2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

3D Neural Networks for Visceral and Subcutaneous Adipose Tissue Segmentation using Volumetric Multi-Contrast MRI

Sevgi Gokce Kafali[1,⋆], Shu-Fu Shih[1], Xinzhou Li[1], Shilpy Chowdhury[2], Spencer Loong[3], Samuel Barnes[2], Zhaoping Li[4], Holden H. Wu [1,†]

Abstract— **Individuals with obesity have larger amounts of visceral (VAT) and subcutaneous adipose tissue (SAT) in their body, increasing the risk for cardiometabolic diseases. The reference standard to quantify SAT and VAT uses manual annotations of magnetic resonance images (MRI), which requires expert knowledge and is time-consuming. Although there have been studies investigating deep learning-based methods for automated SAT and VAT segmentation, the performance for VAT remains suboptimal (Dice scores of 0.43 to 0.89). Previous work had key limitations of not fully considering the multi-contrast information from MRI and the 3D anatomical context, which are critical for addressing the complex spatially varying structure of VAT. An additional challenge is the imbalance between the number and distribution of pixels representing SAT/VAT. This work proposes a network based on 3D U-Net that utilizes the full field-of-view volumetric T_1-weighted, water, and fat images from dual-echo Dixon MRI as the multi-channel input to automatically segment SAT and VAT in adults with overweight/obesity. In addition, this work extends the 3D U-Net to a new Attention-based Competitive Dense 3D U-Net (ACD 3D U-Net) trained with a class frequency-balancing Dice loss (FBDL). In an initial testing dataset, the proposed 3D U-Net and ACD 3D U-Net with FBDL achieved 3D Dice scores of (mean ± standard deviation) 0.99±0.01 and 0.99±0.01 for SAT, and 0.95±0.04 and 0.96±0.04 for VAT, respectively, compared to manual annotations. The proposed 3D networks had rapid inference time (<60 ms/slice) and can enable automated segmentation of SAT and VAT.**

Clinical relevance— **This work developed 3D neural networks to automatically, accurately, and rapidly segment visceral and subcutaneous adipose tissue on MRI, which can help to characterize the risk for cardiometabolic diseases such as diabetes, elevated glucose levels, and hypertension.**

I. INTRODUCTION

As increased adipose tissue (AT) is highly correlated with escalating incidences of obesity [1] and cardiometabolic diseases [2], the need for a rapid and accurate technique to quantify body composition is clear. Magnetic resonance imaging (MRI) can quantify body composition, notably visceral (VAT) and subcutaneous adipose tissue (SAT), non-invasively and without ionizing radiation. VAT and SAT are potential imaging biomarkers for detecting future risks for cardiometabolic diseases [2], [3]. In particular, VAT plays a key role in metabolic activity via secretion of inflammatory markers [4]. Currently, the reference standard to quantify SAT and VAT is manual annotations [3], [5], which require expert knowledge, may have inter-observer variability, is time-consuming (especially for VAT), and challenging for large-scale and longitudinal studies.

Machine learning approaches have been proposed to segment SAT and VAT [6] [7] [8]. Recently, deep learning-based methods have also been proposed, using architectures such as 2-dimensional (2D) U-Net [9], [10] , 2D Dense U-Net [11], 2.5D competitive Dense U-Net (FatSegNet) [12], 3D V-Net [10], and 3D Densely Connected (DC) Net [13]. These deep learning studies reported excellent performance for SAT segmentation (Dice scores of 0.97 to 0.99). However, the performance for VAT remains suboptimal (Dice scores of 0.43 to 0.89) due to the complex spatially varying and disconnected nature of VAT. These previous studies had key limitations of: (1) using only T_1-weighted (T1W) images, or only fat and water images, but not combining them to fully exploit the multi-contrast information from MRI, and/or (2) using only 2D slices or cropped 3D image patches, which do not capture the 3D anatomical context and global associations across the full imaging volume. An additional challenge is the imbalance between the number and distribution of pixels representing SAT/VAT.

In this work, our first objective is to develop a network based on 3D U-Net that combines multi-contrast images (T1W, water, and fat images) as the input and exploits global associations across the entire field-of-view (FOV) for volumetric dual-echo Dixon MRI data. Second, we extend 3D U-Net and propose a new Attention-based Competitive Dense 3D U-Net trained with a class frequency-balancing, boundary-emphasizing Dice loss to achieve rapid and accurate automated segmentation of SAT and VAT.

II. METHODS

A. Datasets

The Habitual Diet and Avocado Trial (HAT) is an ongoing longitudinal multisite clinical trial involving four universities. All participants underwent an abdominal 3T MRI exam, without injection of contrast agents. An axial 3D dual-echo Dixon sequence was acquired in a single breath-hold to measure SAT and VAT. T1W images at the out-of-phase and in-phase echoes (TE_{OP} / TE_{IP} = 1.23 / 2.46 ms) from Dixon MRI were used to calculate fat and water images [14].

[1] SG Kafali, S Shih, X Li and HH Wu are with the Departments of Radiological Sciences and Bioengineering, University of California, Los Angeles, CA, USA
[2] S Chowdhury and S Barnes are with the Department of Radiology, Loma Linda University Medical Center, CA, USA
[3] S Loong is with the Department of Psychology, Loma Linda University School of Behavioral Health, CA, USA
[4] Z Li is with the Department of Medicine, University of California, Los Angeles, CA, USA
[†] For the Habitual Diet and Avocado Trial (HAT) investigators
[⋆] Correspondence to: skafali@mednet.ucla.edu

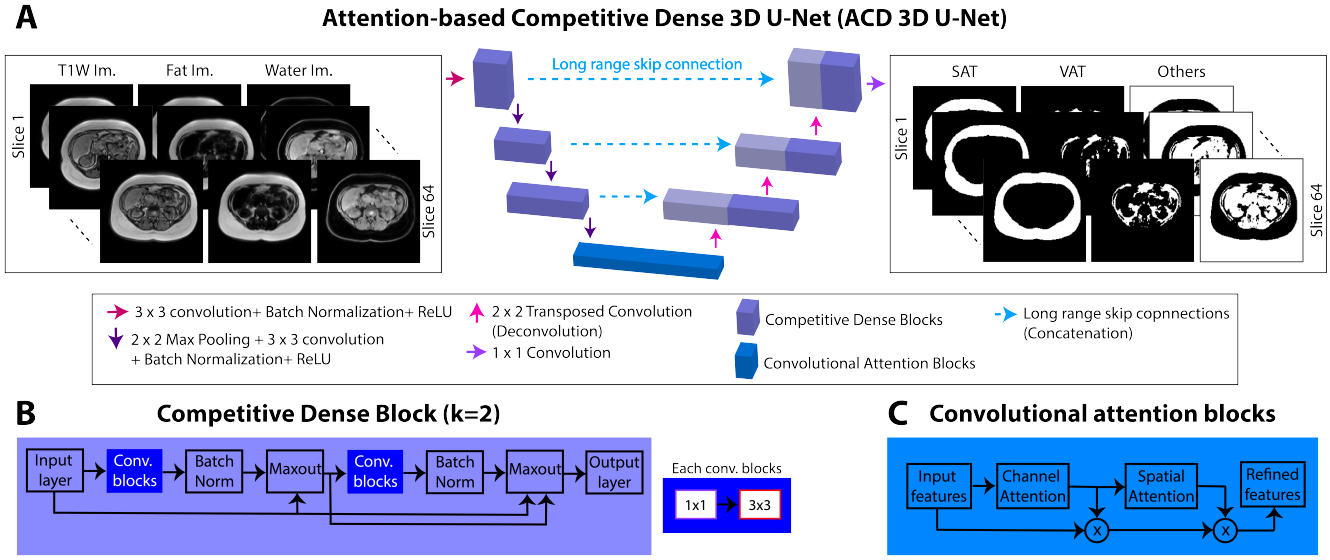

Fig. 1: The proposed Attention-based Competitive Dense 3D U-Net (ACD 3D U-Net).

Slice coverage spanned from 4 cm above the liver dome to 7 cm below the top of the iliac crest. Additional parameters were: matrix size of 192×192, 400-mm in-plane FOV, 96 slices, 3.5-mm slice thickness, TR = 5 ms, and flip angle = 9^o. Semi-automated manual annotation was performed by trained researchers in 51 slices of the Dixon fat images using SliceOmatic (TomoVision, Montreal, Canada) for segmentation with a watershed algorithm and ImageJ (National Institute of Health) was used for thresholding using intermodes. The number of slices was interpolated to 64 to be compatible with the filter sizes in the 3D networks. The same interpolated slices were used in 2D networks to facilitate comparisons.

	Mean±SD BMI (kg/m^2)	Sex
Overall (n=164)	31.89±4.45	68F, 19M
Train./Val. (n=146)	31.86±4.56	60F, 18M
Testing (n=18)	32.07±3.71	8F, 1M

TABLE I: Dataset characteristics. The data were stratified to yield similar distributions of body mass index (BMI) for the training/validation and testing datasets. For subjects with 2 exams, both exams were assigned to either training/validation or testing. n indicates number of exams. See text for details.

This work included the first 87 participants in the HAT study with completed annotations. Of these 87 participants, 77 had two MRI exams performed six months apart. The total number of MRI exams was 164. The dataset was split into separate training (90%; n=146 exams [68 subjects with 2 exams, 10 subjects with 1 exam]) and testing (10%; n=18 exams from 9 subjects) sets, with similar body mass index (BMI) distributions for the subjects (**Table I**). Eighteen exams were randomly selected from the training set and set aside as a validation set for hyperparameter tuning. Once the hyperparameters were selected using a grid search, the training and validation sets were combined for final training. The two exams for each subject often had notable differences in SAT/VAT and effectively augmented the dataset, thus we did not perform any additional augmentation.

B. Neural Networks

To take advantage of multi-contrast MRI information, we combined the T1W TE_{OP}, water, and fat images from dual-echo Dixon MRI as the multi-channel input and then generated output segmentation masks for SAT, VAT, and "other" pixels (**Figure 1A**).

We first implemented 2D U-Net with 4 encoding, 4 decoding, and 1 bottom layer(s) as the baseline network [9]–[12]. Next, we developed a network based on 3D U-Net [15] with multi-contrast full-FOV volumetric MRI as the input. Our implementation of 3D U-Net used 3 encoding, 3 decoding, and 1 bottom layer(s) [15]. The 2D U-Net and 3D U-Net were trained with weighted Dice loss (WDL) with class weights, ω_l (N: number of pixels; l: classes; r_{ln} and p_{ln}: reference and output segmented pixel n, respectively) [16], where the class weights were calculated as $\omega_l = 1/(\sum_{n=1}^{N} r_{ln})^2$

$$WDL = 1 - 2\frac{\sum_{l=1}^{3}\omega_l \sum_n r_{ln}p_{ln}}{\sum_{l=1}^{3}\omega_l \sum_n r_{ln} + p_{ln}} \quad (1)$$

We extended 3D U-Net to a new Attention-based Competitive Dense 3D U-Net (ACD 3D U-Net) (**Figure 1A**). Each purple box represents a competitive densely connected block, shown in **Figure 1B** with growth rate $k = 2$. Competitive dense blocks are more computationally efficient [17] than the popular densely connected blocks [18]. At the bottom layer, we added convolutional attention blocks (blue blocks) to consider the more informative spatial and channel features (**Figure 1C**) [19]. In order to account for the complex spatially varying and disconnected VAT, we trained the ACD

3D U-Net using a novel loss function. We used logistic weights for ω_l in WDL, where the first term considers the class frequency imbalance and the second term emphasizes the boundaries (I: indicator function, f: class frequencies, ∇: gradient). $\omega_0 = 2\frac{median(f)}{f_{min}}$ gives higher priority to boundaries. We refer to this new loss function as frequency-balancing boundary-emphasizing Dice loss (FBDL).

$$\omega_l = \sum_{n=1}^{N} I(p_{ln} == l)\frac{median(f)}{f_l} + \omega_0 I(|\nabla p_{ln}| > 0) \quad (2)$$

The 3D networks are subject-based (i.e., use the full volumetric MRI dataset as input), while the 2D U-Net is a slice-by-slice segmentation method. All neural networks were implemented in PyTorch 1.8.0 and trained on a NVIDIA Quadro RTX 8000 with 48 GB RAM.

C. Evaluations

The segmentation performance was assessed with respect to the reference manual annotations using an overlap-based metric (3D Dice score, range 0 to 1), and a probability-based metric (Cohen's Kappa coefficient, κ, range 0 to 1) to account for different image features. The Dice score is a standard metric for evaluating medical image segmentation and is sensitive to boundary errors [20]. κ is a measure of agreement between two samples, which is robust since it accounts for the agreement caused by chance [20]. The number of false negative (FN) and false positive (FP) pixels (SAT and VAT combined) were normalized to the total number of SAT and VAT pixels in the manual annotations (range 0 to 100%). All evaluation metrics were calculated per exam in the testing set and reported as mean and standard deviation (SD) across all exams in the testing set.

III. RESULTS

The inference time, training time, and number of trainable parameters are listed in **Table II**. The inference times for all networks were comparable. As expected, 2D U-Net had more trainable parameters and the longest training duration due to its deeper structure than its 3D counterparts. ACD 3D U-Net had fewer trainable parameters than 3D U-Net.

	2D U-Net	3D U-Net	ACD 3D U-Net
Inf. time	36 ms/slice	44 ms/slice	55 ms/slice
Train. time	9.8 hr	5.1 hr	5.8 hr
Train. param.	31043651	22403011	20798835

TABLE II: Characteristics of the neural networks. Inf.: inference. Train.: training. Param.: parameters.

Figure 2 shows inputs (T1W TE_{OP}, fat, and water images), manual annotations, and output segmentation masks from the neural networks in 2 representative subjects. The 2D U-Net yielded the poorest SAT and VAT segmentation performance for these 2 subjects. For 3D U-Net, the 3D Dice score and κ for subject 1 were 0.95 and 0.94, respectively. ACD 3D U-Net achieved higher 3D Dice score = 0.98 and higher κ = 0.97. For subject 2, ACD 3D U-Net again achieved higher 3D Dice and κ scores than 3D U-Net (3D Dice scores: 0.98 vs 0.96; κ: 0.97 vs. 0.96). False positives and false negatives are illustrated in **Figure 3** for chosen slices from the same subjects. For both subjects, 2D U-Net resulted in the greatest number of false positive and false negative pixels. The ACD 3D U-Net results had the smallest number of false positives and comparable number of false negatives as 3D U-Net for the chosen slices from both subjects.

Table III reports the overall segmentation performance using 3D Dice score, κ, FN and FP. 2D U-Net yielded the lowest segmentation performance. Both proposed 3D neural networks achieved excellent segmentation performance for SAT and VAT, as evidenced by the high mean 3D Dice and κ scores and low mean FN and FP across all exams in the testing set. The ACD 3D U-Net outperformed 3D U-Net for VAT segmentation.

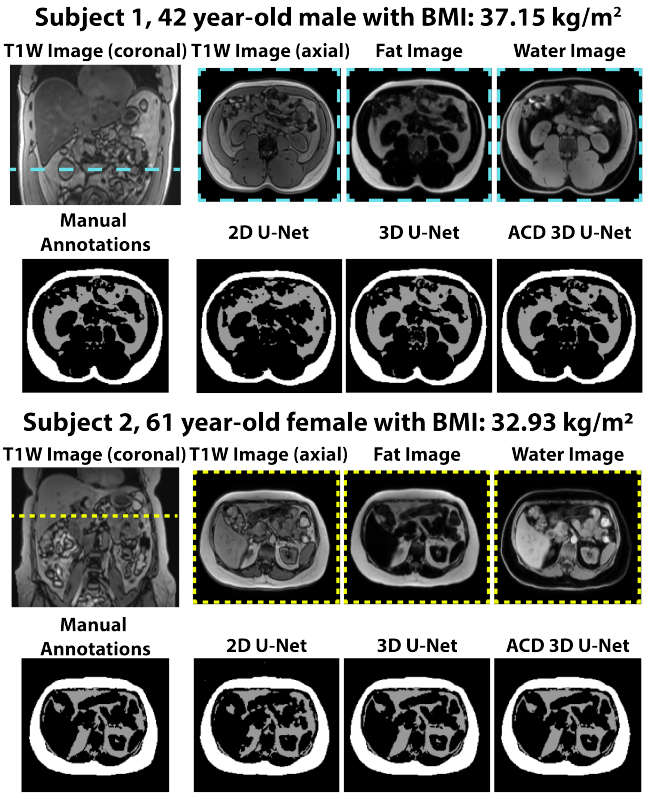

Fig. 2: Input images (T1W TE_{OP}, fat, and water), manual annotations, and output SAT (white) and VAT (gray) masks from 2D U-Net, 3D U-Net, and ACD 3D U-Net in 2 representative subjects. For subject 1, 3D Dice scores for VAT from 2D U-Net, 3D U-Net and ACD 3D U-Net were (0.71, 0.95, and 0.98), respectively. The κ for VAT were: 0.67 (2D U-Net), 0.94 (3D U-Net) and 0.97 (ACD 3D U-Net). The 3D Dice scores for VAT in subject 2 were higher for the ACD 3D U-Net (2D U-Net: 0.81, 3D U-Net: 0.96, ACD 3D U-Net: 0.98). κ for VAT in subject 2 were: 0.79 (2D U-Net), 0.96 (3D U-Net), and 0.97 (ACD 3D U-Net).

IV. DISCUSSION

We developed 3D neural networks for rapid and accurate automated segmentation of VAT and SAT. For both of the proposed 3D networks, we used multi-contrast MRI data

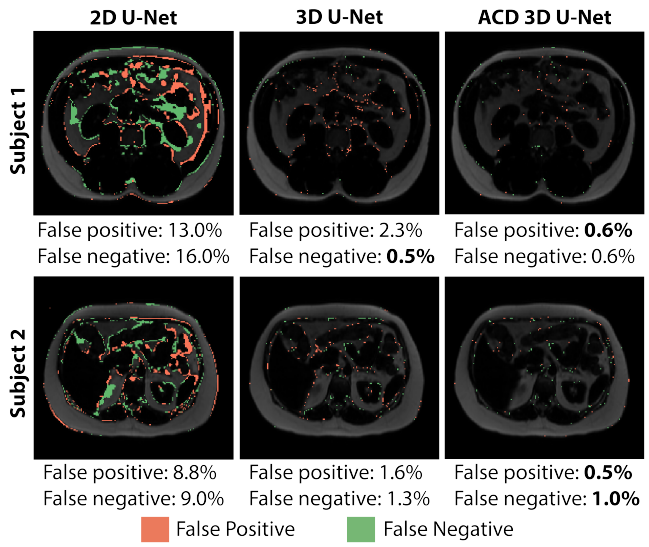

Fig. 3: The false positive (orange) and false negative (green) pixels from 2D U-Net, 3D U-Net, and ACD 3D U-Net, overlaid on fat images in representative slices. The proposed 3D U-Net and ACD 3D U-Net reduced false negatives and false positives compared to 2D U-Net. The ACD 3D U-Net achieved the smallest number of false positives.

	3D Dice (mean±SD)	
	SAT	VAT
2D U-Net+WDL	0.96 ± 0.02	0.77 ± 0.06
3D U-Net+WDL	$\mathbf{0.99 \pm 0.01}$	0.95 ± 0.04
ACD 3D U-Net+FBDL	$\mathbf{0.99 \pm 0.01}$	$\mathbf{0.96 \pm 0.04}$
	κ (mean±SD)	
	SAT	VAT
2D U-Net+WDL	0.95 ± 0.03	0.74 ± 0.07
3D U-Net+WDL	0.98 ± 0.01	0.94 ± 0.04
ACD 3D U-Net+FBDL	$\mathbf{0.99 \pm 0.01}$	$\mathbf{0.96 \pm 0.04}$
	False neg./pos. (mean±SD)	
	FN	FP
2D U-Net+WDL	$9.4\% \pm 4.1\%$	$12.0\% \pm 5.1\%$
3D U-Net+WDL	$\mathbf{1.5\% \pm 2.0\%}$	$1.3\% \pm 1.9\%$
ACD 3D U-Net+FBDL	$\mathbf{1.5\% \pm 2.4\%}$	$\mathbf{1.2\% \pm 1.7\%}$

TABLE III: Segmentation performance metrics 3D Dice scores, κ, false negatives (FN), and false positives (FP) reported as mean $\pm$ standard deviation (SD) across all 18 exams in the testing set. Note the major improvements of 3D networks compared to 2D U-Net. ACD 3D U-Net achieved the highest mean 3D Dice scores and κ and lowest mean FN and FP.

from a single Dixon acquisition (T1W TE_{OP}, fat, and water images) and exploited the full-FOV volumetric information. This allowed us to successfully capture and address the spatially varying and disconnected nature of VAT. Our new ACD 3D U-Net utilized attention mechanisms in addition to competitive dense blocks and a frequency-balancing boundary-emphasizing Dice loss to further improve the segmentation of VAT without major differences in training/inference time. Both the proposed 3D U-Net and ACD 3D U-Net resulted in superior segmentation performance for SAT and VAT compared to 2D U-Net, which was not able to capture the through-plane associations for VAT. The proposed 3D networks yielded excellent segmentation performance for SAT, and importantly, VAT. In particular, ACD 3D U-Net achieved higher 3D Dice scores and κ, and comparable number of false positives and false negatives, for VAT segmentation when compared to 3D U-Net. Having fewer trainable parameters with higher performance levels could mean that the proposed ACD 3D U-Net generalizes the features better than the 3D U-Net and 2D U-Net.

DL-based approaches can achieve rapid inference time (e.g., ms/slice) [10], [12], [13] and can potentially make AT segmentation and body composition analysis a practical tool in research and clinical settings. A recently proposed 2.5D FatSegNet was trained with T1W images at 2 echo times (TE_{OP}, TE_{IP}) from 32 subjects and improved VAT segmentation compared to 2D U-Net (mean 2D Dice scores: 0.850 vs. 0.837) [12] in a cross-validation set of 6 subjects. This network addressed the complex structure of VAT by using a composite Dice and logistic loss function. However, this 2.5D network required training in 3 different image plane orientations separately, and then used an aggregation network to combine outputs from all orientations, which increased the complexity and the training time. Another recent study focused on automated AT segmentation based on 3D U-Net [13] with 3D patches (32×32×32) of Dixon water and fat images as the inputs. This study enhanced the 3D U-Net by incorporating attention mechanisms and standard densely connected block, and then trained the 3D network in 210 subjects using a composite Dice and logistic loss function. The 3D Dice scores for VAT segmentation in a testing set of 60 subjects were 0.43±0.17 for patch-based 3D U-Net and 0.89±0.05 for the proposed patch-based 3D network.

Compared to these representative previous studies, our ACD 3D U-Net had distinct advantages of: (1) using T1W images (TE_{OP}) combined with fat and water images as inputs to take advantage of multi-contrast MRI information, (2) using full-FOV volumetric images, instead of limited slices (2.5D) or 3D patches, to exploit 3D context and global features for resolving VAT, and (3) using a novel loss function to address the class imbalance of VAT/SAT pixels. The VAT segmentation performance of ACD 3D U-Net (3D Dice scores of 0.96±0.04) in an initial testing set demonstrates its potential to advance the state-of-the-art.

Our study had limitations. First, there is relatively high memory requirement for the proposed 3D U-Net and ACD 3D U-Net (∼31 GB) as they require tensor multiplications with larger dimensions due to full-FOV volumetric inputs. The increasing availability of high-performance computing hardware partially alleviates this challenge. Nevertheless, future research can compare trade-offs regarding segmentation performance and memory requirements for different patch-

based and full-FOV models, and investigate strategies to combine the relative advantages. Second, we had limited sample size in our training/testing sets. While the size of our training dataset is already comparable to previous studies on VAT/SAT segmentation, having a larger number of subjects in the testing dataset can strengthen the evaluation and enable statistical analyses. We will include more subjects and exams from the HAT study as more data becomes available. Third, for the comparison between 3D U-Net and ACD 3D U-Net, we have not yet performed a detailed ablation study to investigate the relative contributions of each component (competitive dense connections, attention mechanism, FBDL) in our networks. This will be the topic of further investigation. Fourth, we have not performed head-to-head comparisons of our proposed networks with other 2.5D or 3D networks proposed in recent studies. We acknowledge that the differences in our performance compared to previous work could be due to differences in the underlying subject characteristics and evaluation methods.

V. CONCLUSIONS

Taking advantage of full-FOV volumetric multi-contrast MR images, we demonstrated the superior performance of 3D U-Net to segment SAT/VAT compared to the baseline 2D U-Net. Then, we extended our work in 3D U-Net and proposed a new Attention-based Competitive Dense 3D U-Net (ACD 3D U-Net) and a novel class frequency-balancing boundary emphasizing Dice loss function. 3D U-Net and ADC 3D U-Net yielded excellent SAT and VAT segmentation performance. ADC 3D U-Net achieved higher mean 3D Dice scores (0.96 for VAT; 0.99 for SAT) and κ (0.96 for VAT; 0.99 for SAT) than the other tested networks. Our proposed ADC 3D U-Net could be used as an automated tool to rapidly ($<$60 ms per slice) analyze body composition in adults with overweight/obesity.

VI. COMPLIANCE WITH ETHICAL STANDARDS

The study was approved by the Institutional Review Board and written informed consent was obtained from all participants prior to research procedures.

VII. ACKNOWLEDGMENT

This study was supported in part by the Hass Avocado Board and the NIH (R01-DK124417).

REFERENCES

[1] Attanasio, S and Forte, SM and Restante, G and Gabelloni, M and Guglielmi, G and Neri, E, "Artificial Intelligence, Radiomics and Other Horizons in Body Composition Assessment," *Quantitative Imaging in Medicine and Surgery*, vol. 10, no. 8, pp. 1650, 2020.

[2] Shuster, A and Patlas, M and Pinthus, JH and Mourtzakis, M, "The Clinical Importance of Visceral Adiposity: a Critical Review of Methods for Visceral Adipose Tissue Analysis," *The British Journal of Radiology*, vol. 85, no. 1009, pp. 1–10, 2012.

[3] Linge, J and Borga, M and West, J and Tuthill, T and Miller, MR and Dumitriu, A and Thomas, EL and Romu, T and Tunón, P and Bell, JD and others, "Body Composition Profiling in the UK Biobank Imaging Study," *Obesity*, vol. 26, no. 11, pp. 1785–1795, 2018.

[4] Fox, CS and Massaro, JM and Hoffmann, U and Pou, K and Maurovich-Horvat, P and Liu, CY and others, "Abdominal Visceral and Subcutaneous Adipose Tissue Compartments: Association with Metabolic Risk Factors in the Framingham Heart Study," *Circulation*, vol. 116, pp. 39–48, 2007.

[5] Ly, KV and Armstrong, T and Yeh, J and Ghahremani, S and Kim, GH and Wu, HH and Calkins, KL, "Free-breathing Magnetic Resonance Imaging Assessment of Body Composition in Healthy and Overweight Children: an Observational Study," *Journal of Pediatric Gastroenterology and Nutrition*, vol. 68, no. 6, pp. 782–787, 2019.

[6] Kullberg, J and Karlsson, AK and Stokland, E and Svensson, P and Dahlgren, J, "Adipose Tissue Distribution in Children: Automated Quantification Using Water and Fat MRI," *Journal of Magnetic Resonance Imaging*, vol. 32, no. 1, pp. 204–210, 2010.

[7] Shen, N and Li, X and Zheng, S and Zhang, L and Fu, Y and Liu, X and Li, M and Li, J and Guo, S and Zhang, H, "Automated and Accurate Quantification of Subcutaneous and Visceral Adipose Tissue from Magnetic Resonance Imaging Based on Machine Learning," *Magnetic Resonance Imaging*, vol. 64, pp. 28–36, 2019.

[8] Hui, SCN and Zhang, T and Shi, L and Wang, D and Ip, CB and Chu, WCW, "Automated Segmentation of Abdominal Subcutaneous Adipose Tissue and Visceral Adipose Tissue in Obese Adolescent in MRI," *Magnetic Resonance Imaging*, vol. 45, pp. 97–104, 2018.

[9] Ronneberger, O and Fischer, P and Brox, T, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[10] Langner, T and Hedström, A and Mörwald, K and Weghuber, D and Forslund, A and Bergsten, P and Ahlström, H and Kullberg, J, "Fully Convolutional Networks for Automated Segmentation of Abdominal Adipose Tissue Depots in Multicenter Water–Fat MRI," *Magnetic Resonance in Medicine*, vol. 81, no. 4, pp. 2736–2745, 2019.

[11] Kafali, SG and Shih, S and Li, X and Armstrong, T and Ly, KV and Ghahremani, S and Calkins, KL and Wu, HH, "Fully Convolutional Networks for Adipose Tissue Segmentation Using Free-Breathing Abdominal MRI in Healthy and Overweight Children," *Proceedings of the 28th Annual Meeting of International Society of Magnetic Resonance in Medicine*, 2020.

[12] Estrada, S and Lu, R and Conjeti, S and Orozco-Ruiz, X and Panos-Willuhn, J and Breteler, MMB and Reuter, M, "FatSegNet: A Fully Automated Deep Learning Pipeline for Adipose Tissue Segmentation on Abdominal Dixon MRI," *Magnetic Resonance in Medicine*, vol. 83, no. 4, pp. 1471–1483, 2020.

[13] Küstner, T and Hepp, T and Fischer, M and Schwartz, M and Fritsche, A and Häring, HU and Nikolaou, K and Bamberg, F and Yang, B and Schick, F and others, "Fully Automated and Standardized Segmentation of Adipose Tissue Compartments via Deep Learning in 3D Whole-Body MRI of Epidemiologic Cohort Studies," *Radiology: Artificial Intelligence*, vol. 2, no. 5, pp. e200010, 2020.

[14] Eggers, H and Börnert, P, "Chemical Shift Encoding-based Water–Fat Separation Methods," *Journal of Magnetic Resonance Imaging*, vol. 40, no. 2, pp. 251–268, 2014.

[15] Çiçek, Ö and Abdulkadir, A and Lienkamp, SS and Brox, T and Ronneberger, O, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[16] Sudre, CH and Li, W and Vercauteren, T and Ourselin, S and Cardoso, MJ, "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 240–248. Springer, 2017.

[17] Estrada, S and Conjeti, S and Ahmad, M and Navab, N and Reuter, M, "Competition vs. Concatenation in Skip Connections of Fully Convolutional Networks," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2018, pp. 214–222.

[18] Huang, G and Liu, Z and Van Der Maaten, L and Weinberger, KQ, "Densely Connected Convolutional Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.

[19] Woo, S and Park, J and Lee, JY and Kweon, IS, "CBAM: Convolutional Block Attention Module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[20] Taha, AA and Hanbury, A, "Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool," *BMC Medical Imaging*, vol. 15, no. 1, pp. 1–28, 2015.