Label efficient segmentation of single slice thigh CT with two-stage pseudo labels

Qi Yang,[a,*] Xin Yu,[a] Ho Hin Lee,[a] Yucheng Tang,[b] Shunxing Bao,[a] Kristofer S. Gravenstein,[c] Ann Zenobia Moore,[c] Sokratis Makrogiannis,[d] Luigi Ferrucci,[c] and Bennett A. Landman[a,b]

[a]Vanderbilt University, Department of Computer Science, Nashville, Tennessee, United States
[b]Vanderbilt University, Department of Electrical and Computer Engineering, Nashville, Tennessee, United States
[c]National Institute on Aging, Longitudinal Study Section, Baltimore, Maryland, United States
[d]Delaware State University, Division of Physics, Engineering, Mathematics and Computer Science, Dover, Delaware, United States

Abstract

Purpose: Muscle, bone, and fat segmentation from thigh images is essential for quantifying body composition. Voxelwise image segmentation enables quantification of tissue properties including area, intensity, and texture. Deep learning approaches have had substantial success in medical image segmentation, but they typically require a significant amount of data. Due to the high cost of manual annotation, training deep learning models with limited human label data is desirable, but it is a challenging problem.

Approach: Inspired by transfer learning, we proposed a two-stage deep learning pipeline to address the thigh and lower leg segmentation issue. We studied three datasets, 3022 thigh slices and 8939 lower leg slices from the BLSA dataset and 121 thigh slices from the GESTALT study. First, we generated pseudo labels for thigh based on approximate handcrafted approaches using CT intensity and anatomical morphology. Then, those pseudo labels were fed into deep neural networks to train models from scratch. Finally, the first stage model was loaded as the initialization and fine-tuned with a more limited set of expert human labels of the thigh.

Results: We evaluated the performance of this framework on 73 thigh CT images and obtained an average Dice similarity coefficient (DSC) of 0.927 across muscle, internal bone, cortical bone, subcutaneous fat, and intermuscular fat. To test the generalizability of the proposed framework, we applied the model on lower leg images and obtained an average DSC of 0.823.

Conclusions: Approximated handcrafted pseudo labels can build a good initialization for deep neural networks, which can help to reduce the need for, and make full use of, human expert labeled data.

 [DOI: 10.1117/1.JMI.9.5.052405]

Keywords: thigh segmentation; CT image; pseudo labels; transfer learning.

Paper 21288SSRR received Oct. 29, 2021; accepted for publication May 2, 2022; published online May 19, 2022.

1 Introduction

Estimating volumes and masses of total body components is important for research of cancer, joint replacement, and exercise physiology.[1] Full body CT scans can be used to calculate whole body compositions directly. However, it is hard to acquire a typical full body CT in the usual medical context due to the intense radiology dose. Mourtzakis et al.[2] proposed that body components measured on abdomen or thigh slices are highly correlated with the mass of whole-body

*Address all correspondence to Qi Yang, qi.yang@vanderbilt.edu

2329-4302/2022/$28.00

tissues. Thus, accurate segmentation of thigh slices can quantify tissue area properties to estimate body composition without requiring additional irradiation or examinations. Thus, this paper aims to segment muscle, fat, and bones from 2D thigh and lower leg CT slices.

Several recent techniques have been proposed to address thigh and lower leg segmentation on CT images. Senseney et al.[3] proposed an automatic region growing method using a morphology operation and threshold to extract bone muscle and fat in CT thigh and abdomen images. Tan et al.[4] proposed using a variational Bayesian Gaussian mixture model to cluster fat, marrow, muscle bone, and air on three-dimensional (3D) CT scans. Felinto et al.[5] proposed using a Gaussian mixture model and relative position to cluster similar tissues for intermuscular fat and muscles segmentation. With impressive performance of the deep neural network-based segmentation, Zhu et al.[6] applied the H-DenseU-Net on MRI lower leg data of children with and without cerebral palsy. Rohm et al.[7] created a 3D heterogeneous MRI lower leg dataset and trained a convolution network to segment muscle.

Deep learning methods show impressive performance in segmentation tasks. However, this performance depends on sufficient human annotation.[8] In the medical imaging field, human annotation requires professional knowledge, which is very time-consuming and thus expensive. To avoid annotating new data, many researchers used common data augmentation methods such as rotation, intensity shift, and scaling to artificially enhance the diversity and quality of the training data.[9] Image synthesis is another data augmentation method. Generative adversarial networks (GANs)[10] have been utilized to synthesize new labeled data for segmentation. However, GAN is notorious for training and is hard to implement in practical tasks.[11] The main limitation of data augmentation is data bias generated during data augmentation process. To preserve original data distribution, leveraging the power of unannotated data is another solution to train a model with limited annotation data. Chen et al. proposed using self-supervised learning with image context restoration to achieve brain tumor segmentation with a limited dataset.[12] Instead of self-supervised learning, transfer learning is another way to train with limited label data. First, a model is trained from scratch on a large-scale dataset with a similar task. Then, the model is fine-tuned with human annotated data. Tajbakhsh et al.[13] showed that a fine-tuned network could outperform networks that were trained from scratch with better robustness.

To better segment muscle, cortical bone, internal bone, subcutaneous fat, and intermuscular fat with limited annotated data, we propose a two-stage transfer learning-based framework. We use an approximate handcrafted method to generate pseudo labels for 1883 thighs to train the model in the first stage and fine-tune it with 125 human label thighs in the second stage to achieve segmentation. We test the model on the thigh slice and use the lower leg slice as external data to demonstrate the generalizability of the proposed framework. The target tissue and corresponding legend are shown in Fig. 1. This paper is the significant extension of our prior accepted work[14] of *SPIE* 2022.

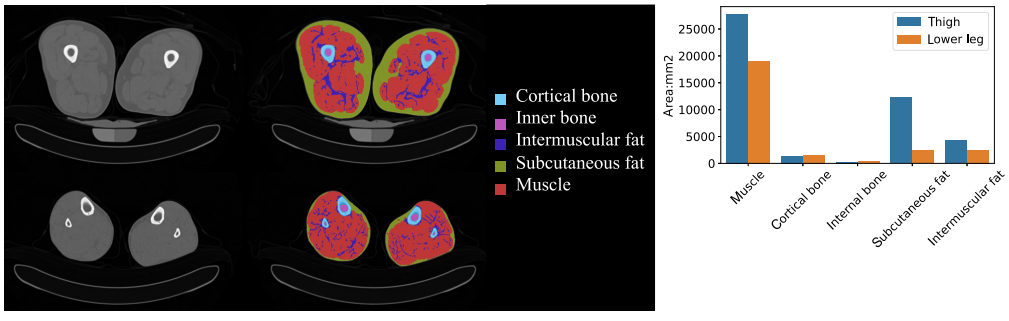

Fig. 1 The first row and second row of (a) represent the middle thigh and lower leg from the same subject, respectively. The left column is the original CT image, and the right column is the target tissues label. Each tissue has a different area, and the imbalance of area makes the segmentation of sparse tissue (intermuscular fat) challenging. The area of each tissue is shown in (b).

2 Material and Methods

We designed a two-stage coarse-to-fine deep learning method to achieve thigh and lower leg segmentation on low-dose CT slices with deep learning. We first split a CT thigh slice into single left and right thigh images. In the first stage, we train the deep neural network with approximate handcrafted labels. In the second stage, we fine-tune the model from the first stage with human expert labels to recover more details.

2.1 *Preprocessing*

The preprocessing pipeline works for both thigh and lower leg images with minor difference. For each thigh image, we first set the field of view of the CT thigh slice to include the left thigh, right thigh, table, and phantom. Next, we use the threshold of -500 Hounsfield unit (HU) to binarize the input thigh slice. We use a square kernel 25×25 to erode binary images and create three independent eroded masks. Then, we choose the left thigh and right thigh according to the area size (the area of the table mask should be smaller than that of the thigh mask) and center position (the centers of the left thigh mask and the right thigh mask should be at an approximate horizontal axis). After picking the eroded mask of two thighs, we dilate the chosen mask with the same kernel size. Based on those two masks, we find the maximal bounding box for each thigh and crop the original CT slice from 512×512 to 256×256 without changing the pixel resolution and intensity range of the whole CT slice. Different from preprocessing on the thigh, we use a kernel size of 10×10 to erode and dilate the mask of the lower leg. Finally, we manually review all of the thighs and lower legs and exclude cropped image including other tissue (e.g., Table 1).

2.2 *Create Pseudolabel for Thigh*

Each CT slice has specific intensity units for each tissue. We use a CT window of $[-190, -30]$ HU for fat, [30,80] HU for muscle, and $[1000,\infty]$ HU for bones.[15] We propose the following pipeline with seven steps to extract five target tissues coarsely using CT intensity and morphology.

(1) Create a cortical bone binary mask image with a threshold of 1000.
(2) Invert the cortical bone mask and find the connected region with an area that is smaller than the half image $(256 \times 256/2)$ as the internal bone mask.
(3) Use a threshold of 0 HU to binarize the thigh image and create a muscle mask.
(4) Fill the holes for result (3).
(5) Subtract the muscle mask from step (4) to create an intermuscular fat mask based on the assumption that intermuscular fat is within the muscle.
(6) Binarize the thigh image with a threshold of -500 HU.

Table 1 The number of slices, thighs, lower legs, labeled thighs, and lower legs for each cohort.

Study name	Cohort	Subjects	Slices	All thighs or lower legs	Pseudo labels/ human labels
BLSA	First stage thigh training	671	944	1883	1883/0
BLSA and GESTALT	Second stage thigh training	86	117	125	0/125
BLSA and GESTALT	Second stage thigh validation	22	26	31	0/31
BLSA and GESTALT	Second stage thigh test	47	65	73	0/73
BLSA	Second stage external lower leg test	1435	8939	17878	0/39
BLSA and GESTALT	Second stage external thigh test	800	1991	3982	0/0

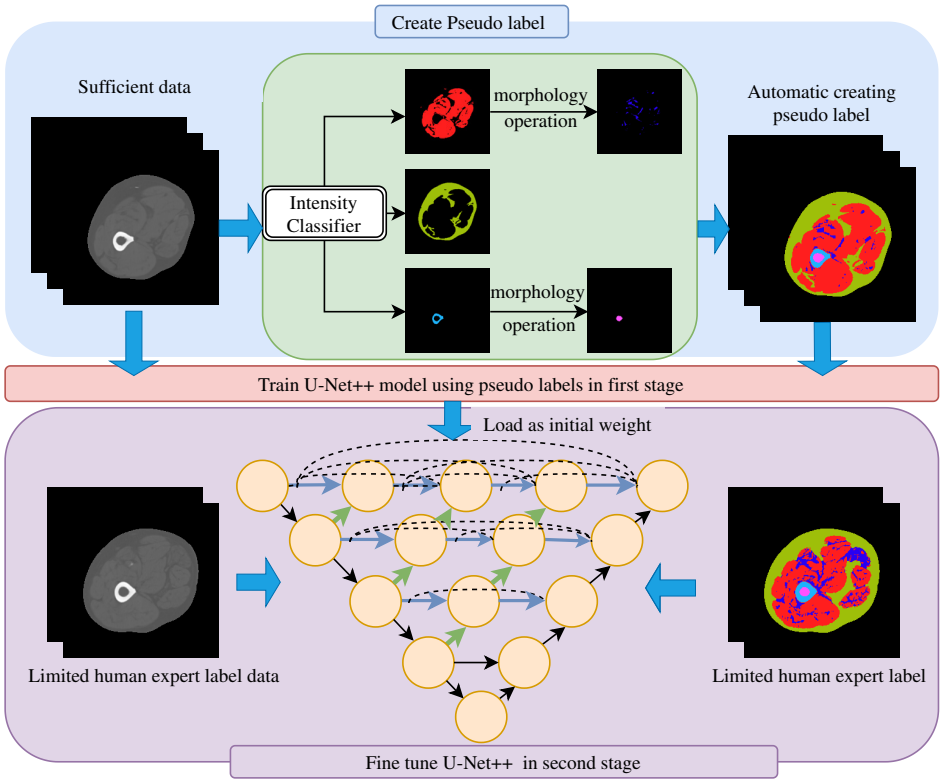

Fig. 2 The proposed hierarchical coarse-to-fine thigh segmentation including three parts: (1) Use the threshold and morphology to generate coarsely segmented pseudo labels. (2) Feed pseudo labels into deep learning model, and train the model from scratch. (3) Use the optimized model from previous stage as initialization, and fine-tune the model with limited expert labels. The model from the first and second stages is optimized separately.

(7) Subtract the result of step (4) from step (6) to create a subcutaneous fat mask. Five coarse approximate segmentation masks are shownin Fig. 2. They are fused into one mask before fed into the deep neural network.

2.3 *Two Stage Training*

U-Net++[16] is an encoder–decoder network in which the encoder and decoder are connected through a series of nested, dense skip connections. The nested skip connections can help bridge the semantic gap between the feature maps of the encoder and decoder, which is helpful for segmenting fine-grain details of target tissues such as intermuscular fat in our case. Thus, we use U-Net++ as our backbone for inference segmentation results. Transfer learning refers to reusing a model developed for a task as the starting point for a model on a different or same task, which alleviates the challenge of limited training data. Thus, we design a two-stage transfer learning strategy. In the first stage, we use approximate pseudo labels to train U-Net++ from scratch and choose the best model according to the performance on the validation dataset. Then, the best model is loaded as the initialization. Human expert labeled data are used to fine-tune the model until it converges. The whole pipeline is shown in Fig. 2.

2.4 *Data Distribution*

We use 3022 de-identified CT thigh slices from the Baltimore Longitudinal Study of Aging (BLSA)[17] and 121 de-identified thigh slices from the Genetic and Epigenetic Signatures of Translational Aging Laboratory Testing (GESTALT) study as well as 8939 de-identified lower leg slices from the BLSA. All data are under Institute Review Board approval. The image size is 512×512. In the preprocessing and quality assurance stages, five thigh images are discarded since they include other structures (e.g., the table). Note that, for some thigh slices, only the left

thigh is manually labeled instead of both thighs. For labels of the lower leg image, a single experienced graduate student labeled the lower leg image from the CT image under supervision of a muscle physiologist with more than 10-years of research experience. The graduate student traced each lower leg CT image using a computer visualization tool (FSLeyes[17]). The student annotated the contour of each tissue by human eye from scratch (i.e., starting on an image with no labels). The process minimized potential visual bias from a prior computer-aided algorithm. Region filling was used once contours were closed. This process was repeated for each of the following with labeling in the order of internal bone, cortical bone, muscle, subcutaneous fat, and intermuscular fat. We perform pseudo labels creation on 1883 thighs as the first stage thigh training cohort. We divide 229 labeled thigh images into training (125 thighs), validation (31 thighs), and testing (73 thighs) cohorts for stage 2, respectively. No subject had images in both the training and validation or testing cohorts.

2.5 *Implementation Details*

Our experiments are implemented in Python 3.7 with PyTorch 1.7. We apply a window of [−150, 100] HU to normalize each input image. In the first stage, the initial learning rate for U-Net and U-Net++ is 0.002 and 0.0002, respectively. In the fine-tuning stage, the initial learning rate for both U-Net and U-Net ++ is 0.0001. We conduct the experiment to train only with human expert labels using U-Net and U-Net++. The learning rate for U-Net is 0.01, and the learning rate for U-Net++ is 0.001. The learning rate decayed to 0 linearly until the end of the training epoch in both stages. Resize and crop are used as online data augmentation. The max-training-epoch is set to 200 with the batch size of 8. The optimizer used in training is stochastic gradient descent.

2.6 *Baseline Method and Metrics*

The U-Net[18] is considered to be an alternate architecture since its impressive performance on medical image segmentation. To validate the effectiveness of the transfer learning strategy, both U-Net and U-Net++ training with human labels only are regarded as baseline methods.

To evaluate the accuracy of our proposed method, we compare the segmentation results against the ground truth provided by expert labels. To quantify the agreement between segmentation and truth, we use the Dice similarity coefficient (DSC) as the main evaluation measurement for inference results by comparing each binary tissue against the ground truth voxel-by-voxel:

$$\mathrm{DSC} = \frac{2|R \cap T|}{|R| + |T|}, \tag{1}$$

where R represents the segmentation result generated by the deep learning model and T represents the corresponding ground truth.

3 Results

Figure 3 compares the DSC of the muscle, cortical bone, inner bone, subcutaneous fat, and intermuscular fat between U-Net++ and U-Net using only human labels, in stage 1 and stage 2. The boxplots presented are evaluated across 73 single thighs. Table 2 shows the mean DSC of each tissue for all six methods. Overall, the average DSC across all five tissues of U-Net ++ in the second stage is significantly better than the other five methods. Except for subcutaneous fat, the proposed method has the highest mean DSC of the remaining tissues. The proposed method makes the largest improvement from 0.681 to 0.782 on mean DSC for sparse and small intermuscular fat compared with U-Net trained only with human labels. Figure 4 compares the qualitative result produced by all six methods. Compared with U-Net in stage 2, the proposed method yields superior performance and segments more details of intermuscular fat.

To test the generalizability of the proposed methods, we applied the model on preprocessed lower leg images. The experimental setting is the same as in the thigh experiment. Figure 5 compares the performance of all six methods on lower leg images. Table 3 shows the mean

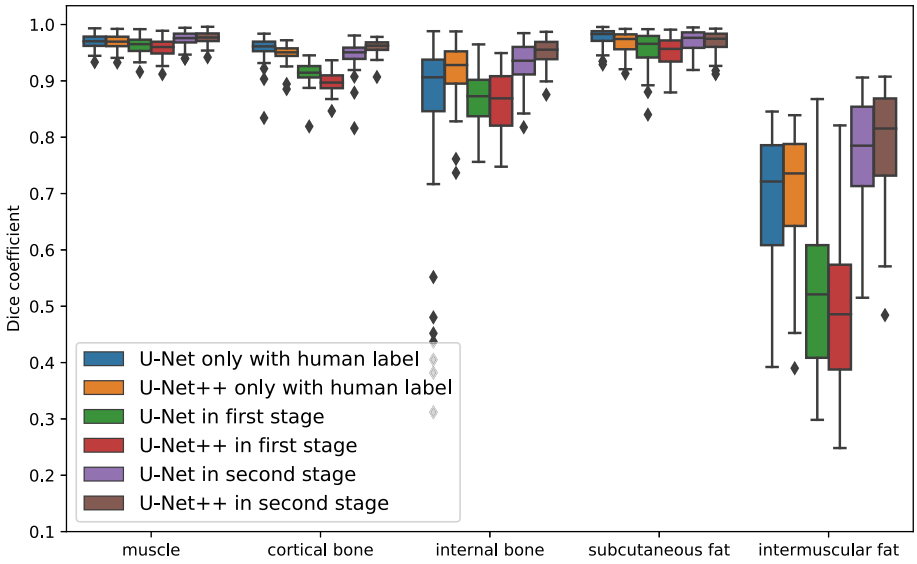

Fig. 3 The DSC comparison of thigh images using U-Net trained only with human labels, U-Net++ trained only with human label, U-Net in stage 1, U-Net++ in stage 1, U-Net in stage 2, and U-Net++ in stage 2 in boxplots of five target tissues.

Table 2 The mean DSC for each tissue of each method for the thigh CT image. The highest result is bolded.

Method	Muscle	Cortical bone	Internal bone	Subcutaneous fat	Intermuscular fat	Average
U-Net only with human labels	0.967*	0.958	0.852*	**0.977***	0.681*	0.887*
U-Net++ only with human label	0.966*	0.947*	0.919*	0.967*	0.695*	0.899*
U-Net in the first stage	0.960*	0.915*	0.868*	0.955*	0.609*	0.841*
U-Net++ in the first stage	0.957*	0.898*	0.865*	0.949*	0.481*	0.830*
U-Net in the second stage	0.971	0.946*	0.932*	0.970	0.762*	0.916*
U-Net++ in the second stage	**0.973**	**0.960**	**0.951**	0.969	**0.782**	**0.927**

*The method is significantly different from U-Net++ in the second stage (p value < 0.05, Wilcoxon signed rank test).

DSC of each tissue of all six methods on lower leg images. The average DSC of all tissues of the proposed method decreased from 0.927 to 0.823 when compared with the thigh experiment, but significantly outperformed all other methods except U-Net in the second stage. The U-Net trained only with human labels has the highest mean DSC 0.922 in cortical bones, and the U-Net++ trained only with human labels has the highest mean DSC 0.9 in internal bones. Figure 6 shows the confidence level of the bone result with qualitative representations.

After training on pseudo labels, we want to investigate the relationship between the number of human expert data and performance of the model in the second (fine-tune) stage. We fed 1, 5, 10, 20, 30, 60, 90 and all expert human label thighs to fine-tune the model from the first stage. The fine-tuning process is repeated 10 times. Each time the data are randomly picked from

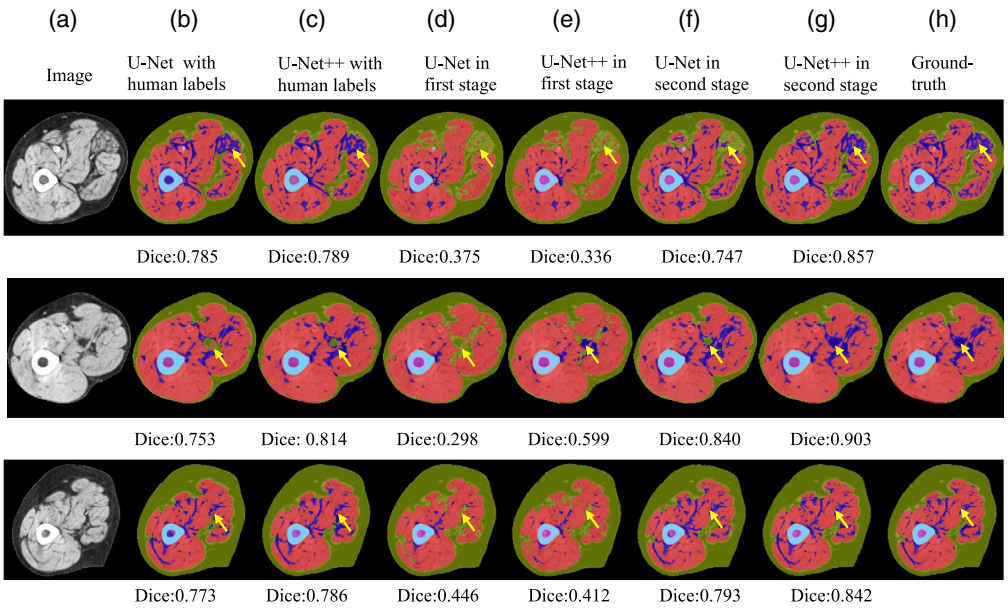

Fig. 4 Qualitative representation of the thigh slice segmentation. (a) Three randomly selected source CT images after applying window $[-150, 100]$. (b) The segmentation from U-Net only trained with human labels. (c) The segmentation from U-Net++ only trained with human labels. (d) and (e) The segmentation using network U-Net and U-Net++ in stage 1, respectively. (f) and (g) The segmentation by network U-Net and U-Net++ in stage 2, respectively. (h) The ground truth. The yellow arrow points to the large difference across the methods and ground truth. The DSC values only show intermuscular fat segmentation performance for reference.

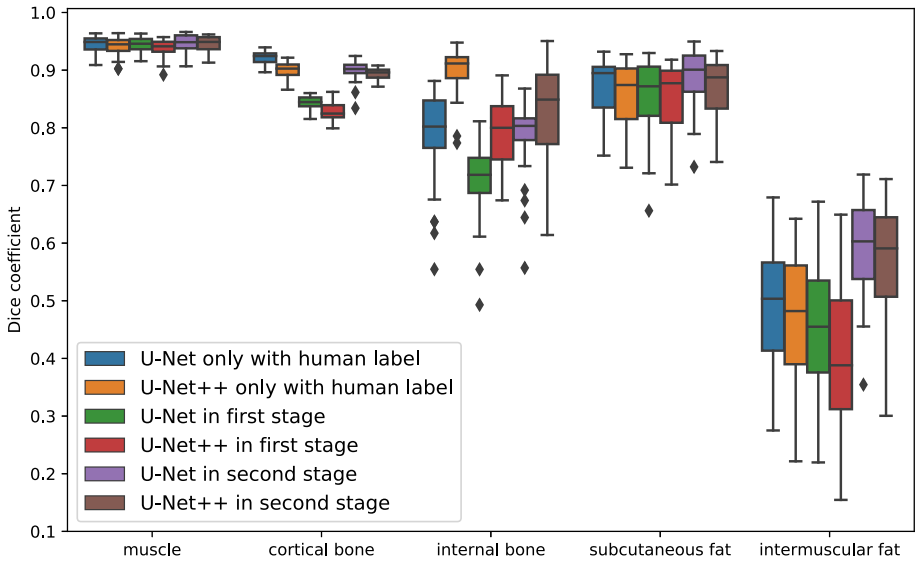

Fig. 5 The DSC comparison of the lower leg image among using U-Net trained only with human labels, U-Net++ trained only with human labels, U-Net in stage 1, U-Net++ in stage 1, U-Net in stage 2, and U-Net++ in stage 2 of five target tissues.

training cohort (125 thighs) and inference on the test cohort (73 thighs). The distribution of the mean dice of each retraining is shown in Fig. 7. When only feeding one thigh to the model, the variance of the mean DSC is largest compared with the others. Also, the variance becomes smaller when increasing the feeding data. When feeding 30 thighs, the mean DSC of 10 retraining is 0.924, almost equivalent to the mean DSC 0.931 of feeding all training data.

Table 3 The mean DSC for each tissue of each method for the lower leg CT image. The highest result is bolded.

Method	Muscle	Cortical bone	Internal bone	Subcutaneous fat	Intermuscular fat	Average
U-Net only with human labels	0.944*	**0.922***	0.786*	0.876*	0.494*	0.805*
U-Net++ only with human label	0.941*	0.900*	**0.901***	0.857*	0.469*	0.814*
U-Net in the first stage	0.945	0.843*	0.708*	0.850*	0.443*	0.758*
U-Net++ in the first stage	0.939*	0.828*	0.791*	0.852*	0.393*	0.761*
U-Net in the second stage	**0.946***	0.901*	0.787*	**0.891***	**0.590***	**0.823**
U-Net++ in the second stage	0.945	0.893	0.836	0.870	0.573	**0.823**

*The method is significantly different from U-Net++ in the second stage (p value <0.05, Wilcoxon signed rank test).

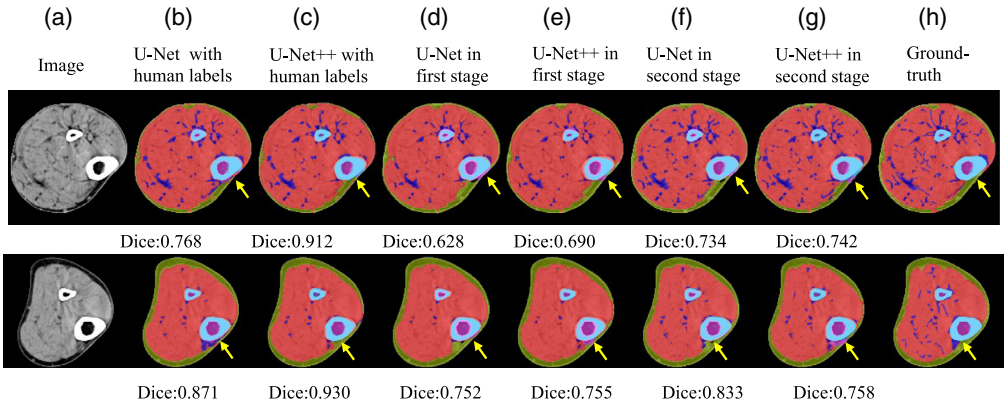

Fig. 6 Qualitative representation of the lower leg slice segmentation. (a) The source CT image after applying window [−150, 100]. (b) The segmentation from U-Net only trained with human labels. (c) The segmentation from U-Net++ only trained with human labels. (d) and (e) The segmentation using network U-Net and U-Net++ in stage 1, respectively. (f) and (g) The segmentation by network U-Net and U-Net++ in stage 2, respectively. (h) The ground truth. The text below each image is the internal bone DSC.

We apply the proposed models on additional CT scans of thighs and lower legs, which do not have human generated label maps for comparison. We overlay the segmentation results on CT images with a colormap of the human review of each segmentation result to find outliers. Ten out of 3982 thigh images and 136 out of 17,878 lower leg images fail human review and are regarded as outliers. Figure 8 shows four outliers from thigh and lower leg images, respectively.

4 Conclusion and Discussion

Herein, we proposed a transfer learning-based method to achieve accurate and robust thigh tissue segmentation, focusing on muscle, cortical bone, internal bone, subcutaneous fat, and intermuscular fat. The proposed framework achieved accurate segmentation on thigh CT slices with limited human labels. Compared with the method only trained with human expert labels, the superior performance of the proposed framework demonstrates the effectiveness of two-stage training. We applied the model only trained on the thigh slice to the lower leg image. The results show that our model still recognizes muscle and fat with high agreement, which demonstrates

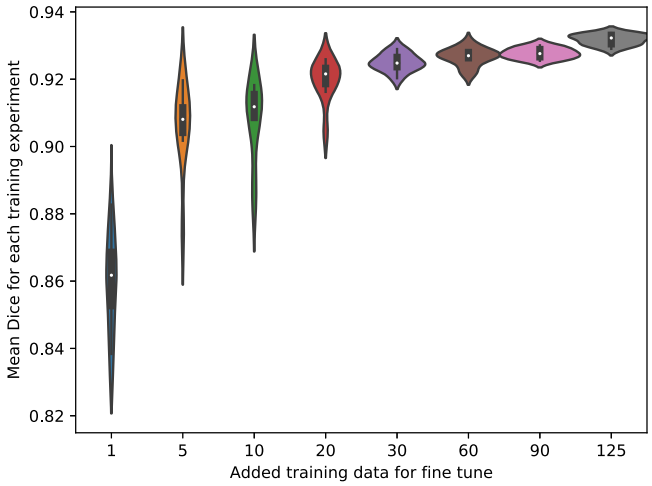

Fig. 7 The relationship between the mean DSC and added data for each fine-tuning. The violin plot includes 10 data points. Each data point represents the mean DSC across all tissues of the test cohort in one fine-tuning process.

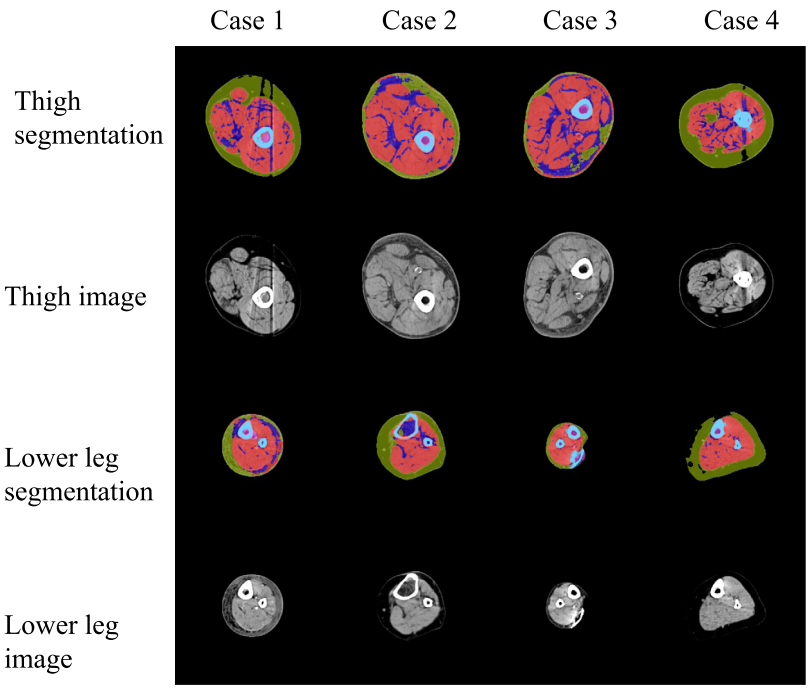

Fig. 8 The outliers from the thigh and lower leg. The first row and third row are segmentation result on the thigh and lower leg, respectively. The second and fourth row are the CT image after applying the window. Each column represents an outlier from the thigh and lower leg, respectively.

that the proposed framework has the generalizability to similar anatomical structures of CT images. Then, we analyzed the relationship between the size of the fine-tuning dataset and performance. The result shows that the proposed framework can use a smaller size of training cohort to keep the same performance as all training data, which indicates that the proposed framework has the potential to fully exploit the data and increase data efficiency.

In the medical image segmentation, label annotation is always a challenging problem for the deep learning methods. Similar to Ref. 18, the proposed work leverages the power of cheaper-to-acquire pseudo labels to achieve high-performance segmentation. Both methods work on the assumption that deep learning models can explore general patterns from pseudo labels and

inference on unseen data. The primary difference between the approaches is that the proposed method used limited human labels to fine-tune model to improve performance of tissue with an appearance that is hard for pseudo labels to capture.

The proposed method can be extended to 3D CT images with target tissue that is strongly related to intensity. The automatic method based on intensity can create pseudo labels for target tissue. Different from 2D slices, 3D volumes require more data to train more parameters in the deep neural network. We expect that the number of pseudo labels and human provided labels needs to be increased for successful application of the method.

One major limitation of the proposed method is that the performance of bone structure is inferior to models only trained with human labels. As shown in Fig. 7, the proposed method might regard subcutaneous fat around cortical bone as internal bone, which leads to inferior performance. Except for the domain shift between the thigh and lower leg, the reason could be that the boundary of the pseudo label between cortical bone and internal bone is not clearly defined, which made the model misclassify subcutaneous fat around cortical bones. Thus, how to improve quality of the pseudo label is an important future research direction.

The training dataset is all mid-thigh images. The model is adapted to the anatomy of mid-thigh and applies well on lower leg images because of the similar anatomy. The CT scans including the femoral head have a different anatomy, bringing varied CT unit scales, particularly in the bone part. The model may segment muscle, intermuscular fat, and subcutaneous fat. However, it cannot recognize targeted tissue from the femoral head correctly. Thus, how to develop a model that can handle different positions of the lower limb will be an important future research direction.

Normal training images have a fixed anatomy pattern and classes of target tissues. The pseudo labels are created with the assumption that there is no abnormal bone intensity or shape. When applying the model on abnormal bones, the method is biased to recognize the bone pathology as the normal bone tissues and segment them into incorrect labels. In the future, the pseudo labels method may be improved to segment the abnormal bone part and feed more abnormal cases into the model, so the model may recognize the pathology part.

In summary, the proposed pipeline has potential to be applied on other medical scenes with low human effort, which makes better use of human expert labels.

Disclosures

The authors of the paper are directly employed by institutes or companies provided in this paper. No conflicts of interest exist in the submission of this paper.

Acknowledgments

This research was supported by NSF CAREER 1452485 and the National Institutes of Health (NIH) under Award Nos. R01EB017230, R01EB006136, R01NS09529, T32EB001628, 5UL1TR002243-04, 1R01MH121620-01, and T32GM007347; by ViSE/VICTR VR3029; and by the National Center for Research Resources, Grant No. UL1RR024975-01 and is now at the National Center for Advancing Translational Sciences, Grant No. 2UL1TR000445-06. This project was also supported by the National Science Foundation under Award Nos. 1452485 and 2040462. This research was conducted with the support from the Intramural Research Program of the National Institute on Aging of the NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This study was in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University, Nashville, TN. The identified datasets used for the analysis described were obtained from the Research Derivative (RD), database of clinical and related data.

References

1. A. L. Jeanson et al., "Body composition estimation from selected slices: equations computed from a new semi-automatic thresholding method developed on whole-body CT scans," *PeerJ* **5**, e3302 (2017).

2. M. Mourtzakis 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.* **33**(5), 997–1006 (2008).
3. J. Senseney, P. F. Hemler, and M. J. McAuliffe, "Automated segmentation of computed tomography images," in *22nd IEEE Int. Symp. Comput.-Based Med. Syst.*, pp. 1–7 (2009).
4. C. Tan et al., "An automated and robust framework for quantification of muscle and fat in the thigh," in *22nd Int. Conf. Pattern Recognit.*, pp. 3173–3178 (2014).
5. J. de Carvalho Felinto et al., "Automatic segmentation and quantification of thigh tissues in CT images," in *Int. Conf. Comput. Sci. and Its Appl.*, pp. 261–276 (2018).
6. J. Zhu et al., "Deep learning methods for automatic segmentation of lower leg muscles and bones from MRI scans of children with and without cerebral palsy," *NMR Biomed.* **34**, e4609 (2021).
7. M. Rohm et al., "3D automated segmentation of lower leg muscles using machine learning on a heterogeneous dataset," *Diagnostics* **11**(10), 1747 (2021).
8. K. A. Philbrick et al., "RIL-contour: a medical imaging dataset annotation tool for and with deep learning," *J. Digital Imaging* **32**(4), 571–581 (2019).
9. C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data* **6**(1), 60 (2019).
10. I. Goodfellow et al., "Generative adversarial nets," in *Adv. Neural Inf. Process. Syst.*, Vol. **27** (2014).
11. W. Nie and A. B. Patel, "Towards a better understanding and regularization of GAN training dynamics," in *Proc. Uncertain. in Artif. Intell.*, pp. 281–291 (2020).
12. L. Chen et al., "Self-supervised learning for medical image analysis using image context restoration," *Med. Image Anal.* **58**, 101539 (2019).
13. N. Tajbakhsh et al., "Convolutional neural networks for medical image analysis: full training or fine tuning?" *IEEE Trans. Med. Imaging* **35**(5), 1299–1312 (2016).
14. Q. Yang et al., "Quantification of muscle, bones and fat on single slice thigh CT," *Proc. SPIE* **12032**, 120321K (2022).
15. K. Engelke et al., "Quantitative analysis of skeletal muscle by computed tomography imaging—state of the art," *J. Orthop. Transl.* **15**, 91–103 (2018).
16. Z. Zhou et al., "Unet++: a nested U-net architecture for medical image segmentation," *Lect. Notes Comput. Sci.* **11045**, 3–11 (2018).
17. M. Jenkinson et al., "FSL," *Neuroimage* **62**, 782–790 (2012).
18. O. Çiçek et al., "3D U-net: learning dense volumetric segmentation from sparse annotation," *Lect. Notes Comput. Sci.* **9901**, 424–432 (2016).

Qi Yang is a graduate student at Vanderbilt University. He received his BS and MS degrees from Beijing Institute of Technology in 2016 and 2019, respectively. He is a member of SPIE.

Biographies of the other authors are not available.