

Segmentación de Imágenes Médicas

1^{ro} Alex Sanchez

Informática

Pontificia Universidad Católica del

Perú

Lima, Perú

a20194323@pucp.edu.pe

2^{do} Daniel Silva

Informática

Pontificia Universidad Católica del

Perú

Lima, Perú

dsilvar@pucp.edu.pe

3^{ro} Daniel Ibáñez

Ingeniería Informática

Pontificia Universidad Católica del

Perú

Lima, Perú

daniel.ibanez@pucp.edu.pe

Abstract— En la actualidad, se tiene muchas técnicas de aprendizaje para diferentes arquitecturas de redes neuronales, las cuales pueden ser usadas en procesamiento de imágenes o procesamiento de texto. Muchas de ellas requieren mucho tiempo de cómputo para ser entrenadas o utilizan enormes datasets. Esto dificulta su entrenamiento en PC/Laptops básicas. En este trabajo de investigación se presenta a la red U-Net++ y se utiliza para la segmentación de imágenes médicas. Para ello, se hará uso de un dataset con las anotaciones correspondientes a cada una de las imágenes. Esta red puede ser entrenada con facilidad de extremo a extremo. Sin embargo, los resultados dependen mucho de la resolución de la imagen. Considerando la segmentación de una imagen de 512x512 píxeles, el cómputo toma menos de un segundo en GPUs de alto rendimiento.

Keywords— FCN, CNN, Image Segmentation, Image Processing U-Net, U-Net++

I. INTRODUCCIÓN

La segmentación de imágenes médicas está fuertemente dominada por redes neuronales convolucionales profundas (CNN). Asimismo, muchas investigaciones siguen siendo publicadas en este campo de estudio. Sin embargo, debemos considerar que según el escenario de segmentación, se requiere modificar o crear nuevas arquitecturas especializadas para así obtener el mejor resultado considerando los mejores tiempos. Además, considerando la limitación en los datasets de entrenamiento y validación, las implementaciones pueden verse restringidas, por ejemplo en algún centro de salud pública o privada.

La identificación automática o semi-automática de secciones de interés en imágenes médicas que ayuden a diagnosticar enfermedades es una tarea de suma importancia y por el impacto que tiene es necesario que estas sean precisas y confiables, la precisión en la identificación será clave para los tratamientos relacionados a tales enfermedades, incluso a salvar vidas. Con las redes neuronales se busca realizar esta tarea, la segmentación semántica de imágenes médicas es el área de la computación visual que se encarga de estudiar esto. Existen investigaciones sobre modelos de redes neuronales que puedan conseguir con mayor precisión la identificación de secciones de interés relacionadas a enfermedades; un mejor modelo dependerá de su estructura y de la data con la que es entrenada, y aquí es donde se ha visto la escasez de dataset respecto a tópicos médicos puntuales[6].

En el siguiente informe, revisaremos a una de las más famosas Fully Convolutional Networks (FCN), la red U-Net++ que es utilizada en segmentación de imágenes biomédicas, la cual fue publicada en 2020. Considerando

que en la anotación de imágenes biomédicas, siempre se necesita de expertos con conocimiento adquirido en los años de trabajo, este proceso aún sigue siendo manual y toma mucho tiempo. Si el proceso de anotación de imágenes se realiza de manera automática, se logra un menor esfuerzo y se reduce el error por parte de los expertos, esto también conlleva a la reducción de los costos.

El informe está compuesto por Estado de Arte, donde revisaremos trabajos o investigaciones que abordan el tema de interés y nos permitan ahondar en experiencias relacionadas a nuestra investigación; Metodología, en el cual se explicará el proceso de segmentar una imagen en sus diferentes regiones de interés; Experimentación y Resultados, donde detallamos el proceso de entrenamiento de la red propuesta (U-Net++), descripción del dataset utilizado y los resultados obtenidos, así mismo interpretamos el resultado y presentaremos nuestras conclusiones.

II. ESTADO DEL ARTE

Las Fully Convolutional Networks (FCN) fueron introducidas por [1] J. Long, E. Shelhamer and T. Darrell, a su vez que U-Net fue presentado por [2] Ronneberger, O., Fischer, P., & Brox, ambas en el año 2015, las 2 arquitecturas omiten conexiones. En FCN, los mapas de características muestreados arriba se enumeran con los mapas de características omitidos del codificador, en U-Net, estos son concatenados, agrega convoluciones y no-linealidad entre cada paso de muestreo ascendente. La omisión de conexiones ayudan a recuperar la resolución espacial en la salida de la red, lo que hace que los métodos FCN sean adecuados para la segmentación semántica, este tipo de redes fueron inspirados por la arquitectura DenseNet [3]. A pesar de las pequeñas diferencias entre la FCN y U-Net, estas tienden a fusionar mapas de características semánticamente diferentes de las subredes del codificador y decodificador, llegando en algunos casos a degradar el rendimiento de la segmentación.

Las redes [4] nnU-Net propone resolver nuevos problemas, los cuales eran complejos de resolver usando la red U-Net, esta nueva arquitectura es auto-adaptable tomando como base las redes U-Net, esto permitiendo obtener mejores resultados en la segmentación de imágenes. Las redes [5] U-Net++ fueron propuestas por Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, esta arquitectura de red es una evolución de las redes U-Net, con mejoras en la arquitectura, esto permitiendo mejorar los resultados en la segmentación de imágenes médicas. En esta nueva arquitectura, se rediseñó las rutas de omisión, con el fin de reducir la brecha

semántica entre los mapas de características de las subredes del codificador y decodificador. Esta nueva arquitectura es esencialmente una red de codificador-decodificador profundamente supervisada donde las subredes del codificador y el decodificador están conectadas a través de una serie de vías de salto densas anidadas.

III. METODOLOGÍA

A. Descripción del dataset

En este caso se está considerando el dataset de Hipocampo, el cual fue publicado en la “Medical Segmentation Decathlon” en la siguiente url: <http://medicaldecathlon.com/>:

El conjunto de datos está compuesto de Imágenes por Resonancia Magnética (IRM), el cual se subdivide de la siguiente manera:

TABLE I. COMPOSICIÓN DEL DATASET

Descripción	# Imágenes
Adultos Sanos	90
Adultos Enfermos	105
Total Imágenes	205

Las imágenes de los adultos enfermos se subdividen según la enfermedad que sufren, los cuales se muestran en la siguiente manera:

TABLE II. IMÁGENES DE ADULTOS ENFERMOS

Descripción	# Imágenes
trastorno psicótico no afectivo (esquizofrenia)	56
trastorno esquizo-afectivo	32
trastorno esquizo-freniforme	17
Total Imágenes	105

Todas las personas consideradas estaban libres de enfermedades médicas o neurológicas significativas, lesiones en la cabeza y/o uso dependencia de sustancias activas.

B. Características de las imágenes

TABLE III. CARACTERÍSTICAS DE IMÁGENES

Descripción	Valor	Observación
Tipo de Secuencia	MPRAGE	ponderada en 3D T1 (TI / TR / TE, 860 / 8.0 / 3.7 ms; 170 cortes sagitales; tamaño de vóxel, 1.0 mm3)
Tipo de Escáner	Philips Achieva	Philips Healthcare, Inc

Para obtener las definiciones de TI, TR y TE, se revisó las definiciones encontradas en [7].

- *TI*: Medium Inversion Time
- *TR*: Long Repetition Time
- *TE*: Rapidly Acquired Gradient Echoes

C. Descripción de la Arquitectura

El modelo en la que se basa esta investigación es la [5] U-Net++, propuesta en 2020, la red es una evolución de la red [2] U-Net, presentada en 2015. Para hablar de [5]U-Net++ es necesario hablar de U-Net como red base.

El Modelo [2]U-Net cuenta con una arquitectura simétrica, a este modelo lo podríamos dividir en 2 secciones, la sección de la contracción (down-sampling) y de expansión (Up-convolution). En la contracción, se aplica dos convoluciones de 3x3 (sin padding), cada una seguida de un ReLU, la capa ReLU está seguida de un max pooling de 2x2 en sus niveles, para cada paso de down-sampling duplica el número de canales. Para la sección de la expansión, aplica convoluciones de 3x3 (seguidas por un ReLU) y una up-convolution de 2x2, en esta parte el número de canales es dividida a la mitad para cada paso. Después de cada up-convolution de 2x2 (de la sección de expansión) se añade el mapa de características de la capa correspondiente de la contracción para contrarrestar la pérdida de información por la reducción de los canales en cada convolución.

Para [5]U-Net++, la novedad está es que entre cada nivel de la red se implementan redes ocultas con conexiones densas bajo supervisión profunda, todas las convoluciones internas son de 3x3, las capas densas aseguran que las características se acumulen y al llegar al último nodo lo enriquezcan, la supervisión profunda ayuda a encontrar un equilibrio entre velocidad y performance del modelo.

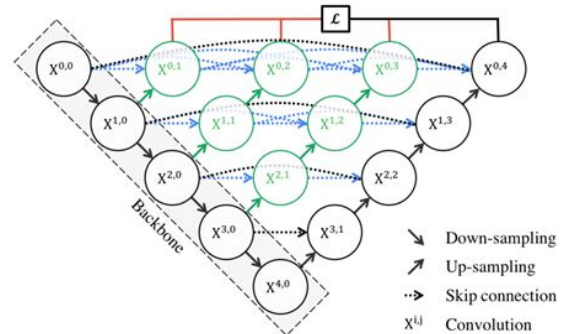


Fig. 1. Arquitectura del modelo U-Net++.

D. Métricas

Para la evaluación del modelo utilizaremos las siguientes métricas:

- 1) *Binary Cross Entropy*: Métrica usada para medir la pérdida en cuanto a la clasificación binaria del modelo.
- 2) *Coefficiente de Sorensen-Dice*: Métrica usada para medir la superposición, traslapo de la sección predecida sobre el área etiquetada de la imagen. El cálculo de esta métrica se basa en dos veces el área de intersección entre la imagen predecida y la etiquetada, esto, dividido entre la suma de ambas áreas.
- 3) *Intersection over Union (IoU)*: Métrica similar a la de coeficiente Dice; el cálculo está en base al área interceptada dividida entre la unión del área predecida y la etiquetada.

E. Estrategia de validación a emplear para el ajuste de hiper-parámetros

La optimización de hiper-parámetros es considerada como parte importante en el aprendizaje profundo, ya que esto nos puede permitir obtener mejores resultados al momento de realizar el entrenamiento de una red neuronal.

El motivo de buscar los parámetros es porque las redes neuronales son complejas de configurar y tienen muchos parámetros a establecer, también debemos considerar que son lentos al momento de entrenar.

1) *Grid Search*: Es una técnica de optimización de hiperparámetros para un modelo, es usado en modelos de aprendizaje supervisado y no supervisado; para hacer uso de esta técnica, se puede utilizar la implementación de scikit-learn, el parámetro $K=5$ o $K=10$ son los más recomendados.

2) *Batch Size y Número de Épocas*

a) *Batch Size*: Se define como el tamaño del lote en el descenso de gradiente iterativo es el número de patrones que se muestran a la red antes de actualizar los pesos, este parámetro también indica cuántos patrones se va leer y guardar en la memoria.

b) *Número de Épocas*: Se define como el número de veces que se muestra todo el conjunto de datos de entrenamiento a la red durante el entrenamiento. Tener en cuenta que algunos modelos de redes neuronales, como las redes neuronales recurrentes LSTM y las redes neuronales convolucionales.

3) *Training Optimization Algorithm*: Entrenar el modelo con diferentes algoritmos de optimización, cada framework tiene implementado diferentes algoritmos de optimización. Para el caso de Keras, se tiene los siguientes algoritmos de optimización: SGD, RMSprop, Adam, Adadelta, Adagrad, Adamax, Nadam.

4) *Learning Rate and Momentum*: El Learning Rate (tasa de aprendizaje) controla cuánto actualizar el peso al final de cada batch y el Momentum (impulso) controla cuánto dejar que la actualización anterior influya en la actualización de peso actual. Learning Rate and Momentum comunes son de 0.2 a 0.8 en pasos de 0.2, así como 0.9.

5) *Network Weight Initialization*: La inicialización del peso de la red neuronal se puede realizar utilizando diferentes técnicas, Para el caso de Keras, se tiene los siguientes técnicas de inicialización: RandomNormal, RandomUniform, TruncatedNormal, Zeros, Ones, GlorotNormal, GlorotUniform, Identity, Orthogonal, Constant, VarianceScaling.

6) *Neuron Activation Function*: La función de activación controla la no linealidad de las neuronas individuales y cuándo activarse. Para el caso de Keras, se tiene los siguientes técnicas de inicialización: relu, sigmoid, softmax, softplus, softsign, tanh, selu, elu, exponential.

7) *Dropout Regularization*: Dropout permite regularizar el sobreajuste y mejorar la capacidad de generalización del modelo.

F. Explicación y diagrama del proceso propuesto

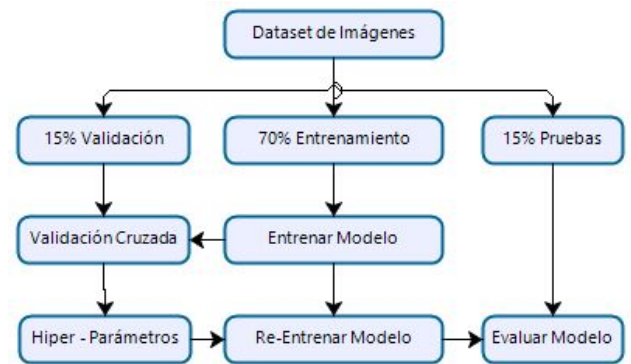


Fig. 2. Arquitectura del modelo U-Net++

En el diagrama de proceso se muestra lo siguiente:

1) El dataset de imágenes será particionado de la siguiente manera:

- 70% para el entrenamiento
- 15% para Validación
- 15% para Pruebas

2) Se valida el modelo usando validación cruzada con el set de validación

3) Para encontrar los hiper-parámetros se utilizará las estrategias indicadas en la sección “Estrategia de validación a emplear para el ajuste de hiper-parámetros”

4) El modelo se vuelve a entrenar con los hiper-parámetros encontrados en el paso anterior

5) Se realiza la evaluación final del modelo utilizando el dataset de datos pruebas y el modelo con los hiper-parámetros re-entrenado.

IV. EXPERIMENTACIÓN Y RESULTADOS

texto

V. DISCUSIÓN

texto

REFERENCIAS

- [1] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 3431-3440, doi: 10.1109/CVPR.2015.7298965.
- [2] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation.
- [3] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, volume 1, page 3, 2017.
- [4] Isensee, Fabian & Petersen, Jens & Klein, André & Zimmerer, David & Jaeger, Paul & Kohl, Simon & Wasserthal, Jakob & Koehler, Gregor & Norajitra, Tobias & Wirkert, S. & Maier-Hein, Klaus. (2018). nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation.
- [5] Zhou, Zongwei & Rahman Siddiquee, Md Mahfuzur & Tajbakhsh, Nima & Liang, Jianming. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation.
- [6] Amber Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram van Ginneken, AnnetteKopp-Schneider, Bennett A. Landman, Geert Litjens, BjoernMenze, Olaf Ronneberger, Ronald M. Summers, Patrick Bilic, Patrick F. Christ, Richard K. G. Do, Marc Gollub, JenniferGolia-Pernicka, Stephan H. Heckers, William R. Jarnagin, Maureen K. McHugo, Sandy Napel, Eugene Vorontsov, LenaMaier-Hein, and M. Jorge Cardoso (2019). A large

annotated medical image dataset for the development and evaluation of segmentation algorithms.

- [7] What is the difference between MPRAGE and MP2RAGE. Consulta: 24 de junion de 2020. <http://mriquestions.com/mp-rage-v-mr2rage.html>
- [8] <https://github.com/shailensobhee/medical-decathlon> [Referencia por revisar]