IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Fully automated body composition analysis in routine CT imaging using 3D semantic segmentation convolutional neural networks

Sven Koitka[1] · Lennard Kroll[1] · Eugen Malamutmann[2] · Arzu Oezcelik[2] · Felix Nensa[1]

Abstract

Objectives Body tissue composition is a long-known biomarker with high diagnostic and prognostic value not only in cardiovascular, oncological, and orthopedic diseases but also in rehabilitation medicine or drug dosage. In this study, the aim was to develop a fully automated, reproducible, and quantitative 3D volumetry of body tissue composition from standard CT examinations of the abdomen in order to be able to offer such valuable biomarkers as part of routine clinical imaging.
Methods Therefore, an in-house dataset of 40 CTs for training and 10 CTs for testing were fully annotated on every fifth axial slice with five different semantic body regions: abdominal cavity, bones, muscle, subcutaneous tissue, and thoracic cavity. Multi-resolution U-Net 3D neural networks were employed for segmenting these body regions, followed by subclassifying adipose tissue and muscle using known Hounsfield unit limits.
Results The Sørensen Dice scores averaged over all semantic regions was 0.9553 and the intra-class correlation coefficients for subclassified tissues were above 0.99.
Conclusions Our results show that fully automated body composition analysis on routine CT imaging can provide stable biomarkers across the whole abdomen and not just on L3 slices, which is historically the reference location for analyzing body composition in the clinical routine.
Key Points
• *Our study enables fully automated body composition analysis on routine abdomen CT scans.*
• *The best segmentation models for semantic body region segmentation achieved an averaged Sørensen Dice score of 0.9553.*
• *Subclassified tissue volumes achieved intra-class correlation coefficients over 0.99.*

Keywords Abdomen · Body composition · Computer-assisted image analysis · Deep learning

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
CT	Computer tomography
GPU	Graphics processing unit
HU	Hounsfield units
L3	Third vertebra of the lumbar spine
PDF	Portable document format
SAT	Subcutaneous adipose tissue
TAT	Total adipose tissue
VAT	Visceral adipose tissue

✉ Sven Koitka
sven.koitka@uk-essen.de

[1] Institute of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Essen, Essen, Germany

[2] Department of General, Visceral and Transplantation Surgery, University Hospital Essen, Essen, Germany

Introduction

Thanks to advances in computer-aided image analysis, radiological image data are now increasingly considered a valuable source of quantitative biomarkers [1–6]. Body tissue composition is a long-known biomarker with high diagnostic and prognostic value not only in cardiovascular, oncological, and orthopedic diseases but also in rehabilitation medicine or drug dosage. As obvious and simple as a quantitative determination of tissue composition based on modern radiological sectional

imaging may seem, the actual extraction of this information in clinical routine is not feasible, since a manual assessment requires an extraordinary amount of human labor. A recent study has shown that some anthropometric measures can be estimated from simple and reproducible 2D measurements in CT using linear regression models [7]. Another study showed that a fully automated 2D segmentation of CT sectional images at the level of L3 vertebra into subcutaneous adipose tissue, muscle, viscera, and bone was possible using a 2D U-Net architecture [8]. The determination of the tissue composition at the level of L3 is often used as a reference in clinical routine to limit the amount of work required for the assessment. However, even here, this is only a rough approximation, since the inter-individual variability between patients is large and the section at the level of L3 does not necessarily have to be representative of the entire human anatomy. Other dedicated techniques for analyzing body composition using dual-energy X-ray absorptiometry or magnetic resonance imaging exist [9] but require additional potentially time-consuming or expensive procedures to be performed.

The aim of our study was therefore to develop a fully automated, reproducible, and quantitative 3D volumetry of body tissue composition from standard CT examinations of the abdomen in order to be able to offer such valuable biomarkers as part of routine clinical imaging.

Materials and methods

Dataset

A retrospective dataset was collected, consisting of 40 abdominal CTs for training and 10 abdominal CTs for testing (Table 1). The included scans were randomly selected from abdominal CT studies performed between 2015 and 2019 at the University Hospital Essen. The indication of the studies was not considered. According to the distribution of clinical studies in our department, more than 50% should have been examined for oncological indications. Each CT volume has a slice thickness of 5 mm and was reconstructed using a soft tissue convolutional reconstruction kernel. The data was annotated with five different labels: background (= outside the human body), muscle, bones, subcutaneous tissue, abdominal cavity, and thoracic cavity. For annotation, the ITK Snap [10] software (version 3.8.0) was used. Region segmentation was performed manually with a polygon tool. In order to reduce the annotation effort, every fifth slice was fully annotated. Remaining slices were marked with an ignore label, as visualized in Fig. 1. The final dataset contains 751 fully annotated slices for training and 186 for testing.

Network architectures

Many different architectural designs exist implementing semantic segmentation, some utilizing pre-trained classification networks trained on ImageNet; others are designed to be trained from scratch. For this study, two different network architectures were chosen for training, namely the commonly used U-Net 3D [11] and a more recent variant multi-resolution U-Net 3D [12]. The latter is shown in Fig. 2; however, U-Net 3D is very similar to residual path blocks replaced by identity operations and multi-resolution blocks replaced by two successive convolutions. In this case, volumetric data limits the batch size to a single example per batch due to a large memory footprint. Therefore, instance normalization [13] layers were utilized in favor of batch normalization layers [14]. In the original architectures, transposed convolutions were employed to upsample feature maps back to the original image size. However, transposed convolutions tend to generate checkerboard artifacts [15]. This is why trilinear upsampling followed by a $3 \times 3 \times 3$ convolution was used instead, which is computationally more expensive, but more stable during optimization. Additionally, different choices for the initial number of feature maps n_f are evaluated: 16, 32, and 64. After each pooling step, the number gets doubled, resulting in 256, 512, and 1024 feature maps in the lowest resolution, respectively.

Training details

The implementation of network architectures and training was done in Python using Tensorflow 2.0 [16] and the Keras API. Nvidia Titan RTX GPUs with 24-GB VRAM were used, which enable the training of more complex network architectures when using large volumetric data.

Adam [17] with decoupled weight decay regularization [18] was utilized, configured with beta_1 = 0.9, beta_2 = 0.999, eps = 1e-7, and weight decay of 1e-4. An exponentially decaying learning rate with an initial value of 1e-4, multiplied by 0.95 every 50 epochs, helped to stabilize the optimization process at the end of the training. For selecting the best model weights during training, fivefold cross-validation was used on the training set and the average dice score was monitored on the respective validation splits. Since the training dataset consists of 40 abdominal CTs, each training run was performed using 32 CTs for training and 8 CTs for validation.

During training, several data augmentations were applied in order to virtually increase the unique sample size for training a generalizable network. For example, in [11, 12, 19], it has been shown that aggressive data augmentation strategies can prevent overfitting on small sample sizes by capturing expectable variations in the data. First, random scale augmentation was applied with a scaling factor sampled uniformly between 0.8 and 1.2. Since this factor was sampled independently for both x- and y-axis, it also acts as an aspect ratio augmentation. Second,

Table 1 Patient characteristics and acquisition parameters of the collected cohort

	Training	Test
Gender	M(24), F (16)	M (5), F (5)
Age (years)	62.6 ± 9.5	65.6 ± 11.4
Contrast Agent		
i.v.	Venoues Phase	
oral	Yes (28), No (12)	Yes (9), No (1)
Convolution Kernel	I30f (24) I31f (16)	I30f (8), I31f (2)
CT Model		
Siemens SIOMATOM Definition		
AS	6	2
AS+	2	1
Edge	31	1
Flash	1	1
CTDI Phanyom Type	IEC Body Dosimentry Phantom	
CTDI Volume (mGy)	7.27 ± 3.0	8.9 ± 3.5
Data Collection Diameter (mm)	500	
Exposure (mAs)	167.2 ± 61.0	200.1 ± 65.1
Exposure Modulation Type	XYZ_EC	
Exposure Time (ms)	500	
Reconstruction Diameter (mm)	389.0 ± 33.1	402.7 ± 24.0
Revolution Time (ms)	806.2 ± 88.5	831.8 ± 1.4
Single Collimation Width (mm)	0.6 (38), 1.2 (2)	0.6 (10)
Slice Thickness (mm)	5.0	
Spiral Pitch Factor	0.6 (37), 1.2 (3)	0.6 (10)
Total Collimation Width (mm)	19.2 (6,) 38.4 (34)	19.2 (2), 38.4 (8)
Tube Current (mA)	211.4 ± 95.0	240.5 ± 78.0
Tube Voltage (kV)	100 (36), 120 (3), 140 (1)	100 (9), 120 (1)

random flipping was utilized to mirror volumes on the x-axis. Third, subvolumes of size $32 \times 256 \times 256$ were randomly cropped from the full volume with size $n \times 512 \times 512$. During inference, the same number of slices was used, but with x- and y-dimension kept unchanged, and the whole volume was processed using a sliding window approach with a 75% overlap. To improve segmentation accuracy, predictions for overlapping subvolumes were aggregated in a weighted fashion, giving the central slices more weight than the outermost.

Besides random data augmentations, additional preprocessing steps were performed before feeding the image data into the neural networks. Volumes were downscaled by factor 2 to 128×128 on the x-/y-axes, retaining a slice thickness of 5 mm on the z-axis. CT images are captured as Hounsfield units (HU), which capture fine details and allow for different interpretations depending on which transfer function is used to map HUs to a color (e.g., black/white). Normally, when using floating-point values, the typical scanner quantization of 12 bits can be stored lossless and a network should be able to process all information without any problems. In this work, multiple HU windows [−1024, 3071], [−150, 250], and [−95, 155] were applied to the 16-bit integer

Fig. 1 Exemplary annotation of an abdominal CT, with subcutaneous tissue (red), muscle (yellow), bones (blue), abdominal cavity (green), thoracic cavity (purple), and ignore regions (white)
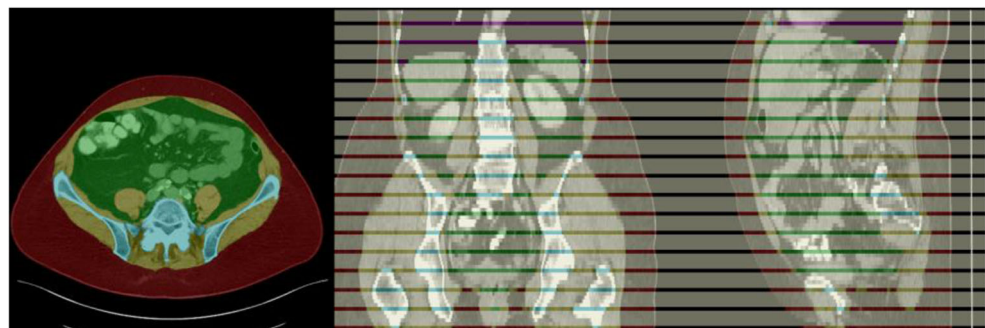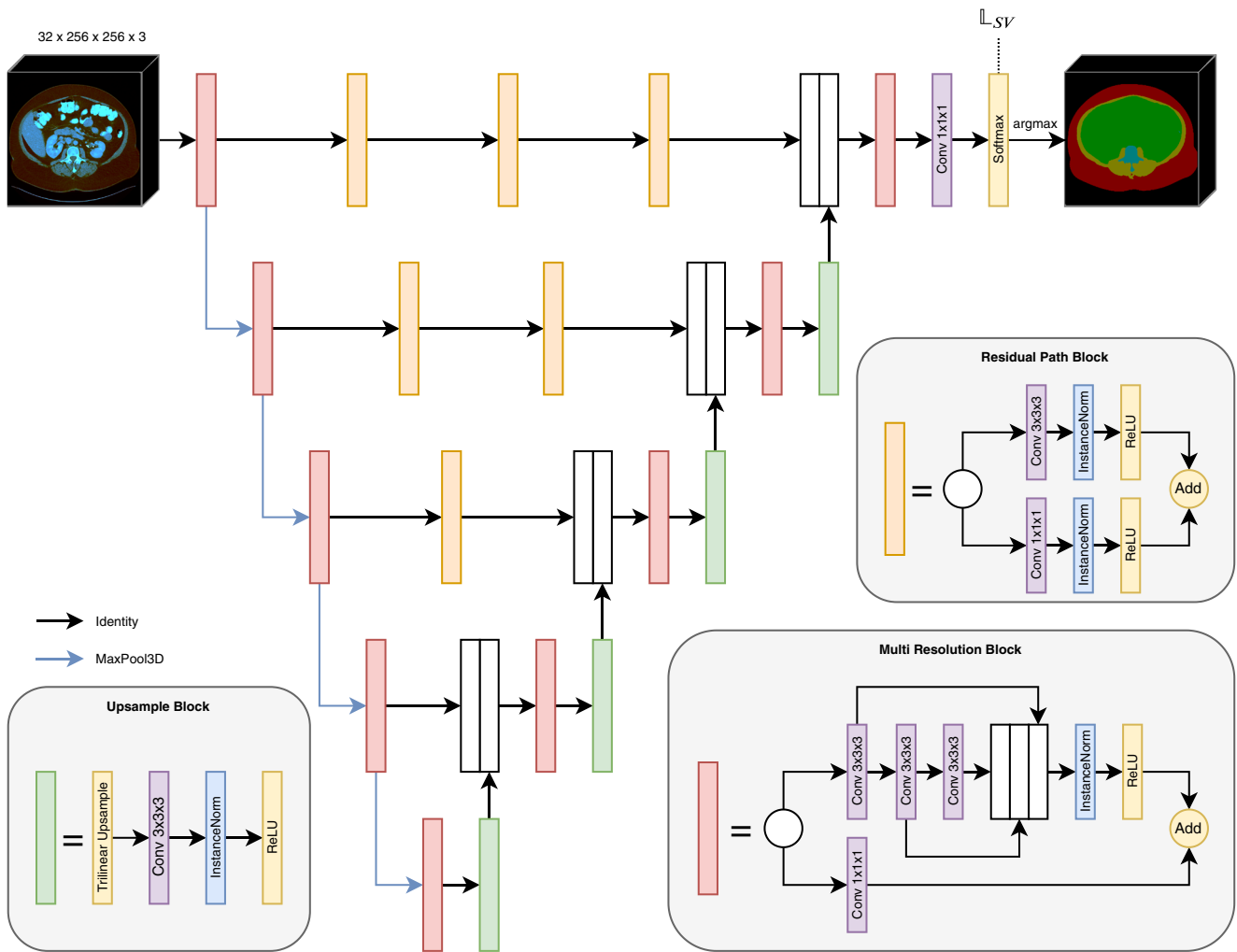
Fig. 2 Schematic overview of the multi-resolution U-Net 3D architecture (red box: multi-resolution block; orange box: residual path block; green box: upsampling block; blue arrow: max-pooling layer; black arrow: identity data flow)

data in order to map to [0, 1] with clipping outliers to the respective minimum and maximum values and stacked as channels. Lastly, the network inputs were centered around zero with a minimum value at − 1 and maximum value at + 1.

For supervision, a combination of softmax cross-entropy loss and generalized Sørensen Dice loss [20] was chosen, similar to [19]. Voxels marked with an ignore label do not contribute to the loss computation. Both losses are defined as below:

$$\mathbb{L}_{\text{XCE}} = -\frac{1}{N} \cdot \sum_{n-1}^{N} \sum_{c=1}^{C} y_{c,n} \cdot \log\left(\widehat{y}_{c,n}\right)$$

$$\mathbb{L}_{\text{Dice}} = 1.0 - \frac{1}{C-1} \cdot \sum_{c=2}^{C} \frac{\sum\limits_{n=1}^{N} 2 \cdot \widehat{y}_{c,n} \cdot y_{c,n} + \epsilon}{\sum\limits_{n=1}^{N} \widehat{y}_{c,n} + y_{c,n} + \epsilon}$$

C stands for the total number of classes, which equals six for the problem at hand. $\widehat{y}_{c,n}$ and $y_{c,n}$ represent the prediction respectively groundtruth label for class c at voxel location n.

The background class is in this work explicitly not covered by the dice loss in order to give the foreground classes more weight in the optimization process. This choice is well known for class imbalanced problems where the foreground class only covers little areas compared with the background class.

The final loss is an equally weighted combination of both losses:

$$\mathbb{L}_{\text{SV}} = 0.5 \cdot \mathbb{L}_{\text{XCE}} + 0.5 \cdot \mathbb{L}_{\text{Dice}}$$

Tissue quantification

Various materials can be extracted from a CT by thresholding the HU to a specific intensity range. For quantifying tissues, the reporting system uses a mixture of classical thresholding and modern semantic segmentation neural networks for building semantic relationships. During training, fivefold cross-

validation [21] was employed to measure the generalization performance of the selected model configuration, which in the end produced five trained model weights per configuration. For inference, those five models were used to build an ensemble system [21] by averaging the probabilities of all individual predictions, which a common method for increasing the stability and accuracy of a machine learning model. The final output of the quantification system is a report about subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), and muscle volume. Muscular tissue is identified by thresholding the HU between − 29 and 150 [22]. Adipose tissue is identified by thresholding the HU between − 190 and − 30 [22]. If an adipose voxel is within the abdominal cavity region, it is counted as VAT. If it is within the subcutaneous tissue region, it is counted as SAT. Automatically subclassified tissue volumes were validated against the tissue volumes derived from groundtruth annotations using the intraclass correlation method on a slice by slice basis.

Results

Model evaluation

As described in the "Network architecture" and "Training details" sections, two different network architectures with the varying initial number of feature maps were systematically evaluated using a fivefold cross-validation scheme on the training dataset. The results are stated in Table 2 (additional complementary evaluation metrics are available for the interested reader in Table A.1, A.2, and A.3). First of all, all networks delivered promising results with average dice scores over 0.93. Second, multi-resolution U-Net variants achieved constantly higher scores compared with their respective U-Net

counterparts. It is interesting to note that the improvements in scores were small compared with the increase in trainable parameters and thus required time to train and test the networks. A single optimization step took 294 ms, 500 ms, and 1043 ms on a NVIDIA Titan RTX for the initial feature map count of 16, 32, and 64, respectively.

For visual inspection of the ensemble segmentations, a few exemplary slices are shown in Fig. 3. Most slices show almost perfect segmentation boundaries; however, especially the ribs are problematic due to the partial volume effect. In 5-mm CTs, it is even sometimes hard for human readers to correctly assign one or the other region.

Ablation study

During model development, it was observed that the choice of HU window has an impact on optimization stability and final achieved scores. Therefore, a small ablation study was conducted in order to systematically evaluate the influence of different HU limits. Additional models were trained using the same training parameters, but only with changed input pre-processing. The results are stated in Table 3.

Increasing the HU intensity range consistently improves dice scores. By combining multiple HU windows as separate input channels, the dice scores can be even more improved to over 0.95 dice score on average on both cross-validation and test set. The lowest scores of 0.829 dice on average for cross-validation and 0.875 for the test set were achieved by an abdominal HU window ranging from − 150 to 250.

Tissue quantification report

As described in the "Tissue quantification" section, the segmentation models are intended to be used for assigning

Table 2 Evaluation for the fivefold cross-validation runs (stated as mean overall runs) and ensemble predictions on the test set. *AC*, abdominal cavity; *B*, bones; *M*, muscle; *ST*, subcutaneous tissue; *TC*, thoracic cavity

				Dice score					
Fivefold CV	Model	n_f	n_{param}	AC	B	M	ST	TC	Average
	U-Net 3D	16	5.34 M	0.9509	0.9462	0.9266	0.9432	0.8823	0.9299
		32	21.36 M	0.9669	0.9540	0.9379	0.9574	0.9336	0.9500
		64	85.43 M	0.9682	0.9561	0.9403	0.9582	0.9481	0.9542
	Multi-res U-Net 3D	16	5.82 M	0.9589	0.9484	0.9328	0.9531	0.9211	0.9429
		42	21.24 M	0.9680	0.9554	0.9399	0.9596	0.9414	0.9529
		64	85.10 M	0.9692	0.9564	0.9414	0.9605	0.9452	0.9545
Test set	U-Net 3D	16	5.34 M	0.9609	0.9340	0.9229	0.9553	0.9172	0.9381
		32	21.36 M	0.9731	0.9390	0.9309	0.9610	0.9598	0.9528
		64	85.43 M	0.9739	0.9406	0.9316	0.9623	0.9641	0.9545
	Multi-res U-Net 3D	16	5.82 M	0.9667	0.9355	0.9272	0.9593	0.9518	0.9481
		32	21.24 M	0.9736	0.9409	0.9328	0.9627	0.9629	0.9546
		64	85.10 M	0.973	0.9423	0.9334	0.9623	0.9652	0.9553

Fig. 3 Comparison of different slices, their respective groundtruth annotation, and predictions of the ensemble formed from five trained models on cross-validation splits
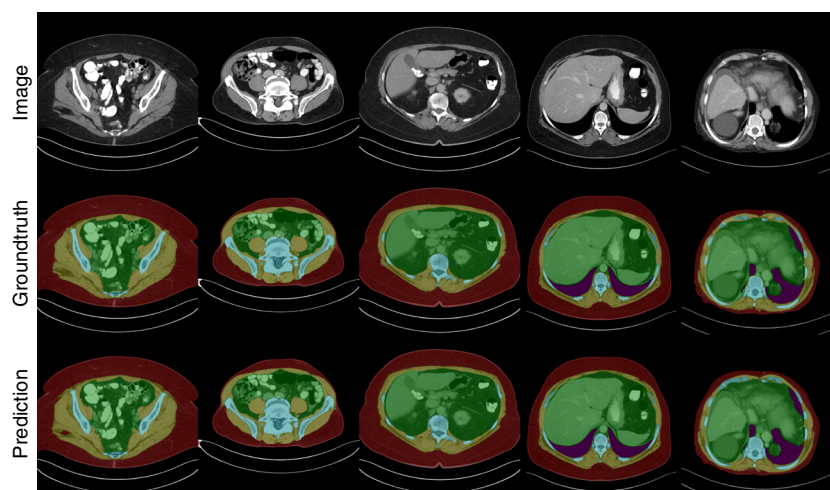
thresholded tissues to different regions, which is technically a logical conjunction. The achieved intra-class correlation coefficients for the derived SAT, VAT, and muscle volumes measured per slice on the test set are 0.999, 0.998, and 0.991, respectively ($p < 0.001$), and corresponding Bland-Altman plots are shown in Fig. 4. In order to visually inspect the quality of the tissue segmentation, a PDF report with sagittal and coronal slices is generated, in conjunction with a stacked bar plot showing the volumes of segmented muscle, SAT, and VAT per axial slice (see Fig. 5). This is only intended to give the human reader a first visual impression on the system output. For analysis, an additional table with all numeric values per slice is generated. The PDF file is encapsulated into DICOM and automatically sent back to the PACS, in order to make use of existing DICOM infrastructure.

Discussion

Our study aimed to develop a fully automated, reproducible, and quantitative 3D volumetry of body tissue composition from standard abdominal CT examinations in order to provide valuable biomarkers as part of routine clinical imaging.

Our best approach using a multi-resolution U-Net 3D with an initial feature map count of 64 was able to fully automatically segment abdominal cavity, bones, muscle, subcutaneous tissue, and thoracic cavity with a mean Sørensen Dice coefficient of 0.9553 and thus yielded excellent results. The derived tissue volumetry had intra-class correlation coefficients of over 0.99. Further experiments showed a high performance with heavily reduced parameter counts which enables considering speed/accuracy trade-offs depending on the type of application. Choosing the transfer function to map from HU to a normalized value range for feeding images into neural networks was found to have a huge impact on segmentation performance.

In a recent study, manual single-slice CT measurements were used to build linear regression models for predicting stable anthropometric measures [7]. As the authors suggest, these measures may be important as biomarkers for several diseases like e.g. sarcopenia, but could also be used where the real measurements are not available. However, manual single-

Table 3 Evaluation of multi-resolution U-Nets with $n_f = 32$ trained on different mappings from Hounsfield units to the target intensity value range of [−1, 1]. Multi-window stands for a combination of theoretical value range of 12-bit CT scans, abdomen window, and liver window. *AC*, abdominal cavity; *B*, bones; *M*, muscle; *ST*, subcutaneous tissue; *TC*, thoracic cavity

| | HU window | Dice score | | | | | |
		AC	B	M	ST	TC	Average
Fivefold CV	Multi-window	0.9680	0.9554	0.9399	0.9596	0.9414	0.9529
	[− 1024, 3071]	0.9561	0.9403	0.9217	0.9494	0.9254	0.9386
	[− 1024, 2047]	0.9533	0.9410	0.9144	0.9412	0.9303	0.9360
	[− 1024, 1023]	0.8731	0.8778	0.7875	0.6959	0.8696	0.8208
	[− 150, 250]	0.8598	0.8687	0.7632	0.7772	0.8759	0.8289
Test set	Multi-window	0.9736	0.9409	0.9328	0.9627	0.9629	0.9546
	[− 1024, 3071]	0.9682	0.9392	0.9261	0.9606	0.9532	0.9495
	[− 1024, 2047]	0.9644	0.9331	0.9174	0.9560	0.9569	0.9455
	[−1024, 1023]	0.9329	0.9002	0.8412	0.8879	0.9066	0.8938
	[− 150, 250]	0.8950	0.8997	0.8004	0.8482	0.9311	0.8749
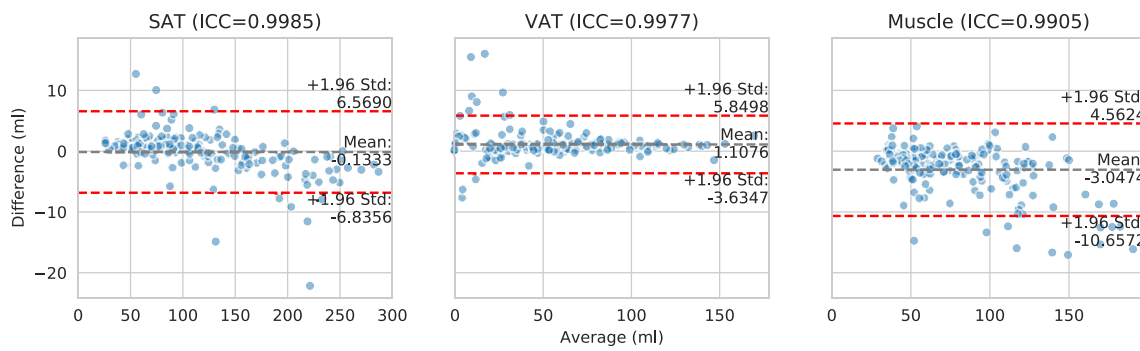
Fig. 4 Bland-Altman plot of SAT, VAT, and muscle volumetry with data points for every annotated slice in the test set

slice CT measurements are still prone to intra-patient variability and inter- and intra-rater variability. By using a fully automated approach, derived anthropometric measures from more than a single CT slice should in theory be more stable.

Fully automated analysis of body composition has been attempted many times in the past. Older methods utilize classical image processing and binary morphological operations [23–25] in order to isolate the SAT and VAT from total adipose tissue (TAT). Other studies use prior knowledge about contours and shapes and actively fit a contour or template to a given CT image [26–30]. Those methods are prone to variations in intensity values and assume certain body structures for algorithmic separation between SAT and VAT. Apart from purely CT imaging–based studies, there have been efforts to apply similar techniques to magnetic resonance imaging (MRI) [31–33]. However, MRI procedures are more cost and time expensive than CT imaging in the clinical routine. Specific MRI procedures exist for body fat assessment, but have to be performed explicitly. Our approach can be used on routine CT imaging and may be used as supplementary material for diagnosis or screening purposes.

Recently, deep learning–based methods have been proposed [8, 34]. In both studies, models were trained solely on single L3 CT slices. However, Weston et al [8] visually showed that their model was able to generalize for other abdominal slices well without being trained on such data. Nonetheless, they mentioned that extending the training and evaluation data to the whole abdomen would be beneficial for stability but also analysis capabilities. Our study uses annotated data for training and evaluation across the whole abdomen and thus is a true volumetric approach to body composition analysis. In addition, they segmented SAT and VAT directly, whereas in our study, the semantic body region was segmented and adipose tissue was subclassified using known HU thresholds.

One major disadvantage of the collected dataset is the slice thickness of 5 mm. Several tissues, materials, and potentially air can be contained within a distance of 5 mm; the resulting HU at a specific location is an average of all components. This is also known as partial volume effect and can be counteracted by using a smaller slice thickness, ideally with isometric voxel sizes. However, a reconstructed slice thickness of 5 mm is
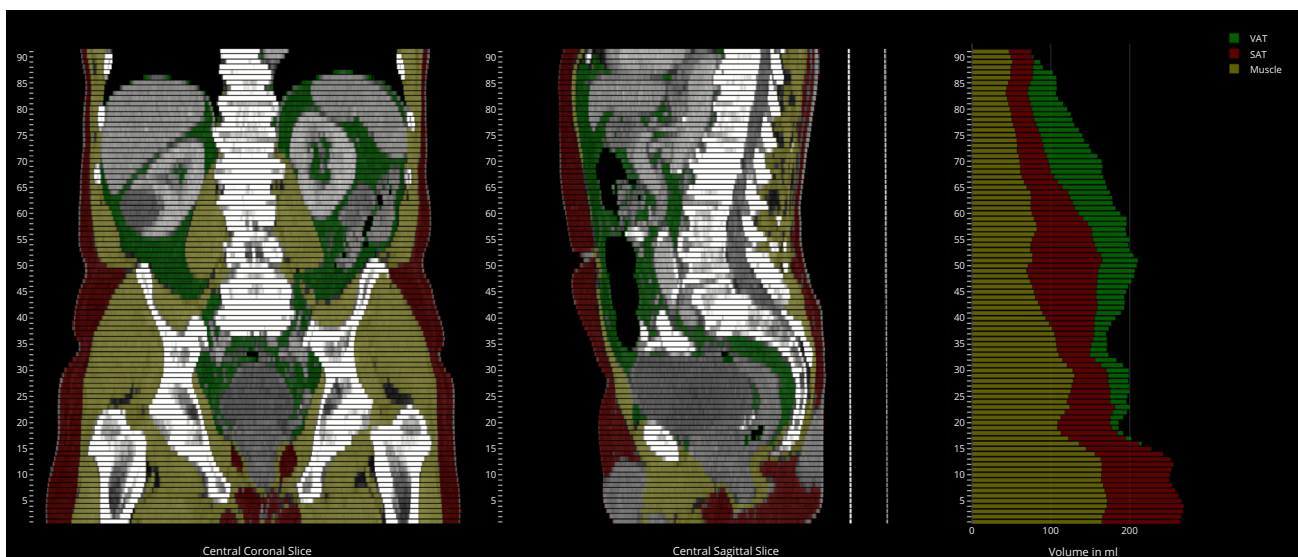

Fig. 5 Final visual report of the tissue quantification system output. SAT is shown in red, VAT is shown in green, and muscle tissue is shown yellow

Fig. 6 Beam hardening artifacts may not only harm segmentation quality (top) but also prevent accurate identification of tissues (bottom). Strong beam hardening artifacts with faults in the segmentation output (left). Beam hardening artifacts with mostly accurate segmentation, but streaking artifacts prevent accurate muscle and SAT identification (middle). No beam hardening artifacts at all, but metal foreign object detected (right)
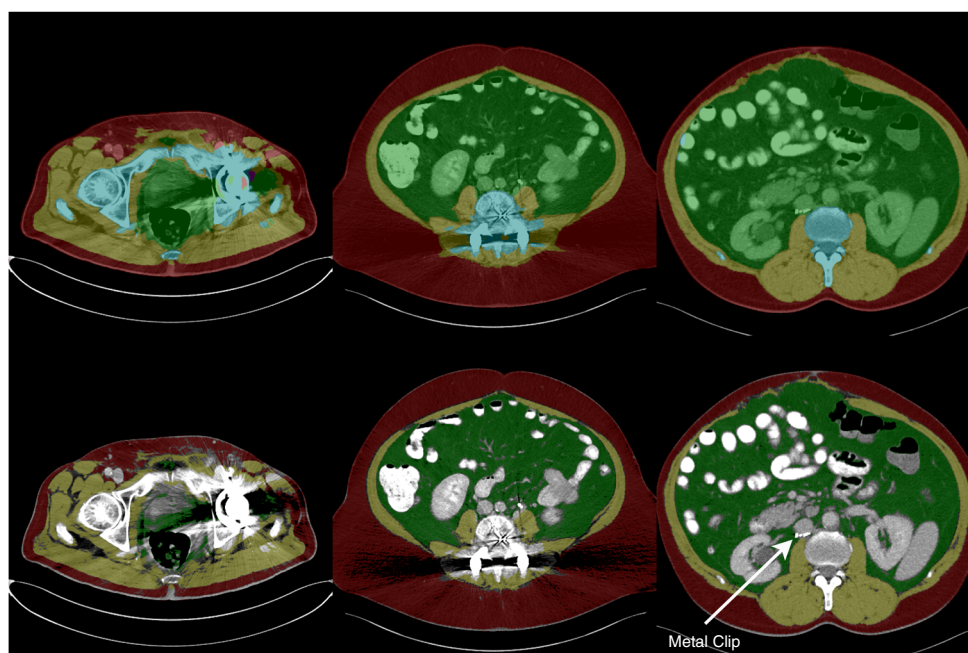
common in clinical routine CT and it is questionable whether the increased precision of calculating the tissue composition on 1-mm slices would have clinical relevance. Nevertheless, we plan to investigate the influence of thinner slices in further studies, as the reading on thin slices is becoming routine in more and more institutions.

Another limitation is the differentiation between visceral fat and fat contained within organs. Currently, every voxel with HU in the fat intensity value range, which is contained within the abdominal cavity region, is counted as VAT. However, per definition, fat cells within organs do not count as VAT and thus should be excluded from the final statistics. Public datasets like [35, 36] already exist for multi-organ semantic segmentation and could be utilized to postprocess the segmentation results from this study by masking organs in the abdominal cavity.

It is quite common to find metal foreign objects like implants in abdominal CTs and thus to encounter beam hardening artifacts. Those artifacts, depending on how strong they are, may affect the segmentation quality, as shown in Fig. 6. Even if the segmentation model is able to predict the precise boundary of the individual semantic regions, streaking and cupping artifacts make it impossible to threshold fatty or muscular tissue based on HU intensities potentially invalidating quantification reports. In a future version of our tool, we are therefore planning functionality for automatic detection and handling of image artifacts.

In future works, we plan to extend the body composition analysis system to incorporate other regions of the body as well. For example, [24] already showed an analysis of adipose tissue and muscle for thighs. Ideally, the system should be capable of analyzing the whole body in order to derive stable biomarkers. Furthermore, an external validation is required in order to prove the stability and generalizability of the developed system. This includes data from different scanners as well as a large variety of body composition cases.

Conclusion

In the present study, we presented a deep learning–based, fully automated volumetric tissue classification system for the extraction of robust biomarkers from clinical CT examinations of the abdomen. In the future, we plan to extend the system to thoracic examinations and to add important tissue classes such as pericardial adipose tissue and myocardium.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Felix Nensa.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
• retrospective
• diagnostic or prognostic study
• performed at one institution

References

1. Sam S (2018) Differential effect of subcutaneous abdominal and visceral adipose tissue on cardiometabolic risk. Horm Mol Biol Clin Invest 33. https://doi.org/10.1515/hmbci-2018-0014

2. Peterson SJ, Braunschweig CA (2016) Prevalence of sarcopenia and associated & outcomes in the clinical setting. Nutr Clin Pract 31:40–48

3. Mraz M, Haluzik M (2014) The role of adipose tissue immune cells in obesity and low- grade inflammation. J Endocrinol 222:R113–R127

4. Kent E, O'Dwyer V, Fattah C, Farah N, O'Connor C, Turner MJ (2013) Correlation between birth weight and maternal body composition. Obstet Gynecol 121:46–50

5. Hilton TN, Tuttle LJ, Bohnert KL, Mueller MJ, Sinacore DR (2008) Excessive adipose tissue infiltration in skeletal muscle in individuals with obesity, diabetes mellitus, and peripheral neuropathy: association with performance and function. Phys Ther 88:1336–1344

6. Mazzali G, Di Francesco V, Zoico E et al (2006) Interrelations between fat distribution, muscle lipid content, adipocytokines, and insulin resistance: effect of moderate weight loss in older women. Am J Clin Nutr 84:1193–1199

7. Zopfs D, Theurich S, Große Hokamp N et al (2020) Single-slice CT measurements allow for accurate assessment of sarcopenia and body composition. Eur Radiol 30:1701–1708

8. Weston AD, Korfiatis P, Kline TL et al (2019) Automated abdominal segmentation of CT scans for body composition analysis using deep learning. Radiology 290:669–679

9. Seabolt LA, Welch EB, Silver HJ (2015) Imaging methods for analyzing body composition in human obesity and cardiometabolic disease. Ann N Y Acad Sci 1353:41–59

10. Yushkevich PA, Piven J, Hazlett HC et al (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31:1116–1128

11. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3D U-net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W (eds) Medical image computing and computer-assisted intervention – MICCAI 2016. Springer International Publishing, Cham, pp 424–432. https://doi.org/10.1007/978-3-319-46723-8_49

12. Ibtehaz N, Rahman MS (2020) MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw 121:74–87

13. Ulyanov D, Vedaldi A, Lempitsky V (2017) Improved texture networks: maximizing quality and diversity in feed-forward stylization and texture synthesis. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) https://doi.org/10.1109/CVPR.2017.437

14. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Bach F, Blei D (eds) Proceedings of the 32nd international conference on machine learning. PMLR, Lille, pp 448–456

15. Odena A, Dumoulin V, Olah C (2016) Deconvolution and checkerboard artifacts. Distill. https://doi.org/10.23915/distill.00003

16. Abadi M, Barham P, Chen J, et al (2016) TensorFlow: a system for large-scale machine learning. 12th USENIX symposium on operating systems design and implementation (OSDI 16). USENIX Association, Savannah, GA, pp 265–283

17. Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. In: 3rd international conference on learning representations (ICLR). San Diego, CA, USA

18. Loshchilov I, Hutter F (2019) Decoupled weight decay regularization. In: seventh international conference on learning representations (ICLR). Ernest N. Morial Convention Center, New Orleans, USA

19. Isensee F, Petersen J, Klein A et al (2019) nnU-Net: self-adapting framework for U-net-based medical image segmentation. In: Handels H, Deserno TM, Maier A, Maier-Hein KH, Palm C, Tolxdorff T (eds) Bildverarbeitung für die Medizin 2019. Springer Fachmedien Wiesbaden, Wiesbaden, pp 22–22. https://doi.org/10.1007/978-3-658-25326-4_7

20. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M (2017) Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, JMRS T, Moradi M, Bradley A, Greenspan H, Papa JP, Madabhushi A, Nascimento JC, Cardoso JS, Belagiannis V, Lu Z (eds) Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer International Publishing, Cham, pp 240–248. https://doi.org/10.1007/978-3-319-67558-9_28

21. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press https://www.deeplearningbook.org

22. Aubrey J, Esfandiari N, Baracos VE et al (2014) Measurement of skeletal muscle radiation attenuation and basis of its biological variation. Acta Physiol (Oxf) 210:489–497

23. Kim YJ, Lee SH, Kim TY, Park JY, Choi SH, Kim KG (2013) Body fat assessment method using CT images with separation mask algorithm. J Digit Imaging 26:155–162

24. Kullberg J, Hedström A, Brandberg J et al (2017) Automated analysis of liver fat, muscle and adipose tissue distribution from CT suitable for large-scale studies. Sci Rep 7:10425

25. Mensink SD, Spliethoff JW, Belder R, Klaase JM, Bezooijen R, Slump CH (2011) Development of automated quantification of visceral and subcutaneous adipose tissue volumes from abdominal CT scans. In: M.D RMS, Ginneken B van (eds) Medical imaging 2011: computer-aided diagnosis. SPIE, pp 799–810. https://doi.org/10.1117/12.878017

26. Agarwal C, Dallal AH, Arbabshirani MR, Patel A, Moore G (2017) Unsupervised quantification of abdominal fat from CT images using Greedy Snakes. In: Styner MA, Angelini ED (eds) Medical Imaging 2017: Image processing. SPIE, pp 785–792. https://doi.org/10.1117/12.2254139

27. Ohshima S, Yamamoto S, Yamaji T et al (2008) Development of an automated 3D segmentation program for volume quantification of body fat distribution using CT. Nihon Hoshasen Gijutsu Gakkai Zasshi 64:1177–1181

28. Parikh AM, Coletta AM, Yu ZH et al (2017) Development and validation of a rapid and robust method to determine visceral

adipose tissue volume using computed tomography images. PLoS One 12:1–11

29. Pednekar A, Bandekar AN, Kakadiaris IA, Naghavi M (2005) Automatic segmentation of abdominal fat from CT data. In: 2005 seventh IEEE workshops on applications of computer vision (WACV/MOTION'05), pp 308–315. https://doi.org/10.1109/ACVMOT.2005.31

30. Popuri K, Cobzas D, Esfandiari N, Baracos V, Jägersand M (2016) Body composition assessment in axial CT images using FEM-based automatic segmentation of skeletal muscle. IEEE Trans Med Imaging 35:512–520

31. Joshi AA, Hu HH, Leahy RM, Goran MI, Nayak KS (2013) Automatic intra-subject registration-based segmentation of abdominal fat from water–fat MRI. J Magn Reson Imaging 37:423–430

32. Positano V, Gastaldelli A, Sironi AM, Santarelli MF, Lombardi M, Landini L (2004) An accurate and robust method for unsupervised assessment of abdominal fat by MRI. J Magn Reson Imaging 20:684–689

33. Zhou A, Murillo H, Peng Q (2011) Novel segmentation method for abdominal fat quantification by MRI. J Magn Reson Imaging 34:852–860

34. Bridge CP, Rosenthal M, Wright B et al (2018) Fully-automated analysis of body composition from CT in cancer patients using convolutional neural networks. In: Stoyanov D, Taylor Z, Sarikaya D, McLeod J, González Ballester MA, NCF C, Martel A, Maier-Hein L, Malpani A, Zenati MA, De Ribaupierre S, Xiongbiao L, Collins T, Reichl T, Drechsler K, Erdt M, Linguraru MG, Oyarzun Laura C, Shekhar R, Wesarg S, Celebi ME, Dana K, Halpern A (eds) OR 2.0 Context-aware operating theaters, computer assisted robotic endoscopy, clinical image-based procedures, and skin image analysis. Springer International Publishing, Cham, pp 204–213. https://doi.org/10.1007/978-3-030-01201-4_22

35. Gibson E, Giganti F, Hu Y et al (2018) Automatic multi-organ segmentation on abdominal CT with dense V-networks. IEEE Trans Med Imaging 37:1822–1834

36. Gibson E, Giganti F, Hu Y et al (2018) Multi-organ abdominal CT reference standard segmentations. Zenodo. https://doi.org/10.5281/zenodo.1169361