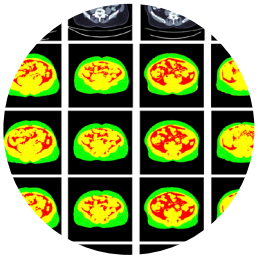

Estudio comparativo de modelos de aprendizaje profundo para segmentar tejido adiposo abdominal en tomografía axial computarizada

Comparative study of Deep Learning models for segmenting abdominal adipose tissue in CT scans

Juan Pablo Reyes[1]
Cristian Mateo Amaya Porras[2]
Leonardo Mejía Bustos[2]
Luis Felipe Uriza Carrasco[3]
Álvaro Ruiz Morales[4]
Diego Ortiz Santos[5]
Catalina Barragán[6]
Carlos José Castro[7]
Marcela Hernández Hoyos[8]

https://doi.org/10.53903/01212095.280

Palabras clave (DeCS)

Grasa abdominal
Grasa intraabdominal
Tomografía computarizada por rayos X

Key words (MeSH)

Abdominal fat
Intra-abdominal fat
Tomography, X-Ray computed

Resumen

Propósito: El análisis de composición corporal sirve como indicador de ciertas condiciones médicas como el síndrome metabólico, el cáncer, la diabetes o las enfermedades cardiovasculares. Tradicionalmente, estos análisis se realizan mediante métodos antropométricos o herramientas clínicas que proporcionan un resultado aproximado. Usando la familia de arquitecturas de Aprendizaje Profundo U-NET, se realizó una segmentación completamente automática del tejido adiposo abdominal visceral y subcutáneo. Se estudiaron estos resultados de segmentación y se compararon con el patrón de oro generado por segmentación manual de expertos. ***Materiales y métodos:*** Se emplearon cuatro variaciones de la arquitectura de Aprendizaje Profundo de U-Net: U-Net, R2U-Net, Attention U-Net y Attention R2U-Net. Estos métodos se entrenaron en un conjunto de datos que consta de 554 imágenes recopiladas entre 2015 y 2017 en el Hospital Universitario San Ignacio y en el Instituto IDIME en Bogotá, Colombia. Esta base de imágenes contiene anotaciones para tres tejidos diferentes: grasa visceral, grasa subcutánea y otros tejidos, generadas a través de herramientas de segmentación semiautomáticas. ***Resultados***: El índice de Sørensen-Dice se utiliza como métrica de evaluación al comparar con los datos obtenidos del patrón de oro, que consiste en segmentaciones manuales realizadas por expertos. Se obtuvo que la arquitectura U-Net fue la más precisa en términos de segmentación de la composición corporal general, con un puntaje promedio de Dice de 93,0 %, seguida de cerca por la arquitectura Attention U-Net con un puntaje promedio de Dice de 92,0 %. ***Conclusiones:*** Según los resultados, se descubrió que las arquitecturas U-Net y Attention U-Net son las más adecuadas para el análisis de la composición corporal. Los resultados de segmentación producidos por estos métodos podrían usarse para obtener métricas precisas y ayudar a los médicos a comprender la condición física del paciente.

Summary

Purpose: Body composition analysis is a test that measures the proportion of various tissues of a person's body. It serves as an indicator for certain medical conditions such as metabolic syndrome, cancer, diabetes, or cardiovascular disease. Traditionally, these analyses are done using anthropometric methods or clinical tools that provide an approximated result. Using the family of U-NET Deep Learning architectures, we perform a fully automatic segmentation of visceral and subcutaneous abdominal adipose tissues. We study these segmentation results and compare them against semiautomatic and manual generated ground truths. ***Materials and methods:*** We employ several variations of the U-Net Deep Learning architecture: U-Net, R2U-Net, Attention U-Net, and Attention R2U-Net. These methods were trained on a dataset, which consists of 554 images from the Hospital Universitario San Ignacio and IDIME Institute in Bogota, Colombia, collected from 2015 to 2017. This dataset contains annotations for three different tissues: visceral fat, subcutaneous fat and other tissue generated through semiautomatic segmentation tools. ***Results:*** Sørensen-Dice index is used as the evaluation metric against the ground truth which consists of manual segmentations performed by experts. We obtained that the U-Net architecture was the most accurate in terms of overall body composition segmentation, with a mean Dice score of 93.0%, followed closely by the

[1]PhD, Departamento de Ingeniería de Sistemas y Computación, Universidad de los Andes. Bogotá, Colombia.

[2]Departamento de Ingeniería de Sistemas y Computación, Universidad de los Andes. Bogotá, Colombia.

[3]Radiólogo, Departamento de Radiología, Hospital Universitario San Ignacio, Pontificia Universidad Javeriana. Bogotá, Colombia.

[4]Profesor titular, Departamento de Epidemiología, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, Colombia.

[5]Radiólogo, Hospital Manuel Uribe Ángel, Clínica Sagrado Corazón. Bogotá, Colombia.

[6]*Fellow* Neurorradiología, Fundación Santa Fe de Bogotá. Bogotá, Colombia.

[7]Radiólogo, Clínica del Country, Clínica de la Colina, Clínica los Nogales. Bogotá, Colombia.

[8]PhD, profesora titular, Departamento de Ingeniería de Sistemas y Computación, Universidad de los Andes. Bogotá, Colombia.

Attention U-Net architecture. ***Conclusions:*** We found that the U-Net and Attention U-Net architectures are more suited for body composition analysis. The segmentation results produced by these methods could be used to obtain precise metrics and help physicians understand the patient's physical condition.

Introducción

El estudio de la composición corporal es importante en la medida en que permite obtener un indicador para detectar diversas condiciones de salud. Mediante el análisis de la cantidad y la distribución de diferentes tipos de tejidos en el cuerpo, en particular tejido adiposo y muscular, es posible determinar la probabilidad de desarrollar ciertas condiciones médicas, como las enfermedades cardiovasculares, la diabetes, la enfermedad renal crónica, los trastornos musculoesqueléticos (1), el cáncer (2) o el síndrome metabólico (3).

A pesar de los avances en la medicina moderna, los métodos de análisis de composición corporal contemporáneos están fuertemente basados en métodos antropométricos, como el cálculo del índice de masa corporal (IMC), el índice de cintura-cadera (WHR, por sus siglas en inglés) o el porcentaje de grasa corporal (BFP, por sus siglas en inglés). Para algunos de estos índices, como el IMC, se ha descubierto una relación entre los mismos y el incremento de riesgo de muerte. Dicho incremento además afecta especialmente a países con niveles sociodemográficos bajos y medios (1).

Una manera de aumentar la precisión de estos métodos de análisis de composición corporal consiste en medir la proporción de la distribución de tejido adiposo para cada paciente. El cálculo de estos indicadores ha sido recientemente incorporado en herramientas clínicas y a través de métodos de computación gráfica y aproximaciones basadas en segmentación de tejidos, la medición de la proporción de estos tejidos se ha hecho más precisa.

Más recientemente se han propuesto nuevas alternativas automáticas y semiautomáticas para la cuantificación de tejido adiposo y muscular, usando en particular métodos de Aprendizaje Profundo. Las arquitecturas U-Net (4) han ganado popularidad para segmentar semánticamente imágenes del torso; muchos de estos avances todavía requieren validación de expertos, pues son susceptibles de aparición de sesgos y requieren aún trabajo para volverse alternativas confiables completamente automatizadas. Los sesgos más importantes en este tipo de trabajos provienen de los instrumentos usados y los operadores, por ejemplo para la obtención de las tomografías (TAC), la selección de conjuntos de datos no-representativos, el uso de datos de entrenamiento, e imágenes con problemas de anotación.

Para abordar dicha brecha, este estudio busca probar en detalle un subconjunto de las arquitecturas de Aprendizaje Profundo U-Net usando como entrada un grupo de imágenes mixtas para entrenar, comparar y evaluar el desempeño de cada una al segmentar tejido abdominal adiposo y tejido muscular.

Estado del arte

Trabajos previos han utilizado umbralización y operaciones de morfología básica para la detección de diferentes tipos de tejidos. Con el tiempo, varios estudios han incorporado también la cuantificación de tejido corporal para obtener información acerca de las condiciones de salud que un paciente pueda tener. Los avances más recientes han refinado las segmentaciones usando técnicas más avanzadas como crecimiento de regiones dinámico o modelos de contorno activo (5): algoritmos basados en imágenes para detectar las regiones estrechas que conectan los tejidos subcutáneo y visceral en el tronco (6) o algoritmos basados en análisis morfológico a partir de umbralizaciones sobre TAC para calcular los volúmenes de las regiones que contienen grasa visceral (7).

Una revisión de trabajos previos revela que los métodos basados en Aprendizaje Profundo se aproximan bastante al desempeño de expertos humanos para la segmentación. Algunos ejemplos en el área han empleado redes neuronales convolucionales (CNN, por sus siglas en inglés) para clasificar tejido adiposo visceral (VAT, por sus siglas en inglés) y tejido adiposo subcutáneo (SAT, por sus siglas en inglés) (8) o redes convolucionales neuronales automatizadas (A-CNN, por siglas en inglés) con el mismo fin (9).

Entre los métodos de Aprendizaje Profundo se destacan las arquitecturas basadas en redes neuronales convolucionales profundas (RNCP). En particular, la familia de arquitecturas U-Net ha ganado prevalencia. La arquitectura original U-Net fue desarrollada como una RNCP modificada cuya aplicación se demostró inicialmente para segmentación de células (4). Posteriormente, otros trabajos han demostrado el alcance de esta arquitectura. Versiones extendidas de la arquitectura que usan módulos residuales y de atención han tenido éxito en la segmentación del músculo paravertebral en TAC abdominales (10). Otras adaptaciones que emplean una arquitectura denominada U-Net densa han servido para detectar grasa, huesos y músculos en RM a través de un flujo de trabajo, o *"pipeline",* de segmentación (11).

También se han desarrollado variaciones del modelo de arquitectura U-Net en busca de mejorar su eficacia. Se destacan las arquitecturas RU-Net y R2U-Net que fueron presentadas y probadas en conjuntos de datos de prueba para segmentar vasos sanguíneos en la retina, cáncer de la piel y lesiones pulmonares (12). Otra variación es la arquitectura Attention U-Net, probada en la segmentación de múltiples tejidos en TAC abdominales (13). Finalmente, está la arquitectura Attention R2U-Net, que ha sido probada en los mismos conjuntos de datos que la RU-Net y la R2U-Net (14). Más adelante se explican con detalle estas arquitecturas.

Las U-Net se han usado para análisis de composición corporal en TAC para segmentar músculo y SAT. En particular, esta arquitectura se ha usado para segmentar VAT, músculo y los órganos abdominales a la altura de la vértebra L3 (15). La U-Net también se ha usado para segmentar, además de VAT, tejido intermuscular adiposo a la misma altura (16) y contenido mixto intrapélvico a la altura supraacetabular (17). Finalmente, se han utilizado otras variaciones de la arquitectura, como la U-Net 3D multirresolución, para generar una volumetría de la composición corporal de tejidos en TAC a partir de un conjunto de datos anotados (18).

Todos estos trabajos sugieren que las arquitecturas basadas en U-Net pueden servir como base para la construcción de un método completamente automatizado de segmentación y cuantificación de tejido adiposo y muscular. En ese sentido, este trabajo busca explorar el conjunto de las arquitecturas de Aprendizaje Profundo U-Net, R2U-Net, Attention U-Net y Attention R2U-Net, para estudiar su desempeño y determinar cuál de ellas es la más indicada para resolver el problema de segmentación de VAT y SAT en el análisis de composición corporal.

Métodos y materiales

Imágenes

El conjunto de datos empleado consta de 513 imágenes abdominales de TAC tomadas por el Hospital Universitario San Ignacio y el instituto IDIME en Bogotá, Colombia. Estas imágenes fueron obtenidas del 2015 al 2017 y cada una corresponde a un paciente diferente. De estas imágenes, 185 pertenecen a pacientes de sexo masculino en un rango de edad de entre 17 y 71 años, con una media de edad 48,1 años; 328 de las imágenes pertenecen a pacientes de sexo femenino en un rango de edad de 17 a 87 años, con una media de edad de 46,6 años. Las imágenes tienen una resolución de 0,482×0,482 a 0,953×0,953 mm^2, con tamaño matricial de 512×512 y espaciado entre los cortes de entre 1 y 6 mm. Un radiólogo seleccionó un corte axial para cada paciente de forma manual, a la altura de la cuarta vértebra lumbar, con base en la definición de exceso de grasa visceral; el corte poseía un área mayor a 100 cm^2 de grasa visceral medida en este nivel. La variedad de la población y el uso de imágenes provenientes de dos hospitales permiten mitigar algunos sesgos propios de los conjuntos de datos.

Patrón de oro

Para generar el patrón de oro, o verdad terreno, del tejido adiposo abdominal, tanto subcutáneo como visceral, se usó un conjunto de 41 imágenes adicionales: 25 de hombres y 16 de mujeres. A partir de las segmentaciones manuales de tres expertos se creó un primer conjunto de validación con la intersección de las áreas identificadas por los mismos en cada imagen. Posteriormente, se generó un segundo conjunto de validación a partir de la revisión de las intersecciones por un cuarto experto, y la segmentación resultante fue definida por consenso de los cuatro radiólogos. En este texto se menciona al primer conjunto como *intersección* y al segundo como *consenso*, respectivamente. La conformación del patrón de oro usando esta estrategia permite también mitigar los sesgos cognitivos presentes en las anotaciones de los conjuntos de datos entrenamiento y validación.

Arquitecturas de Aprendizaje Profundo U-Net

En este trabajo se evaluó el desempeño de cuatro arquitecturas de RNCP: U-Net, R2U-Net, Attention U-Net y Attention R2U-Net. Típicamente las RNCP se especializan en la identificación de patrones en imágenes y su arquitectura básica está dividida en diferentes capas de procesamiento: convolución, agrupamiento, activación y completamente conectada.

La capa de convolución es la más importante en este tipo de arquitecturas y está basada en una operación comúnmente utilizada en procesamiento de imágenes llamada *convolución*, la cual consiste en tomar pequeñas partes de una imagen (usando una ventana denominada filtro o *kernel* de pixeles) y combinarlas con la imagen original para resaltar ciertas características. Esto se hace repetidamente en toda la imagen para producir una nueva imagen filtrada que muestra cómo las características se relacionan con diferentes partes de la imagen original. En el aprendizaje automático, las redes neuronales convolucionales utilizan este proceso para realizar tareas como la segmentación y la clasificación de objetos.

La capa de agrupamiento reduce la resolución de la imagen (submuestreo) al tiempo que preserva características. La capa de activación introduce no-linearidad en la arquitectura y normalmente se implementa usando una Unidad Lineal Rectificada (ReLU) que elimina los valores negativos de la imagen filtrada para reemplazarlos por cero. Finalmente, la capa completamente conectada recibe como entrada las salidas de las capas anteriores y determina la probabilidad de cada clase o etiqueta en la imagen; en el caso particular de las escanografías abdominales, la probabilidad de que un pixel pertenezca a las clases VAT, SAT, a otros tipos de tejidos o al fondo de la imagen.

La arquitectura U-Net se basa en la arquitectura de la RNCP, pero añade operadores de interpolación para incrementar la resolución de las imágenes (supermuestreo) y para extraer características de alto nivel; además de no emplear capas completamente conectadas. Se denomina *U-Net* debido a su forma en U, que consiste en un camino que se contrae o que desciende (red de codificación) y un camino que se expande o que asciende (red de decodificación). La arquitectura busca aprender diferentes filtros a lo largo de ambas redes para extraer las características de la imagen que le permitan determinar qué pixeles pertenecen a alguna de las clases usadas para entrenar (SAT, VAT, otros tejidos y fondo).

La figura 1 ilustra las capas presentes en la arquitectura, incluyendo la red de codificación (segmento descendiente) y la red de decodificación (segmento ascendente). La red de codificación de la arquitectura U-Net consiste en cuatro bloques de submuestreo compuestos por las siguientes capas: dos convoluciones con *kernels* de 3×3 con una función de activación ReLU, cuya salida es un mapa de características, y una operación de agrupamiento máximo de 2×2 con una longitud de paso 2 para submuestreo. La red de decodificación está formada por cuatro bloques de supermuestreo con las siguientes capas: una capa de convolución transpuesta para supermuestreo del mapa de características, una concatenación del mapa de características que proviene del correspondiente bloque de la red de codificación (ubicado a la misma altura en el segmento descendiente de la U) y dos convoluciones de 3×3, seguidas de una ReLU. El bloque inferior que une las dos redes (ubicado en la base de la U) consiste en dos convoluciones con una función ReLU de activación. El bloque final de la red decodificadora aplica una convolución con un *kernel* de 1×1 para producir el mapa de segmentación final con el número de clases indicado (4).

La arquitectura R2U-Net está basada en la arquitectura Recurrent U-Net o RU-Net. La arquitectura RU-Net es una versión modificada de la arquitectura U-Net que usa capas convolucionales recurrentes hacia adelante (RCL) en lugar de capas convolucionales tradicionales. A través de RCL, una RU-Net permite una acumulación más efectiva de características entre las dos subredes. Una R2U-Net es, justamente, una variante de la RU-Net que usa unidades residuales junto con RCL para transmitir salidas a las capas posteriores de la red y evitar la degradación de la eficiencia al momento de mantener características (12). La arquitectura R2U-Net se ilustra en la figura 2.

La arquitectura Attention U-Net está también basada en la U-Net, pero incorpora varias compuertas de atención (AG) para filtrar las características propagadas por las conexiones de salto de la arquitectura original. Esto ayuda a progresivamente eliminar segmentaciones en regiones asociadas con el fondo de la imagen (13).

Finalmente, la última arquitectura estudiada fue la Attention R2U-Net. Esta arquitectura mezcla las unidades residuales con RCL de la

R2U-Net con las AG de la Attention U-Net (14). Tanto la Attention U-Net como la Attention R2U-Net se ilustran en las figuras 1 y 2, respectivamente.

En resumen, la arquitectura U-Net se basa en el cálculo de sucesivos filtros durante la etapa de submuestreo con el fin de reducir la imagen para resaltar a través de convoluciones, y a medida que se avanza en la etapa de aumento de muestreo, estas características se usan para ayudarle a la red a detectar los diferentes tipos de tejidos. Las variantes que se exploraron añaden etapas que permiten que la red intente aprender a través de diferentes mecanismos (AG, RCL y similares) en qué zonas de la imagen debería concentrarse para mejorar la calidad de la segmentación.

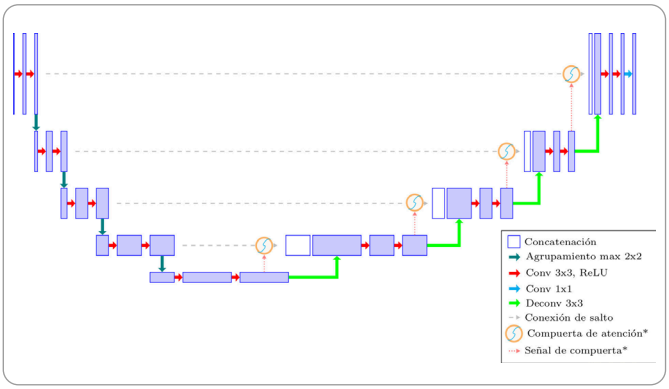

Figura 1. Arquitecturas U-Net y Attention U-Net. Los componentes marcados con (*) corresponden a la versión Attention de la arquitectura.

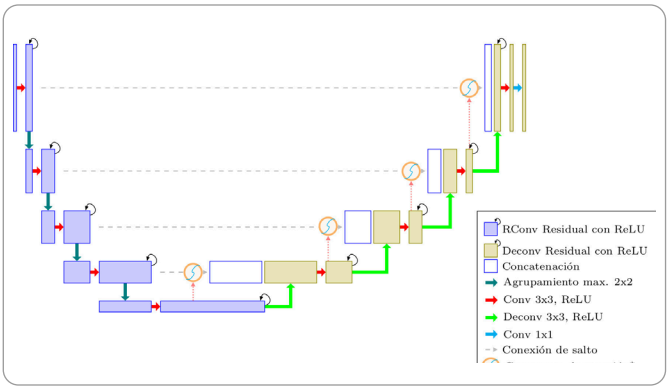

Figura 2. Arquitecturas R2U-Net y Attention R2U-Net. Los componentes marcados con (*) corresponden a la versión Attention de la arquitectura.

Medición del desempeño

Para medir el desempeño, se usó como métrica principal el índice de Dice (DSC):

$$DSC(A,B) = \frac{2|A| \cap |B|}{|A| + |B|}$$

Donde A está dado por la máscara de segmentación de la verdad terreno a través del patrón de oro, B es la máscara de segmentación generada para cada clase por cada una de las arquitecturas; A y B son el número de elementos positivos en cada máscara. Para deducir el índice de Dice total, se calculó el puntaje para cada clase y luego se obtuvo el promedio.

Durante el entrenamiento de las redes se mide la discrepancia entre las predicciones del modelo y los valores reales a través de una función de pérdida. El objetivo de cada iteración es que se minimice la pérdida calculada a través de la respectiva función para la segmentación. En este caso se usaron, en particular, dos funciones para calcular la función de pérdida total: la función de pérdida enfocada y la función de pérdida Dice.

La función de pérdida enfocada FL(P_c) está definida como:

$$FL(p_c) = -\alpha_t(1-p_c)^{\gamma} log(p_c)$$

$$p_c = \begin{cases} p \text{ si } y = c \\ 1-p \text{ de lo contrario} \end{cases}$$

Donde P_c es la probabilidad estimada para la clase c por el modelo, α es el factor de peso y γ es el factor de enfoque. La pérdida total para todas las clases, con C como el número total de clases, es calculada así:

$$Pérdida\ Total\ Enfocada = \sum_{c=1}^{C} -\alpha(1-p_c)^{\gamma} log(p_c)$$

Mientras tanto, la función de pérdida de Dice está basada en el índice de Dice y se define como:

$$Pérdida\ Dice(\hat{y}, y) = \sum_{c=1}^{C} 1 - DSC(\hat{y_c}, y_c)$$

Curva ROC

Adicionalmente, para medir el desempeño se usó la curva de característica operativa del receptor (ROC, por sus siglas en inglés) para cada clase de manera independiente. La curva ROC se construye a partir de la proporción de verdaderos positivos contra la proporción de falsos positivos. La curva ROC es, en últimas, una curva de probabilidad, y calcular el área bajo la curva (AUC, por sus siglas en inglés) permite entender qué tan bueno es el modelo prediciendo o segmentando una clase contra todas las demás. Entre más cercano el valor de AUC esté a 1 para una clase en particular, mejor será la arquitectura prediciendo dicha clase (100 % de sus predicciones son correctas), e inversamente, entre más cercano esté el valor del AUC a 0, peor será el modelo.

Conjunto de entrenamiento

Uno de los desafíos más importantes al momento de entrenar modelos de redes neuronales consiste en la capacidad de generar conjuntos de datos de entrenamiento suficientemente grandes para garantizar que la red sea capaz de generalizar el aprendizaje a otras imágenes sin provocar sobreajuste. La cantidad de segmentaciones manuales dependerá siempre de los expertos; sin embargo, en aras de facilitar e incrementar la cantidad de imágenes utilizadas, se propone una estrategia diferente complementaria al uso del patrón del oro.

Para aumentar el número de imágenes segmentadas, se usó el *software* CAAVAT (Computer Assisted Analysis of Visceral Adipose Tissue). CAVAAT es una herramienta que permite estimar la cantidad de tejido adiposo y subcutáneo en TAC. Esta herramienta realiza una segmentación a través de una umbralización sobre la imagen del cuerpo del paciente con un intervalo de –500, 100 UH, y luego aísla el tejido adiposo a través de una umbralización con intervalo de –150, 50 UH. Finalmente, un contorno activo inicial se adhiere al cuerpo y un segundo contorno se adhiere al perineo para ayudar a diferenciar VAT y SAT. CAAVAT segmenta la imagen en cuatro clases de tejido: VAT, SAT, otros tejidos y fondo (figura 3). Se usó CAAVAT para segmentar las 513 imágenes presentes en el conjunto de datos para medir la precisión de las arquitecturas.

Se validó la eficiencia de CAAVAT realizando una segmentación de las 41 imágenes del patrón de oro y comparándola contra los resultados de los expertos. CAAVAT obtuvo una eficiencia levemente inferior a la de los expertos en segmentación manual, como se muestra en la tabla 1, cuando se usó el índice de DICE para comparar la segmentación individual de cada experto con relación al consenso en el patrón de oro.

Tabla 1. Comparación de desempeño de expertos y CAAVAT contra el consenso en el patrón de oro

Segmentación	Puntaje DICE
Experto 1	0,852 +- 0,002
Experto 2	0,858 +- 0,002
Experto 3	0,866 +- 0,002
CAAVAT	0,834 +- 0,002

El objetivo de usar CAAVAT como experto adicional permite entrenar un modelo de Aprendizaje Profundo con mayor capacidad de generalización que no dependa únicamente de una estrategia de aumento de datos a partir de transformaciones sobre un conjunto reducido de datos. En este caso, CAAVAT permitió a entrenar y validar los modelos con 513 imágenes en total, en lugar de 41 que hacían parte del patrón de oro.

Aumento de datos

El aumento de datos es una técnica esencial para el entrenamiento de redes basadas en Aprendizaje Profundo, incluyendo la familia de arquitecturas U-Net estudiadas. Aumentar datos permite generalizar mejor y con más diversos ejemplos para incrementar la robustez de estas. Se hizo un aumento del 40 % de los datos de entrenamiento usando operaciones de redimensionamiento aleatorio, rotaciones aleatorias, recorte central, desplazamiento de intervalo y distorsión de color.

Evaluación y resultados

Para el entrenamiento de las arquitecturas, se dividió el conjunto de datos de la siguiente manera:

- Un conjunto de entrenamiento de 353 imágenes, usadas para entrenar a la red neuronal tras cada época de entrenamiento.
- Un conjunto de validación de 58 imágenes, usadas para validar la precisión de cada época de entrenamiento.
- Un conjunto de prueba de 102 imágenes, usadas para ajustar los hiperparámetros de la red.

Adicional a estos conjuntos, se emplearon las 41 imágenes del patrón de oro para determinar el desempeño final de la red.

Todos los modelos fueron entrenados durante 200 épocas con una tasa de aprendizaje de $2\times10^{-}4$ y con un tamaño del lote de entrenamiento de 4. Adicionalmente, se empleó el optimizador de gradiente estocástico descendiente con una tasa de decaimiento de $\beta1 = 0,5$ y $\beta2 = 0,999$.

El objetivo del entrenamiento es reducir la función de pérdida, que mide la diferencia entre la predicción $\hat{y}$ y el correspondiente mapa de etiquetas y para una entrada x. La predicción $\hat{y}$ es un tensor de tamaño R^{HxWxC} con altura H, ancho W y C clases.

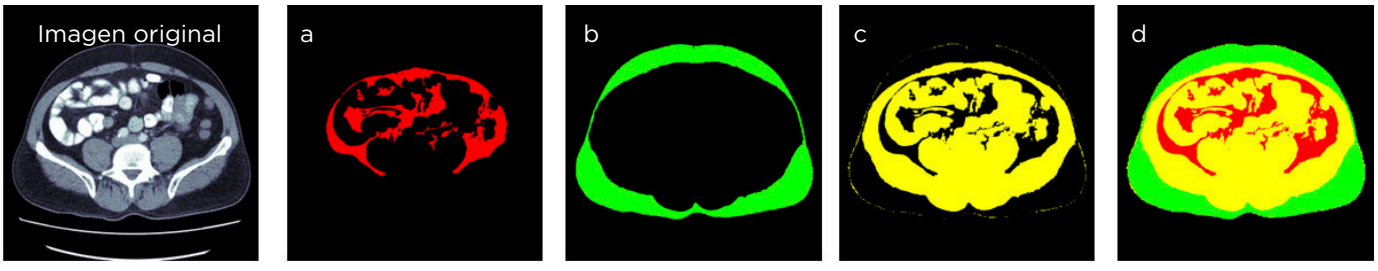

Figura 3. Resultados de la segmentación de tejidos usando CAAVAT: a) VAT, b) SAT, c) otros tejidos y d) unión de todas las segmentaciones.

Para cada pixel de la entrada x, se obtiene un vector de estimación de probabilidad de clase. Para comparar la salida $\hat{y}$ con el respectivo mapa de etiquetas, las clases de segmentación son codificadas usando One-hot para asimilar las dimensiones R^{HxWxC} de la predicción. Para la evaluación final con el conjunto de prueba, el mejor modelo es seleccionado basado en el desempeño obtenido en el conjunto de validación.

Las tablas 2 y 3 muestran los resultados de la segmentación con cada arquitectura para las cuatro clases, así como el promedio por clase. Los resultados muestran que la arquitectura U-Net obtuvo el mejor valor de índice de Dice para VAT/SAT y otros tejidos, seguida muy de cerca por la arquitectura Attention U-Net. Ambas tuvieron un desempeño similar sobre todas las clases. Para analizar en mayor detalle, se obtuvo el índice Dice para cada clase en tres escenarios de prueba.

Para la tarea de análisis de composición corporal, las arquitecturas U-Net y Attention U-Net tuvieron un mejor desempeño para segmentar ambos tipos de tejido adiposo. En todos los casos, la clase más difícil para segmentar fue el VAT debido a la alta variabilidad de ubicación y tamaño por cada paciente, así como la presencia de otros tejidos musculares y de contenido intestinal en las regiones aledañas o internas. Esta dificultad se hizo bastante notoria en las arquitecturas U-Net que emplean RCL con unidades residuales.

Con relación a otros trabajos, la tabla 4 muestra los puntajes de referencia de los estudios (15-18). Si bien los puntajes obtenidos están por debajo de las referencias, la mayor limitación actual consiste en el conjunto de imágenes de partida y, si bien los resultados para SAT se acercan, la segmentación de VAT tiene mucho menor puntaje que el obtenido por Dabiri et al (16). En términos del número de pacientes empleados y el puntaje de Dice para SAT, los resultados obtenidos se acercan en parámetros y resultados a los de Hemke et al. (17).

Tabla 2. Índices Dice obtenidos para la segmentación de VAT y SAT en el conjunto de datos para cada arquitectura con relación a la intersección

Arquitectura / clase	Fondo	Otros tejidos	VAT	SAT	Promedio
U-Net	0,99 +- 0,00	0,92 +- 0,02	0,84 +- 0,04	0,95 +- 0,01	0,93 +- 0,02
Attention U-Net	0,99 +- 0,00	0,91 +- 0,02	0,83 +- 0,05	0,94 +- 0,01	0,92 +- 0,02
R2U-Net	0,92 +- 0,02	0,72 +- 0,06	0,66 +- 0,11	0,55 +- 0,14	0,71 +- 0,08
Attention R2U-Net	0,885 +- 0,044	0,85 +- 0,04	0,01 +- 0,14	0,38 +- 0,14	0,53 +- 0,06

Tabla 3. Índices Dice obtenidos para la segmentación de VAT y SAT en el conjunto de datos para cada arquitectura con relación al consenso

Arquitectura / clase	Fondo	Otros tejidos	VAT	SAT	Promedio
U-Net	0,99 +- 0,00	0,92 +- 0,02	0,84 +- 0,05	0,95 +- 0,01	0,93 +- 0,02
R2U-Net	0,92 +- 0,02	0,72 +- 0,06	0,66 +- 0,11	0,56 +- 0,16	0,72 +- 0,08
Attention U-Net	0,99 +- 0,00	0,91 +- 0,02	0,83 +- 0,05	0,94 +- 0,01	0,92 +- 0,02
Attention R2U-Net	0,88 +- 0,04	0,85 +- 0,04	0,02 +- 0,01	0,38 +- 0,14	0,53 +- 0,06

Tabla 4. Comparación de resultados de puntaje Dice para la segmentación de VAT y SAT de cuatro estudios de referencia y el mejor resultado del trabajo actual

Autores	Número de pacientes	Puntaje Dice para VAT	Puntaje DICE para SAT
Weston et al. (15)	2.369	-	0,98
Dabiri et al. (16)	2.529	0,98	0,98
Hemke et al. (17)	200	-	0,95
Koitka et al. (18)	50	-	0,99
Estudio actual	513	0,84	0,95

La figura 4 muestra algunos resultados cualitativos del análisis de composición corporal en comparación con la segmentación con relación al consenso. Para el caso de las arquitecturas U-Net y Attention U-Net, las clases segmentadas tienen aspectos similares con relación a la verdad terrena para la intersección y el consenso. En el caso de la arquitectura R2U-Net, esta obtuvo resultados menos consistentes y cualitativamente tendió a equivocarse más en el abdomen superior, alterando la segmentación de VAT y SAT en dichas regiones. Finalmente, la arquitectura Attention R2U-Net tuvo un desempeño muy bajo, ignorando en muchos casos las fronteras de los tejidos y clasificando la mayoría de las regiones de la imagen como VAT.

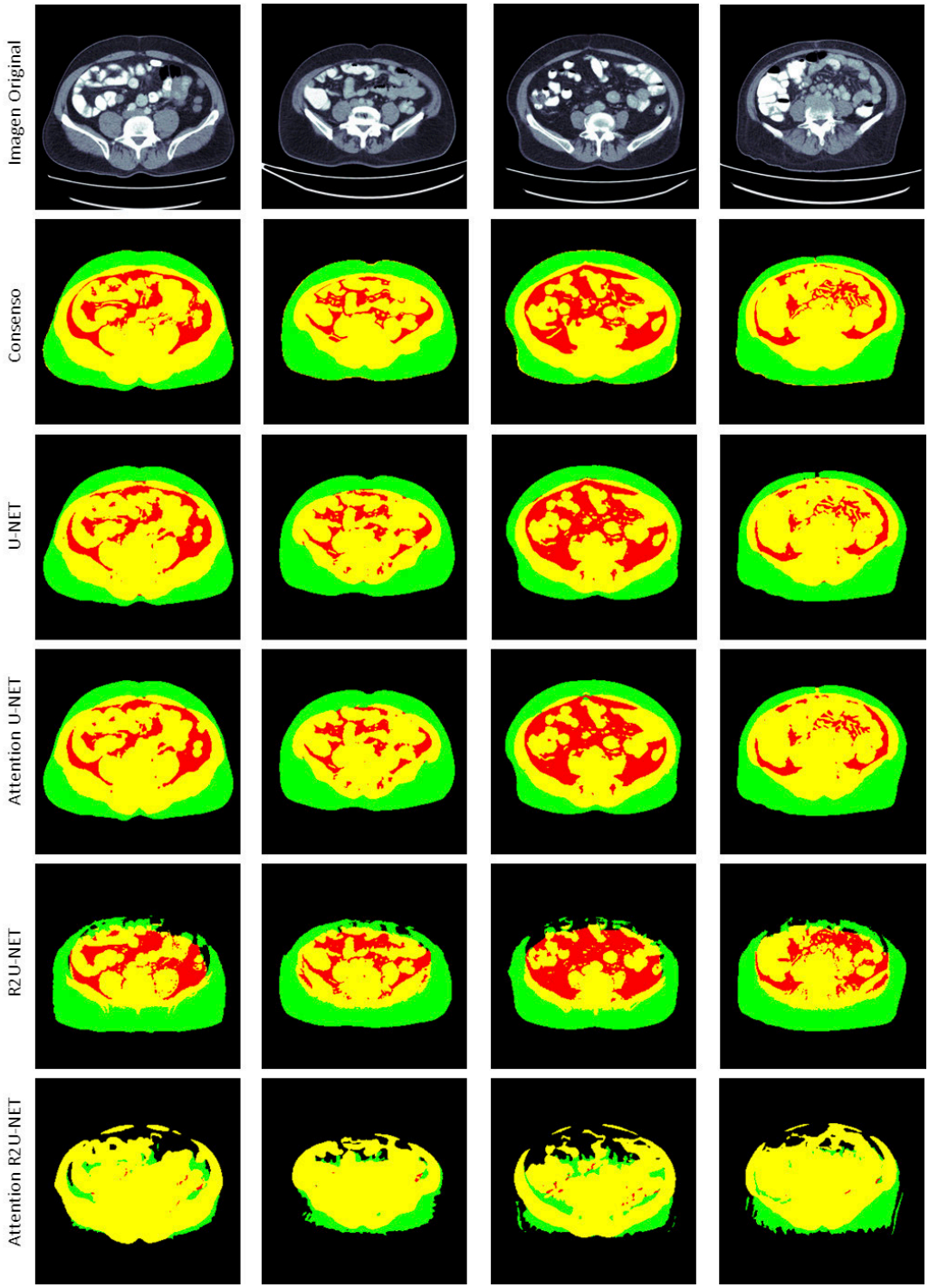

Figura 4. Resultados cualitativos de la segmentación en diferentes pacientes usando las distintas arquitecturas.

Para entender mejor el bajo desempeño de las arquitecturas R2U-Net y Attention R2U-Net, se hizo una validación con únicamente dos clases: otros tejidos y tejido adiposo; es decir, VAT y SAT conjuntamente. Se usó como punto de comparación la arquitectura U-Net con el mismo número de clases. En la tabla 5 se resumen los resultados para cada segmentación calculados con la media y la desviación estándar del índice de Dice para cada arquitectura. En ambos casos el índice Dice obtenido con R2U-Net y Attention R2U-Net tuvo un aumento considerable, en el caso de la arquitectura Attention R2U-Net, esta se acercó bastante al desempeño de U-Net, mientras que R2U-Net apenas si tuvo variación.

Tabla 5. Índices Dice obtenidos en el conjunto de datos para otros tejidos y tejido adiposo para cada arquitectura

Arquitectura / clase	Otros tejidos	Tejido adiposo
U-Net	0,91 +- 0,03	0,92 +- 0,15
R2U-Net	0,71 +- 0,07	0,62 +- 0,08
Attention R2U-Net	0,91 +- 0,03	0,90 +- 0,04

Curva ROC

Para calcular el AUC de cada una de las arquitecturas, se trazó la curva ROC por cada una de las clases contra las demás clases existentes, analizando el problema de la segmentación para cada tejido como si se tratara de múltiples clasificaciones binarias, usando el consenso como verdad terrena para calcular la proporción de verdaderos positivos y falsos positivos para cada arquitectura.

Las figuras 5 y 6 muestran los resultados de AUC para las arquitecturas U-Net y Attention U-Net. Ambas arquitecturas muestran un valor de AUC aproximadamente de 1,00 para las clases asociadas a otros tejidos y de 0,99 para las clases asociadas a SAT. Con relación a la clase asociada a VAT, la arquitectura U-Net tiene un valor levemente menor (0,97) que la arquitectura Attention U-Net (0,98). Estos valores indican que ambas arquitecturas son capaces de clasificar correctamente hasta un 97 % y 98 % de pixeles pertenecientes al tejido adiposo abdominal de cada paciente.

Por otro lado, de acuerdo con las figuras 7 y 8, las arquitecturas R2U-Net y Attention R2U-Net tienen valores de AUC más bajos. El valor de AUC asociado a otros tejidos es menor que en las arquitecturas U-Net (0,99 y 0,97). Los valores de AUC asociados a VAT son similares (0,96 y 0,98); sin embargo, el valor de SAT presenta problemas importantes: para la arquitectura R2U-Net es de cerca de 0,92 y para la arquitectura Attention R2U-Net es de 0,35. Si se examinan todos los valores tomando en cuenta el índice de Dice, pareciera indicar que estas arquitecturas son capaces de aislar al menos una parte del VAT, aunque no corresponda a la región total. En particular, la arquitectura Attention R2U-Net tiene problemas importantes con la detección de SAT, como confirma el análisis cualitativo de las segmentaciones.

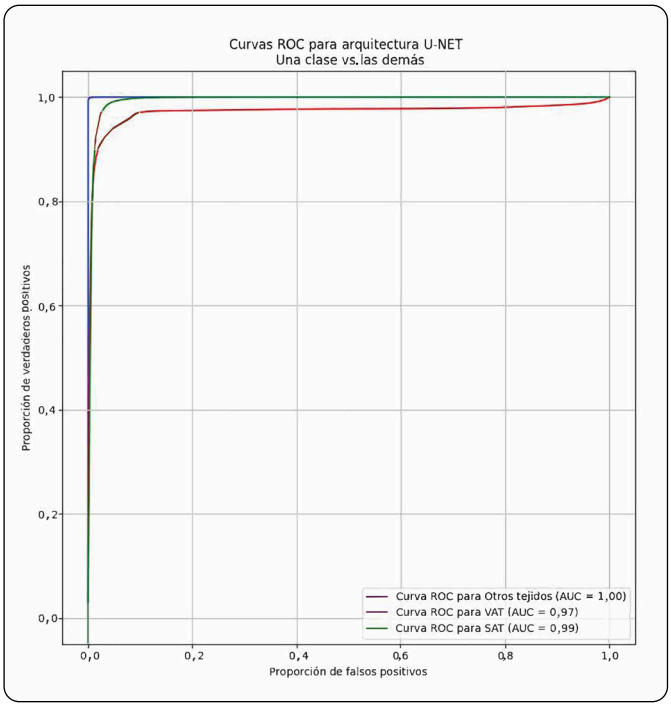

Figura 5. Curva ROC para la arquitectura U-Net para las clases Otros tejidos, VAT y SAT, junto con su respectivo de AUC.

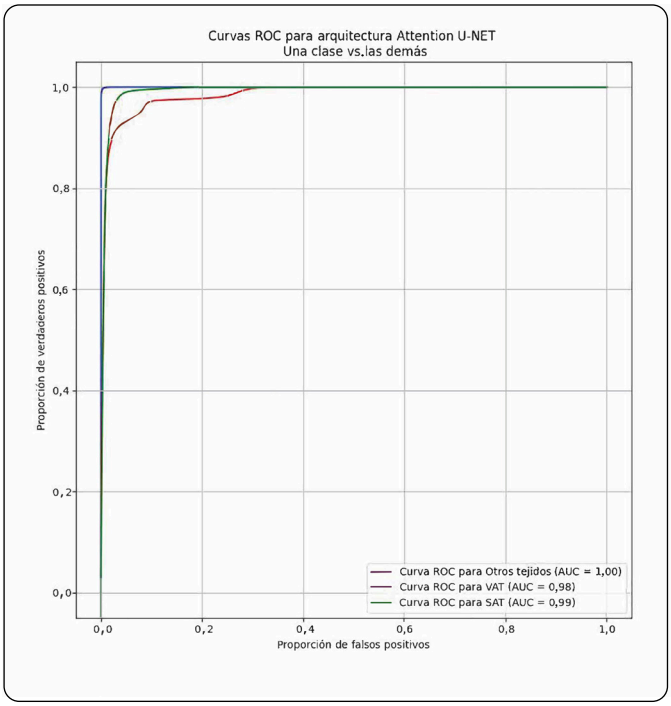

Figura 6. Curva ROC para la arquitectura Attention U-Net para las clases Otros tejidos, VAT y SAT, junto con su respectivo de AUC.

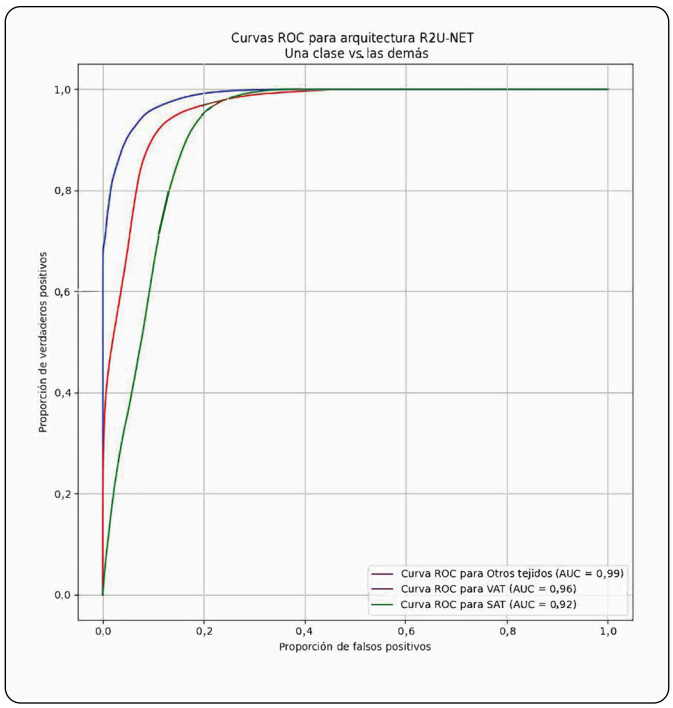

Figura 7. Curva ROC para la arquitectura R2U-Net para las clases Otros tejidos, VAT y SAT, junto con su respectivo de AUC.

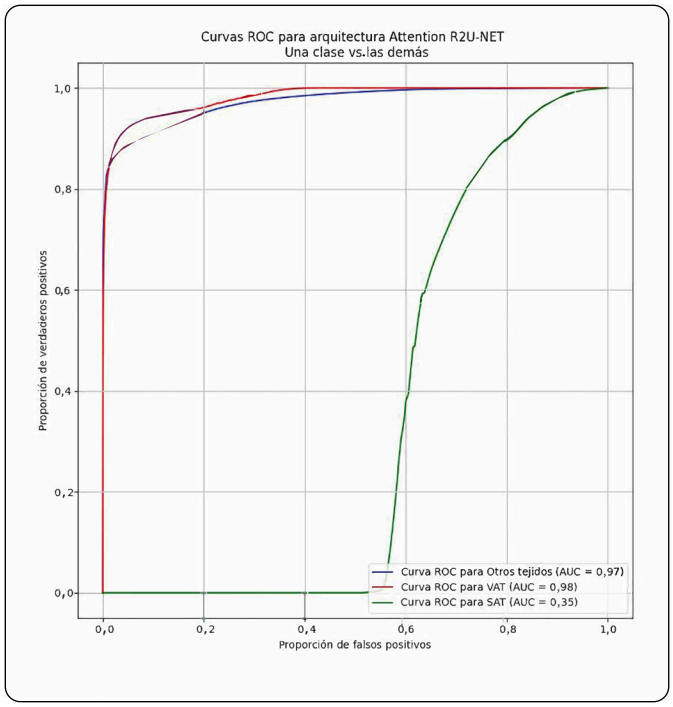

Figura 8. Curva ROC para la arquitectura Attention R2U-Net para las clases Otros tejidos, VAT y SAT, junto con su respectivo de AUC.

Discusión

En este artículo se describe cómo se entrenó, se probó y se estudió un subconjunto de arquitecturas de Aprendizaje Profundo basadas en U-Net con el propósito de realizar análisis de composición corporal. Estas arquitecturas funcionan bastante bien para la segmentación de las regiones asociadas a VAT y a SAT. En particular, la arquitectura U-Net y Attention U-Net obtuvieron el mejor desempeño para las clases individuales con los escenarios de prueba, demostrando su relevancia como uno de los métodos de segmentación semántica a la vanguardia del área de procesamiento de imágenes médicas. Entender y explorar en mayor detalle este par de arquitecturas podría resultar en aún mejores resultados para los análisis de composición corporal.

El bajo desempeño de las arquitecturas con unidades residuales y RCL sigue siendo sorprendente. Aunque las implementaciones originales aplicadas a la segmentación de lesiones mostraron el potencial de las arquitecturas R2U-Net y Attention R2U-Net, su eficacia es cuestionable para problemas con múltiples clases desbalanceadas. Sin embargo, se logró incrementar considerablemente la eficiencia de la Attention R2U-Net cuando se redujo el problema a una clasificación con dos clases más balanceada: tejido adiposo y otros tejidos, sin obtener una mejora sustancial en el desempeño de la arquitectura R2U-Net. Esto lleva a pensar que estas últimas arquitecturas pueden funcionar mejor en otros tipos de problemas, particularmente, segmentación binaria de tejidos con clases mejor balanceadas.

Los resultados aquí documentados muestran que las arquitecturas basadas en U-Net sin unidades residuales ni RCL pueden servir de base para la implementación de métodos automáticos de segmentación que sean capaces de aliviar la dependencia de cálculos exclusivamente antropométricos para analizar los tejidos de un paciente y detectar tempranamente sus condiciones de salud. Estas arquitecturas son invaluables y pueden apoyar enormemente el trabajo de los expertos y el desarrollo de herramientas completamente automáticas en el campo.

Con relación a la comparación con otros estudios similares (tabla 4), los resultados señalan la necesidad de ampliar el conjunto de entrenamiento y aumentar la mayor variedad en términos de tiempo de obtención e instrumentos, para lograr una mejor generalización de las arquitecturas. Si bien este conjunto de datos mitiga los sesgos asociados a las arquitecturas, son necesarios más datos de entrenamiento. Adicionalmente, el enfoque en segmentación de SAT además de VAT añade una capa de complejidad que solo otros estudios abordan directamente, o emplean directamente clases diferentes, concentrándose particularmente en la segmentación de VAT.

A futuro, los autores consideran importante continuar recopilando escanografías abdominales para enriquecer los conjuntos de entrenamiento, haciendo énfasis en obtener variedad en términos de la obtención: máquinas, operadores y población. Del mismo modo, prevén la continuación del uso de CAVAAT como herramienta de anotación para facilitar el trabajo de los expertos al momento de generar patrones de oro que permitan probar y validar los entrenamientos de las arquitecturas. Finalmente, la segmentación del músculo paravertebral es otro de los desafíos por abordar, que permitirá mejorar los resultados para discriminar los tejidos de interés para estas arquitecturas con mayor precisión.

Referencias

1. Afshin A, Vos T, Murray C, Fernandes J, Silverberg J, Bjertness E, et al. Health effects of overweight and obesity in 195 countries over 25 years. New Eng J Med. 2017;377: 13-27.
2. González MC, Pastore C, Orlandi S, Heymsfield S. Obesity paradox in cancer: New insights provided by body composition. Am J Clin Nut. 2014;99.
3. Seabolt LA, Welch EB, Silver HJ. Imaging methods for analyzing body composition in human obesity and cardiometabolic disease. Ann New York Acad Sci. 2015;1353:41-59.
4. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [Presentación]. En: Computación de imágenes médicas e intervención asistida por computadora (MICCAI). 2015. Disponible en: https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28
5. Lee S, Liu J, Yao J, Kanarek A, Summers R, Pickhardt P. Fully automated segmentation and quantification of visceral and subcutaneous fat at abdominal CT: Application to a longitudinal adult screening cohort. Br J Radiol. 2018;91:20170968.
6. Hui SCN, Zhang T, Shi L, Wang D, Ip CB, Chu WCW. Automated segmentation of abdominal subcutaneous adipose tissue and visceral adipose tissue in obese adolescent in MRI. Magn Resonan Imag. 2018;45:97-104.
7. Nemoto M, Yeernuer T, Masutani Y, Nomura Y, Hanaoka S, Miki S, et al. Development of automatic visceral fat volume calculation software for CT volume data. J Obesity. 2014;2014:495084.
8. Wang Z, Hounye A, Zhang J, Hou M, Qi M. Deep learning for abdominal adipose tissue segmentation with few labelled samples. International J Comp Assis Radiol Surg. 2021;17.
9. Micomyiza C, Zou B, Li Y. An effective automatic segmentation of abdominal adipose tissue using a convolution neural network. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2022;16:102589.
10. Li H, Luo H, Liu Y. Paraspinal muscle segmentation based on deep neural network. Sensors. 2019;19:2650.
11. Estrada S, Lu R, Conjeti S, Orozco-Ruiz X, Panos-Willuhn J, Breteler MMB, et al. FatSegNet: A fully automated deep learning pipeline for adipose tissue segmentation on abdominal dixon MRI. Magn Reson Med. 2020;83:1471-83.
12. Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK. Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation [internet]. 2018 [citado: 2023 feb. 15]. Disponible en: https://arxiv.org/abs/1802.06955
13. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: learning where to look for the pancreas [internet]. 2018 [citado: 2023 feb. 15]. Disponible en: https://arxiv.org/abs/1804.03999
14. Zhang L, Zuo Q, Chen S, Wang Z. R2AU-Net: Attention recurrent residual convolutional neural network for multimodal medical image segmentation. Sec Commun Net. 2021;2021:6625688.
15. Weston AD, Korfiatis P, Kline TL, Philbrick KA, Kostandy P, Sakinis T, et al. Automated abdominal segmentation of CT scans for body composition analysis using deep learning. Radiology. 2019;290:669-79.
16. Dabiri S, Popuri K, Cespedes E, Caan B, Baracos V, Beg MF. Deep learning method for localization and segmentation of abdominal CT. Comput Med Imag Graph. 2020;85:101776.
17. Hemke R, Buckless C, Tsao A, Wang B, Torriani M. Deep learning for automated segmentation of pelvic muscles, fat, and bone from CT studies for body composition assessment. Skel Radiol. 2019;49.
18. Koitka S, Kroll L, Malamutmann E, Oezcelik A, Nensa F. Fully automated body composition analysis in routine CT imaging using 3D semantic segmentation convolutional neural networks. Eur Radiol. 2020;31.

Correspondencia

Marcela Hernández Hoyos
Carrera 1 Este # 19A-40
Bogotá, Colombia
marc-her@uniandes.edu.co

Recibido para evaluación: 18 de marzo de 2023
Aceptado para publicación: 15 de mayo de 2023