

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

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

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

Key words (MeSH)

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

Palabras clave (DeCS)

Grasa abdominal
Grasa intraabdominal
Tomografía computarizada por rayos X

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 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.

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

[1]PhD, Department of Systems and Computer Engineering, Universidad de los Andes. Bogotá, Colombia.

[2]Department of Systems and Computer Engineering, Universidad de los Andes. Bogotá, Colombia.

[3]Radiologist, Department of Radiology, Hospital Universitario San Ignacio, Pontificia Universidad Javeriana. Bogotá, Colombia.

[4]Professor, Department of Epidemiology, School of Medicine, Pontificia Universidad Javeriana. Bogotá, Colombia.

[5]Radiologist, Hospital Manuel Uribe Angel, Clinica Sagrado Corazon. Bogotá, Colombia.

[6]Fellow Neuroradiology, Fundación Santa Fe de Bogotá. Bogota, Colombia.

[7]Radiologist, Clínica del Country, Clínica de la Colina, Clínica los Nogales. Bogota, Colombia.

[8]PhD, Full Professor, Department of Systems and Computer Engineering, Universidad de los Andes. Bogotá, Colombia.

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.

Introduction

The study of body composition is important in that it provides an indicator to detect various health conditions. By analyzing the amount and distribution of different types of tissues in the body, in particular adipose and muscle tissue, it is possible to determine the probability of developing certain medical conditions, such as cardiovascular disease, diabetes, chronic kidney disease, musculoskeletal disorders (1), cancer (2) or metabolic syndrome (3).

Despite advances in modern medicine, contemporary body composition analysis methods are heavily based on anthropometric methods, such as the calculation of body mass index (BMI), waist-hip ratio (WHR), or body fat percentage (BFP). For some of these indices, such as BMI, a relationship has been found between them and increased risk of death. Such an increase also particularly affects countries with low and middle socio-demographic levels (1).

One way to increase the accuracy of these methods of body composition analysis is to measure the proportion of adipose tissue distribution for each patient. The calculation of these indicators has recently been incorporated into clinical tools and through computer graphics methods and approaches based on tissue segmentation, the measurement of the proportion of these tissues has become more accurate.

More recently, new automatic and semiautomatic alternatives have been proposed for the quantification of adipose and muscle tissue, using in particular Deep Learning methods. U-Net architectures (4) have gained popularity for semantically segmenting torso images; many of these advances still require expert validation, as they are susceptible to biases and still require work to become fully automated reliable alternatives. The most important biases in this type of work come from the instruments used and the operators, for example in obtaining CT scans, the selection of non-representative datasets, the use of training data, and images with annotation problems.

To address such a gap, this study seeks to test in detail a subset of the U-Net Deep Learning architectures using as input a set of mixed images to train, compare, and evaluate the performance of each when segmenting abdominal adipose tissue and muscle tissue.

State of the art

Previous work has used thresholding and basic morphology operations for the detection of different tissue types. Over time, several studies have also incorporated body tissue quantification to obtain information about the health conditions a patient may have. More recent advances have refined segmentations using more advanced techniques such as dynamic region growing or active contour models (5): image-based algorithms to detect the narrow regions connecting subcutaneous and visceral tissues in the trunk (6) or algorithms based on morphological analysis from thresholding on CT to calculate the volumes of regions containing visceral fat (7).

A review of previous work reveals that methods based on Deep Learning are very close to the performance of human experts for segmentation. Some examples in the field have employed convolutional neural networks (CNN) to classify visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT) (8) or automated convolutional neural networks (A-CNN) for the same purpose (9).

Among the Deep Learning methods, architectures based on deep convolutional neural networks (DCNN) stand out. In particular, the U-Net family of architectures has gained prevalence. The original U-Net architecture was developed as a modified RNCP whose application was initially demonstrated for cell segmentation (4). Subsequently, other work has demonstrated the scope of this architecture. Extended versions of the architecture using residual and attention modules have been successful in segmenting paravertebral muscle in abdominal CT scans (10). Other adaptations using a so-called dense U-Net architecture have been used to detect fat, bone and muscle in MRI through a segmentation workflow, or "pipeline" (11).

Variations of the U-Net architecture model have also been developed to improve its efficiency. Of particular note are the RU-Net and R2U-Net architectures that were presented and tested on test datasets for segmenting retinal blood vessels, skin cancer, and lung lesions (12). Another variation is the Attention U-Net architecture, tested in segmenting multiple tissues in abdominal CT scans (13). Finally, there is the Attention R2U-Net architecture, which has been tested on the same datasets as RU-Net and R2U-Net (14). These architectures are explained in detail below.

U-Net has been used for body composition analysis in CT to segment muscle and SAT. In particular, this architecture has been used to segment VAT, muscle and abdominal organs at the L3 vertebra (15). The U-Net has also been used to segment, in addition to VAT, adipose intermuscular tissue at the same height (16) and mixed intrapelvic contents at the supraacetabular height (17). Finally, other variations of the architecture, such as multiresolution 3D U-Net, have been used to generate a volumetry of tissue body composition in CT from an annotated dataset (18).

All these works suggest that U-Net-based architectures can serve as a basis for the construction of a fully automated method for adipose and muscle tissue segmentation and quantification. In that sense, this work seeks to explore the set of Deep Learning architectures U-Net, R2UNet, Attention U-Net and Attention R2U-Net, to study their performance and determine which of them is best suited to solve the VAT and SAT segmentation problem in body composition analysis.

Methods and materials

Images

The dataset used consists of 513 abdominal CT images taken by the Hospital Universitario San Ignacio and the IDIME institute in Bogota, Colombia. These images were obtained from 2015 to 2017 and each one corresponds to a different patient. Of these images, 185

belong to male patients in an age range of 17 to 71 years, with a mean age of 48.1 years; 328 of the images belong to female patients in an age range of 17 to 87 years, with a mean age of 46.6 years. The images have a resolution of 0.482×0.482 to 0.953×0.953 mm^2, with matrix size of 512×512 and inter-slice spacing between 1 and 6 mm. A radiologist selected an axial slice for each patient manually, at the level of the fourth lumbar vertebra, based on the definition of excess visceral fat; the slice possessed an area greater than 100 cm2 of visceral fat measured at this level. The variety of the population and the use of images from two hospitals allow us to mitigate some biases inherent to the data sets.

Gold standard

To generate the gold standard, or ground truth, of abdominal adipose tissue, both subcutaneous and visceral, a set of 41 additional images was used: 25 from men and 16 from women. From the manual segmentations of three experts, a first validation set was created with the intersection of the areas identified by the experts in each image. Subsequently, a second validation set was generated from the review of the intersections by a fourth expert, and the resulting segmentation was defined by consensus of the four radiologists. In this text, the first set is referred to as intersection and the second as consensus, respectively. The conformation of the gold standard using this strategy also allows mitigating the cognitive biases present in the annotations of the training and validation datasets.

U-Net Deep Learning Architectures

In this work we evaluated the performance of four RNCP architectures: U-Net, R2U-Net, Attention U-Net and Attention R2U-Net. Typically RNCPs specialize in pattern identification in images and their basic architecture is divided into different processing layers: convolution, clustering, activation and fully connected.

The convolution layer is the most important in this type of architectures and is based on an operation commonly used in image processing called convolution, which consists of taking small parts of an image (using a window called a filter or pixel kernel) and combining them with the original image to highlight certain features. This is done repeatedly throughout the image to produce a new filtered image that shows how the features relate to different parts of the original image. In machine learning, convolutional neural networks use this process to perform tasks such as segmentation and object classification.

The grouping layer reduces image resolution (undersampling) while preserving features. The activation layer introduces nonlinearity into the architecture and is typically implemented using a Rectified Linear Unit (ReLU) that removes negative values from the filtered image to replace them with zero. Finally, the fully connected layer receives as input the outputs of the previous layers and determines the probability of each class or label in the image; in the particular case of abdominal scans, the probability that a pixel belongs to the VAT, SAT, other tissue types or to the background of the image.

The U-Net architecture is based on the RNCP architecture, but adds interpolation operators to increase image resolution (supersampling) and to extract high-level features, and does not use fully connected layers. It is called U-Net because of its U-shape, which consists of a path that contracts or descends (encoding network) and a path that expands or ascends (decoding network). The architecture seeks to learn different filters along both networks to extract image features that allow it to determine which pixels belong to one of the classes used for training (SAT, VAT, other tissues and background).

Figure 1 illustrates the layers present in the architecture, including the coding network (downstream segment) and the decoding network (upstream segment). The coding network of the U-Net architecture consists of four subsampling blocks composed of the following layers: two convolutions with 3×3 kernels with a ReLU activation function, whose output is a feature map, and a 2×2 max pooling operation with a step length of 2 for subsampling. The decoding network consists of four supersampling blocks with the following layers: a transposed convolution layer for supersampling the feature map, a feature map concatenation coming from the corresponding block of the coding network (located at the same height in the downstream segment of the U), and two 3×3 convolutions followed by a ReLU. The bottom block linking the two networks (located at the base of the U) consists of two convolutions with an activating ReLU function. The final block of the decoder network applies a convolution with a 1×1 kernel to produce the final segmentation map with the indicated number of classes (4).

The R2U-Net architecture is based on the Recurrent U-Net or RU-Net architecture. The RU-Net architecture is a modified version of the U-Net architecture that uses recurrent forward convolutional layers (RCL) instead of traditional convolutional layers. Through RCL, an RU-Net allows for a more effective accumulation of features between the two subnets. An R2U-Net is, rightly, a variant of the RU-Net that uses residual units along with RCL to transmit outputs to the downstream layers of the network and avoid efficiency degradation when maintaining features (12). The R2U-Net architecture is illustrated in Figure 2.

The Attention U-Net architecture is also based on the U-Net, but incorporates several attention gates (AG) to filter the features propagated by the hop connections of the original architecture. This helps to progressively eliminate segmentations in regions associated with the image background (13).

Finally, the last architecture studied was Attention R2UNet. This architecture mixes the residual units with RCL of the R2U-Net with the GAs of the Attention U-Net (14). Both Attention U-Net and Attention R2U-Net are illustrated in Figures 1 and 2, respectively.

In summary, the U-Net architecture relies on the computation of successive filters during the undersampling stage to reduce the image to highlight through convolutions, and as the upsampling stage progresses, these features are used to help the network detect different tissue types. The variants that were explored add stages that allow the network to try to learn through different mechanisms (GA, RCL and the like) which areas of the image it should concentrate on to improve the quality of the segmentation.

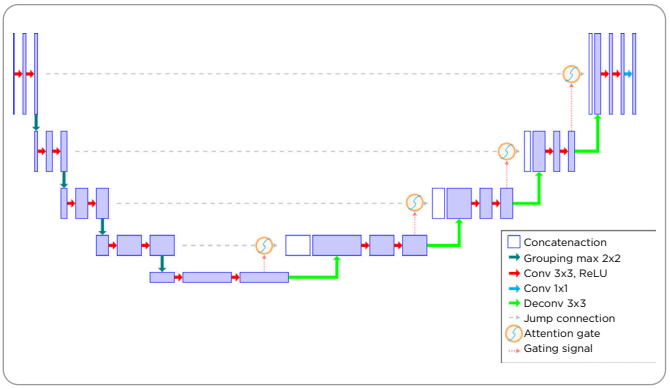


Figure 1. U-Net and Attention U-Net architectures. The components marked with (*) correspond to the Attention version of the architecture.

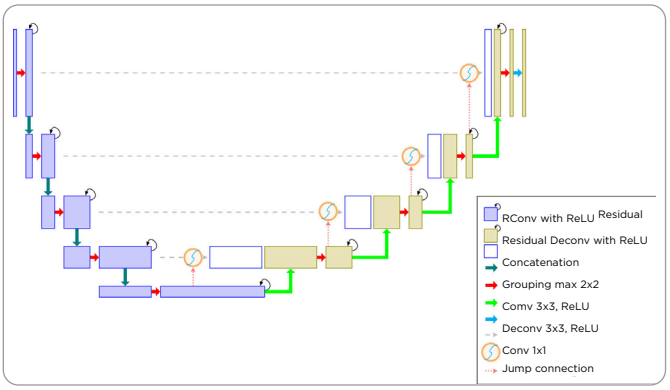


Figure 2. R2U-Net and Attention R2U-Net architectures. Components marked with (*) correspond to the Attention version of the architecture.

Performance measurement

To measure performance, the main metric used was the Dice index (DSC):

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

Where A is given by the ground truth segmentation mask through the gold standard, B is the segmentation mask generated for each class by each of the architectures; A and B are the number of positive elements in each mask. To derive the total Dice index, the score for each class was calculated and then the average was obtained.

During the training of the networks, the discrepancy between the model predictions and the real values is measured through a loss function. The objective of each iteration is to minimize the loss calculated through the respective function for segmentation. In this case, in particular, two functions were used to calculate the total loss function: the focused loss function and the Dice loss function.

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

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

Where P_c is the probability estimated for class c by the model, α is the weight factor and γ is the focus factor. The total loss for all classes, with C as the total number of classes, is calculated as follows:

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

Meanwhile, the Dice loss function is based on the Dice index and is defined as:

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

ROC Curve

Additionally, the receiver operating characteristic (ROC) curve for each class was used independently to measure performance. The ROC curve is constructed from the proportion of true positives versus the proportion of false positives. The ROC curve is ultimately a probability curve, and calculating the area under the curve (AUC) provides insight into how good the model is at predicting or segmenting one class against all others. The closer the AUC value is to 1 for a particular class, the better the architecture is at predicting that class (100% of its predictions are correct), and conversely, the closer the AUC value is to 0, the worse the model is.

Training set

One of the most important challenges when training neural network models is the ability to generate sufficiently large training data sets to ensure that the network is able to generalize learning to other images without overfitting. The number of manual segmentations will always depend on the experts; however, to facilitate and increase the number of images used, a different strategy complementary to the use of the gold standard is proposed.

To increase the number of segmented images, CAAVAT (Computer Assisted Analysis of Visceral Adipose Tissue) *software* was used. CAVAAT is a tool that allows estimating the amount of adipose and subcutaneous tissue in CT. This tool performs a segmentation through a thresholding on the patient's body image with an interval of -500, 100 HU, and then isolates the adipose tissue through a thresholding with an interval of -150, 50 HU. Finally, an initial active contour is attached to the body and a second contour is attached to the perineum to help differentiate VAT and SAT. CAAVAT segments the image into four tissue classes: VAT, SAT, other tissues and background (Figure 3). CAAVAT was used to segment the 513 images present in the dataset to measure the accuracy of the architectures.

The efficiency of CAAVAT was validated by performing a segmentation of the 41 gold standard images and comparing it against expert results. CAAVAT obtained a slightly lower efficiency than the experts in manual segmentation, as shown in Table 1, when the DICE index was used to compare the individual segmentation of each expert relative to the consensus on the gold standard.

Table 1. Comparison of expert and CAAVAT performance against the gold standard consensus

Segmentation	Score DICE
Expert 1	0,852 +- 0,002
Expert 2	0,858 +- 0,002
Expert 3	0,866 +- 0,002
CAAVAT	0,834 +- 0,002

The objective of using CAAVAT as an additional expert allows training a Deep Learning model with greater generalization capability that does not rely solely on a data augmentation strategy from transformations on a reduced dataset. In this case, CAAVAT allowed to train and validate the models with 513 images in total, instead of 41 that were part of the gold standard.

Data augmentation

Data augmentation is an essential technique for training Deep Learning based networks, including the family of U-Net architectures studied. Data augmentation allows to generalize better and with more diverse examples to increase their robustness. A 40 % augmentation of the training data was performed using random resizing, random rotations, center clipping, interval shifting and color distortion operations.

Evaluation and results

For the training of the architectures, the data set was divided as follows:

- A training set of 353 images, used to train the neural network after each training epoch.
- A validation set of 58 images, used to validate the accuracy of each training epoch.
- A test set of 102 images, used to adjust the hyperparameters of the network.

In addition to these ensembles, the 41 gold standard images were used to determine the final network performance.

All models were trained for 200 epochs with a learning rate of 2×10-4 and with a training batch size of 4. Additionally, the downward stochastic gradient optimizer with a decay rate of $\beta 1 = 0.5$ and $\beta 2 = 0.999$ was employed.

The goal of the training is to reduce the loss function, which measures the difference between the prediction $\hat{y}$ and the corresponding label map y for an input x. The prediction $\hat{y}$ is a tensor of size R^{HxWxC} with height H, width W and C classes.

For each pixel of the input x, a class probability estimate vector is obtained. To compare the output $\hat{y}$ with the respective label map, the segmentation classes are coded using One-hot to assimilate the R^{HxWxC} dimensions of the prediction. For the final evaluation with the test set, the best model is selected based on the performance obtained in the validation set.

Tables 2 and 3 show the segmentation results with each architecture for the four classes, as well as the average per class. The results show that the U-Net architecture obtained the best Dice index value for VAT/SAT and other fabrics, followed closely by the Attention U-Net architecture. Both had similar performance over all classes. For further analysis, the Dice index was obtained for each class in three test scenarios.

For the body composition analysis task, the U-Net and Attention U-Net architectures performed best for segmenting both types of adipose tissue. In all cases, the most difficult class to segment was VAT due to the high variability of location and size per patient, as well as the presence of other muscle tissues and intestinal contents in the surrounding or internal regions. This difficulty was quite noticeable in U-Net architectures using RCL with residual units.

In relation to other works, Table 4 shows the reference scores of the studies (15-18). Although the scores obtained are below the references, the major current limitation consists of the starting image set and, although the results for SAT are close, VAT segmentation scores much lower than that obtained by Dabiri et al (16). In terms of the number of patients used and Dice's score for SAT, the results obtained are close in parameters and results to those of Hemke et al (17).

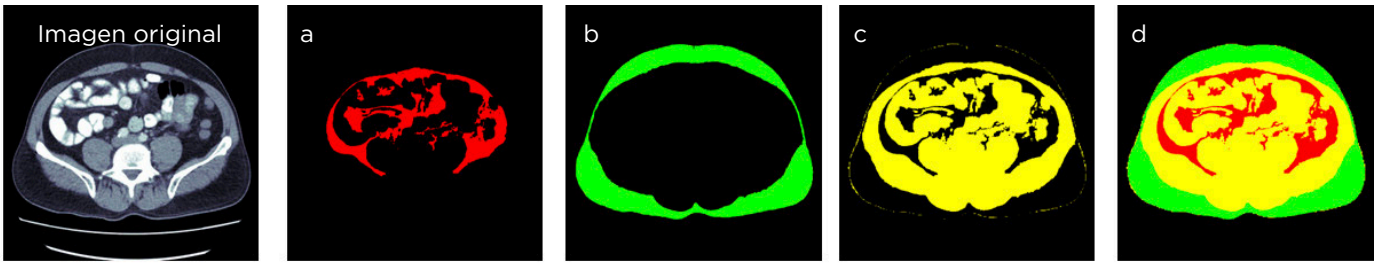


Figure 3. Tissue segmentation results using CAAVAT: a) VAT, b) SAT, c) other tissues and d) union of all segmentations.

Table 2. Dice indices obtained for VAT and SAT segmentation in the data set for each architecture in relation to the intersection

Architecture / class	Background	Other fabrics	VAT	SAT	Average
U-Net	0,99 +- 0,00	0,92 +- 0,02	0,84 +- 0,04	0,95 +- 0,01	0,93 +- 0,02
R2U-Net	0,99 +- 0,00	0,91 +- 0,02	0,83 +- 0,05	0,94 +- 0,01	0,92 +- 0,02
Attention U-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

Table 3. Dice indices obtained for VAT and SAT segmentation in the dataset for each architecture in relation to the consensus

Architecture / class	Background	Other fabrics	VAT	SAT	Average
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

Table 4. Comparison of Dice score results for VAT and SAT segmentation of four reference studies and the best result of the current work

Authors	Number of patients	DICE score for VAT	DICE score for 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
Current study	513	0.84	0.95

Figure 4 shows some qualitative results of the body composition analysis in comparison with segmentation in relation to consensus. For the case of the U-Net and Attention U-Net architectures, the segmented classes have similar aspects in relation to ground truth for intersection and consensus. In the case of the R2U-Net architecture, it obtained less consistent results and qualitatively tended to err more in the upper abdomen, altering the segmentation of VAT and SAT in those regions. Finally, the Attention R2U-Net architecture performed very poorly, in many cases ignoring tissue boundaries and classifying most image regions as VAT.

To better understand the poor performance of the R2UNet and Attention R2U-Net architectures, a validation was performed with only two classes: other tissues and adipose tissue; i.e., VAT and SAT together. The U-Net architecture with the same number of classes was used as a point of comparison. Table 5 summarizes the results for each segmentation calculated with the mean and standard deviation of the Dice index for each architecture. In both cases the Dice index obtained with R2U-Net and Attention R2U-Net had a considerable increase, in the case of the Attention R2U-Net architecture, it was quite close to the performance of U-Net, while R2U-Net had hardly any variation.

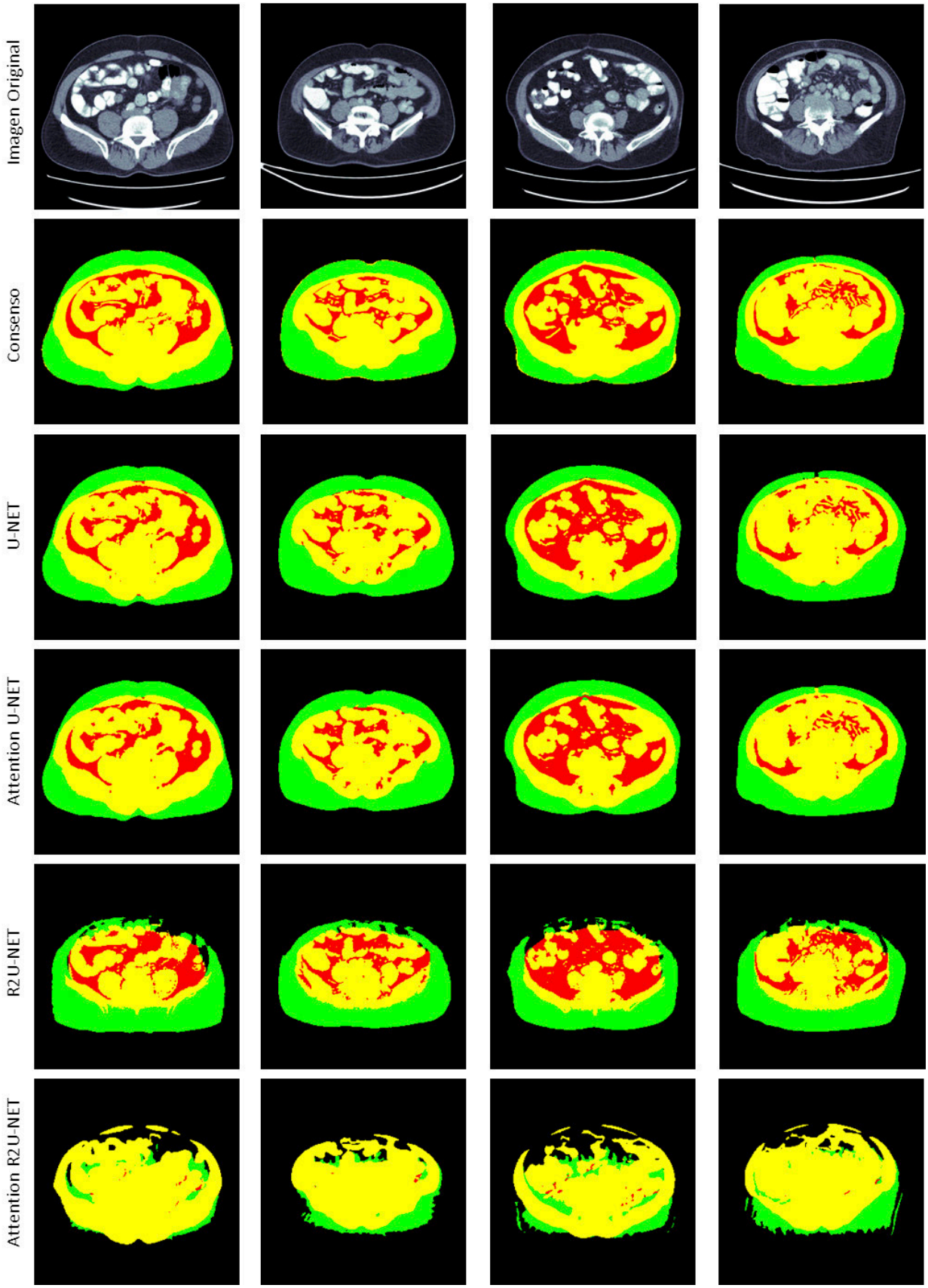


Figure 4. Qualitative segmentation results in different patients using the different architectures.

Table 5. Dice indices obtained in the dataset for other tissues and adipose tissue for each architecture

Architecture / class	Other tissues	Adipose tissue
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

ROC curve

To calculate the AUC for each of the architectures, the ROC curve for each of the classes was plotted against the other existing classes, analyzing the segmentation problem for each tissue as if it were multiple binary classifications, using consensus as ground truth to calculate the ratio of true positives to false positives for each architecture.

Figures 5 and 6 show the AUC results for the U-Net and Attention U-Net architectures. Both architectures show an AUC value of approximately 1.00 for the classes associated with other tissues and 0.99 for the classes associated with SAT. Regarding the class associated with VAT, the U-Net architecture has a slightly lower value (0.97) than the Attention U-Net architecture (0.98). These values indicate that both architectures are able to correctly classify up to 97% and 98% of pixels belonging to the abdominal adipose tissue of each patient.

On the other hand, according to Figures 7 and 8, the R2U-Net and Attention R2U-Net architectures have lower AUC values. The AUC value associated with other tissues is lower than in the U-Net architectures (0.99 and 0.97). The AUC values associated with VAT are similar (0.96 and 0.98); however, the SAT value presents significant problems: for the R2U-Net architecture it is about 0.92 and for the Attention R2U-Net architecture it is 0.35. If all values are examined taking into account the Dice index, it seems to indicate that these architectures are able to isolate at least a part of the VAT, even if it does not correspond to the total region. In particular, the Attention R2U-Net architecture has significant problems with SAT detection, as confirmed by the qualitative analysis of the segmentations.

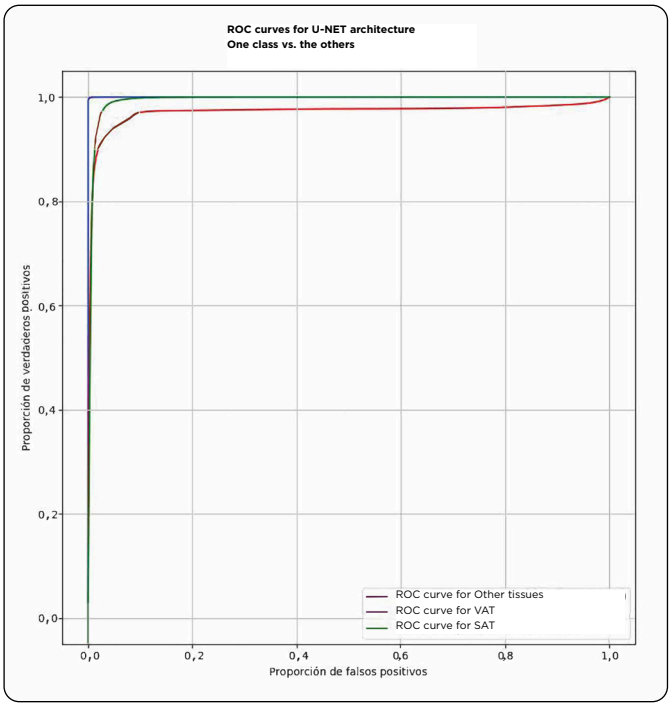


Figure 5. ROC curve for the U-Net architecture for the classes Other Tissues, VAT and SAT, together with their respective AUC.

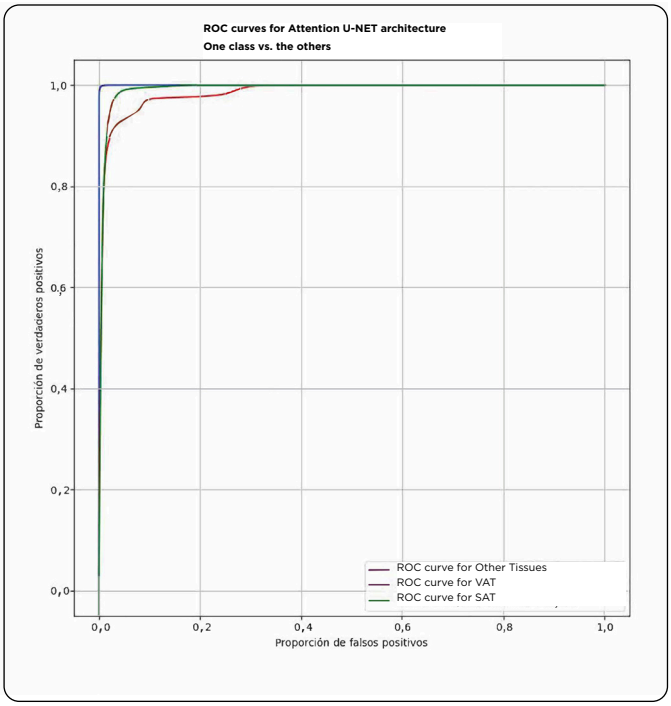


Figure 6. ROC curve for the Attention U-Net architecture for the classes Other Tissues, VAT and SAT, together with their respective AUC.

Discussion

This paper describes how a subset of U-Net based Deep Learning architectures were trained, tested and studied for the purpose of performing body composition analysis. These architectures perform quite well for the segmentation of regions associated with VAT and SAT. In particular, the U-Net and Attention U-Net architecture obtained the best performance for individual classes with the test scenarios, demonstrating their relevance as one of the semantic segmentation methods at the forefront of the medical image processing area. Further understanding and exploration of this pair of architectures could result in even better results for body composition analysis.

The poor performance of architectures with residual and RCL units remains surprising. Although the original implementations applied to lesion segmentation showed the potential of the R2U-Net and At-

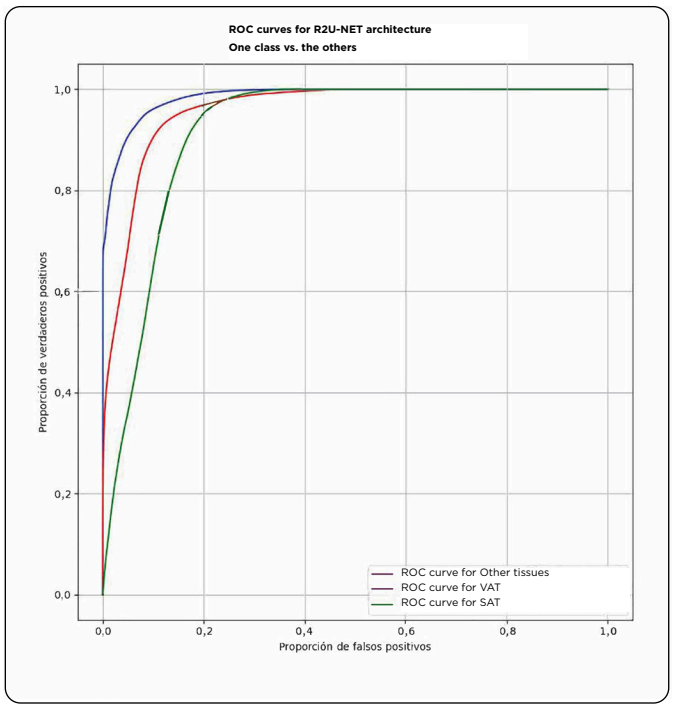


Figure 7. ROC curve for the R2U-Net architecture for the classes Other Tissues, VAT and SAT, together with their respective AUC.

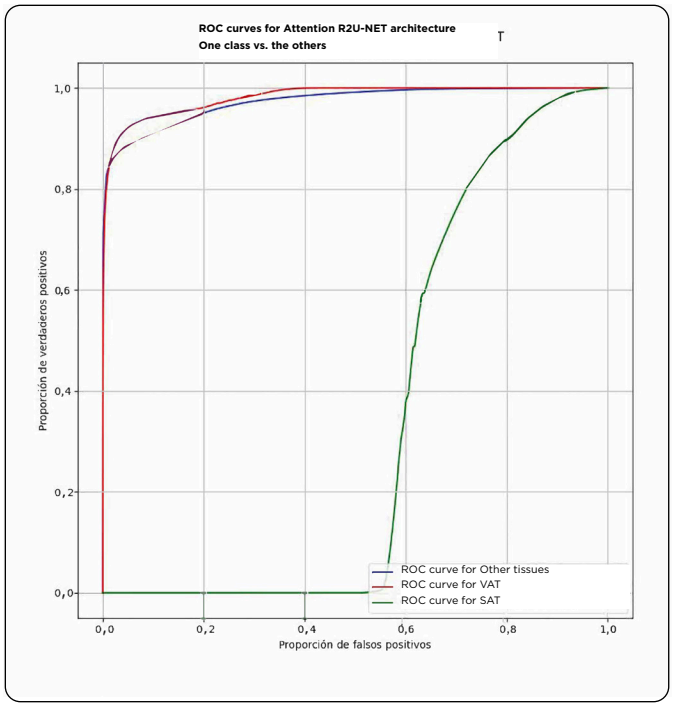


Figure 8. ROC curve for the Attention R2U-Net architecture for the classes Other Tissues, VAT and SAT, together with their respective AUC.

tention R2U-Net architectures, their effectiveness is questionable for problems with multiple unbalanced classes. However, a considerable increase in the efficiency of Attention R2U-Net was achieved when the problem was reduced to a classification with two more balanced classes: adipose tissue and other tissues, without obtaining a substantial improvement in the performance of the R2U-Net architecture. This leads one to believe that the latter architectures may perform better on other types of problems, particularly, binary tissue segmentation with better balanced classes.

The results documented here show that U-Net-based architectures without residual units and RCLs can serve as a basis for the implementation of automatic segmentation methods that are able to alleviate the dependence on purely anthropometric calculations to analyze a patient's tissues and detect their health conditions early. These architectures are invaluable and can greatly support the work of experts and the development of fully automatic tools in the field.

Regarding the comparison with other similar studies (Table 4), the results point to the need to expand the training set and increase the greater variety in terms of elicitation time and instruments, in order to achieve better generalization of the architectures. While this data set mitigates the biases associated with the architectures, more training data are needed. Additionally, the focus on SAT segmentation in addition to VAT adds a layer of complexity that only other studies directly address, or directly employ different classes, focusing particularly on VAT segmentation.

Going forward, the authors consider it important to continue collecting abdominal scans to enrich the training sets, with an emphasis on obtaining variety in terms of elicitation: machines, operators, and population. Similarly, they foresee the continued use of CAVAAT as an annotation tool to facilitate the work of experts when generating gold standards to test and validate training architectures. Finally, the segmentation of paravertebral muscle is another challenge to be addressed, which will allow improving the results to discriminate the tissues of interest for these architectures with greater precision.

References

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.

Correspondence

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

Received for evaluation: March 18, 2023
Accepted for publication: May 15, 2023