra en la figura [1]. Estos pasos pueden ser modificados de acuerdo a los algoritmos que tenemos.

Adquisición de Imagen

Para trabajar con cualquier tipo de tarea de procesamiento de imágenes, es esencial adquirir primero la imagen para aplicar el procesamiento.

Pre-procesamiento

La generación de imágenes a partir de diversas técnicas de imágenes médicas puede generar ruidos innecesarios en la imagen. Por lo tanto, cualquier resonancia magnética, tomografía computarizada, imagen mamográfica, etc., generalmente viene con mucho ruido. Este ruido puede actuar como un obstáculo al segmentar la región tumoral de la imagen de entrada dada. Para eliminar este problema, primero se procesa una imagen para eliminar

los valores atípicos no deseados y luego se envía para su posterior procesamiento. Este paso de pre-procesamiento puede incluir técnicas como eliminación de ruido, aplicación de filtros, mejora de imagen, normalización, etc.

Segmentación

Este es el paso más crucial y fundamental en aplicaciones como la detección de tumores. Es un fenómeno simple de dividir una imagen en diferentes segmentos significativos para interpretarlos mejor. La segmentación de una imagen se puede lograr de numerosas maneras como ya se discutió en la literatura.

Extracción de características

Una vez que la imagen se ha dividido en segmentos, se necesita un paso de procesamiento posterior para afilar los bordes y desenfocar los detalles no deseados. Este paso se llama extracción de características, donde algunas de las características de la imagen se extraen para su análisis. Esto mejorará la región del tumor para que el cálculo del área u otras operaciones se puedan aplicar sobre ella para obtener resultados más efectivos. Las operaciones morfológicas, las técnicas de detección de bordes o la ecualización de histogramas son los pasos de extracción de características más utilizados.

Capítulo 3

Estado del Arte

En la literatura existen dos métodos de segmentación de tumores cerebrales automáticos y semiautomáticos, esto se conoce como los métodos basados en modelos generativos y modelos discriminativos [Menze et al., 2015].

Los modelos generativos dependen en gran medida del conocimiento previo específico del dominio sobre la apariencia de los tejidos sanos y tumorales. La apariencia del tejido es difícil de caracterizar, y los modelos generativos existentes generalmente identifican un tumor como una forma o una señal que se desvía de un cerebro normal. Por lo general, estos métodos se basan en modelos anatómicos obtenidos después de alinear la imagen 3D MRI en un atlas. Un modelo generativo típico de imágenes cerebrales MRI se puede encontrar con el atlas cerebral ICBM, este método alinea el cerebro con el atlas y calcula las probabilidades posteriores de tejidos sanos (materia blanca, materia gris y líquido cefalorraquídeo). Luego se encuentran regiones tumorales localizando vóxeles cuya probabilidad posterior es inferior a un cierto umbral. Luego se aplica un paso de procesamiento posterior para garantizar una buena regularidad espacial. [Prastawa et al., 2003], también registra imágenes cerebrales en un atlas para obtener un mapa de probabilidad de anomalías. Luego se inicializa un contorno activo en este mapa y se itera hasta que el cambio en la probabilidad posterior esté por debajo de cierto umbral. Se han propuesto muchos otros métodos de contorno activo en la misma línea, todos los cuales dependen de las características de simetría cerebral izquierda-derecha y/o características basadas en la alineación.

El otro enfoque para la segmentación de tumores cerebrales, es el modelo discriminativo. A diferencia de los enfoques de modelado generativo, estos enfoques explotan poco conocimiento previo sobre la anatomía del cerebro y en su lugar se basan principalmente en la extracción de una gran cantidad de características de imagen de bajo nivel, modelando directamente la relación entre estas características y la etiqueta de un vóxel dado. Estas características pueden ser valores de píxeles de entrada sin procesar [Havaei et al., 2016], histogramas locales, características de textura como los bancos de filtros Gabor, o características basadas en la alineación, como el gradiente entre imágenes, la diferencia de forma de la región y el análisis de simetría. También se han utilizado técnicas de aprendizaje discriminatorias clásicas como *Support Vector Machine* (SVM) [Ruan et al., 2007] y *Decision Random Forest* [Götz et al., 2014]. Los resultados de las ediciones de

2012 y 2013 del Desafío *Brain Tumor Segmentation Challenge* (BRATS) sugieren que los métodos que dependen de Random Forest se encuentran entre los más precisos [Menze et al., 2015] [Götz et al., 2014].

Un aspecto común con los modelos discriminatorios es su implementación de un pipeline de aprendizaje automático convencional que se basa en características diseñadas a mano. Para estos métodos, el clasificador está entrenado para separar los tejidos sanos de los no sanos, suponiendo que las características de entrada tengan un poder discriminativo suficientemente alto ya que el comportamiento del clasificador es independiente de la naturaleza de esas características. Una dificultad con los métodos basados en características diseñadas a mano es que a menudo requieren el cálculo de una gran cantidad de características para ser precisos cuando se usan con muchas técnicas tradicionales de aprendizaje automático. Esto puede hacerlos lentos de calcular y costosos en cuanto a memoria. Las técnicas más eficientes emplean un número menor de características, utilizando métodos de reducción de dimensionalidad o selección de características, pero la reducción en el número de características a menudo tiene el costo de una precisión reducida.

Por lo tanto, el rendimiento de estos modelos consiste en las características de la imagen y los algoritmos de clasificación. Recientemente, se han adoptado técnicas de aprendizaje profundo en estudios de segmentación de tumores cerebrales, debido a su éxito en los campos de análisis de imágenes. Para BRATS 2014, las CNN se han utilizado en general [Zikic et al., 2014] [Urban et al., 2014], en BRATS 2015 se presentaron varias redes neuronales profundas [Menze et al., 2015] [Havaei et al., 2016] [Pereira et al., 2015], Autoencoders Stacked Denoising [Vaidhya et al., 2016].

Entre los métodos de segmentación tumoral basados en el aprendizaje profundo, los métodos basados en CNN han logrado un mejor rendimiento. En particular, los modelos 3D-CNN [Myronenko, 2018] [Urban et al., 2014] [Kamnitsas et al., 2017] y 2D-CNN [Myronenko, 2018] [Zikic et al., 2014] [Menze et al., 2015] [Havaei et al., 2016] [Pereira et al., 2015]. Estos se adoptaron para construir métodos de segmentación. Aunque 3D-CNN puede aprovechar al máximo la información 3D de los datos de MRI, el costo de procesar la información también aumenta. Por lo tanto, 2D-CNN se ha utilizado en mayores cantidades para la segmentación de tumores cerebrales. [Havaei et al., 2016] propuso un método de aprendizaje profundo con dos rutas CNN, una ruta convolucional y una totalmente conectada. [Menze et al., 2015] modeló la tarea como 3 subtarefas de segmentación binaria y cada subtarea se resolvió usando CNN. [Pereira et al., 2015] adoptó una CNN muy profunda.

La mayoría de los métodos de segmentación tumoral entrenan las CNN utilizando parches de imágenes, es decir, regiones de resonancia magnética local, clasificando cada parche en diferentes clases, como tejido sano, necrosis, edema, núcleo potenciador y núcleo no potenciador. El resultado de la clasificación de cada parche de imagen se utiliza para etiquetar su vóxel central para lograr la segmentación del tumor. La mayoría de los métodos de segmentación anteriores suponían que la etiqueta de cada vóxel es *independiente*, y no tenían en cuenta la apariencia y la consistencia espacial. Para tener en cuenta las dependencias locales de las etiquetas, [Havaei et al., 2016] [Havaei et al., 2017] construyeron una arquitectura en cascada tomando los resultados de segmentación de probabilidad en píxeles obtenidos por los CNN entrenados en las etapas iniciales como una contribución adicional a su siguiente CNN.

Para el año 2017, en el concurso de BRATS de ese mismo año, las presentaciones de mayor rendimiento incluyeron a [Kamnitsas et al., 2017], que propuso ensamblar varios modelos para una segmentación robusta llamada *Ensembles of Multiple Models and Architectures* (EMMA), y [Wang et al., 2018] quienes propusieron segmentar subregiones tumorales en cascada usando convoluciones anisotrópicas.

EMMA aprovecha un conjunto de varias arquitecturas formadas independientemente. En particular, EMMA combinó los modelos DeepMedic [Kamnitsas et al., 2017], FCN [Long et al., 2015] y U-net [Ronneberger et al., 2015] y reunió sus predicciones de segmentación. Durante el entrenamiento usaron un batch size de 8 y un recorte de patch 3D de 64x64x64. El conjunto de diferentes modelos de EMMA demostró un buen rendimiento de generalización al ganar el desafío BRATS 2017. [Wang et al., 2018], quien quedó en segundo lugar adoptó un enfoque diferente, al entrenar 3 redes para cada subregión tumoral en cascada, con cada red posterior tomando la salida de la red anterior como su entrada. Cada red era similar en estructura y constaba de una gran parte del codificador y un decodificador básico. También descomponen el kernel de convolución 3x3x3 en kernel intra-slice (3x3x1) e inter-slice (1x1x3) para optimizar tanto en la memoria de la GPU como en el tiempo de cálculo.

En BRATS 2018, se presentaron varias propuestas, entre la presentación de alto rendimiento, tenemos a [Myronenko, 2018], que ganó el primer puesto, teniendo a [Isensee et al., 2018] en segundo lugar, [McKinley et al., 2019] y [Zhou et al., 2019], que compartió el 3er lugar. [Isensee et al., 2018] demostró que una arquitectura genérica de U-net con algunas modificaciones menores es suficiente para lograr un rendimiento competitivo. Los autores utilizaron un tamaño de batch de 2 y un crop size de 128x128x128. [McKinley et al., 2019] propuso una CNN de segmentación en la que una estructura DenseNet con convoluciones dilatadas estaba incrustada en una red tipo U-net. Los autores también introducen una nueva función de pérdida, una generalización de la entropía cruzada binaria, para tener en cuenta la incertidumbre de la etiqueta. Finalmente, [Zhou et al., 2019] propuso utilizar un conjunto de redes diferentes: teniendo en cuenta la información de contexto a múltiples escalas, segmentando 3 subregiones tumorales en cascada con una columna vertebral compartida y agregando un bloqueo de atención.

En comparación con los trabajos mencionados, [Myronenko, 2018], utilizó un crop size más grande de 160x192x128, pero compromete el tamaño del batch a 1 para poder encajar la red en los límites de memoria de la GPU. También muestra las 3 subregiones tumorales anidadas directamente después del sigmoide. Finalmente, agrega una rama adicional para regularizar el codificador compartido, que se usa solo durante el entrenamiento. Haciendo uso de VAE, tenemos la parte del Encoder, que extrae características en bloques inicial de 4, 2 y 1, hasta obtener un mapa de características del tamaño 256x20x24x16, en la parte del Decoder, tenemos la misma estructura del Encoder, pero con un bloque por cada nivel espacial. VAE, comienza desde la salida del punto final del codificador, primero reduce la entrada a un espacio dimensional bajo de 256 (128 para representar la media y 128 para representar el estándar). Luego, se extrae una muestra de la distribución gaussiana con la media y el estándar proporcionados, y se reconstruye en las dimensiones de la imagen de entrada siguiendo la misma arquitectura que el decodificador, excepto que aquí no usamos las conexiones de salto entre niveles desde el codificador. El método propuesto en [Myronenko, 2018], es el que mejor resultados a traído en la actualidad, alcanzando un accuracy de 0.87.

En muchos casos las nuevas propuestas, como es el caso de [Myronenko, 2018] [Isensee et al., 2018] [Kamnitsas et al., 2017], tienen un desafío aún inerte a ellos, el cual es uso de los recursos, no se trata de usar gran cantidad de estas, sino en una manera eficiente e inteligente, tanto como memoria, procesamiento, etc, siendo este un problema muy común en gran parte de las técnicas existentes. Por ello el trabajo esta centrado en reducir el tiempo y hacer un uso eficiente de memoria en nuestra propuesta.

Capítulo 4

Propuesta

Como vimos en el estado del arte, en los últimos años se propusieron un gran número de soluciones algorítmicas para resolver el problema de segmentación de tumores cerebrales, trabajos como [Myronenko, 2018] [Kamnitsas et al., 2017] [Havaei et al., 2016]. Gran parte de estos métodos de segmentación automática se basan en una CNN entrenada con parches de imágenes MRI de distintos pacientes. Estos con suficiente datos de entrenamiento, aprenden a generalizar a través de los pacientes y permiten identificar el tumor y su diseminación en datos MRI que no se probaron. Sin embargo, hay al menos dos desafíos relacionas con estos métodos basados en CNN. Primero, tienden a sobre-ajustarse (overfit) a los datos de entrenamiento, por lo que nos lleva a necesariamente tener grandes conjuntos de datos, o utilizar una variedad de técnicas de alimentación de datos para hacerlos generalizar de manera más solida. en segundo lugar, muchas implementaciones actuales basadas en CNN requieren potentes recursos computacionales para ser utilizados en un tiempo razonable.

Para resolver estos desafíos y promover el desarrollo de métodos de segmentación automática, BRATS proporciona estos conjuntos de datos MRI segmentados manualmente para que los usuarios prueben nuevas implementaciones. Inspirado en el trabajos ganador presentado en BRATS 2018 y el trabajo de [Ronneberger et al., 2015], presentamos una nueva red CNN Autoencoder Regularization U-Net para la segmentación de tumores cerebrales, pretendemos usar el conjunto de datos de BRATS 2018 para implementar nuestra red neuronal. Como entrada de nuestro modelo, utilizamos imágenes MRI T1, T1ce, T2 y FLAIR de 285 pacientes.

4.1. Arquitectura

Como mencionamos brevemente, nuestro modelo está inspirado en trabajos anteriores [Myronenko, 2018] [Ronneberger et al., 2015]. Ajustaremos la arquitectura presentada por [Myronenko, 2018] para incorporar una segmentación en parches de la imagen de entrada, ya que la entrada completa con una resolución de 240x240x155 voxel, como se usa en el modelo original este es demasiado grande para adaptarse a la mayoría de las tarjetas gráficas disponibles. Esto es cierto incluso con un tamaño de batch de 1. Por lo

tanto, planteamos el uso de bloques 3D de tamaño 80x80x80 y ajustamos la cantidad de filtros para hacer uso completo de la memoria del GPU disponible, lo que lleva a 32 filtros en la primera capa convolucional con un ratio de 2 en las siguientes capas. También reemplazamos las funciones de activación ReLU con LeakyReLU presentada en [Xu et al., 2015] para un mejor rendimiento del modelo original simplificado.

Para ello, también planteamos varios factores de mejora en el rendimiento del modelo:

- Variar el downsampling en las capas convolucionales desde el de stride a max o average pooling.
- Ajustar la proporción entre el número de filtros entre las capas convolucionales.
- Aumentar el número de downsampling y los siguientes pasos de up-sampling.
- Reemplazar GroupNorm por BatchNorm.

La arquitectura general del modelo sigue una estructura muy similar a U-Net presentado por [Ronneberger et al., 2015], con un modulo Variational Autoencoder adicional para regularizar la segmentación de las mascararas tumorales. Como funciones de perdida para la parte del Encoder automático, pretendemos usar el error cuadrático medio entre las imágenes de entrada reconstruida y la real, también la función de pérdida de Kullback-Leibler para garantizar una distribución normal.

4.1.1. Optimización

Para entrenar el modelo, utilizamos una versión ajustada de Dice loss en [Myronenko, 2018]:

$$L_{Dice} = 1 - \left(2 * \frac{\sum(y_{true} * y_{pred}) + s}{(\sum y_{true}^2 + \sum y_{pred}^2) + s} \right) \quad (4.1)$$

donde y_{true} es la mascara 3D real y y_{pred} la predicción 3D correspondiente. Está versión de Dice loss, asegura que la estimación de pérdida se encuentre dentro del intervalo $[0, 1]$. El término de suavidad s , asegura que el modelo pueda predecir 0 voxeles tumorales sin incurrir en una perdida alta en si estimación general. Basándonos en la investigación [Cahall et al., 2019], usamos un s con valor 100.

Autoencoder

La parte del codificador automático de nuestro modelo consta de dos funciones de pérdida. Como pérdida de reconstrucción, utilizamos el error cuadrático medio entre la imagen de entrada reconstruida y la real:

$$L_{L2} = ||y_{true} - y_{pred}||_2^2 \quad (4.2)$$

Además, utilizamos una pérdida Kullback-Leibler para asegurar una distribución normal, siendo N el número de vóxeles en la entrada:

$$L_{KL} = \frac{1}{N} \sum \mu^2 + \sigma^2 - \log \omega^2 - 1 \quad (4.3)$$

con μ y σ^2 la media y la varianza de la distribución estimada. Siguiendo [Myronenko, 2018], ponderamos las pérdidas del autoencoder en 0.1, lo que resulta en una función de pérdida general:

$$L = 0.1 * L_{L2} + 0.1 * L_{KL} + 0.33 * (L_{Dice_{wt}} + L_{Dice_{tc}} + L_{Dice_{et}}) \quad (4.4)$$

Para entrenar el modelo, utilizamos el optimizador Adam [Kingma and Ba, 2015], comenzando con una tasa de aprendizaje de $1e - 4$ y disminuyéndolo de acuerdo con:

$$\alpha = \alpha_0 * \left(1 - \frac{\epsilon}{N_\epsilon}\right)^{0.9} \quad (4.5)$$

con ϵ la época y N_ϵ el número de épocas totales.

4.1.2. Arquitectura

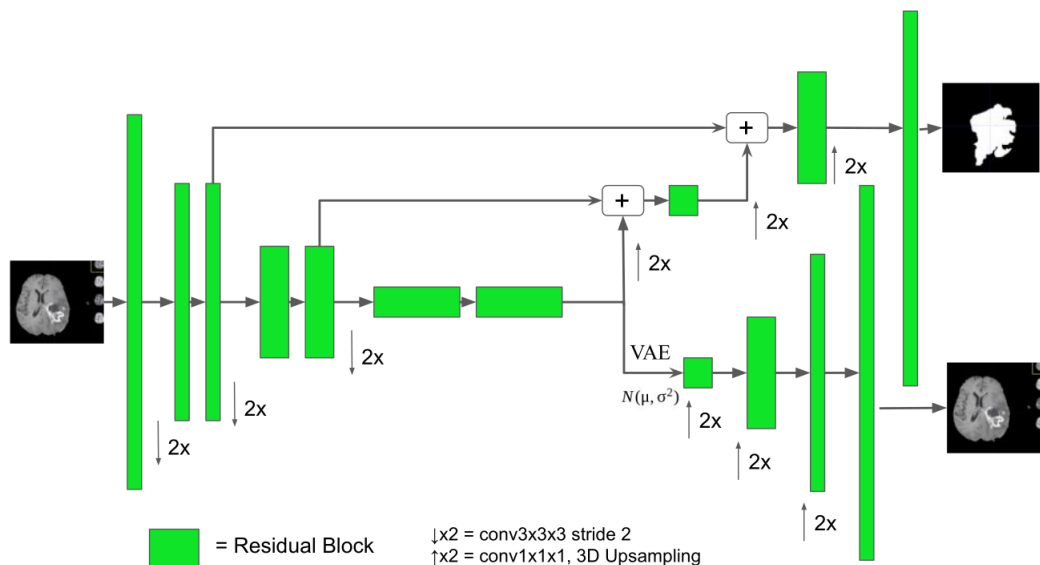


Figura (4.1) Arquitectura propuesta

4.2. Entrenamiento

El entrenamiento modelo consta de tres partes. Primero, en la etapa de codificación, el modelo aprende una representación de baja dimensión de la entrada. En segundo lugar, la etapa del VAE reconstruye la imagen de entrada de un espacio de baja dimensión a la imagen real. Tercero, parte de U-Net crea segmentaciones reales. En esta arquitectura del modelo, se supone que la parte de VAE debe actuar como un fuerte regularizador en todo los pesos del modelo [Myronenko, 2018] y, por lo tanto, para evitar el sobre-ajuste en los datos de entrenamiento.

Abordamos los desafíos presentados anteriormente presentando una solución eficiente en memoria y ampliamente accesible para la segmentación de tumores cerebrales en línea con los objetivos de la iniciativa Green Artificial Intelligence.

Capítulo 5

Resultados

Aquí, presentamos los resultados de nuestro modelo Autoencoder regularizado para la segmentación de tumores cerebrales. El modelo fue entrenado en los datos de entrenamiento BRATS 2018, que consistieron en 285 pacientes separados en casos de glioma de alto grado y glioma de bajo grado. La forma inicial de los datos de entrada fue de 240x240x155, con máscaras de segmentación de etiquetas que corresponden a 1 (CT), 2 (WT) y 4 (ET), esta máscara del mismo tamaño de la imagen de resonancia magnética. Creamos nuestro corte inicial que resultó en dimensiones de imagen de 120x120x96.

5.1. Interpretación de los resultados

El problema que estamos tratando de resolver es la segmentación de tumores cerebrales en imágenes MRI, así que el resultados que nos presenta, es un voxel que contiene las etiquetas (1,2,4) del tumor. Estas etiquetas tienen una representación del tipo de tumor que se ha segmentado.

5.1.1. Etiquetas

- **Whole Tumor (WT):** Clase que incluye toda las etiquetas visibles, (WT, TC, ET) donde el tumor se encuentra, el cual es detectado como un edema. *2*
- **Core Tumor (TC):** Clase que contiene las etiquetas TC y ET, esta representa necrosis y tumor no potenciador. *1*
- **Enhancing Tumor Core (ET):** Clase que representa el tumor activo/potenciador. *4*

5.1.2. Métricas

Para evaluar el perfomance de nuestra segmentación, tenemos dos métricas: Coeficiente Dice y la Distancia Hausdorff.

- Coeficiente Dice: Este coeficiente a menudo es usado para cuantificar el rendimiento de los métodos de segmentación de imágenes. Para hacer uso de esta métrica, necesitamos una región de verdad (GroundTruth) y la parte segmentada con nuestro algoritmo. Se valida el algoritmo calculando el coeficiente Dice, el cual es una medida de cuan similar son las dos segmentaciones. Por lo tanto es el tamaño de la superposición de las dos segmentaciones dividido por el tamaño total de los dos objetos.
- Distancia Hausdorff: Generalmente es sensible a los valores atípicos. Debido que el ruido y los valores atípicos son comunes en segmentación, no se recomienda su uso directamente. Sin embargo, el método cuantil propuesto por [Huttenlocher et al., 1993] es una forma de manejar los valores atípicos. Según el método de cuantiles de Hausdorff, la distancia es definida como el cuantil $q_i h$ de distancias en lugar del máximo, de modo que se excluyen los posibles valores atípicos, donde se selecciona q dependiendo de la aplicación y la naturaleza de los conjuntos de puntos medidos .

5.2. Resultados

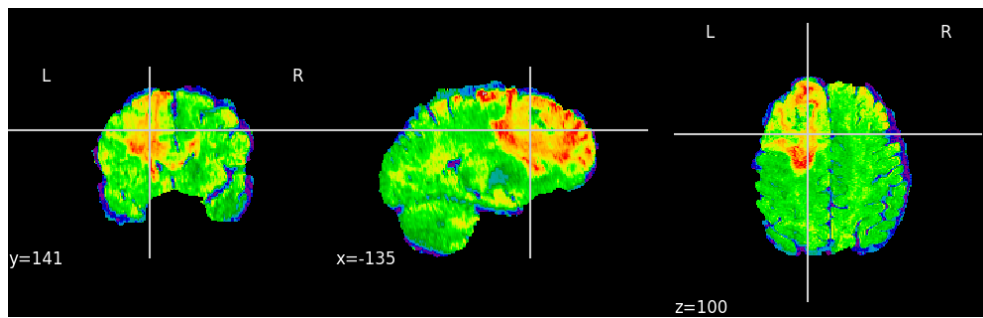
Utilizamos el conjunto de datos de entrenamiento provisto por BRATS 2018. Las métricas promedio alcanzadas se presenta en la tabla 5.1

BRATS 2015	Dice			Hausdorff		
	WT	TC	ET	WT	TC	ET
[Kamnitsas et al., 2017]	0.85	0.67	0.63	0.85	0.86	0.63
[Kayalibay et al., 2017]	0.85	0.72	0.61	0.82	0.77	0.61
[Isensee et al., 2018]	0.85	0.74	0.64	0.83	0.80	0.63
BRATS 2018	WT	TC	ET	WT	TC	ET
[Myronenko, 2018]	0.904	0.859	0.814	4.483	8.278	3.805
[Isensee et al., 2019]	0.878	0.806	0.778	6.03	5.08	2.90
[McKinley et al., 2019]	0.901	0.854	0.795	4.26	5.37	3.61
Propuesta	0.836	0.760	0.737	8.472	8.327	6.048

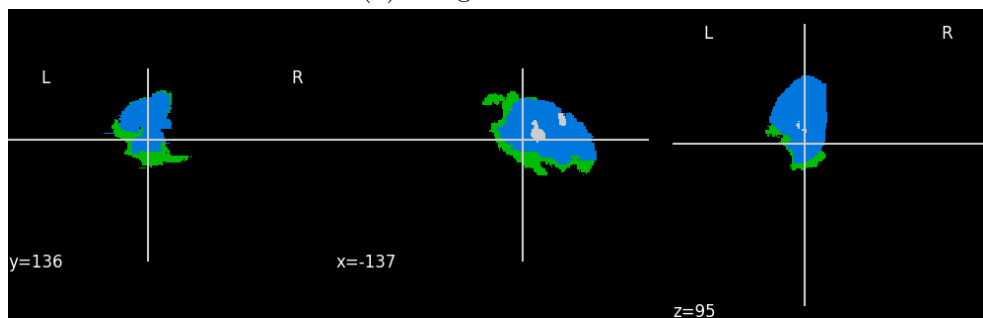
Tabla (5.1) Promedio de las métricas Coeficiente Dice y la Distancia Hausdorff (WT, TC, ET)

5.3. Segmentación

En esta sección mostramos los resultados con nuestra propuesta de manera visual, tenemos segmentaciones de los posibles tumores en 3 distintos pacientes. En las figuras 5.3 5.1 5.2, se muestran las segmentaciones del modelo para pacientes del conjunto de validación.

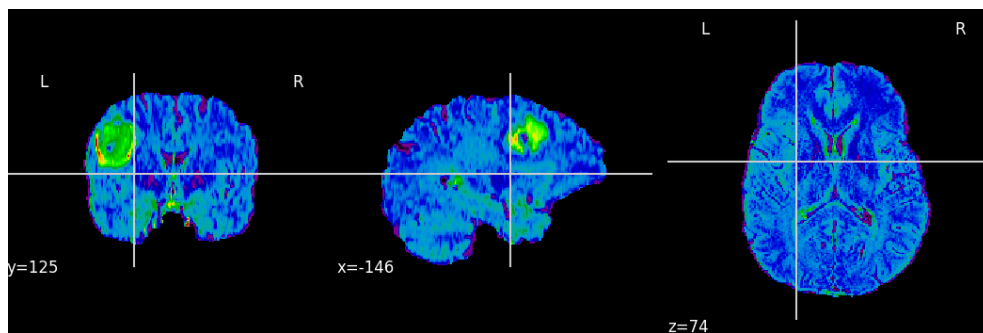


(a) Imagen MRI Flair

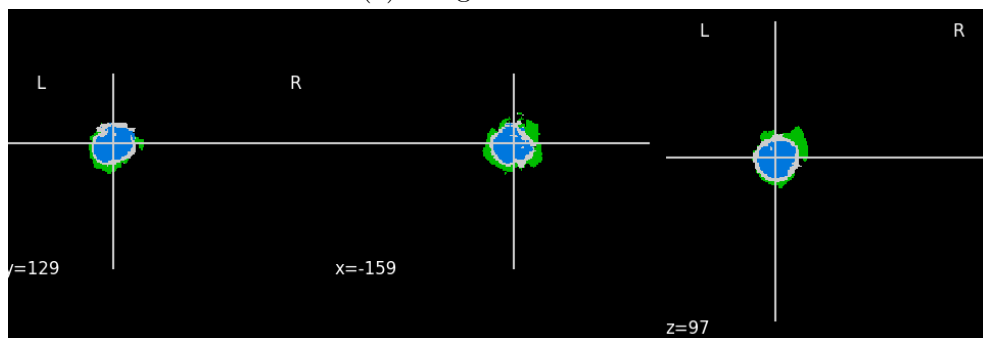


(b) Segmentación del tumor

Figura (5.1) Segmentación en un corte axial, sagital y coronal en el tipo Flair de la imagen MRI.

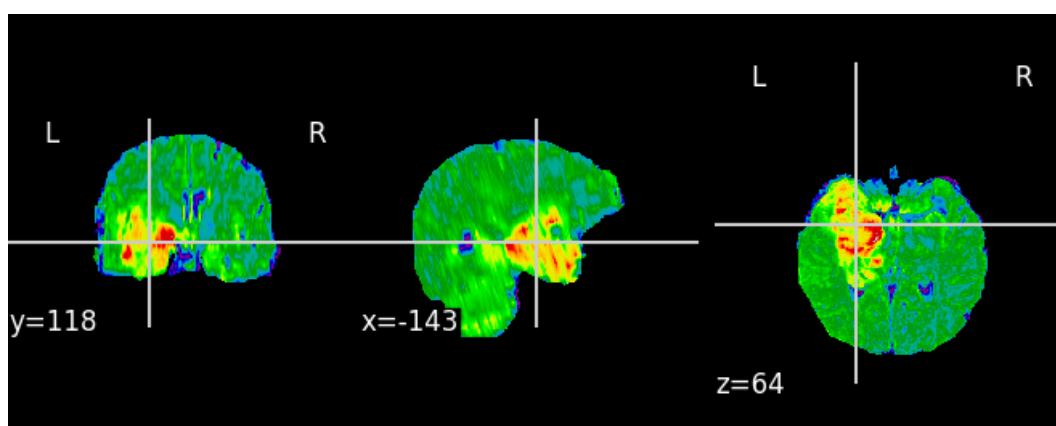


(a) Imagen MRI T1ce

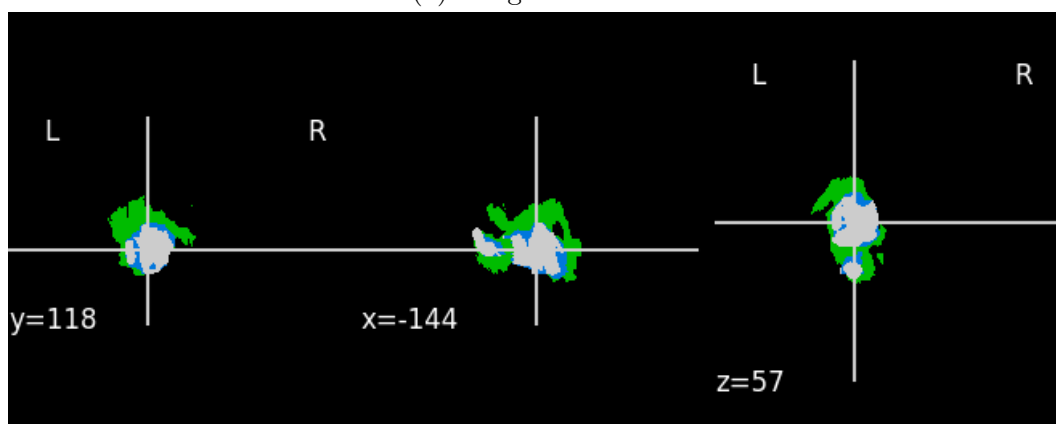


(b) Segmentación del tumor

Figura (5.2) Segmentación en un corte axial, sagital y coronal en el tipo T1ce de la imagen MRI.



(a) Imagen MRI T1



(b) Segmentación del tumor

Figura (5.3) Segmentación en un corte axial, sagital y coronal en el tipo T1 de la imagen MRI. WT representa la unión de azul, verde y blanco, TC está representado por la unión de azul y blanco, y finalmente tenemos ET, que está representado por azul.

Capítulo 6

Conclusiones

La segmentación de tumores cerebrales a menudo se basa en la segmentación manual realizada por personal médico capacitado. En este documento se presenta una solución rápida, automatizada y precisa para este problema. Nuestras segmentaciones se pueden utilizar para informar a los médicos y ayudar al proceso de diagnóstico. Hemos segmentado con éxito tumores cerebrales en tres tipos de tejido: Whole Tumor (WT), Enhancing Tumor Core (ET) y Tumor Core (TC) en 66 pacientes proporcionados por BRATS 2018.

Para el entrenamiento, utilizamos Google Colaboratory, que es una máquina virtual ofrecida por Google para Deep Learning. Esta máquina cuenta con una tarjeta gráfica Tesla K80 14GB RAM. Utilizamos el conjunto de datos [Menze et al., 2014] 2018 para entrenar la red convolucional, que presentan 285 imágenes de resonancia magnética, de las cuales 210 son del tipo HGG y 75 del tipo LGG. Para satisfacer estas bajas demandas de memoria, recortamos la imagen inicial en parches de 80x80x80. El modelo iteró a través de estos parches y convergió en la segmentación cerebral más probable dada todas las iteraciones al final. Por lo tanto, presentamos una solución de memoria eficiente y asequible para la segmentación del cerebro.

Naturalmente, una limitación de este enfoque de bajo costo es que el modelo todavía es relativamente lento. Naturalmente, más recursos computacionales aliviarían este problema. Además, una mayor cantidad de memoria gráfica permitiría aumentar aún más el tamaño del parche de entrada, lo que a su vez también beneficiaría enormemente el rendimiento del modelo. Con toda estas limitaciones, aún nuestros resultados están lejos del estado del arte [Myronenko, 2018], vemos que se aproximan, pero aún así es mucha la diferencia, alrededor de un 15% es la pérdida que existe entre nuestros y los resultados de [Myronenko, 2018], [Isensee et al., 2019] y [McKinley et al., 2019], los tres primeros puestos en el concurso BRATS 2018.

Además, implementamos el modelo utilizando los datos proporcionados solo para el desafío BRATS 2018. No se utilizaron otros datos. El trabajo de [Myronenko, 2018], incluidos los desafíos anteriores de BraTS, mostraron que incorporar datos adicionales, por lo tanto, aumentar el conjunto de datos de entrenamiento, reduce en gran medida el sobreajuste y mejora drásticamente el rendimiento del modelo. Aquí, nuestro objetivo era optimizar la segmentación de los tumores cerebrales explícitamente en vista de estas limitaciones comunes de recursos informáticos y de datos. Una observación interesante fue

que el modelo funcionó bien en la mayoría de los pacientes, pero falló por completo en unos pocos. Las razones de esto siguen sin estar claras y deben explorarse en el futuro.

En conjunto, nuestros resultados demuestran la amplia aplicabilidad de las U-Nets para mejorar la segmentación de tejidos y el diagnóstico médico. Demostramos que las arquitecturas de modelos eficientes en memoria dedicada pueden superar las limitaciones de recursos informáticos y de datos y que la segmentación de tumores cerebrales rápida y eficiente se puede lograr en un hardware ampliamente accesible.

Bibliografía

- [Armstrong et al., 2004] Armstrong, T. S., Cohen, M. Z., Weinberg, J., and Gilbert, M. R. (2004). Imaging techniques in neuro-oncology. *Seminars in Oncology Nursing*, 20(4):231 – 239.
- [Cahall et al., 2019] Cahall, D. E., Rasool, G., Bouaynaya, N. C., and Fathallah-Shaykh, H. M. (2019). Inception modules enhance brain tumor segmentation. *Frontiers in computational neuroscience*, 13:44.
- [Finzel, 2017] Finzel, M. (2017). Convolutional neural networks in medical imaging. *Morris Undergraduate Journal*, 4.
- [Götz et al., 2014] Götz, M., Weber, C., Blöcher, J., Stieltjes, B., Meinzer, H.-P., and Maier-Hein, K. (2014). Extremely randomized trees based brain tumor segmentation. In *MICCAI BraTS (Brain Tumor Segmentation) Challenge. Proceedings, winning contribution*, pages 6–11.
- [Havaei et al., 2017] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., and Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35:18 – 31.
- [Havaei et al., 2016] Havaei, M., Dutil, F., Pal, C., Larochelle, H., and Jodoin, P.-M. (2016). A convolutional neural network approach to brain tumor segmentation. In *MICCAI BraTS (Brain Tumor Segmentation) Challenge. Proceedings, winning contribution*, volume 9556, pages 195–208.
- [Huttenlocher et al., 1993] Huttenlocher, D. P., Klanderman, G. A., and Rucklidge, W. J. (1993). Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863.
- [Isensee et al., 2018] Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., and Maier-Hein, K. (2018). Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In Menze, B., Crimi, A., Kuijf, H., Reyes, M., and Bakas, S., editors, *Brainlesion, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 287–297. Springer Verlag.
- [Isensee et al., 2019] Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., and Maier-Hein, K. H. (2019). No new-net. In Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., and van Walsum, T., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 234–244, Cham. Springer International Publishing.

-
- [Kamnitsas et al., 2017] Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., and Glocker, B. (2017). Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61 – 78.
- [Kaur et al., 2017] Kaur, T., Saini, B. S., and Gupta, S. (2017). Quantitative metric for mr brain tumour grade classification using sample space density measure of analytic intrinsic mode function representation. *IET Image Processing*, 11(8):620–632.
- [Kayalibay et al., 2017] Kayalibay, B., Jensen, G., and van der Smagt, P. (2017). Cnn-based segmentation of medical imaging data. *CoRR*, abs/1701.03056.
- [Kingma and Ba, 2015] Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.
- [Long et al., 2015] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440.
- [McKinley et al., 2019] McKinley, R., Meier, R., and Wiest, R. (2019). Ensembles of densely-connected cnns with label-uncertainty for brain tumor segmentation. In Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., and van Walsum, T., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 456–465, Cham. Springer International Publishing.
- [Menze et al., 2014] Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, L., Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Antonio Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S. J., Riklin-Raviv, T., Reza, S., Ryan, M., Schwartz, L., Shin, H.-C., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Hye Ye, D., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., and Van Leemput, K. (2014). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, page 33.
- [Menze et al., 2015] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., and Kirby, J. (2015). The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024.
- [Myronenko, 2018] Myronenko, A. (2018). 3d MRI brain tumor segmentation using autoencoder regularization. *CoRR*, abs/1810.11654.
- [Pereira et al., 2015] Pereira, S., Pinto, A., Alves, V., and Silva, C. A. (2015). Deep convolutional neural networks for the segmentation of gliomas in multi-sequence mri. In *MICCAI BraTS (Brain Tumor Segmentation) Challenge. Proceedings, winning contribution*, pages 52–55.

-
- [Prastawa et al., 2003] Prastawa, M., Bullitt, E., Ho, S., and Gerig, G. (2003). Robust estimation for brain tumor segmentation. In Ellis, R. E. and Peters, T. M., editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003*, pages 530–537, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*.
- [Ruan et al., 2007] Ruan, S., Lebonvallet, S., Merabet, A., and Constans, J. (2007). Tumor segmentation from a multispectral mri images by using support vector machine classification. In *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1236–1239.
- [Siegel et al., 2019] Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, 69(1):7–34.
- [Urban et al., 2014] Urban, G., Bendszus, M., Hamprecht, F. A., and Kleesiek, J. (2014). Multi-modal brain tumor segmentation using deep convolutional neural networks. In *MICCAI BraTS (Brain Tumor Segmentation) Challenge. Proceedings, winning contribution*, pages 31–35.
- [Vaidhya et al., 2016] Vaidhya, K., Thirunavukkarasu, S., Alex, V., and Krishnamurthi, G. (2016). Multi-modal brain tumor segmentation using stacked denoising autoencoders. In *MICCAI BraTS (Brain Tumor Segmentation) Challenge. Proceedings, winning contribution*, volume 9556, pages 181–194.
- [Wang et al., 2018] Wang, G., Li, W., Ourselin, S., and Vercauteren, T. (2018). Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In Crimi, A., Bakas, S., Kuijf, H., Menze, B., and Reyes, M., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 178–190, Cham. Springer International Publishing.
- [Watanabe et al., 2009] Watanabe, H., Okada, M., Kaji, Y., Satouchi, M., Sato, Y., Yamabe, Y., Onaya, H., Endo, M., Sone, M., and Arai, Y. (2009). New response evaluation criteria in solid tumours - revised recist guideline (version 1.1). *Gan to kagaku ryoho. Cancer & chemotherapy*, 36:2495–501.
- [Xu et al., 2015] Xu, B., Wang, N., Chen, T., and Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*.
- [Zhao et al., 2018] Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., and Fan, Y. (2018). A deep learning model integrating fcnn and crfs for brain tumor segmentation. *Medical Image Analysis*, 43:98 – 111.
- [Zhou et al., 2019] Zhou, C., Chen, S., Ding, C., and Tao, D. (2019). Learning contextual and attentive information for brain tumor segmentation. In Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., and van Walsum, T., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 497–507, Cham. Springer International Publishing.

-
- [Zikic et al., 2014] Zikic, D., Ioannou, Y., Brown, M., and Criminisi, A. (2014). Segmentation of brain tumor tissues with convolutional neural networks. In *MICCAI BraTS (Brain Tumor Segmentation) Challenge. Proceedings, winning contribution*, pages 33–39.