DON'T MISS OUR SPECIAL SESSION AT ASTRO

Performing QA Remotely in the Age of COVID

October 26, 11:45 AM EST

At home. On site. **In sync.**

SunCHECK™ enables complete, collaborative **remote QA coverage** for COVID-19 & beyond.

Click to explore:

- Advantages of a centralized Patient & Machine QA solution
- How SunCHECK eased the transition to remote work for users worldwide
- Three new ways we're simplifying Platform adoption

Go to: **sunnuclear.com/getprepared**

SUN NUCLEAR corporation

Patient Safety Starts Here

ABCNet: A new efficient 3D dense-structure network for segmentation and analysis of body tissue composition on body-torso-wide CT images

Tiange Liu
School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China

Junwen Pan
School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China
College of Intelligence and Computing, Tianjin University, Tianjin 300072, China

Drew A. Torigian
Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia 19104PA, USA

Pengfei Xu
School of Information Science and Technology, Northwest University, Xi'an 710127, China

Qiguang Miao
School of Computer Science and Technology, Xidian University, Xi'an 710126, China

Yubing Tong and Jayaram K. Udupa[a)]
Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia 19104PA, USA,

(Received 2 September 2019; revised 2 March 2020; accepted for publication 3 March 2020; published 21 April 2020)

Purpose: Quantification of body tissue composition is important for research and clinical purposes, given the association between the presence and severity of several disease conditions, such as the incidence of cardiovascular and metabolic disorders, survival after chemotherapy, etc., with the quantity and quality of body tissue composition. In this work, we aim to automatically segment four key body tissues of interest, namely subcutaneous adipose tissue, visceral adipose tissue, skeletal muscle, and skeletal structures from body-torso-wide low-dose computed tomography (CT) images.
Method: Based on the idea of residual Encoder–Decoder architecture, a novel neural network design named ABCNet is proposed. The proposed system makes full use of multiscale features from four resolution levels to improve the segmentation accuracy. This network is built on a uniform convolutional unit and its derived units, which makes the ABCNet easy to implement. Several parameter compression methods, including Bottleneck, linear increasing feature maps in Dense Blocks, and memory-efficient techniques, are employed to lighten the network while making it deeper. The strategy of dynamic soft Dice loss is introduced to optimize the network in coarse-to-fine tuning. The proposed segmentation algorithm is accurate, robust, and very efficient in terms of both time and memory.
Results: A dataset composed of 38 low-dose unenhanced CT images, with 25 male and 13 female subjects in the age range 31–83 yr and ranging from normal to overweight to obese, is utilized to evaluate ABCNet. We compare four state-of-the-art methods including DeepMedic, 3D U-Net, V-Net, Dense V-Net, against ABCNet on this dataset. We employ a shuffle-split fivefold cross-validation strategy: In each experimental group, 18, 5, and 15 CT images are randomly selected out of 38 CT image sets for training, validation, and testing, respectively. The commonly used evaluation metrics — precision, recall, and F1-score (or Dice) — are employed to measure the segmentation quality. The results show that ABCNet achieves superior performance in accuracy of segmenting body tissues from body-torso-wide low-dose CT images compared to other state-of-the-art methods, reaching 92–98% in common accuracy metrics such as F1-score. ABCNet is also time-efficient and memory-efficient. It costs about 18 h to train and an average of 12 sec to segment four tissue components from a body-torso-wide CT image, on an ordinary desktop with a single ordinary GPU.
Conclusions: Motivated by applications in body tissue composition quantification on large population groups, our goal in this paper was to create an efficient and accurate body tissue segmentation method for use on body-torso-wide CT images. The proposed ABCNet achieves peak performance in both accuracy and efficiency that seems hard to improve any more. The experiments performed demonstrate that ABCNet can be run on an ordinary desktop with a single ordinary GPU, with practical times for both training and testing, and achieves superior accuracy compared to other state-of-the-art segmentation methods for the task of body tissue composition analysis from low-dose CT images.
© *2020 American Association of Physicists in Medicine* [https://doi.org/10.1002/mp.14141]

Key words: body tissue composition, computed tomography (CT), low-dose CT, medical image segmentation, convolutional neural networks

1. INTRODUCTION

1.A. Background

Assessment of body tissue composition body-wide is important for various clinical and research applications.[1] It is widely accepted that body composition can independently influence health.[2] Bone, adipose tissue, and muscle occupy more than three-quarters of whole-body weight,[3] which makes them particularly amenable to body composition analysis. For instance, the mass and distribution of adipose tissues can significantly influence the incidence of various cardiovascular and metabolic disorders and of various cancers,[4] as well as the clinical outcome of patients with lung transplantation.[5,6] Obesity is strongly associated with the risk of acute kidney injury in trauma[7] and obstructive sleep apnea syndrome.[8] Muscle mass has been shown to correlate with important clinical outcomes such as postoperative mortality, survival after chemotherapy, and non-ventilator status.[9] The quality of bone tissue is directly related to osteoporosis as well as to noncancer death in men with prostate cancer.[10] Emerging research[11,12] also shows that the same type of tissue but distributed in different anatomical locations, such as the subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT), may have different effects on health and in the setting of disease states, which make it necessary to analyze those tissues separately. Therefore, an accurate, efficient, practical, and production-mode method of segmentation and quantification of body tissue composition has potentially far-reaching consequences. Our work aims at finding such a method to quantify the four main tissues of interest, including adipose tissues (SAT and VAT separately), skeletal muscle tissue, and skeletal structures (including cortical bone, trabecular bone, and bone marrow) on body-torso-wide CT images.

1.B. Related work

There have been many previously published approaches to quantify body tissue composition in various applications, as reviewed in Refs. [2,13-15]. Anthropometry, including body mass index (BMI), skinfold thickness, waist circumference, etc., is the easiest technique to perform and is widely used to assess obesity, as discussed.[13] However, this kind of method does not provide information about the individual contributions of each tissue type to body composition. Bioelectrical impedance analysis (BIA) and air displacement plethysmography (ADP) are two other noninvasive methods with better accuracy than anthropometric methods. Although BIA has been considered as a simple and reliable method for assessment of body composition, as discussed in Ref. [14], its accuracy has been questioned.[2,15] Air displacement plethysmography has a strict requirement for the subject to fully exhale, which requires patient coaching and which may be difficult to achieve in children and in other patients who are unable to cooperate. Regardless of the degree of accuracy, none of the above methods permit regional body tissue quantification, such as quantification of the SAT and VAT components of adipose tissue.

Medical imaging techniques, including dual-energy x-ray absorptiometry (DXA), magnetic resonance imaging (MRI), and computed tomography (CT), make the in vivo imaging of anatomic organs and tissues possible. Assessment of body composition on these 2D or 3D images is more intuitive, flexible, and accurate compared to other noninvasive methods. DXA has been regarded as the reference standard for body composition analysis.[16] However, DXA is not useful for assessment of most clinical diagnoses. Therefore, if applied for body composition quantification, it requires additional radiation exposure.[14,17] The accuracy of assessment based on DXA is also being questioned since it is difficult to evaluate 3D volume from a 2D projected image.[18] Computed tomography and MRI are routinely acquired in many clinical scenarios, and thus can be utilized opportunistically to quantify body composition with little-added healthcare cost. However, compared with CT, MRI is more expensive, slower in terms of image acquisition time, and is less widely available.[18]

Moreover, the signal intensities of cortical bone and other connective tissues such as ligaments and tendons overlap on MRI and pose challenges for accurate segmentation, which makes it difficult to quantify bone tissues accurately. Therefore, to accurately assess SAT, VAT, skeletal muscle tissue, and skeletal structure tissue, CT is an ideal modality. Therefore, our work will focus on CT images, and in particular low-dose CT images. Different from the diagnostic CT technique, the low-dose CT technique utilized in PET/CT allows for a reduced radiation exposure, which facilitates whole-body CT imaging and makes this an attractive modality for performing direct whole-body composition analysis.

Manual segmentation of tissues of interest on CT images by experienced readers is the commonly used method for quantifying body composition.[6–8] Unfortunately, it is labor-intensive, time-consuming, and prone to inter-reader variability, limiting its practical application for assessment of large numbers of datasets. Owing to these limitations, most existing reports of body composition assessment on CT use a single slice or a few slices to estimate the whole mass or volume of tissues of interest.[19–23] These methods assume a strong correlation and predictive ability between tissue properties in the selected slices and the whole body. As such, they are prone to inaccuracy since they assume that all subjects have a uniform and same body composition distribution. These assumptions are tenuous and cannot be guaranteed to be valid in most applications. Furthermore, the optimal selection of slices of interest to be assessed for particular tissues of interest is still controversial.[24] Additionally, even when the correlation is strong, it does not imply high accuracy of predicting tissue composition in the whole body.

Popuri et al.[25] proposed a finite element method to automatically recognize the region of interest (ROI) in CT images. This algorithm cannot separately quantify SAT and VAT in abdominal and thoracic regions. Irmakci et al.[26] proposed a segmentation framework based on fuzzy connectivity utilizing three different MRI contrasts, separately. Although this method is fully automated and data-driven, the SAT and VAT components cannot be separated and skeletal muscle cannot be extracted.

As compared to 2D and interactive methods, automated 3D methods have been reported recently.[27,28] Kim et al. [27] used the Convex Hull algorithm and a coordinate correction strategy to detect closed paths strictly surrounding muscle and bone. Although the accuracy reported of this method is relatively high, it is still assessed on a single 2D slice in each CT image. Hussein et al.[28] proposed an unsupervised method to segment SAT and VAT from CT images based on appearance and geometric information. The methodology reported in this study is the state of the art; however, it is focused on segmentation of only adipose tissues in the abdominal region, where other tissue types and other body regions were not assessed.

Recent advances in deep learning, in particular, convolutional neural networks (CNN), have shown dominant performance in the medical image analysis field. U-Net[29] brought the residual Encoder–Decoder technique, which was first presented by fully convolutional network (FCN)[30] for the purpose of semantic image segmentation, into medical image segmentation and added shortcut connections between the corresponding downsampling and upsampling layers. Those connections can help to transform feature information and gradients in multiple paths, during forward and backward propagations. There are some emerging studies employing U-Net[31–33] and FCN[34] for body composition tissue segmentation in CT and MR images. However, those networks still maintain 2D segmentation on slices, which makes directly applying them for capturing 3D spatial information difficult. As discussed in Ref. [31], 3D analysis of body composition is more accurate than 2D approximations. Besides the above methods, there have been a number of CNNs that were initially designed for segmenting medical images, which have an objective similar to body tissue composition assessment and can be potentially directly applied to this task.[35–38] But these networks have the disadvantages of heavy computational cost or high model complexity and their utility on body tissue composition analysis needs to be demonstrated.

1.C. Contributions

With the goal of addressing the above technical gaps, in this paper, we present a practical residual Encoder–Decoder network, named ABCNet*, for body composition quantification on body-torso-wide low-dose CT images. The contribution of this work is threefold: (a) This network is basically built on a basic convolutional unit named BasicConv as well as its variations. One particular type of BasicConv with kernel size of 1^3, called Bottleneck, is frequently employed to compress parameters; a special Dense Block, which contains multiple Dense Layers with feature map linearly increased, is utilized to make the network deeper, and an memory-efficient technique is used to further decrease the storage requirement. These designs make ABCNet become a very deep network with relatively a small number of parameters. (b) The strategy of dynamic soft Dice loss is applied to optimize the network, requiring two steps of adjustment where coarse-to-fine tuning helps to improve the segmentation accuracy. (c) ABCNet is the first 3D CNN designed to automatically and practically perform body composition analysis. In summary, the above innovations have led to a segmentation algorithm that is accurate, robust, and very efficient in terms of both time and memory, as we demonstrate in this paper.

2. MATERIALS AND METHODS

Adipose tissues (SAT and VAT separately), skeletal muscles, and skeletal structures, especially body-wide, are different from typical organs as 3D objects in that they have much larger volumes, more complex shapes (such as VAT which is amorphous), and much larger spatial extents. In our earlier work,[17] we used precise definitions of the craniocaudal extent of the body torso and the different body regions, namely thorax, abdomen, and pelvis and the SAT, VAT, skeletal muscle, and skeletal structure regions in each of them. Without such definitions, the quantification of SAT and VAT and other tissues in a standardized manner will not be possible and comparison among different methods becomes meaningless. We employ the same definitions and abbreviations in this paper as illustrated in Table I.

Figure 1 illustrates some sample low-dose axial CT slices from three different (Plv, Abd, and Thx) levels in the body torso of a subject with color overlays that depict manual segmentations of the four tissues of interest. It can be seen that different tissue types and organs appear very similar and are tightly packed within the body region. More importantly, some of them have similar or even the same intensity values. For instance, SAT and VAT are normally very similar to each other in terms of attenuation values on CT. Msl and other solid organs (such as the liver and spleen shown in the middle image in the upper row of Fig. 1) have fairly similar soft tissue attenuation values on CT. As such, these challenges for segmentation of the four tissues of interest cannot be overcome through simply utilizing intensity or texture information. The information from localization and context also needs to be utilized.

The residual Encoder–Decoder type of CNN has been proven to be effective in medical image segmentation.[30,37,39] Its success ensues largely from the wide receptive field of the downsampled feature maps. Our new network is inspired by

*ABCNet denotes "A Body Composition Network." The expression "ABC" also connotes "basic" to reflect our belief that, for parcellation of other types of tissues in a different body region (e.g., within the calvarium) or within a single anatomic object ("bone," e.g., consisting of cortical bone, trabecular bone, and bone marrow), such an approach can be utilized.

TABLE I. Definitions of body regions and tissue regions (objects) in the body torso.

Abbreviation	Definition
Thx	Thoracic region extending from 5 mm inferior to the bases of the lungs to 15 mm superior to the lung apices.
Abd	Abdominal region extending from the point of bifurcation of the abdominal aorta into common iliac arteries to the superior aspect of the liver.
Plv	Pelvic region extending from the inferior aspect of the ischial tuberosities to the point of bifurcation of the abdominal aorta into common iliac arteries.
BT	Body torso extending from the inferior aspect of the pelvic region to the superior aspect of the thoracic region.
Msl	All skeletal musculature in the body torso region.
Sk	All skeletal structures in the body torso region.
SAT	Subcutaneous adipose tissue in the body torso region.
VAT	Visceral adipose tissue (internal to Msl) in the body torso region.

those previous studies[30,37,39] but with practical improvements. ABCNet can achieve a very deep structure, a large receptive field, and therefore an accurate segmentation performance even for these objects of very complex shape and confounding appearance, but with a relatively low number of parameters. In this network, Bottleneck and feature map recomputing techniques are widely utilized.

The architecture of ABCNet is shown in Fig. 2. It consists of mainly five types of components, including normal convolution, Bottleneck, Dense Block, downsampling, and upsampling. Downsampling is utilized to create lower resolution layers of feature maps, and upsampling is used for recovering all feature maps into the original resolution. At each resolution level of the feature map, the Dense Block (which is composed of Dense Layers) is employed to extract deep features. Bottlenecks are frequently used since they can compress the feature maps and the parameters. After each upsampling, Bottlenecks are also employed to fuse the feature maps from different resolution levels. Subsequently, the softmax classifier is employed to make the final decision. The details of ABCNet and its implementation are described in the following sections.

2.A. Basic convolution block

ABCNet is built on a basic convolution block, named BasicConv. Except for the first convolutional layer at the beginning of this network and the upsampling, all other components of ABCNet are derivatives of the BasicConv. As shown in Fig. 3, those components in Fig. 2 are formed from the BasicConv with different kernel sizes or from the concatenation of derivatives of it. The BasicConv consists of four modules, including optional concatenation, batch normalization, activation, and convolution, in sequence. The structure of BasicConv is shown in Fig. 3. Although convolution is the key operation in ABCNet, other assistant modules are also indispensable. Concatenation is the first operation in each BasicConv since the input feature maps are likely obtained from multiple previous BasicConv operations such as the Bottlenecks after each upsampling and the Dense Layers in Dense Blocks, as can be observed from the multiple connections in Fig. 2. Yet, it is not always necessary to have all layers, which make concatenation an optional operation in BasicConv. Batch normalization can avoid the vanishing gradient and exploding gradient issues,[40] which make this operation necessary for BasicConv. It should be noted that both batch normalization and activation are performed before convolution. We apply pre-activation instead of post-activation, which has previously shown a superior ability to improve the performance of networks.[41,42] The employed activation function is ReLU.[43] Then, a conventional convolution layer with padding is utilized to obtain the deeper feature maps with the same size as the inputs.

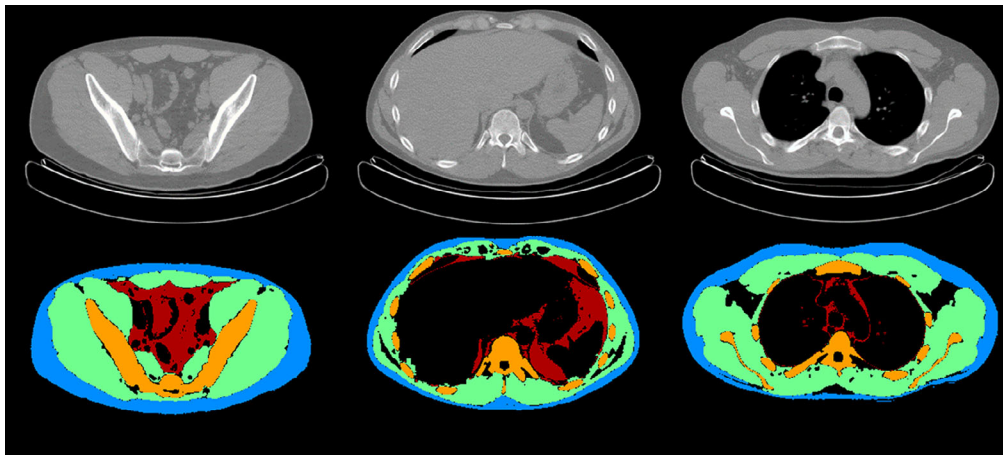

FIG. 1. From left to right are sample axial slices selected from the Plv, Abd, and Thx aspects of the body torso in one subject. Both original low-dose CT images (upper row) and manually delineated masks of the four tissue regions of interest (lower row) are shown (where blue = SAT; red = VAT; green = Msl; orange = Sk). [Color figure can be viewed at wileyonlinelibrary.com]

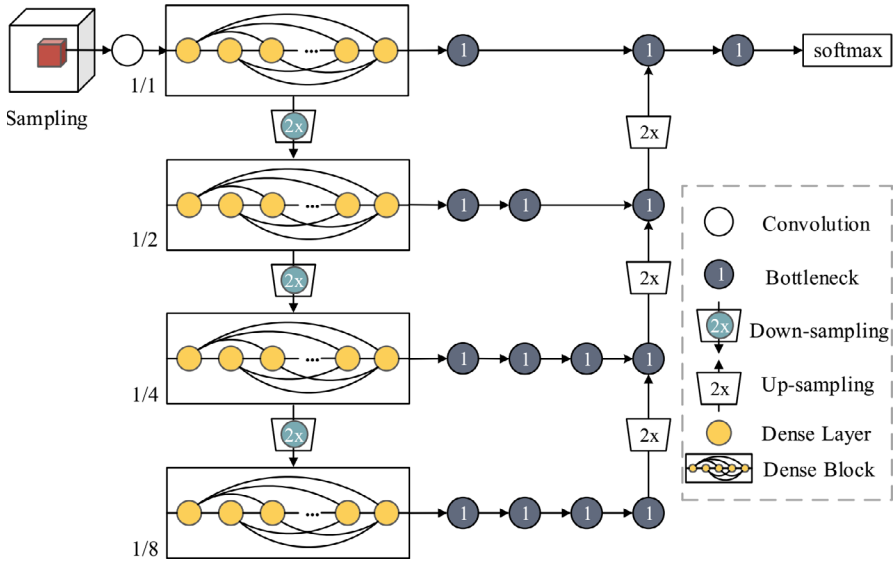

FIG. 2. The architecture of ABCNet. [Color figure can be viewed at wileyonlinelibrary.com]

2.B. Bottleneck

Regular convolution has the drawback of a large number of parameters. Especially for 3D image processing, the number of parameters will increase dramatically when the size of the convolution kernel becomes large. Hence, with only BasicConv, the network will become very complex while it gets deep. The popularly utilized Bottleneck, which is a convolution with the kernel size of 1×1, can reduce the dimensionality of feature maps.[44,45] Recently, it has been reported to help CNNs to become increasingly deeper.[46] We employ this convolution in ABCNet to compress the network, except that it is a derivative of BasicConv with a kernel of 1^3, as seen from Fig. 3. Because of the relatively complex structure, Bottleneck in ABCNet has more powerful and wider usages. It can not only reduce the number of parameters of the network but also fuse the feature maps which come from different pathways. Combined with the advantages from batch normalization and pre-activation, this Bottleneck fusion is comparable to normal size convolution. This point will be further discussed in the experimental and discussion sections. With these advantages, Bottlenecks are widely utilized in ABCNet, including embedding in the Dense Layer and down-sampling, as well as independently, as seen from Fig. 2.

2.C. Dense layer and dense block

The Dense Block idea was put forward by Huang et al. in DenseNet,[39] which is applied to alleviate the vanishing gradient problem. In a Dense Block, the prior layer connects with each of the following layers. This architecture can strengthen feature propagation and encourage feature reuse. Gibson et al.[37] utilized Dense Block in their Dense V-Net for medical image segmentation. Since the medical tomographic images are 3D, the feature maps, as well as convolution kernels, are usually much larger than in networks for 2D imagery. Thus, directly utilizing the Dense Block of Dense V-Net will pose computational and memory challenges.

In ABCNet, we present a special component to form Dense Block, which is named *Dense Layer*. As shown in Fig. 2, the proposed Dense Layer consists of a Bottleneck followed by a BasicConv with a kernel size of 3^3. Bottleneck is employed to compress the parameters and reduce feature maps (Fig. 4). It concatenates the feature maps from previous layers, reduces

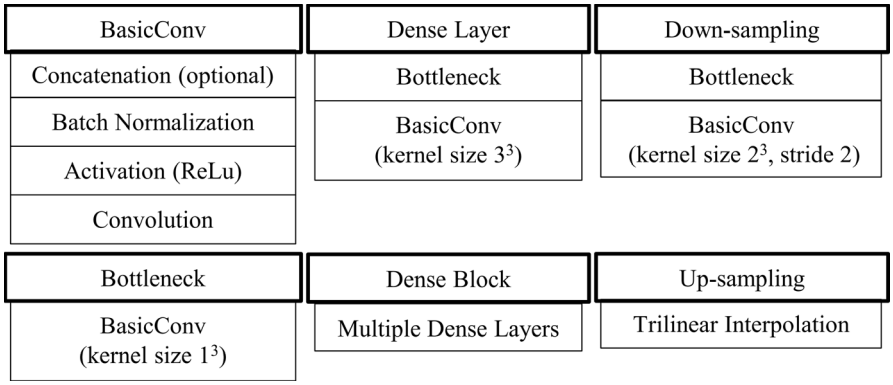

FIG. 3. Five components in ABCNet.

the number of feature maps, and then inputs them into subsequent BasicConv. Bottleneck is a crucial step since both parameters and feature maps can be compressed in the Dense Layer, which improves efficiency and memory requirements. The union structure of Dense Layer allows the Dense Block to contain more Dense Layers and thus makes the network go deeper. Different from previous strategies for going deeper,[37,45] the Bottlenecks in ABCNet contain normalization and activation operations as well, which make them no longer just simple compress tools but also feature extraction layers. As we demonstrate in our experiments, the feature extraction capability of Bottelenecks is comparable to that of BasicConvs with normal size convolution. Furthermore, this character makes Bottleneck an independent convolutional layer in ABCnet.

Assume that k_0 feature maps will input into a Dense Block and that each layer will output k feature maps, where k is known as the growth rate of this Dense Block. Thus, for the l_{th} layer in a Dense Block, $k_0 + k \times (l\text{-}1)$ feature maps will be input. If we apply only normal convolution in each layer, such as the 3^3 convolution,[37] then $[k_0 + k \times (l\text{-}1)] \times k \times 3^3$ convolutional parameters will be needed in this layer. However, when Bottleneck is used, assuming k_{bn} feature maps are being output from the Bottleneck, we need $[k_0 + k \times (l\text{-}1)] \times k_{bn} + k_{bn} \times k \times 3^3$ or $k_{bn} \times [k_0 + k \times (l\text{-}1) + k \times 3^3]$ convolutional parameters. Obviously, Bottlenecks can significantly reduce the number of parameters, especially when Dense Blocks get deeper.

2.D. Downsampling and Upsampling

Similar to other FCN methods, ABCNet also has downsampling to capture features at different levels of feature resolution, followed by upsampling to recover those downsampled feature maps into the initial resolution. Downsampling consists of a Bottleneck followed by a BasicConv with a kernel size of 2^3 and a stride of 2. Similar to the Dense Layer, the Bottleneck is applied to concatenate feature maps obtained from the upper resolution level and more importantly to compress the parameters. Different from the often used pooling operations, convolution with a stride of 2 is employed to downsample the feature maps. For upsampling, the untrained trilinear interpolation is applied, which can further reduce the number of parameters. This free-of-training technique has been reported as an alternative upsampling method and has been used in previous studies.[30,37] Benefiting from downsampling, ABCNet extracts extensive features from various resolution levels. Benefiting from upsampling, ABCNet can achieve accurate segmentation at the voxel level. The components of downsampling and upsampling are illustrated in Fig. 3.

2.E. Dynamic soft dice loss

Recently, Dice Coefficient-based loss functions have become popular in medical image segmentation.[36–37,47] We use a probabilistic Dice-based loss function for multiclass segmentation, named Soft Dice Loss (*SDL*). For C classes in the segmentation problem, the Softmax layer will output C probability maps which have the same size as that of the input patch/image. Assume the input patch/image contains N voxels. Let $p_n^c \in [0, 1]$ denotes the output probability of the n^{th} voxel belonging to class c and $g_n^c \in \{0,1\}$ denotes whether or not the n^{th} voxel belongs to class c in ground truth labels. Then, the *SDL* is defined as

$$SDL = 1 - \frac{1}{C}\left[\sum_{c=1}^{C} \frac{\left[2\sum_{n=1}^{N} p_n^c \times g_n^c\right] + \varepsilon}{\left[\sum_{n=1}^{N} p_n^c + \sum_{n=1}^{N} g_n^c\right] + \varepsilon}\right], \quad (1)$$

where ε is a small constant to avoid the numerical issue of dividing by 0. Essentially, *SDL* is equivalent to 1 minus the mean of the probabilistically evaluated Dice similarity coefficients over all segmented classes.

In ABCNet, we apply a dynamic *SDL* strategy for coarse-to-fine tuning of the network. For the first several epochs, the background is considered as an additional object class, which, in our case of four tissue types, implies $C = 5$. When the model becomes relatively steady, the background will cease to be counted as an object, which means C will be treated as 4 then on. This strategy is employed to prevent the network from going toward an undesirable state of getting trapped at a local optimum at the beginning and to assist the network to approach the optimal model in the end. If the background (including air and other uninteresting tissues) is not included at the beginning of training, then it becomes easy for the network to seek an undesirable state where all four tissues of interest are over-segmented (resulting in high false-positive rates). This occurs because of the inherent bias in the Dice metric, which favors over-segmentation over under-segmentation due to its known higher sensitivity to false negatives than false positives.[48] Therefore, we compute the *SDL* based on all five objects to roughly train the network, and then exclude the background in the *SDL* for refined training. In our implementation, the number of objects will be changed while *SDL* is lower than 0.2 (or when mean of all Dice values is > 0.8). This threshold is set based on our observation that the Dice values for one or more classes will tend to 0 when the network is trapped at a local optimum. Therefore, $SDL < 0.2$ can ensure that for none of the five classes, Dice is 0.

2.F. Implementation

Because of the large size of CT images, it is difficult for typical FCN structures to be directly implemented. Thousands of feature maps will be generated during the forward and backward propagations, which create challenges for both memory and computational complexity. Therefore, we apply patch-wise training and memory-efficient techniques in ABCNet. The learning rate is changed according to the cosine annealing method.[49]

2.F.1. Patch-wise training

Similar to V-Net (which uses patches of $128 \times 128 \times 64$ voxels) and Dense V-Net (which uses patches of $72 \times 72 \times 72$ voxels), we also apply patch-wise training with a patch size of $72 \times 72 \times 72$ voxels (which

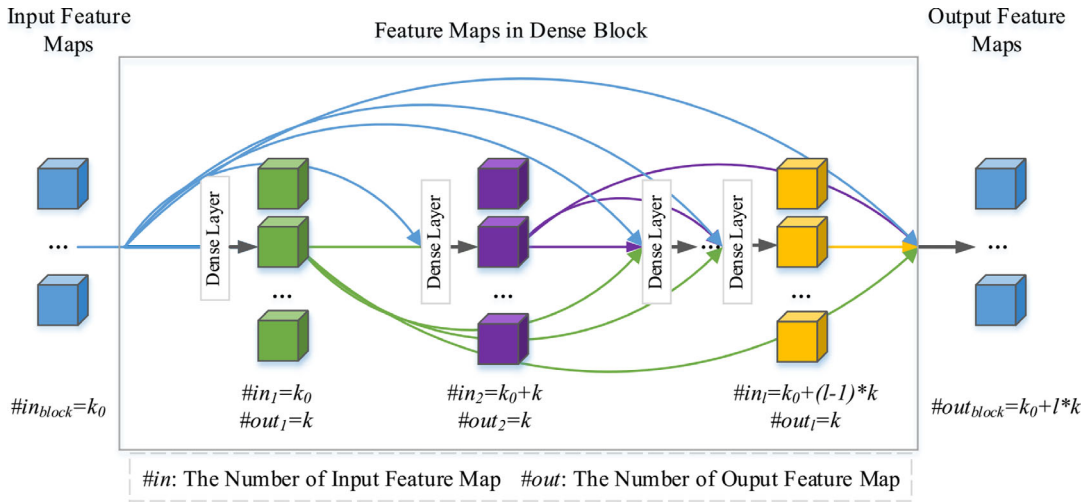

FIG. 4. Architecture of Dense Block. [Color figure can be viewed at wileyonlinelibrary.com]

equals the size of $8.64 \times 8.64 \times 28.8$ cm^3 under the spatial resolution of our dataset). We randomly harvest those patches from the training dataset on-the-fly. There is no ROI specified in the training stage, which means that all patches are randomly selected from the original CT images. Mini-batch gradient descent with a batch size of 4 is utilized to reduce the data bias effect in optimization. Similar to other residual Encoder–Decoder models, after training, ABCNet can achieve an end-to-end segmentation for much larger patches. The network is trained for 10,000 iterations (50 epochs), and the initial learning rate is 0.01, which is reduced by the cosine annealing strategy[49] for each epoch with a minimum learning rate of 0.00001.

2.F.2. Memory-efficient technique

The cost of the memory-efficient technique is a little extra computational time in order to save significantly in memory storage.[41] The main idea of this technique is to drop some of the middle results during forward propagation and to recompute them while executing backward propagation. The selected intermediate results have to satisfy two conditions, namely, they are easy to compute but expensive storage-wise. In ABCNet, the concatenation and batch normalization operations are easy to compute but consume a lot of memory to keep the feature maps for the subsequent backward propagation. Thus, we choose these two layers for dropping middle results. Assume a Dense Block has a growth rate of k and has l Dense Layers in it. Without memory conservation, at least $k \times l \times (l + 1)/2$ feature maps need to be stored in a Dense Block. However, it has to store only $k \times l$ feature maps if we adopt this frugal technique. Through efficient use of memory in this manner, we can handle much larger patches and make the network much deeper, at the cost of a little additional computational time.

2.F.3. Detailed network structure

The majority of the architecture of ABCNet is already presented in Fig. 2. The only missing details are the structures of the Dense Blocks. As we mentioned previously, localization (positional information) is very important in body tissue composition segmentation. The Dense Blocks of low-resolution feature maps are crucial to providing global locational information. Therefore, Dense Blocks at lower resolution should contain deeper Dense Layers to extract deep features. Fortunately, this is easy to realize since the lower the feature map resolution, the smaller will be the feature map size, which consequently allows more layers and feature maps to be contained in the Dense Block. In our implementation, from top to bottom, four Dense Blocks contain 6, 12, 14, and 16 Dense Layers and have growth rates of 8, 12, 14, and 16, respectively. The Bottleneck components produce $\min(input, 4k)$ feature maps to compress the network, where k is the growth rate in Dense Blocks. It means that if the input to Bottleneck is less than $4k$ (the first several layers in a Dense Block) or the Bottleneck is independent, it will output as many as input feature maps; otherwise, the Bottleneck will compress the feature maps to $4k$. Table II demonstrates the detailed layout of ABCNet through defined components shown in Fig. 2.

3. EXPERIMENTS AND RESULTS

To demonstrate the superior performance of ABCNet for body composition analysis, several state-of-the-art medical image segmentation methods, including DeepMedic,[35] V-Net,[36] and Dense V-Net,[37] are employed for comparison. Although those methods were not primarily developed for quantifying body composition, the similar application for segmentation makes them the most suitable comparison methods, since, to the best of our knowledge, there are few algorithms that can automatically segment all four tissue regions of interest from whole-body-wide low-dose CT images for assessment of body composition. Although Weston et al.[31] employed the U-Net structure to segment all four body composition tissues, only one slice at the level of the L3 vertebra was selected to quantitatively evaluate their method.

As they reported in their article through qualitative analysis, the segmentation of the 3D image is more accurate than the 2D approximation for tissue quantification. Thus, we directly employ the 3D U-Net,[38] which is known as the 3D expansion version of U-Net, as one of the compared methods.

All compared DL methods are tested through their publicly available implementations. For DeepMedic, we use the authors' version at https://biomedia.doc.ic.ac.uk/software/deepmedic/. For the other three methods, we apply their reimplements in Niftynet platform[50] at https://niftynet.io/, which are implemented based on their original presentation. Most of the parameters of these methods are at default setting according to the description in the papers, and we adjust only certain key parameters which may influence performance in our task. In Dense V-Net, the explicit spatial prior information is not applied since the four tissues are distributed body-wide, and the rough spatial prior information brings negative effect to the segmentation accuracy. In addition, we do not utilize the loss function as well as the adjustment strategy presented in Dense V-Net since the four tissues do not present extreme class imbalance. Instead, we apply the Dice loss presented in V-Net but extended to multiple classes, which was also considered in the study of Dense V-Net. In V-Net, the above loss function is also applied so that it can perform segmentation on multiple tissues. For fair comparison, we iteratively train the V-Net and Dense V-Net with the same number of patches as utilized in ABCNet, for example, 40000, which are balanced and selected on-the-fly from the training dataset randomly. For 3D-UNet and DeepMedic, we use 80,000 and 120,000, respectively, to achieve optimal performance. In our experiments, all losses for these networks remain relatively stable under the above strategies.

Recently, there have been some studies focused on particular tissue type,[28] or that have evaluated tissue mass based on specific slices,[20] which are all reported as state-of-the-art algorithms in their domains. In our current study, we also provide a quantitative comparison based on the performance reported in those studies.

In addition, we analyze the performances of different variations of ABCNet illustrated in Table II. ABCNet-Lite is a simplified ABCNet with fewer Dense Blocks as well as fewer Dense Layers in the Dense Block. In ABCNet-Rm-BN, all of the Bottlenecks in Dense Blocks are removed, which will increase the number of parameters in the network. ABCNet-Skip-BN is the network where all Bottlenecks after Dense Blocks are replaced by BasicConv with a kernel size of 3^3.

Table II. The structural details of ABCNet and its variations.

Component	ABCNet	ABCNet-Lite	ABCNet-Rm-BN	ABCNet-Skip-BN
Convolution	Conv(3^3)			
Dense Block (1)	DL × 6, k = 8	DL × 6, k = 5	BC(3^3) × 6, k = 8	DL × 6, k = 8
Downsampling (1)	BN Conv(2^3), St-ride = 2			
Dense Block (2)	DL × 12, k = 12	DL × 8, k = 8	BC (3^3) × 12, k = 12	DL × 12, k = 12
Downsampling (2)	BN Conv(2^3), Stride = 2			
Dense Block (3)	DL × 14, k = 14	DL × 10, k = 16	BC (3^3) × 14, k = 14	DL × 14, k = 14
Downsampling (3)	BN Conv(2^3), Stride = 2	–	BN Conv(2^3), Stride = 2	
Dense Block (4)	DL × 16, k = 16	–	BC (3^3) × 16, k = 16	DL × 16, k = 16
Skip Bottleneck (4)	BN × 4	–	BN × 4	BC (3^3) × 4
Upsampling (4)	Interpolation BN	–	Interpolation BN	BC(3^3)
Skip Bottleneck (3)	BN × 3			BC (3^3) × 3
Upsampling (3)	Interpolation BN			BC(3^3)
Skip Bottleneck (2)	BN × 2			BC (3^3) × 2
Upsampling (2)	Interpolation BN			BC(3^3)
Skip Bottleneck (1)	BN			BC(3^3)
Final Convolution	BN × 2			BC (3^3) × 2

DL denotes Dense Layer, BN denotes Bottleneck, BC(3^3) denotes BasicConv with kernel size 3^3, and Conv(3^3) and Conv(2^3) denote convolutional operation with kernel sizes of 3^3 and 2^3, respectively. k is the growth rate.

3.A. Datasets, evaluation metrics, and experiments

This retrospective study was conducted following approval from the Institutional Review Board at the Hospital of the University of Pennsylvania along with a Health Insurance Portability and Accountability Act waiver. The image dataset included low-dose unenhanced CT images from 38 subjects who previously underwent ^{18}F-2-fluoro-2-deoxy-D-glucose (FDG) PET/CT imaging without administration of intravenous contrast material on a 16-detector row LYSO PET/CT scanner with time-of-flight capabilities (Gemini TF, Philips Healthcare, Bothell, WA). The CT images had been acquired using a kVp of 120, an effective mAs of 50, a gantry rotation time of 0.5 msec, and a voxel size of 1.2 × 1.2 × 4 mm^3. The images were selected from our hospital patient image database by a board-certified radiologist (co-author Torigian). The patient cohort consists of 25 male and 13 female subjects in the age range 31–83 yr and in the BMI range from 17.27 to 38.28 kg/m^2, and is composed of 31 minimally abnormal subjects and 7 cancer patients. Manual ground truth segmentations of the four tissue regions were generated via CAVASS software[51] by well-trained operators and verified by the above-mentioned radiologist. The segmentation methods utilized included iterative live wire,[52] thresholding, and manual painting and correction.

In our experiments, we adopt the commonly used evaluation metrics — precision, recall, and F1-score, to measure the segmentation quality. Precision is the fraction of segmented results that constitute true positives, while recall is the fraction of ground truth that represent true positives. F1-score is the harmonic mean of precision and recall and is equivalent to the other commonly used metric called Dice Coefficient. To compare with existing body composition quantification methods,[20] mean percent prediction error (%PE) as used in that report, calculated as $(|(ground\ truth - predicted)|/|predicted|) \times 100$, where the variables represent sets and $|.|$ denotes the number of voxels in the set, was also employed in our quantitative comparison.

We employ a shuffle-split fivefold cross-validation strategy for evaluating ABCNet, as well as the comparison methods,[35–38] on our dataset. This means that the entire process, including training and testing, is repeated five times. In each experimental group, 18 and 5 CT images are randomly selected out of 38 CT images for training and validation, respectively. The remaining 15 CT images are used for testing. The results reported below are based on those 75 testing CT image sets.

3.B. Results

Sample segmentation results for ABCNet and the other three methods, as well as manual ground truth segmentations, are illustrated in Fig. 5 for one test CT image. We note that it is impossible to create perfect ground truth due to, for example, the presence of microvascular and nerve structures within the adipose regions and similarly subtle intramuscular fat within muscle regions. For creating ground truth, we first manually delineate certain major interfaces like the muscle wall or the interface between SAT and VAT regions, then apply thresholding within such defined regions to separate tissue components, and then make manual corrections as needed. Finally, the ground truth so created is checked meticulously by our expert (Torigian) to make sure that no gross and manually correctable errors remain. We believe that this is the best currently possible way to create ground truth for these tissue components. Errors if any in ground truth at such microscopic level are impossible to decipher and correct manually.

The selected slices in Fig. 5 are from Plv (first row), Abd (second row), and Thx (third row) regions in the body. SAT (blue), VAT (red), Msl (green), and Sk (orange) are together shown in this figure. Quantitative evaluations of the segmentation results are summarized in Table III. Under the same experimental strategy (shuffle-split fivefold cross validation), the segmentation performance of ABCNet's variants is also evaluated and the results are summarized in Table III. Box plots are shown in Fig. 6 to further compare the results of all methods for all four tissues.

It is difficult for us to reproduce the previous methods which estimate tissue mass based on a particular slice(s)[20] since we do not have the precise mass information of body composition for our dataset. Hence, we assume that all tissues are uniformly distributed in each voxel, and therefore, that the tissue mass estimation is equivalent to the voxel amount estimation. Based on this assumption, %PEs of segmentation results of adipose tissues achieved by ABCNet and that reported in a previous study[20] are summarized in Table IV. Furthermore, the means of F1-scores of SAT and VAT for ABCNet and Hussein et al.' s method[28] are also illustrated in this table.

4. DISCUSSION

From the sample display in Fig. 5, we may observe that, compared to other methods, there are distinct advantages of ABCNet. One advantage is the ability to retain detailed information, which can be observed from the small holes and gaps within the Msl and adipose tissues. Those details are accurately segmented, which we believe should be attributed to the high-resolution Dense Block. With direct convolution on the original CT images, all Dense Layers in the first Dense Block can receive the initial information, which makes ABCNet accurately extract those details. We can see in Fig. 5 that some small gaps are falsely labeled in the ground truth within Msl tissue, but are precisely delineated by ABCNet. Those small gaps are very hard to be accurately labeled manually because of the partial volume effect. However, ABCNet seems to correct this problem at some level. These observations demonstrate the ability of ABCNet in keeping details. Another advantage is that the location information can be fully taken into consideration, which can be particularly observed from the results of SAT and VAT. In the state-of-the-art methods shown in Fig. 5, some large areas in SAT and VAT are falsely delineated. Similar situations also occur in the Msl results of V-Net. This type of false-positive segmentation on large areas rarely occurs with ABCNet. This observation should be attributed to the lower resolution Dense Blocks. Although in the other three compared networks, the lower resolution information is also applied, none of them achieves such deep structure as that of ABCNet while keeping the network practically manageable. The particular designed components BasicConv, Bottleneck, and Dense Blocks facilitate ABCNet to embody very deep layers in the lower resolution Dense Blocks. This helps ABCNet to capture global features as well as location information.

The quantitative evaluations in Table III also demonstrate the superior performance of ABCNet. We can see that for all four tissues of interest, all three evaluation metrics are above 0.91. In particular, for SAT, all three assessments are around 0.97. The results for Sk are also excellent with all three metrics exhibiting values above 0.96. The results for the other two tissues achieve relatively less accuracy, where VAT achieves an F1-score (Dice) of 0.94 and Msl tissue achieves an F1-score of 0.92 although still excellent considering that they are very challenging regions to segment given that they are sparsely distributed within the body, have very complex shapes, and we are dealing with low-dose (and resolution) CT images compared to diagnostic quality images.

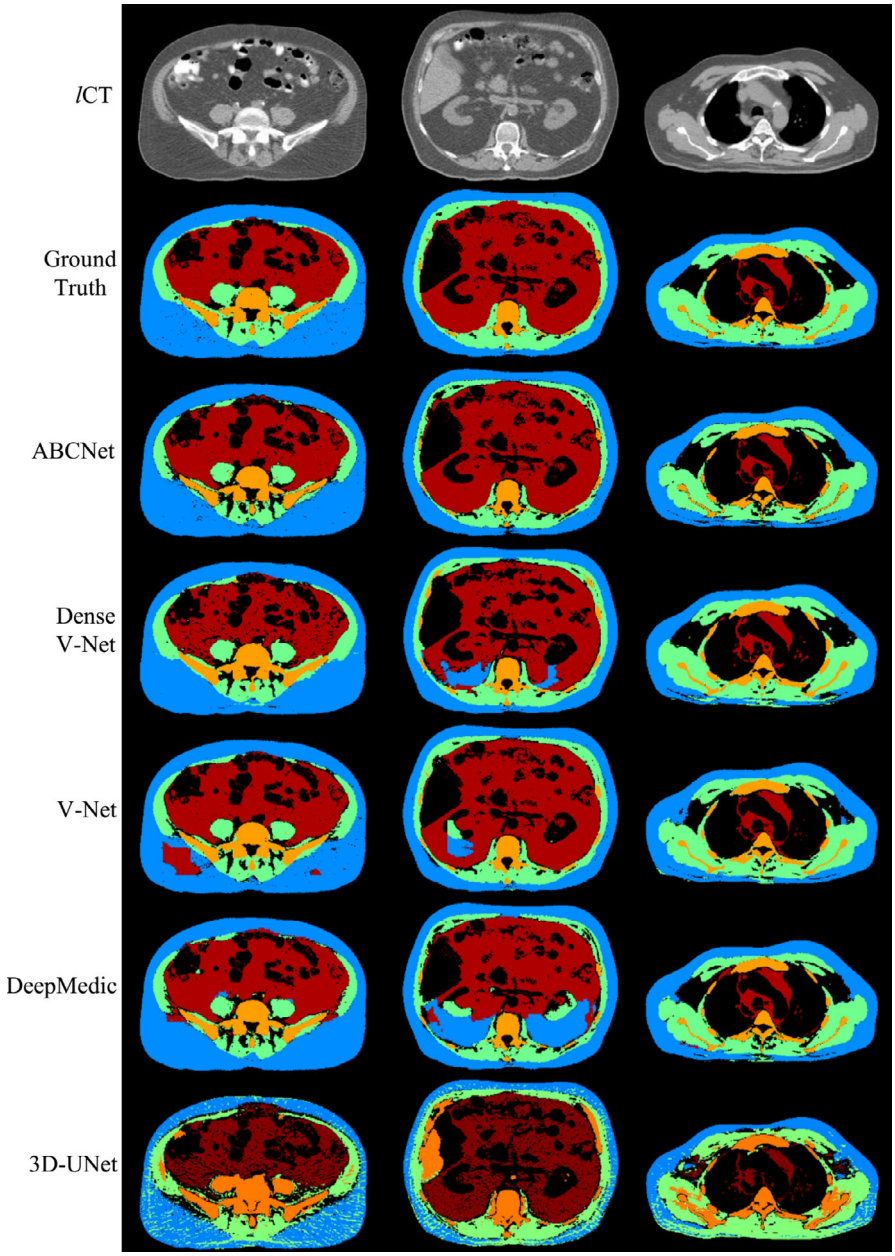

FIG. 5. Sample slices from segmentation results for the compared methods. Blue: SAT, Red: VAT, Green: Msl, Orange: Sk. [Color figure can be viewed at wileyonlinelibrary.com]

Statistically significantly improved performance of ABCNet over other state-of-the-art methods can be seen in Table III. ABCNet achieves statistically significant superior results for all F1-scores. For the other two evaluation components, except the recall of Sk, ABCNet achieves the best or one of the best performance measures. Compared with other specific body composition quantification methods, as shown in Table IV, ABCNet also achieves superior accuracy. As discussed in the previous studies,[20] the accuracy of the evaluation of adipose tissue based on particular slices was with a % PE range from 8.77 to 24.06. In ABCNet, this error for SAT and VAT achieves a mean of 1.8 and 5.1, respectively. Compared with the method designed expressly for quantifying SAT and VAT,[28] our network can obtain more accurate results along with accurately quantifying other two important tissue components of interest. Overall, ABCNet seems to be more accurate and able to handle other and more challenging body composition tissues encountered on CT images throughout the body.

Our previous method AAR-BCA[17] used a model-based strategy to model certain anatomic (some artificial and others natural) objects which best facilitated the delineation of all four component tissues. That method was very accurate in localizing objects but it had deficiencies in precise delineation, particularly of the VAT region. The F1 scores for SAT, VAT, Msl, and Sk were in the range 0.87–0.97, 0.56–

TABLE III. The mean (and standard deviation) of precision, recall, and F1-score of four segmentation algorithms.

Algorithm	SAT	VAT	Msl	Sk
Precision				
ABCNet	0.977 (0.018)	0.950 (0.045)	0.919 (0.066)	0.968 (0.025)
ABCNet-Lite	0.977 (0.017)	0.948 (0.042)	0.917 (0.063)	**0.964 (0.028)**
ABCNet-Rm-BN	**0.980 (0.016)**	0.955 (0.038)	0.923 (0.063)	0.970 (0.025)
ABCNet-Skip-BN	0.973 (0.020)	0.947 (0.040)	0.921 (0.068)	**0.971 (0.022)**
DeepMedic	**0.956 (0.031)**	**0.873 (0.092)**	**0.896 (0.065)**	**0.907 (0.064)**
V-Net	**0.965 (0.021)**	**0.926 (0.052)**	**0.890 (0.066)**	**0.927 (0.031)**
Dense V-Net	0.972 (0.020)	**0.851 (0.081)**	**0.859 (0.078)**	**0.829 (0.039)**
3D-Unet	**0.990(0.006)**	0.946(0.037)	**0.884(0.068)**	**0.941(0.037)**
Recall				
ABCNet	0.972 (0.033)	0.937 (0.035)	0.934 (0.056)	0.967 (0.027)
ABCNet-Lite	0.968 (0.028)	**0.914 (0.041)**	0.936 (0.055)	0.969 (0.030)
ABCNet-Rm-BN	0.968 (0.029)	**0.918 (0.043)**	0.937 (0.053)	0.967 (0.029)
ABCNet-Skip-BN	0.978 (0.025)	0.943 (0.037)	0.937 (0.052)	0.968 (0.032)
DeepMedic	0.968 (0.018)	**0.907 (0.047)**	**0.920 (0.063)**	0.961 (0.047)
V-Net	**0.963 (0.025)**	**0.929 (0.035)**	**0.902 (0.043)**	**0.934 (0.030)**
Dense V-Net	**0.933 (0.030)**	**0.788 (0.063)**	0.942 (0.030)	**0.974 (0.013)**
3D-Unet	**0.910(0.067)**	**0.747(0.099)**	**0.959(0.023)**	**0.979(0.022)**
F1-score				
ABCNet	0.974 (0.024)	0.942 (0.032)	0.924 (0.041)	0.967 (0.023)
ABCNet-Lite	0.973 (0.020)	**0.930 (0.036)**	0.924 (0.037)	0.965 (0.026)
ABCNet-Rm-BN	0.974 (0.021)	**0.936 (0.034)**	0.928 (0.038)	0.968 (0.024)
ABCNet-Skip-BN	0.975 (0.020)	0.943 (0.032)	0.927 (0.043)	0.968 (0.024)
DeepMedic	**0.962 (0.021)**	**0.887 (0.065)**	**0.905 (0.047)**	**0.932 (0.051)**
V-Net	**0.964 (0.020)**	**0.927 (0.037)**	**0.894 (0.036)**	**0.930 (0.028)**
Dense V-Net	**0.952 (0.025)**	**0.818 (0.070)**	**0.896 (0.046)**	**0.895 (0.027)**
3D-Unet	**0.947(0.040)**	**0.832(0.076)**	**0.918(0.041)**	**0.959(0.026)**

Boldface denotes a statistically significant difference of the results from different algorithms compared to those of ABCNet ($P < 0.05$).

0.91, 0.72–0.94, and 0.78–0.97, respectively. In the current paper, we embed the previous idea of global recognition or localization of anatomy into the design of ABCNet through a multiscale and multiresolution approach for encoding and decoding features without compromising its detailed delineation capability and thereby obtain substantially improved accuracy.

Table III also shows the assessments of the variants of ABCNet. From the F1-scores, we can see that ABCNet and its varieties basically have the same levels of accuracy. Through those evaluation values, ABCNet appears to have slightly better performance than ABCNet-Lite and ABCNet-Rm-BN but inferior accuracy compared to ABCNet-Skip-BN. Compared with ABCNet-Lite, the superior accuracy of ABCNet for VAT segmentation is statistically significant although not substantial. We believe that the deeper network of ABCNet has more ability to extract more details of such sparse objects. Although ABCNet-Skip-BN has slightly better segmentation results but are not statistically significant, it contains many more parameters in its network (Table II), which makes it much more complex in terms of both computation and storage space.

Besides having high accuracy, ABCNet is also time-efficient and memory-efficient. The above-described experiments were run on a modern desktop computer with the following specifications: 4-core Intel Xeon 3.3 GHz base CPU with 8 GB RAM and an NVIDIA-GTX-1080 GPU, running on the Linux operating system. The training time of ABCNet (for one fold) is about 18 h, and it costs an average of 12 sec to process a body-torso-wide CT image and output the four tissue components.

ABCNet has a depth of 118 layers but only 1.36 million (M) parameters. The complexity of the network is comparable to those of DeepMedic (12 layers with 0.6M parameters) and Dense V-Net (35 layers with 0.9 M), and much lower than that of V-Net (34 layers with about 71M parameters) and 3D U-Net (31 layers with 19M parameters). In Table V, we provide other information about ABCNet and its variants, including network depth, number of parameters, and segmentation time. It is worth noting that although ABCNet-Lite has fewer parameters than DeepMedic and Dense V-Net, it still has a superior performance, which can be seen from Table III. We believe that this superiority owes to the particular component — Dense Layer, which allows ABCNet to become deeper, therefore perform better, but with relatively lightweight structure.

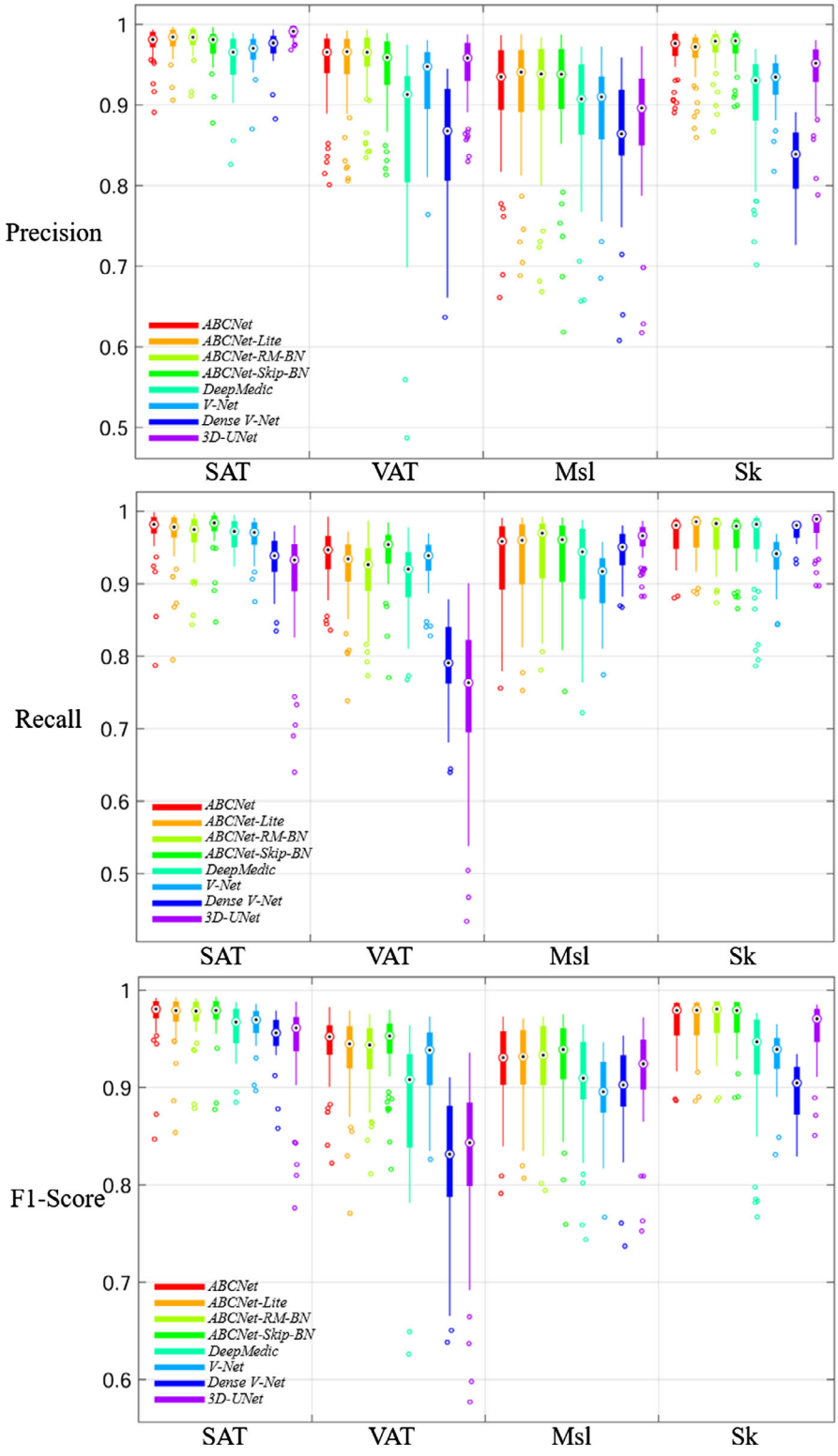

FIG. 6. Box plots of the metrics for the results of comparison summarized in Table III. The lower and upper edges of the boxes represent 25th and 75th percentiles, respectively. The median value is marked by a solid circle in the box. The hollow circle represents outliers beyond the 1.5 interquartile range. [Color figure can be viewed at wileyonlinelibrary.com]

In summary, ABCNet seems to be able to segment all major body composition tissues from CT images more accurately compared to other state-of-the-art methods. It is also a time-efficient and memory-efficient network which can be run on a normal desktop with a single ordinary GPU. It costs only a couple of hours to train the network and takes only a few seconds to segment a body-torso-wide CT image. It is a practical method that can be applied for body tissue

TABLE IV. Mean percent prediction errors (%PE) and mean of F1-scores of adipose tissue segmentation results for ABCNet and other comparable methods.

Algorithm	SAT	VAT	Adipose Tissues
	%PE		
ABCNet	0.015–9.61 (mean = 1.8)	0.007–16.78 (mean = 5.1)	0.017–14.28 (mean = 2.4)
Estimate based on Particular Slices[20]	Not reported	Not reported	8.77–24.06
	Mean of F1-score		
ABCNet	0.974	0.942	0.968
Hussein et al.' s method[28]	0.943	0.919	Not reported

composition quantification routinely on large groups of image datasets.

5. CONCLUSIONS

Motivated by applications in body tissue composition quantification on large population groups, our goal in this paper was to create an efficient and accurate body tissue segmentation method for use on body-torso-wide CT images. In this work, a new residual Encoder–Decoder neural network, named ABCNet, was proposed. ABCNet is mainly formed by a particular processing unit called BasicConv which consists of concatenation, batch normalization, activation, and convolution. Bottleneck, which is one particular type of BasicConv, is widely used in ABCNet to achieve fusion of feature maps, parameter compression, and extraction of deeper features. In each resolution level of the feature map, the Dense Block is used to extract deeper features. With the linear growth in the number of feature maps, the number of parameters and the feature map size can be efficiently controlled in Dense Blocks and therefore in all networks. A recomputing strategy is employed to reduce the memory storage requirement.

With this approach, ABCNet achieves excellent performance in both accuracy and efficiency. The experiments performed demonstrate that ABCNet can be run on an ordinary desktop with a single ordinary GPU, with practical times for both training and testing, and achieves superior accuracy compared to other state-of-the-art medical image segmentation methods for the task of body tissue composition analysis.

One limitation of this study is the small number of datasets utilized and the unicity of the data source. This was necessitated by the difficulty of manually segmenting the intricate and detailed patterns of VAT throughput the body torso especially in the thorax and abdomen and to a lesser extent also the muscle tissues. Even so, the number of studies we have employed is not out of line with the current practice dealing with similar tasks.[26] Although we believe the proposed network system is ready for production-mode use for routine body composition analysis, testing on a larger number of independent datasets is needed, which we are currently in the process of performing.

TABLE V. Network characteristics of ABCNet and its varieties.

Algorithm	Depth	Number of Parameters (million)	Segmenting Time (seconds/subject)
ABCNet	118	1.36	11.6
ABCNet-Lite	61	0.44	8.1
ABCNet-Rm-BN	70	2.27	11.8
ABCNet-Skip-BN	118	3.35	17.8

ACKNOWLEDGMENTS

The research reported in this paper is partly supported by the National Natural Science Foundation of China under grant nos. 61802335, 61973250, and 61973249; the Natural Science Foundation of Hebei Province of China under grant no. F2018203096. The research of Udupa, Tong, and Torigian is partly supported by internal research funds. The authors have no conflicts to disclose.

[a)]Author to whom correspondence should be addressed. Electronic mail: jay@mail.med.upenn.edu

REFERENCES

1. Mazzoccoli G. Body composition: where and when. *Eur J Radiol*. 2016;85:1456–1460.
2. Lemos T, Gallagher D. Current body composition measurement techniques. *Curr Opin Endocrinol Diabetes Obes*. 2017;24:310–314.
3. Duda K, Majerczak J, Nieckarz Z, Heymsfield SB, Zoladz A. Human Body Composition and Muscle Mass. *Muscle and Exercise Physiology*. Salt Lake City: Academic Press, 2019; ch. 1, ch. 1, sec. 1, 3-26.
4. Asghar A, Sheikh N. Role of immune cells in obesity induced low grade inflammation and insulin resistance. *Cell Immunol*. 2017;315:18–26.
5. Anderson MR, Udupa JK, Edwin EA, et al. Adipose tissue quantification and primary graft dysfunction after lung transplantation: The Lung Transplant Body Composition Study. *J Heart Lung Transpl*, 2019;38:1246–1256.
6. Anderson MR, Kolaitis N, Kukreja J, et al. A non-linear relationship between visceral adipose tissue and frailty in adult lung transplant candidates. *Am J Transpl*. 2019;19:3155–3161.
7. Shashaty MGS, Kalkan E, Bellamy Scarlett L, et al. Computed tomography-defined abdominal adiposity is associated with acute kidney injury in critically ill trauma patients. *Crit Care Med*. 2014;42:1619–1628.
8. Tong Y, Udupa JK, Sin S, et al. MR image analytics to characterize upper airway structure in obese children with obstructive sleep apnea syndrome. *PLoS ONE*. 2016;11:e0159327.
9. Fuchs G, Chretien YR, Mario J, et al. Quantifying the effect of slice thickness, intravenous contrast and tube current on muscle segmentation: Implications for body composition analysis. *Eur Radiol*. 2018;28:1–9.
10. Mcdonald AM, Swain TA, Mayhew DL, et al. CT measures of bone mineral density and muscle mass can be used to predict noncancer death in men with prostate cancer. *Radiology*. 2017;282:475–483.
11. Kershaw EE, Flier JS. Adipose tissue as an endocrine organ[J]. *J Clin Endocrinol Metab*. 2004;89:2548–2556.
12. Fox CS, Massaro JM, Hoffmann U, et al. Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study. *Circulation*. 2007;116:39–48.
13. Duren DL, Sherwood RJ, Czerwinski SA, et al. Body composition methods: comparisons and interpretation. *J. Diabetes Sci Technol*. 2008;2:1139–1146.
14. Achamrah N, Jesuus P, Grigioni S, et al. Validity of predictive equations for resting energy expenditure developed for obese patients: impact of body composition method. *Nutrients*. 2018;10:63.

15. Tewari N, Awad S, Macdonald IA, et al. A comparison of three methods to assess body composition. *Nutrition*. 2018;47:1–5.
16. Mourtzakis KMDSM. A critical evaluation of body composition modalities used to assess adipose and skeletal muscle tissue in cancer. *Appl Physiol Nutr Metab*. 2012;37:811–821.
17. Liu T, Udupa JK, Miao Q, et al. Quantification of body-torso-wide tissue composition on low-dose CT images via automatic anatomy recognition. *Med Phys*. 2019;46:1272–1285.
18. Martínez-Martínez F, Kybic J, Lambert L, Mecková Z. Fully automated classification of bone marrow infiltration in low-dose CT of patients with multiple myeloma based on probabilistic density model and supervised learning. *Comput Biol Med*. 2016;71:57–66.
19. Mourtzakis M, Prado CMM, Lieffers JR, et al. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. *Appl Physiol Nutr Metab*. 2008;33:997–1006.
20. Jeanson AL, Dupej J, Villa C, et al. Body composition estimation from selected slices: equations computed from a new semi-automatic thresholding method developed on whole-body CT scans. *Peerj*. 2017;5:e3302.
21. Cheng X, Zhang Y, Wang C, et al. The optimal anatomic site for a single slice to estimate the total volume of visceral adipose tissue by using the quantitative computed tomography (QCT) in Chinese population. *Eur J Clin Nutr*. 2018;72:1567–1575.
22. Srikumar T, Siegel EM, Gu Y, et al. Semiautomated measure of abdominal adiposity using computed tomography scan analysis. *J Surg Res*. 2019;237:12–21.
23. Kullberg J, Hedström A, Brandberg J, et al. Automated analysis of liver fat, muscle and adipose tissue distribution from CT suitable for large-scale studies. *Sci Rep*. 2017;7:10425.
24. Tong Y, Udupa JK, Torigian DA. Optimization of abdominal fat quantification on CT imaging through use of standardized anatomic space: a novel approach. *Med Phys*. 2014;41:063501.
25. Popuri K, Cobzas D, Esfandiari N, et al. Body composition assessment in axial CT images using FEM-based automatic segmentation of skeletal muscle. *IEEE Trans Med Imaging*. 2016;35:512–520.
26. Irmakci I, Hussein S, Savran A, et al. A novel extension to fuzzy connectivity for body composition analysis: applications in thigh, brain, and whole body tissue segmentation. *IEEE Trans Biomed Eng*. 2019;66:1069–1081.
27. Kim YJ, Park JW, Kim JW, et al. Computerized automated quantification of subcutaneous and visceral adipose tissue from computed tomography scans: development and validation study. *JMIR Med Inf*. 2016;4:e2.
28. Hussein S, Green A, Watane A, et al. Automatic segmentation and quantification of white and brown adipose tissues from PET/CT Scans. *IEEE Trans Med Imaging*. 2016;36:734–744.
29. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Proc MICCAI*. 2015; 234-241.
30. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2014;39:640–651.
31. Weston AD, Korfatis P, Kline TL, et al. Automated abdominal segmentation of CT scans for body composition analysis using deep learning. *Radiology*. 2019;290:669–679.
32. Shen N, Li X, Zheng S, et al. Automated and Accurate Quantification of Subcutaneous and Visceral Adipose Tissue from Magnetic Resonance Imaging based on Machine Learning. *Magn Reson Imaging*, to be published.
33. Bridge CP, Rosenthal M, Wright B, et al.Fully-Automated Analysis of Body Composition from CT in Cancer Patients Using Convolutional Neural Networks. in Proc. Workshop MICCAI. 2018; 204–213.
34. Lee H, Troschel FM, Tajmir S, et al. Pixel-level deep segmentation: artificial intelligence quantifies muscle on computed tomography for body morphometric analysis. *J Digital Imaging*. 2017;30:487–498.
35. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*. 2016;36:61–78.
36. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. in Proc 3DV. 2016; 565–571.
37. Gibson E, Giganti F, Hu Y, et al. Automatic multi-organ segmentation on abdominal CT with dense V-networks. *IEEE Trans Med Imaging*. 2018;37:1822–1834.
38. Çiçek Özgün, Abdulkadir A, Lienkamp SS, et al.3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. in Proc MICCAI. 2016; 424–432.
39. Huang G, Liu Z, Weinberger KQ, et al.Densely connected convolutional networks. in Proc. CVPR. 2017; 2261–2269.
40. Ioffe S, Batch Szegedy C.Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. in Proc ICML. 2015, 448–456.
41. Pleiss G, Chen D, Huang G, et al.Memory-Efficient Implementation of DenseNets. 2017, arXiv: 1707.06990v1, [online]. Available: https://arxiv.org/abs/1707.06990.
42. He K, Zhang X, Ren S, et al.Identity Mappings in Deep Residual Networks. in Proc ECCV. 2016; 630–645.
43. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. in *Proc AISTATS*. 2011; 315–323.
44. Lin M, Chen Q, Yan S.Network in Network. 2015, arXiv:1312.4400, [online]. Available: https://arxiv.org/abs/1312.4400.
45. Szegedy C, Liu W, Jia Y, et al. Deeper with Convolutions.in *Proc CVPR*. 2015;1–9.
46. Sze V, Chen YH, Yang TJ, et al. Efficient processing of deep neural networks: a tutorial and survey. *Proc. IEEE*. 2017;105:2295–2329.
47. Sudre CH, Li W, Vercauteren T, et al.Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. in *Proc. Workshops DLMIA and ML-CDS*. 2017; 240–248.
48. Cappabianco FAM, Miranda PAV, Udupa JK. A critical analysis of the methods of evaluating MRI brain segmentation algorithms, in *Proc ICIP*. 2017; 3894–3898.
49. Loshchilov I, Hutter FSGDR.Stochastic gradient descent with warm restarts. in *Proc ICLR*. 2017.
50. Gibson E, Li W, Sudre C, et al. NiftyNet: a deep-learning platform for medical imaging. *Comput Meth Prog Bio*. 2018;158:113–122.
51. Grevera G, Udupa JK, Odhner D, et al. CAVASS: A computer-assisted visualization and analysis software system. *J Digital Imaging*. 2007;20:101–118.
52. Falcão AX, Udupa JK, Samarasekera S, et al. User-steered image segmentation paradigms: live wire and live lane. *Graph Models Image Proc*. 1998;60:233–260.