Radiological Physics and Technology (2019) 12:210–215
https://doi.org/10.1007/s12194-019-00512-y

Automated segmentation of 2D low-dose CT images of the psoas-major muscle using deep convolutional neural networks

Fumio Hashimoto[1] · **Akihiro Kakimoto**[1] · **Nozomi Ota**[2] · **Shigeru Ito**[2,3] · **Sadahiko Nishizawa**[3]

Received: 16 January 2019 / Revised: 26 March 2019 / Accepted: 28 March 2019 / Published online: 1 April 2019

Abstract
The psoas-major muscle has been reported as a predictive factor of sarcopenia. The cross-sectional area (CSA) of the psoas-major muscle in axial images has been indicated to correlate well with the whole-body skeletal muscle mass. In this study, we evaluated the segmentation accuracy of low-dose X-ray computed tomography (CT) images of the psoas-major muscle using the U-Net convolutional neural network, which is a deep-learning technique. Deep learning has been recently known to outperform conventional image-segmentation techniques. We used fivefold cross validation to validate the segmentation performance ($n = 100$) of the psoas-major muscle. For the intersection over union and CSA ratio, segmentation accuracies of 86.0 and 103.1%, respectively, were achieved. These results suggest that the U-Net network is competitive compared with the previous methods. Therefore, the proposed technique is useful for segmenting the psoas-major muscle even in low-dose CT images.

Keywords Psoas-major muscle · Deep learning · Convolutional neural networks · X-ray computed tomography · Automated segmentation · Sarcopenia

1 Introduction

The skeletal muscle mass of a living body is known to be an important factor for maintaining the body's physical function. In particular, the aging of the psoas-major muscle affects the walking speed of a person [1]. In addition, the psoas-major muscle has been reported as a predictive factor of sarcopenia [2]. To measure the muscle mass, X-ray computed tomography (CT) and magnetic resonance imaging (MRI) are often used [3]. Furthermore, various studies have indicated that the cross-sectional area (CSA) of the psoas-major muscle in axial images correlates well with the whole-body skeletal muscle mass [4–6]. As the manual segmentation by specialists is a time-consuming task, a few studies have reported methods for the automated segmentation of this muscle [7–9].

Meesters et al. [7] proposed a multi-atlas-based segmentation method with weighted decision fusion and nonrigid registration. Kamiya et al. [8] proposed a method using a 3D-shape model with an approximated function obtained from test cases. This method can estimate the spindle-like shape of a muscle using a quadratic function with two parameters based on an anatomic centerline connecting two landmarks corresponding to the origin and insertion of the muscle. Inoue et al. [9] proposed a higher order graph-cut method, which enables the utilization of prior knowledge regarding anatomical shapes. Table 1 presents the accuracy of the segmented area in these studies in terms of the intersection over union (IoU), which is also known as the Jaccard similarity coefficient [10].

Deep learning has recently been shown to outperform conventional image-segmentation techniques [10–13]. For example, almost all state-of-the-art approaches for semantic image segmentation have incorporated deep learning through the use of fully convolutional networks (FCNs) [10], SegNet [11], U-Net [12], and pyramid scene-parsing networks [13]. In particular, U-Net is fast and provides the exact segmentation of images. This architecture won the Dental

✉ Fumio Hashimoto
fumio.hashimoto@crl.hpk.co.jp

[1] Central Research Laboratory, Hamamatsu Photonics K.K., Hamamatsu 434-8601, Japan

[2] Global Strategic Challenge Center, Hamamatsu Photonics K.K., Hamamatsu 434-8601, Japan

[3] Hamamatsu Medical Imaging Center, Hamamatsu Medical Photonics Foundation, Hamamatsu 434-8601, Japan

Table 1 Accuracy of the segmented area in terms of the intersection over union (IoU) in previous studies

	Meesters et al. [7]	Kamiya et al. [8]	Inoue et al. [9]
Left	63.4%	–	75.4%
Right	68.6%	–	77.5%
Average	66.0%	72.3%	76.5%

X-Ray Image-Segmentation Challenge and Cell Tracking Challenges at the *International Symposium on BIOMEDICAL IMAGING* (ISBI) in 2015.

In addition, many research groups have studied the application of deep learning to medical image segmentation [14–20]. Zhou et al. [17] presented automatic multiple-organ segmentation in 3D CT images using FCNs. Lee et al. [18] demonstrated the real-time segmentation of skeletal muscles in an axial CT image using FCNs. Ghosh et al. [19] proposed a 3D-segmentation method from MRI images using convolutional neural networks. The method outperformed the conventional model-based approach. Kamiya et al. [20] performed the segmentation of the skeletal muscle and its adhesion area using FCN.

To the best of our knowledge, the segmentation of CT images of the psoas-major muscle through deep learning has not yet been studied. In this paper, we present an automated segmentation technique using U-Net convolutional neural networks. This technique was applied to noisy low-dose CT images obtained for the measurement of visceral fat as a screening test. In addition, we evaluated the IoU and CSA ratio (CSAR), which is the ratio of the CSA of the ground truth to the measured CSA using the proposed method, and showed the effectiveness of the proposed technique.

2 Materials and methods

2.1 Image data set

The current study was approved by the Institutional Review Board of Hamamatsu Medical Photonics Foundation, and written informed consent was obtained from each participant after a detailed explanation of the study (number 111). One hundred participants who underwent medical checkups for cancer screening were enrolled in this study. Among them, 38 "normal" subjects were without lumbar lesions (10 males and 28 females; mean age: 51.0 ± 5.8), and 62 "abnormal" subjects had degenerative lumbar diseases (26 males and 36 females; mean age: 53.5 ± 7.0). One-shot low-dose CT scans for the measurement of visceral fat areas at the L4 level were performed on a LightSpeed Ultra (GE Healthcare), a Biograph mCT (Siemens), or an Aquilion (Toshiba). The X-ray tube current and exposure time product was from 10 to 32 mAs. Full resolution 2D CT images, 512×512 voxels, with DICOM format, were used. On each DICOM image, a segmented-mask image of the psoas-major muscle was created manually by a radiological technician with 10 years of experience in radiation photography and medical image analysis, and a radiologist with 35 years of experience in radiologic interpretations. First, the radiological technician set a segmented-mask image of the psoas-major muscle and superimposed it on an original DICOM image for each subject. Second, the radiologist checked the segmented and superimposed images. Two operators then modified the segmented image in detail. The mean CSAR and IoU between the segmented images of the above-mentioned procedures and segmented images set by another radiological technician, with 15 years of experience in radiation photography, were 102 and 91%, respectively. The segmented images were used to obtain the ground truths during training and testing.

We resized the CT and segmented images from 512×512 to 256×256 voxels for use as acceptable input for the U-Net using a bilinear interpolation. In addition, the training of the U-Net required the preparation of a sufficient amount of training data, because insufficient training data can lead to overfitting. To reduce this overfitting problem, we performed data augmentation for a total of 14 times (i.e., four, two, and eight times by rotating, zooming, and translating the original image, respectively).

2.2 Network architecture

Various state-of-the-art deep-learning algorithms have been validated for image-segmentation applications. We chose to develop our model for the segmentation of the psoas-major muscle based on the U-Net architecture, because this architecture can recognize structural edges, such as soft tissues, in X-ray CTs. Figure 1 shows the architecture of the U-Net network used for segmenting the image of the psoas-major muscle in this study.

The architecture consists of an encoding path (left side) and a decoding path (right side). The encoding path follows the typical architecture of a convolutional neural network comprising the repeated application of two 5×5 convolutions, each followed by a batch normalization and leaky rectified linear unit (LReLU), as well as a 2×2 max pooling operation for downsampling. In addition, at each downsampling step, the number of feature channels is doubled. The decoding path consists of a 5×5 deconvolution, upsampling, concatenation with the corresponding cropped feature map from the encoding path and two 5×5 convolutions, each followed by a batch normalization and LReLU. Moreover, we activated the output layer through sigmoid, and a dice similarity coefficient loss (DSC) ($\mathrm{DSC} = 1 - 2|A \cap B|/|A| + |B|$) was used as the loss function in this study. Here, A and B

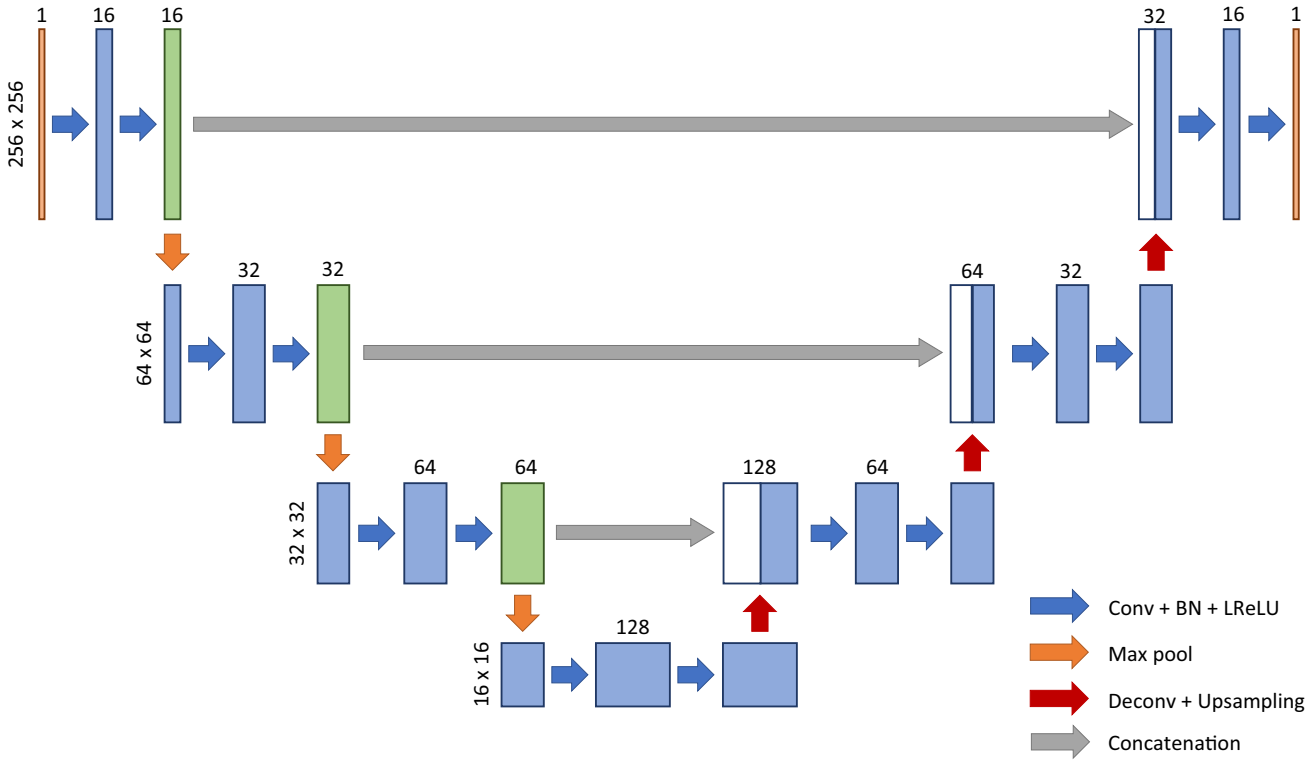

Fig. 1 Architecture of the U-Net used for segmenting the image of the psoas-major muscle. The number of channels is denoted on the top of the box. The voxel size is provided at the left of the figure. The arrows denote the different operations

represent the muscle regions of the ground truth and segmentation results, respectively.

The U-Net was run on a computer with Ubuntu 16.04, a graphic processing unit (NVIDIA GeForce GTX TITAN X with 12 GB of memory), Tensorflow 1.3 [21], and Keras 2.0.8 [22]. The input data were standardized by removing the mean and scaling to unit variance. The number of epochs was 200, and Adam [23] was used as the optimization function. A mini-batch size of 10 images and a learning rate of 10^{-3} were used as the training parameters.

3 Results

We validated the performance of the proposed method for segmentation of the psoas-major muscle through fivefold cross validation. The training and test data sets were randomly divided into five sets. The original number of images in each data set is shown in Table 2, and 280 augmented images were generated for each set.

Figure 2 shows sample images of correctly and incorrectly segmented muscles. The yellow, red, and green regions indicate the regions of overlap between the ground truth and segmentation results, areas of overextraction, and areas of under extraction, respectively.

Table 2 Number of images in each data set for fivefold cross validation

	Set 1	Set 2	Set 3	Set 4	Set 5
Normal	10	8	5	7	8
Abnormal	10	12	15	13	12

Subjects without lumbar lesions and with degenerative lumbar diseases were classified as "normal" and "abnormal", respectively

The images in which the muscle was close to the intestinal tract often resulted in a segmentation error. The segmentation accuracy was calculated between the segmentation results and ground truth for the psoas-major muscle using the IoU (IoU $= A \cap B / A \cup B$) and CSAR, as shown in Fig. 3. The average IoUs were 86.6 and 85.6% for the right and left majors, respectively, and the average IoU of the right and left majors was 86.0%. The average CSARs were 103.6 and 102.9% for the right and left majors, respectively, and the average CSAR of the right and left majors was 103.1%. The segmentation accuracies for "normal" and "abnormal" are shown in Fig. 4. The average IoUs were 87.0 and 85.5% for "normal" and "abnormal", respectively. The average CSARs were 101.8 and 104.1% for "normal" and "abnormal", respectively.

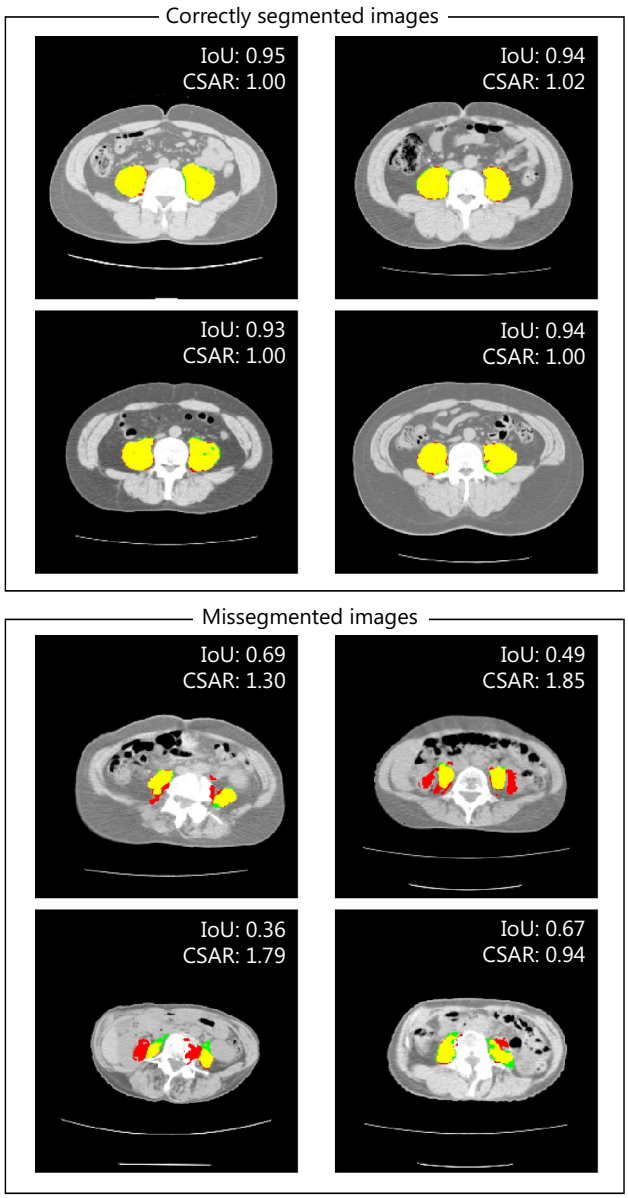

Fig. 2 Sample images of correctly and incorrectly segmented muscles. Yellow, red, and green indicate the areas of overlap; overextraction; and under extraction, respectively

4 Discussion

In this paper, we proposed an automated technique for the segmentation of the psoas-major muscle using the U-Net convolutional neural network and evaluated its performance for low-dose 2D CT images. The segmentation accuracy was 86.0% for IoU and 103.1% for CSAR, and only a slight difference was evident in terms of the distribution of IoU and CSAR, except for the outliers among each validation data set (Fig. 3). These results suggest that the validation data sets were selected without bias. We could not directly compare the U-Net network with the methods used in [7–9], because 2D CT images were used instead of 3D CT images. Nonetheless, we compared the accuracy of our method with those of the methods in [7–9]. The averaged IoU of the other methods varied from 66.0 to 76.5%. In addition, according to our results, an approximate 10% error occurred between the results produced by the two technicians with respect to the IoU. Therefore, we concluded that the U-Net network is sufficiently accurate as an automated segmentation technique and superior to other methods. Furthermore, our trial demonstrated its effectiveness, even for low-dose 2D CT images. For example, Kamiya et al. [8] showed that 3D CT images were required to include two landmarks, corresponding to the origin and insertion of the muscle. In addition, Kamiya et al. [20] showed the possibility of recognizing the attachment site of the muscle by FCNs including the recognition of the landmarks. These studies enhance the performance using the 3D structure of the human body. However, in this study, a user did not need to prepare the 3D CT images. This aspect may prove useful for clinical settings, such as medical checkups.

In the correctly segmented images (Fig. 2), the psoas-major muscle was separated from other organs, such as the blood vessels and intestinal tract, because the smooth continuity of the muscle border and fat layer between them was recognized. However, in the incorrectly segmented images, the muscle was not well separated from other organs, because they had the same CT value. In addition, deformities of the lumbar region caused by degenerative diseases and/or aging obscured the continuity of the muscle border and prevented accurate demarcation. Therefore, if a voxel had values close to the major psoas muscle at the average position of the major psoas muscle, it was possibly recognized as a major psoas muscle, causing a decrease in the segmentation accuracies (as shown in the bottom left of the mis-segmented images, Fig. 2). These voxels created difficulties for the experts when setting a border. Therefore, these "abnormal" cases appeared as low outliers in Fig. 4 compared with the "normal" cases. The U-Net architecture could not correctly recognize a border due to these voxels as well. However, in some cases, even though the muscle was not exactly recognized, a high performance may be obtained in terms of the CSAR. For deep learning, the number of images with degenerative lumbar diseases and aging must be increased in the training data sets.

In the future, we will try to visualize the behavior of the network [24, 25]. Furthermore, we plan to analyze thousands of subjects using the U-Net network and to investigate the relationship between the CSA of the muscle and various diseases or walking ability.

This study had several limitations. First, the number of training data sets was not sufficient for deep learning. Although one data set was augmented 14 times and such data

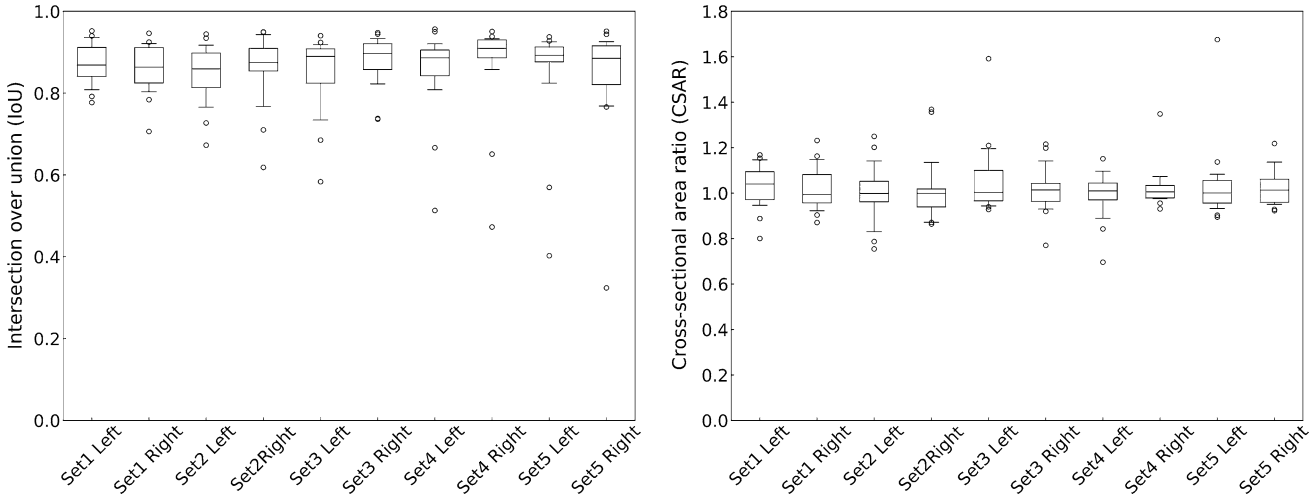

Fig. 3 Accuracy of the segmented area in terms of the IoU (left) and CSAR (right) for each validation data set in the U-Net network. For each plot, the line within the box represents the median. The lower and upper lines of the box represent the 25th and 75th percentiles, and the lower and upper adjacent lines (whiskers) represent the 10th and 90th percentiles, respectively. The circles indicate outliers

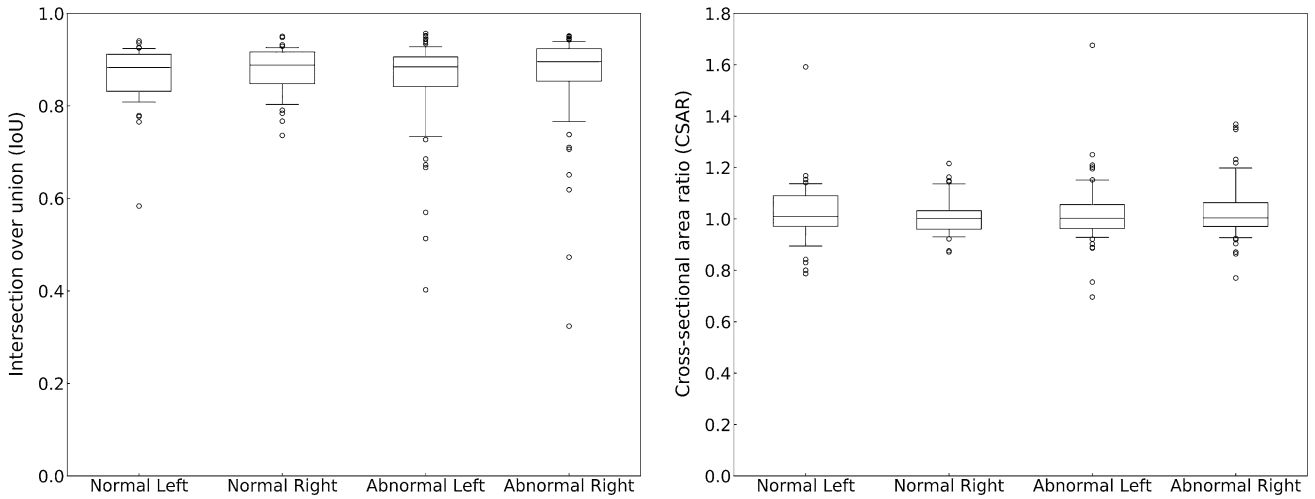

Fig. 4 Accuracy of the segmented area in terms of the IoU (left) and CSAR (right) for "normal" and "abnormal" in the U-Net network. For each plot, the line within the box represents the median. The lower and upper lines of the box represent the 25th and 75th percentiles, and the lower and upper adjacent lines (whiskers) represent the 10th and 90th percentiles, respectively. The circles indicate the outliers

augmentation is often used for deep learning, increasing the number of data sets would be beneficial for future analysis. Second, CT scans were performed using three types of scanners; therefore, an investigation of the influence of different CT scanners may be necessary. Finally, we applied the method to low-dose 2D CT images. Although a sufficiently high accuracy was demonstrated, a higher accuracy may be expected for high-contrast CT and MRI images.

5 Conclusions

This paper presented an automated technique for the segmentation of the psoas-major muscle for 2D low-dose CT images using the U-Net convolutional neural network. The segmentation accuracy was 86.0% for IoU and 103.1% for CSAR, indicating that the U-Net network is superior

to previous methods. In the U-Net network, a user does not need to prepare 3D CT images to set two land marks. Therefore, this method can easily and feasibly be used in clinical settings. A future study will compare the performances of U-Net and other architectures such as FCNs and SegNet. In addition, we will analyze thousands of subjects using the U-Net network and investigate the relationship between the CSA of the muscle and various diseases or walking ability.

Acknowledgements We would like to thank the staff of the Hamamatsu Medical Imaging Center and Hamamatsu Photonics K. K. for their support.

Compliance with ethical standards

Conflict of interest We declare that this work is free from financial limitation or any other relationship that might lead to a conflict of interest. The authors have no conflicts of interest to declare.

Statement of human and animal rights All procedures performed in the studies involving human participants were in accordance with the ethical standards of the Institutional Review Board of Hamamatsu Medical Photonics Foundation and 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies performed with animals.

Informed consent Informed consent was obtained from all study participants.

References

1. Kim JD, Kuno S, Soma R, Masuda K, et al. Relationship between reduction of hip joint and thigh muscle and walking ability in elderly people. Jpn J Phys Fit Sports Med. 2000;49(5):737–8. https://doi.org/10.7600/jspfsm1949.49.589.
2. Yaguchi Y, Kumata Y, Horikawa M, et al. Clinical significance of area of psoas major muscle on computed tomography after gastrectomy in gastric cancer patients. Ann Nutr Metab. 2017;71(3–4):145–9.
3. Fearon K, Strasser F, Anker SD, et al. Definition and classification of cancer cachexia: an international consensus. Lancet Oncol. 2011;12(5):489–95.
4. Shen W, Punyanitya M, Wang Z, et al. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. J Appl Physiol. 2004;97(6):2333–8.
5. Baracos VE, Reiman T, Mourtzakis M, Gioulbasanis I, Antoun S. Body composition in patients with non-small cell lung cancer: a contemporary view of cancer cachexia with the use of computed tomography image analysis. Am J Clin Nutr. 2010;91(4):1133–7.
6. Martin L, Birdsell L, MacDonald N, et al. Cancer cachexia in the age of obesity: skeletal muscle depletion is a powerful prognostic factor, independent of body mass index. J Clin Oncol. 2013;31(12):1539–47.
7. Meesters SPL, Yokota F, Okada T, et al. Multi atlas-based muscle segmentation in abdominal CT images with varying field of view. Paper presented at the International Forum on Medical Imaging in Asia (IFMIA), November 16–17, 2012, Daejon Korea.
8. Kamiya N, Zhou X, Chen H, et al. Automated segmentation of psoas major muscle in X-ray CT images by use of a shape model: preliminary study. Radiol Phys Technol. 2012;5(1):5–14.
9. Inoue T, Kitamura Y, Li Y, Ito W, Ishikawa H. Psoas major muscle segmentation using higher-order shape prior. In: International MICCAI workshop on medical computer vision, 2016; p. 116–24. https://doi.org/10.1007/978-3-319-42016-5_11.
10. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell. 2016;39(4):640–51.
11. Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell. 2017;39(12):2481–95.
12. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention 2015; p. 234–41. https://doi.org/10.1007/978-3-319-24574-4_28.
13. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: IEEE conference on computer vision and pattern recognition (CVPR) 2017; p. 2881–90. https://doi.org/10.1109/cvpr.2017.660.
14. Greenspan H, van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans Med Imaging. 2016;35(5):1153–9.
15. Suzuki K. Overview of deep learning in medical imaging. Radiol Phys Technol. 2017;10(3):257–73.
16. Teramoto A, Tsukamoto T, Kiriyama Y, Fujita H. Automated classification of lung cancer types from cytological images using deep convolutional neural networks. Biomed Res Int. 2017;4067832:1–6.
17. Zhou X, Takayama R, Wang S, Hara T, Fujita H. Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. Med Phys. 2017;44(10):5221–33.
18. Lee H, Troschel FM, Tajmir S, et al. Pixel-level deep segmentation: artificial intelligence quantifies muscle on computed tomography for body morphometric analysis. J Digit Imaging. 2017;30(4):487–98.
19. Ghosh S, Boulanger P, Acton ST, Blemker SS, Ray N. Automated 3D muscle segmentation from MRI data using convolutional neural network. In: IEEE international conference on image processing (ICIP) 2017; p. 4437–41. https://doi.org/10.1109/icip.2017.8297121.
20. Kamiya N, Kume M, Zheng G, et al. Automated recognition of erector spinae muscles and their skeletal attachment region via deep learning in torso CT images. In: International workshop on computational methods and clinical applications in musculoskeletal imaging. 2018. https://doi.org/10.1007/978-3-030-11166-3_1.
21. Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning. In: Proc 12th USENIX conf operating syst design implement. 2016;16:265–83.
22. Keras: The Python Deep Learning library. http://keras.io/. Accessed 30 Mar 2019.
23. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv:1412.6980.
24. Mahendran A, Vedaldi A. Understanding deep image representations by inverting them. In: IEEE international conference on computer vision and pattern recognition (CVPR) 2014; p. 5188–96. https://doi.org/10.1109/cvpr.2015.7299155.
25. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. In: IEEE international conference on computer vision (ICCV) 2017; p. 618–26. https://doi.org/10.1109/iccv.2017.74.