FatSegNet: A Fully Automated Deep Learning Pipeline for Adipose Tissue Segmentation on Abdominal Dixon MRI

Santiago Estrada\textsuperscript{1,2} \textvert \ Ran Lu\textsuperscript{2} \textvert \ Sailesh Conjeti\textsuperscript{1} \textvert 
Ximena Orozco-Ruiz\textsuperscript{2} \textvert \ Joana Panos-Willuhn\textsuperscript{2} \textvert 
Monique M.B. Breteler\textsuperscript{2,3} \textvert \ Martin Reuter\textsuperscript{1,4,5}

\textsuperscript{1}Image Analysis, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany
\textsuperscript{2}Population Health Sciences, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany
\textsuperscript{3}Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), Faculty of Medicine, University of Bonn, Bonn, Germany
\textsuperscript{4}A.A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston MA, USA
\textsuperscript{5}Department of Radiology, Harvard Medical School, Boston MA, USA

Correspondence
Santiago Estrada, German Center for Neurodegenerative Diseases (DZNE), Sigmund-Freud-Straße 27, 53127 Bonn, Germany
Email: santiago.estrada@dzne.de

Funding information
Diet-Body-Brain Competence Cluster in Nutrition Research, Federal Ministry of Education and Research, Grant Numbers: 01EA1410C and FKZ: 01EA1809C; JPI HDHL on Biomarkers for Nutrition and Health, Federal Ministry of Education and Research, Grant Number: 01EA1705B; National Institutes of Health (NIH), Grant Number: R01NS083534 and R01LM012719

Purpose: Development of a fast and fully automated deep learning pipeline (FatSegNet) to accurately identify, segment, and quantify abdominal adipose tissue on Dixon MRI from the Rhineland Study - a large prospective population-based study.

Method: FatSegNet is composed of three stages: (i) consistent localization of the abdominal region using two 2D-Competitive Dense Fully Convolutional Networks (CDFNet), (ii) segmentation of adipose tissue on three views by independent CDFNets, and (iii) view aggregation. FatSegNet is trained with 33 manually annotated subjects, and validated by: 1) comparison of segmentation accuracy against a testing set covering a wide range of body mass index (BMI), 2) test-retest reliability, and 3) robustness in a large cohort study.

Results: The CDFNet demonstrates increased robustness compared to traditional deep learning networks. FatSegNet dice score outperforms manual raters on the abdominal visceral adipose tissue (VAT, 0.828 vs. 0.788), and produces comparable results on subcutaneous adipose tissue (SAT, 0.973 vs. 0.982). The pipeline has very small test-retest absolute percentage difference (APD) and excellent agreement between scan sessions (VAT: APD = 2.957%, ICC=0.998 and SAT: APD = 3.254%, ICC=0.996).

Conclusion: FatSegNet can reliably analyze a 3D Dixon MRI in ~ 1 min. It generalizes well to different body shapes, sensitively replicates known VAT and SAT volume effects in a large cohort study, and permits localized analysis of fat compartments.

Keywords
Subcutaneous adipose tissue, Visceral adipose tissue, Dixon MR imaging, neural networks, deep learning, semantic segmentation
INTRODUCTION

The excess of body fat depots is an increasing major public health issue worldwide and is an important risk factor for the development of metabolic disorders and a reduced quality of life [1,2]. While the body mass index (BMI) is a widely used indicator of adipose tissue accumulation in the body, it does not provide information on fat distribution [3] - neither with respect to different fat tissue types nor with respect to deposit location. Different compartments of adipose tissue are associated with different physiopathological effects [4,5]. Abdominal adipose tissue (AAT), composed of subcutaneous and visceral adipose tissue (SAT and VAT), has long been associated with an increased risk of chronic cardiovascular diseases, glucose impairment and dyslipidemia [4,7]. Recently, several studies have indicated a stronger relation between the accumulation of VAT with an adverse metabolic and inflammatory profile compared to SAT [8,9]. Therefore, an accurate and independent measurement of VAT and SAT volumes (VAT-V and SAT-V) is of significant clinical and research interest.

Currently, the gold standard for measuring VAT-V and SAT-V is the manual segmentation of abdominal fat images from Dixon magnetic resonance (MR) scans - a very expensive and time-consuming process. Thus, especially for large studies, automatic segmentation methods are required. However, achieving good accuracy is challenging due to complex AAT structures, a wide variety of VAT shapes, large anatomical differences across subjects, and the inherent properties of the Dixon images: low intensity contrast between adipose tissue classes, inhomogeneous signals, and potential organ motion. So far, those limitations impeded the wide-spread implementation of automatic and semi-automatic techniques based on intensity and shape features, such as fuzzy-clustering [10], k-means clustering [11], graph cut [12] [13] active contour methods [14] and statistical shape models [15].

Recently, fully convolutional neural networks (F-CNNs) [16,17] have been widely adopted in the computer vision community for pixel/voxel-wise image segmentation in an end-to-end fashion to overcome above-mentioned challenges. With these methods there is no need to extract manual features, divide images into patches, or implement sliding window techniques. F-CNNs can automatically extract intrinsic features and integrate global context to resolve local ambiguities thereby improving the results of the predicted models [17]. Langer et al. [18] proposed a three channel U-Net for AAT segmentation, which is a conventional architecture for 2D medical image segmentation [19]. While this method showed promising results, we demonstrate that our network architecture outperforms the traditional U-Net for segmenting AAT on our images with a wide range of anatomical variation. More recent architectures such as the SD-Net [20] and DenseNet, a densely connected network [21], have the potential to improve generalizability and robustness by encouraging feature re-usability and strengthening information propagation across the network [21]. In prior work, we introduced a competitive dense fully convolutional network (CDFNet) [22] as a new 2D F-CNN architecture that promotes feature selectivity within a network by introducing maximum attention through a maxout activation unit [23]. The maxout boosts performance by allowing the creation of specialized sub-networks that target a specific structure during training [24]. Therefore, this approach facilitates the learning of more complex structures [22,24] with the added benefit of reducing the number of training parameters relative to the aforementioned networks.

In this paper, we propose FatSegNet, a novel fully automated deep learning pipeline based on our CDFNet architecture to localize and segment VAT and SAT on abdominal Dixon MR images from the Rhineland Study, an ongoing large population-based cohort study [25,26]. The proposed pipeline consists of three stages:

1. Localization of the abdominal region using a semantic segmentation approach by implementing CDFNet models on sagittal and coronal planes; we use the lumbar vertebrae positions as reference points for selecting the region of interest.
2. Segmentation of VAT and SAT on the abdominal region through CDFNet models on three different planes (axial, sagittal and coronal).
3. A view aggregation stage where the previous generated label maps are combined to generate a final segmentation.
We initially evaluate and compare the CDFNet (stages 1 and 2) with other deep learning approaches on a testing set with a wide range of BMI. We show that the CDFNet reduces the number of required training parameters and improves the segmentation performance. Afterwards, we validate the whole pipeline (FatSegNet) by first testing the accuracy against the testing set, and then evaluating robustness and reliability against an independent test-retest set. Finally, we present a case study on unseen data comparing the VAT-V and SAT-V calculated from the segmentations generated by FatSegNet against BMI and explore the effect of age and sex on these volumes.

2 | METHODS

2.1 | Data

2.1.1 | MR imaging acquisition

MR image acquisition was performed at two different sites both with identical 3T Siemens MAGNETOM Prisma MR scanners (Siemens Healthcare, Erlangen, Germany). The body coil was used for signal reception of a three-dimensional two-point Dixon sequence (acquisition time = 12 s, echo time TE1=1.23 ms, TE2=2.46 ms, repetition time TR=4.12 ms, axial field of view =500 mm x 437 mm, flip angle = 6 degrees, left-right readout bandwidth = 750 Hz/pixel, partial Fourier factor 6/8 x 5/8). Based on a preceding moving-table abdominal localizer, the field-of-view was centered on the middle of the third lumbar vertebra (L, L3). Data were acquired during a single breath-hold in supine position with arms placed at the sides. The image resolution was finally interpolated from 2.0 mm x 2.7 mm x 10.0 mm to 2.0 mm x 2.0 mm x 5.0 mm (matrix size = 256 x 224 x 72).

2.1.2 | Datasets

The Rhineland Study is an ongoing population-based prospective cohort (https://www.rheinland-studie.de/) which enrolls participants aged 30 years and above at baseline from Bonn, Germany. The study is carried out in accordance with the recommendations of the International Council for Harmonisation (ICH) Good Clinical Practice (GCP) standards (ICH-GCP). Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki. The first 641 subjects from the Rhineland Study with BMI and abdominal MR Dixon scans are included. The group presents a mean age of 54.2 years (range 30 to 95) and 55.2% of the subjects are women. The BMI of the participants range from 17.2 to 47.7 kg/m² with a mean of 25.2 kg/m². Subjects were stratified into two subsets: 38 scans were manually annotated for training and testing; the remaining 603 subjects were segmented using the proposed pipeline. After visual inspection, 16 subjects were excluded due to poor image quality or extreme motion artifacts (e.g. potentially caused by breathing). Thus, 587 participants were used for later statistical analysis. Additionally, an independent set of 17 healthy student volunteers was used to assess test-retest reliability of the automated segmentation and volume estimates.

Ground Truth Data: 38 subjects were randomly selected from sex and BMI strata to ensure a balanced gender and BMI distribution. These scans were manually annotated by two trained raters without any semi-automated support such as thresholding, which can reduce accuracy in the ground truth and lead to overestimation of the performance of the proposed automated method.

Specific label schemes were created for each individual task of the pipeline. For localizing the abdominal region, raters divided the scans into three different blocks defined by the location of the vertebrae as follows: the abdominal region (from lower bound of twelfth thoracic vertebra (Th12) to the lower bound of L5), the thoracic region (all above the lower bound of Th12), and the pelvic region (everything below the lower bound of L5), as illustrated in Fig. 1c). For AAT segmentation, 60 slices per subject were manually labeled into three classes: SAT, VAT, and bone with neighbouring tissues. The bone was labeled to prevent bone marrow from being misclassified as adipose tissue. In order to improve spatial context and prevent missegmentation of the arms, the dataset was complemented by a synthetic class defined as "other tissue" that was composed of any soft tissue inside the abdomen cavity that is not VAT or SAT. The manual annotations are illustrated in
Fig. 1(b) and c). Furthermore, four subjects were labeled by both raters to evaluate the inter-rater variability.

Test-Retest Data: 17 subjects were recruited with the exclusive purpose of measuring the acquisition protocol reliability. The group presents a mean age of 25.5 years (range: 20 to 31) and 65.0% of the participants are women; all of them have a normal BMI (BMI < 25 kg/m²). Subjects were scanned in two consecutive sessions. Before starting the second session subjects were removed from the scanner and re- positioned.

2.2 | FatSegNet Pipeline

The FatSegNet was conceived to be deployed as a post-processing adipose analysis pipeline for the abdominal Dixon MR images acquired in the Rhineland Study. Therefore, it should meet the following requirements: 1) be fully automated, 2) only segment the adipose tissue on the abdominal region inside the abdominal cavity, and 3) be robust to body type variations and generalizable in presence of high population heterogeneity. Following the prior conditions we designed FatSegNet as a fully automated deep learning pipeline for adipose segmentation (Fig. 2).

The proposed pipeline is divided into three stages: (i) the abdominal region is localized by two independent CDFNets on the sagittal and coronal view. A bounding box containing only the abdominal region class is extracted from each predicted label map; the obtained bounding boxes are averaged (mean center and mean dimensions) to produce a final abdominal region. (ii) Afterwards, the adipose tissue is segmented within the abdominal region by three CDFNets, each processes a different view (axial, coronal and sagittal). (iii) Finally, a view aggregation network merges the predicted label maps from the previous stage into a final segmentation; the implemented multi-view scheme is designed to improve segmentation of structures that are not clearly visible due to poor lateral resolution. This 2.5D strategy produces a fully automated pipeline to accurately segment adipose tissue inside a consistent anatomically defined abdominal region.

2.2.1 | Pipeline components

Competitive Dense Fully Convolutional Network (CDFNet): For the segmentation task we selected our CDFNet architecture due to its robustness and generalizability properties as demonstrated in a similar and challenging task of multi-organ segmentation in contrast-enhanced abdominal MRI from the publicly available Visceral segmentation benchmark [27]. The proposed network improves feature selectivity and, thus, boosts the learning of fine-grained anatomies without increasing the number of learned parameters [22]. We implemented the CDFNet by suitably adopting the Dense-UNet architecture proposed
by Roy et al. [28] and extending it towards competitive learning via maxout activations [24].

The Dense-UNet proposed in [28] follows the usual dumb-bell like architecture with 4 dense-block encoders, 4 dense-block decoders and a bottleneck layer. Each dense-block is based on short-range-skip-connections between convolutional layers as introduced for densely-connected neural networks [27]; the dense connection approach stacks multiple convolutional layers in sequence and the input of a layer is iteratively concatenated with the outputs of the previous layers. This type of connectivity improves feature reusability, increases information propagation and alleviates vanishing gradients [27]. The architecture additionally incorporates the traditional long-range-skip-connections between all encoder and decoder blocks of the same spatial resolution as introduced by Ronnenberger et al. [19] which improves gradient flow and spatial information recovery.

Within the network the information aggregation through these connections is performed by concatenation layers. Such a design increases the size of the output feature map along the feature channels, which in turn results in the need to learn filters with a higher number of parameters. Goodfellow et al. introduced the idea of competitive learning through maxout activations [23], which was adapted by Liao and Carneiro [24] for competitive pooling of multi-scale filter outputs. Both [23] and [24] proved that the use of maxout competitive units boosts performance by creating a large number of dedicated sub-networks within a network that learns to target specific sub-tasks and reduces the number of required parameters significantly.

The maxout is a simple feed-forward activation function that chooses the maximum value from its inputs [23]. Within a CNN, a maxout feature map is constructed by taking the maximum across multiple input feature maps (\(X\)) for a particular spatial location (say \((i,j,k)\)). Assuming \(L\) inputs, denoted as \(X = \{x_l\}_{l=1}^L\), with each \(x_l^{H,W,C}_{i,j,k=1}\) where \(H\) is height, \(W\) is width and \(C\) is number of channels for a particular feature map(\(x_l\)). The maxout(\(X\)) output is given by:

\[
\text{maxout}(X) = \left[ y_{ijk} \right]_{i,j,k=1}^{H,W,C}
\]

where \(y_{ijk} = \max\left\{ x_1^{i,j,k}, \ldots, x_L^{i,j,k} \right\} \) (1)

The proposed CDFNet uses competitive layers (maxout activation) instead of concatenation layers. However, in an encoder/decoder architecture there are both short-range and long-range-skip-connections which introduce competition in two distinct ways. We introduce two novel archi-
tectural blocks for dealing with each type of competition (local and global) as follows:

- **Local Competition - Competitive Dense Block (CDB):** The dense convolutional block proposed in [29] introduces feed-forward connections from each layer to every other layer. The dense block concatenates feature-maps of all previous layers as input to the current layer and the output of the current layer is used as input to all subsequent layers within the block (dense connections) (Eq. 2-4). We replace the feature map concatenations with maxout activations to promote local competition among the layers. This is mathematically formulated in Eq. 5-7 and illustrated in Fig. 3.

\[
X_1 = H_3^l(y_2) \\
y_2 = [H_2^l(y_1), y_1, X_{i-1}] \\
y_1 = [H_1^l(X_{i-1}), X_{i-1}] \\
X_i = H_3^l(y_2) \\
y_2 = \text{maxout}(H_2^l(y_1), y_1) \\
y_1 = \text{maxout}(H_1^l(X_{i-1}), X_{i-1})
\]

Here, “[ ]” represents the concatenation operator and \(H_j^l\) a composite function of three consecutive operations: Batch Normalization (BN), followed by ReLU and convolution. However, this sequence of operations does not guarantee normalized inputs to the maxout activation. Therefore, \(\tilde{H}_l^j\) reorders the operations: first a convolution followed by a ReLU and finally the BN to normalize the features. Providing pre-conditioned inputs to the maxout activation simultaneously improves convergence [30] and increases the exploratory span of the created sub-networks [24]. Furthermore, filter co-adaptation (filters having the same convergence point) is implicitly prevented by the dense blocks [24].

- **Global Competition - Competitive Un-pooling Block (CUB):** As mentioned in [19][21], the long-range skip-connections between encoding and decoding paths is performed through the concatenation layer. To induce competition within this layer, a naive solution would be to perform a maxout operation directly between the feature maps of the upsampling path and the skip connection as in the CDB design. However, we empirically observed that such architecture was unstable and resulted in loss of information. To counter this problem, we propose to first learn a joint feature-map (through a \(1 \times 1\) convolutional layer \(\tilde{H}\)), which in turn competes with the features from the skip connection. Such a design (Fig. 3) improves feature selectivity between fine-grained (with local span) and coarser high-context information (with much wider span) coming from the up-sampling path.

In brief, the proposed CDFNet comprises a sequence of four CDBs, constituting the encoder path (down-sampling block), and four CDBs constituting the decoder path (up-sampling block), which is joined via a bottleneck layer. The bottleneck consists of a 2D convolutional layer followed by a Batch Normalization. The skip-connections from each of the encoder blocks feed into the CUB that subsequently forwards features into the corresponding decoder block of the same resolution as illustrated in Fig. 3.

**View Aggregation Network** The proposed view aggregation network is designed to regularize the prediction for a given voxel by considering spatial information from the coronal, axial and sagittal view. The network, therefore, merges the probability maps of the three different CDF-Nets from the previous stage by applying a \(1 \times 1 \times 1\) 3D-convolution (30 filters) followed by a Batch Normalization. Then a \(1 \times 1 \times 1\) 3D-convolution is employed to reduce the feature maps to the desired number of classes \((n=5)\). The final prediction probabilities are obtained via a concluding softmax layer. Our approach learns to weight each view differently compared to standard aggregation schemes, i.e. voting or averaging, which assign constant weights to each view. Such hard-coded weighting schemes are suboptimal in situations with anisotropic voxels (we have \(2 \times 2 \times 5\) mm voxels sizes in the abdominal MRI of the Rhinelander Study).
2.3 | Experimental setup

For training and testing the pipeline, we perform a subject-space split on the ground truth dataset. For selecting the testing set, the ground truth dataset is divided into 3 groups based on their BMI classification (normal [BMI < 25 kg/m²], overweight [25 ≤ BMI < 30 kg/m²] and obese [BMI ≥ 30 kg/m²]). 33 subjects are used for training and 5 held out for testing: One test-subject is randomly selected from each group, and two more subjects are randomly selected without considering group memberships. This selection process ensures that all BMI categories are used for bench-marking the pipeline.

**Baselines and comparative methods:** We validate the FatSegNet by comparing the performance of each stage of the pipeline against the test set using a dice score index (DSC) to measure similarity between the predicted segmentation and the ground truth. Let M (ground truth) and A (automatic segmentation) denote the labels binary segmentation, the dice score index is defined as

\[
DSC = \frac{2 \cdot |M \cap A|}{|M| + |A|}
\]

(8)

Where |M| and |A| represents the number of elements in each segmentation, and |M \cap A| the number of common elements. Therefore, the DSC ranges from 0 to 1 and a higher DSC represents a better agreement between segmentations.

Additionally, we benchmark the proposed CDFNet models for abdominal region localization and AAT delineation with state-of-the-art segmentation F-CNNs such as U-Net [19], SD-Net [20], and Dense-U-Net [28]. We use the probability maps generated from the aforementioned networks to train the view aggregation model and measure performance with or without view aggregation. To permit a fair comparison, all benchmark networks follow the same architecture of 4 encoder blocks, 4 decoders blocks, and a bottleneck layer as illustrated in Fig. 3 with an input image.
size of 224 × 256.

The aforementioned models are implemented in Keras [31] with a TensorFlow back-end and trained until convergence using an NVIDIA Titan Xp GPU with 12 GB RAM and the following parameters: batch size of 4, momentum set to 0.9, constant weight decay of $10^{-06}$, and an initial learning rate of 0.01 decreased by a order of 10 every 20 epochs. The models are trained with a composite loss function of median frequency balanced logistic loss and dice loss [20]. This loss function emphasizes the boundaries between classes and supports learning of unbalanced classes such as VAT. Finally, an online affine data augmentation is performed to increase training set size and improve the networks generalizability.

Pipeline reliability: We assess the FatSegNet reliability by comparing the difference of VAT-V and SAT-V across sessions for each subject of the test-retest data set. Given a predicted label map and $N_i(l)$ the number of voxels classified as $l$ (VAT or SAT) in a session $i$ (1 or 2), the absolute percent difference (APD ($l$)) of a label volume measures variability across sessions and is defined as

$$APD(l) = \frac{2 \cdot |N_1(l) - N_2(l)|}{N_1(l) + N_2(l)} \cdot 100$$

Additionally, we calculate the agreement of the total VAT-V and SAT-V by an intra-class correlation (ICC) using a two-way fixed, absolute agreement and single measures ICC(A,1) [32].

Statistical analysis: We compare the volumes of abdominal adipose tissue (AAT-V, SAT-V and VAT-V) generated from FatSegNet with BMI on the unseen dataset. A fast quality control is performed to identify drastic failure cases. The differences among BMI groups are evaluated with a one-way analysis of variance (ANOVA) with subsequent Tukey’s honest significant difference (HSD) post-hoc comparisons. The associations of volumes of abdominal adipose tissue and BMI is assessed using partial correlation and linear regression after accounting for age, sex, and height of the abdominal region. A separate multiple linear regression is performed to explore the effect of age on SAT-V and VAT-V in men and women. All the statistical analyses are performed in R [33].

3 | RESULTS

3.1 | Method Validation

Localization of abdominal region: Table 1 presents the performance of abdominal region localization evaluated on the held-out test data. The table shows the mean Dice scores of the localization labels on an prediction map calculated by averaging the probability maps from the coronal and sagittal view. We observe that the CDFNet outperforms the comparative baselines on all classes, especially on the pelvic region. Thus, the proposed method performs well in finding the desired abdominal region.

**TABLE 1** Mean and standard deviation of the Dice scores of the different models for localization of the abdominal region

| Models     | Thoracic | Abdominal | Pelvic   |
|------------|----------|-----------|----------|
| UNet [19]  | 0.913 (0.059) | 0.909 (0.050) | 0.827 (0.078) |
| SD-Net [20] | 0.900 (0.054) | 0.908 (0.043) | 0.822 (0.070) |
| Dense-UNet [28] | 0.920 (0.053) | 0.917 (0.043) | 0.826 (0.085) |
| CDFNet     | 0.922 (0.059) | 0.921 (0.045) | 0.842 (0.081) |

**FIGURE 4** FatSegNet prediction on an unseen subject with and without view aggregation:

a) FatSegNet with out view aggregation and AAT only segmented on the abdominal region by the axial CDFNet
b) FatSegNet with view aggregation. The red squares on the coronal and axial plane show where view aggregation model prevents arm mis-classification.
TABLE 2  Mean and standard deviation of the Dice scores for the different models for Abdominal Adipose Tissue Segmentation.

| Models            | Parameters* | Subcutaneous (SAT) | Visceral (VAT) |
|-------------------|-------------|--------------------|----------------|
|                   |             | Axial | Coronal | Sagittal | V. Aggregation | Axial | Coronal | Sagittal | V. Aggregation |
| UNet              | 17          | 21,789.529 | 0.969 (0.024) | 0.964 (0.029) | 0.965 (0.028) | 0.972 (0.026) | 0.802 (0.112) | 0.786 (0.137) | 0.784 (0.125) | 0.806 (0.134) |
| SD-Net            | 20          | 1,299.417  | 0.969 (0.024) | 0.962 (0.030) | 0.963 (0.027) | 0.972 (0.028) | 0.805 (0.110) | 0.788 (0.124) | 0.786 (0.120) | 0.813 (0.129) |
| Dense-UNet        | 28          | 3,370.853  | 0.971 (0.024) | 0.964 (0.031) | 0.966 (0.026) | 0.978 (0.025) | 0.826 (0.106) | 0.804 (0.126) | 0.814 (0.113) | 0.826 (0.117) |
| CDFNet            | 25,651.129  | 0.970 (0.024) | 0.965 (0.029) | 0.964 (0.025) | 0.973 (0.026) | 0.826 (0.105) | 0.807 (0.122) | 0.804 (0.108) | 0.828 (0.121) |
|                   |             |        |        |          | Manual inter-rater variability | 0.962 (0.018) | 0.788 (0.060) |

* The number of parameters reported is for the models without the View Aggregation Network

Segmentation of AAT: In Table 2, we present the mean Dice score for VAT and SAT for each individual view and the view aggregation model using the test data set. Here, we observe that all methods work extremely well for SAT segmentation. However, there is a substantial difference in recognition accuracy of VAT which is a more fine-grained compartment with large shape variation. The proposed CDFNet outperforms the UNet and SD-Net on VAT segmentation; when compared with Dense-UNet there is no significant disparity. Nonetheless, CDFNet achieves the same performance with 34% less parameters, demonstrating that the proposed architecture improves feature selectivity and simplifies network learning.

It must also be noted that view aggregation does not improve the outcome significantly compared to the axial view, which holds a higher resolution in comparison to the sagittal and coronal. However, we empirically observe that the view aggregation model smooths the label maps and prevents the arms from being classified as adipose tissue by introducing spatial information from multiple views. Arm misclassification can otherwise be observed especially on overweight and obese subjects, where arms are located closer to the abdominal cavity, as seen in Fig. 2.

Finally, all CDFNet view models achieve similarly excellent results on the SAT segmentation compared to inter-rater variability and outperform the manual raters for the more challenging VAT segmentation.

FatSegNet reliability: For the test-retest dataset, we observe excellent agreement between sessions: the ICC with a 95% confidence interval are 0.998 (0.995 to 0.999) for VAT-V, and 0.996 (0.986 to 0.999) for SAT-V. Finally, the mean (SD) APD for VAT-V and SAT-V are 2.957% (2.600) and 3.254% (2.524), respectively. Note that these values overestimate variance of the processing pipeline, i.e. AAT segmentation and abdominal region localization, as they also include variance from acquisition noise (e.g. motion artefacts, non-linearities based on different positioning).

3.2  Statistical Analysis Rhineland Data

The characteristics of the study population: The characteristics of the 587 participants with valid data on BMI and volumes of abdominal adipose tissue are presented in Table 3. The mean (SD) age of the subjects is 54.2 (13.3) years, and 54.7% are women. 311 (53.0%) subjects are normal-weight, 209 (35.6%) overweight, and 67 (11.4%) obese. We observed a BMI increase with age ($\beta = 0.03$, $P = 0.007$), and a borderline significance of age difference among BMI groups ($P = 0.052$, ANOVA). Obvious differences are observed in AAT-V, VAT-V and SAT-V across BMI groups ($P < 0.001$, ANOVA). VAT-V to SAT-V ratio is higher in overweight and obese participants compared to those with normal weight ($P < 0.001$), but there is no difference between overweight and obese ($P = 0.505$).

TABLE 3  Characteristics of the participants (n=587) showing mean (SD) for continuous and counts (PCT) for categorical variables

| Characteristic | Normal weight | Overweight | Obesity |
|----------------|---------------|------------|--------|
| n              | 311           | 209        | 67     |
| Age in years   | 53.1 (13.2)   | 55.1 (13.5) | 56.9 (12.4) |
| Women, n (%)   | 202 (65.0)    | 76 (36.4)  | 43 (64.2) |
| BMI in kg/m²   | 22.4 (1.6)    | 27.0 (1.4) | 33.3 (3.4) |
| VAT-V in liters| 4.24 (1.42)   | 7.29 (1.64) | 10.59 (2.27) |
| SAT-V in liters| 1.44 (0.93)   | 3.06 (1.40) | 3.99 (1.71) |
| SAT-V to VAT-V | 2.80 (0.80)   | 4.23 (0.94) | 6.60 (2.04) |
| VAT-V to SAT-V ratio | 0.52 (0.32) | 0.76 (0.40) | 0.70 (0.47) |

BMI, body mass index; AAT-V, abdominal adipose tissue volume; SAT-V, subcutaneous adipose tissue volume; VAT-V, visceral adipose tissue volume.
The association between abdominal adipose tissue volumes and BMI: BMI shows a strong positive correlation with AAT-V and SAT-V (AAT-V: $r = 0.88$, $P < 0.001$; SAT-V: $r = 0.85$, $P < 0.001$), but only a moderate correlation with VAT-V ($r = 0.65$, $P < 0.001$) after adjusting for age, sex, and abdominal region height. As illustrated in Fig. 5, both SAT-V and VAT-V are positively associated with BMI after accounting for age, sex, and abdominal region height ($P < 0.001$). The accumulation of SAT-V is higher than VAT-V as BMI increases.

Influence of age and sex on VAT-V and SAT-V: The influence of age and sex on VAT-V and SAT-V follows different patterns (as illustrated in Fig. 6). Men tend to have lower SAT and higher VAT compared to women ($P < 0.001$). VAT-V significantly increase with age in both men and women. Conversely, SAT-V is weakly associated with age in women ($\beta = 0.02$, $P = 0.012$), but not in men ($\beta = -0.01$, $P = 0.337$).

4 DISCUSSION

In our study, we established, validated, and implemented a novel deep learning pipeline to segment and quantify the components of abdominal adipose tissue, namely, VAT-V, SAT-V, and AAT-V on a fast acquisition abdominal Dixon MR protocol for subjects from the Rhineland Study, a large population-based cohort. The proposed pipeline is fully automated and requires approximately 1 min for analyzing a subject’s whole volume. Moreover, since the pipeline is based on deep learning models, it can be easily updated and retrained as the study progresses and new manual data is generated - which can further improve overall pipeline robustness and generalizability, providing a pragmatic solution for a population-based study.

The proposed pipeline, termed FatSegNet, implements a three stage design with the CDFNet architecture at the core for localizing the abdominal region and segmenting the AAT. The introduction of our CDFNet inside the pipeline boosts the competition among filters to improve feature selectivity within the networks. CDFNet introduces com-
petition at a local scale by substituting concatenation layers with maxout activations that prevent filter co-adaptation and reduce the overall network complexity. It also induces competition at a global scale through competitive unpooling. This network design, in turn, can learn more efficiently.

For the first stage of the pipeline, i.e. localization of the abdominal, CDFNet clearly outperforms the other methods and successfully establishes the region of interest. Furthermore, the localization block is able to identify the abdominal region correctly even in cases with scoliosis (curved spine) as illustrated in Fig. 7e). For the more complicated task of segmenting AAT, we demonstrate that CDFNet recovers significantly better VAT compared to traditional deep learning variants that employ concatenation layers. Additionally, each individual CDFNet view model outperforms manual raters for segmenting the complex VAT and accomplishes equivalent results on SAT for the test set. The selection of an inhomogeneous BMI testing set ensures that our method is evaluated for different body types and avoids biases, as higher segmentation performance can be achieved on subjects with high content of AAT compared to lean subjects [34,35].

On the other hand, images from individuals with high AAT could be accompanied by other type of issues, such as fat shadowing (Fig. 7a and h), or arms located in close proximity to the abdominal cavity (Fig. 7a, b and d). These issues are mitigated by our view aggregation model that regularizes the predicted segmentation by combining the spatial context from different views ultimately improving segmentation of tissue boundaries. Moreover, this approach automatically prevents misclassification of arms whereas previous deep learning AAT segmentation methods required manual removal of the upper extremities in a pre-processing step [18]. Furthermore, FatSegNet has a high test-retest agreement and reliability between the calculated volumes of VAT and SAT without the need of any image pre-processing (bias-correction, image registration, etc.) or manual selection of a slice or region.

Based on the Rhineland Study, the proposed pipeline demonstrates a high robustness and generalizability on a large population covering a wide range of age, BMI, and a variety of body shapes as seen in Fig. 7. FatSegNet successfully identifies the AAT in different abdomen morphologies, spine curvatures, adipose shadowing, arms positioning, or intensity inhomogeneities. However, as is usual with any pipeline, segmentation reliability decreases when input images have low quality as illustrated in Fig. 7i) and j) where the scans present severe motion/breathing artifacts or very low image contrast. In order to detect these problematic cases for AAT analysis of large datasets, an automated or manual quality control system should be implemented.

In accordance with previous studies on smaller data sets [13,36], our data showed a lower correlation of BMI with VAT than with AAT-V and SAT-V. We also observed a sex difference of the SAT-V and VAT-V accumulation as previously reported [37,38]: men were more likely to have higher VAT-V and lower SAT-V compared to women. Moreover, we further explored the association between age with SAT-V and VAT-V and found an obvious age effect on the accumulation of VAT-V in both men and women, and a weak age effect on SAT-V in women but not in men. This discrepancy was previously observed by Machann et al. [37], who assessed the body composition using MRI in 150 healthy volunteers aged 19 to 69 years. They reported a strong correlation between VAT-V and age both in men and women, whereas SAT-V only slightly increased with age in women. The fact that our results replicate these previous findings on a large unseen dataset corroborates stability and sensitivity of our pipeline.

In conclusion, we have developed a fully automated post-processing pipeline for adipose tissue segmentation on abdominal Dixon MRI based on deep learning methods. While reducing the number of required parameters, the pipeline outperforms other deep learning architectures and demonstrates high test-retest reliability. Furthermore, the proposed method was successfully validated in a large population-based cohort, where it demonstrated generalizability across a large range of anatomical differences, both with respect to body shape and fat distribution.
5 | ACKNOWLEDGEMENTS

We would like to thank the Rhineland Study group for supporting the data acquisition and management, as well as Mohammad Shahid for his support on the deployment of the FatSegNet method into the Rhineland Study processing pipeline. Furthermore, we acknowledge Tony Stöcker and his team for developing and implementing the Dixon MRI sequence used in this work. This work was supported by the Diet-Body-Brain Competence Cluster in Nutrition Research funded by the Federal Ministry of Education and Research (grant numbers 01EA1410C and FKZ: 01EA1-809C), by the JPI HDHL on Biomarkers for Nutrition and Health (HEALTHMARK), by the Federal Ministry of Education and Research, Germany (grant number 01EA1705B), the NIH R01NS083534, R01LM012719, and an NVIDIA Hardware Award.

REFERENCES

[1] Padwal R, Leslie WD, Lix LM, Majumdar SR. Relationship among body fat percentage, body mass index, and all-cause mortality: a cohort study. Annals of internal medicine 2016;164(8):532–541.

[2] Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. The lancet 2014;384(9945):766–781.

[3] Tomiyama AJ, Hunger JM, Nguyen-Cuu J, Wells C. Misclassification of cardiometabolic health when using body mass index categories in NHANES 2005–2012. International Journal of Obesity 2016;40(5):883.

[4] Linge J, Borga M, West J, Tuthill T, Miller MR, Dumitriu A, et al. Body composition profiling in the UK biobank imaging study. Obesity 2018;26(11):1785–1795.
[5] Després JP. Body fat distribution and risk of cardiovascular disease: an update. Circulation 2012;126(10):1301–1313.

[6] Kissebah AH, Vydelingum N, Murray R, Evans DJ, Kalkhoff RK, Adams PW. Relation of body fat distribution to metabolic complications of obesity. The Journal of Clinical Endocrinology & Metabolism 1982;54(2):254–260.

[7] Després JP, Moorjani S, Lupien PJ, Tremblay A, Nadeau A, Bouchard C. Regional distribution of body fat, plasma lipoproteins, and cardiovascular disease. Arteriosclerosis, Thrombosis, and Vascular Biology 1990;10(4):497–511.

[8] Després JP, Lemieux I. Abdominal obesity and metabolic syndrome. Nature 2006;444(7121):881.

[9] De Larochellière E, Côté J, Gilbert G, Bibeau K, Ross MK, Dion-Roy V, et al. Visceral/epicardial adiposity in nonobese and apparently healthy young adults: association with the cardiometabolic profile. Atherosclerosis 2014;234(1):23–29.

[10] Zhou A, Murillo H, Peng Q. Novel segmentation method for abdominal fat quantification by MRI. Journal of Magnetic Resonance Imaging 2011;34(4):852–860.

[11] Thörner G, Bertram HH, Garnov N, Peter V, Schütz T, Shang E, et al. Software for automated MRI-based quantification of abdominal fat and preliminary evaluation in morbidly obese patients. Journal of Magnetic Resonance Imaging 2013;37(5):1144–1150.

[12] Christensen AN, Larsen CT, Mandrup CM, Petersen MB, Larsen R, Conradsen K, et al. Automatic segmentation of abdominal fat in MRI-Scans, using graph-cuts and image derived energies. In: Scandinavian Conference on Image Analysis Springer; 2017. p. 109–120.

[13] Sadananthan SA, Prakash B, Leow MKS, Khoo CM, Chou H, Venkataraman K, et al. Automated segmentation of visceral and subcutaneous (deep and superficial) adipose tissues in normal and overweight men. Journal of Magnetic Resonance Imaging 2015;41(4):924–934.

[14] Mosbech TH, Pilgaard K, Vaag A, Larsen R. Automatic segmentation of abdominal adipose tissue in MRI. In: Scandinavian Conference on Image Analysis Springer; 2011. p. 501–511.

[15] Wald D, Teucher B, Dinkel J, Kaaks R, Delorme S, Boeing H, et al. Automatic quantification of subcutaneous and visceral adipose tissue from whole-body magnetic resonance images suitable for large cohort studies. Journal of Magnetic Resonance Imaging 2012;36(6):1421–1434.

[16] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 2017;39(12):2481–2495.

[17] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 3431–3440.

[18] Langner T, Hedström M, Mörwald K, Weghuber D, Forslund A, Bergsten P, et al. Fully convolutional networks for automated segmentation of abdominal adipose tissue depots in multicenter water-fat MRI. Magnetic resonance in medicine 2019;81(4):2736–2745.

[19] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention Springer; 2015. p. 234–241.

[20] Roy AG, Conjeti S, Sheet D, Katouzian A, Navab N, Wachinger C. Error corrective boosting for learning fully convolutional networks with limited data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention Springer; 2017. p. 231–239.

[21] Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on IEEE; 2017. p. 1175–1183.

[22] Estrada S, Conjeti S, Ahmad M, Navab N, Reuter M. Competition vs. Concatenation in Skip Connections of Fully Convolutional Networks. In: International Workshop on Machine Learning in Medical Imaging Springer; 2018. p. 214–222.

[23] Goodfellow IJ, Warde-Farley D, Mirza M, Courville A, Bengio Y. Maxout networks. In: Proceedings of the 30th International Conference on International Conference on Machine Learning-Volume 28 JMLR. org; 2013. p. III–1319.

[24] Liao Z, Carneiro G. A deep convolutional neural network module that promotes competition of multi-size filters. Pattern Recognition 2017;71:94–105.
[25] Breteler MM, Stöcker T, Pracht E, Brenner D, Stirnberg R. MRI in the Rhineland study: a novel protocol for population neuroimaging. Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association 2014;10(4):P92.

[26] Stöcker T. Big data: the Rhineland study. In: Proceedings of the 24th Scientific Meeting of the International Society for Magnetic Resonance in Medicine (Singapore): 2016.

[27] Jimenez-del Toro O, Müller H, Krenn M, Gruenberg K, Taha AA, Winterstein M, et al. Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL anatomy benchmarks. IEEE transactions on medical imaging 2016;35(11):2459–2475.

[28] Roy AG, Conjeti S, Navab N, Wachinger C, Initiative ADN, et al. QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. NeuroImage 2019;186:713–727.

[29] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE; 2017. p. 2261–2269.

[30] Liao Z, Carneiro G. On the importance of normalisation layers in deep learning with piecewise linear activation units. In: Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on IEEE; 2016. p. 1–8.

[31] Chollet F, et al., Keras; 2015. https://keras.io.

[32] McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological methods 1996;1(1):30.

[33] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria: 2013. http://www.r-project.org/.

[34] Kullberg J, Ahlström H, Johansson L, Frimmel H. Automated and reproducible segmentation of visceral and subcutaneous adipose tissue from abdominal MRI. International journal of obesity 2007;31(12):1806.

[35] Addeman BT, Kutty S, Perkins TG, Soliman AS, Wiens CN, McCurdy CM, et al. Validation of volumetric and single-slice MRI adipose analysis using a novel fully automated segmentation method. Journal of Magnetic Resonance Imaging 2015;41(1):233–241.

[36] Sun J, Xu B, Freeland-Graves J. Automated quantification of abdominal adiposity by magnetic resonance imaging. American Journal of Human Biology 2016;28(6):757–766.

[37] Machann J, Thamer C, Schmedt B, Stefan N, Stumvoll M, Haring HU, et al. Age and gender related effects on adipose tissue compartments of subjects with increased risk for type 2 diabetes: a whole body MRI/MRS study. Magnetic Resonance Materials in Physics, Biology and Medicine 2005;18(3):128–137.

[38] Kuk JL, Lee S, Heymsfield SB, Ross R. Waist circumference and abdominal adipose tissue distribution: influence of age and sex–. The American journal of clinical nutrition 2005;81(6):1330–1334.