Fully Automated 3D Segmentation of MR-Imaged Calf Muscle Compartments: Neighborhood Relationship Enhanced Fully Convolutional Network

Zhihui Guo[a,1], **Honghai Zhang**[a], **Zhi Chen**[a], **Ellen van der Plas**[b], **Laurie Gutmann**[c], **Daniel Thedens**[b], **Peggy Nopoulos**[b], **Milan Sonka**[a]

[a]Iowa Institute for Biomedical Imaging, University of Iowa, Iowa City, IA 52242, USA

[b]Department of Psychiatry, University of Iowa, Iowa City, IA 52242, USA

[c]Department of Neurology, University of Iowa, Iowa City, IA 52242, USA.

Abstract

Automated segmentation of individual calf muscle compartments from 3D magnetic resonance (MR) images is essential for developing quantitative biomarkers for muscular disease progression and its prediction. Achieving clinically acceptable results is a challenging task due to large variations in muscle shape and MR appearance. In this paper, we present a novel fully convolutional network (FCN) that utilizes contextual information in a large neighborhood and embeds edge-aware constraints for individual calf muscle compartment segmentations. An encoder-decoder architecture is used to systematically enlarge convolution receptive field and preserve information at all resolutions. Edge positions derived from the FCN output muscle probability maps are explicitly regularized using kernel-based edge detection in an end-to-end optimization framework. Our method was evaluated on 40 T1-weighted MR images of 10 healthy and 30 diseased subjects by 4-fold cross-validation. Mean DICE coefficients of 88.00%–91.29% and mean absolute surface positioning errors of 1.04–1.66 mm were achieved for the five 3D muscle compartments.

Keywords

Calf muscle compartment segmentation; fully convolutional network; edge constraint; magnetic resonance image; 3D

1. Introduction

Calf muscle is a skeletal muscle group in the lower leg between the knee joint and the ankle, primarily supporting weight-bearing activities such as walking, running, and jumping. Anatomically, the group can be divided into five individual muscle compartments: Tibialis

---

[1]Corresponding author. zhihui-guo@uiowa.edu.

Anterior (TA), Tibialis Posterior (TP), Soleus (Sol), Gastrocnemius (Gas), and Peroneus Longus (PL) [1], as shown in Fig. 1(a). Volumetric and structural changes of the muscle are important for evaluating muscular disease severity and progression. For example, myotonic dystrophy type 1 (DM1), the most common form of inherited muscular dystrophy in adults, causes severe fatty degeneration of calf muscle in most of the patients [2]. Magnetic resonance (MR) imaging has been widely used in the clinic for muscular disease diagnosis and follow-up evaluation due to its high sensitivity to dystrophic changes [2, 3]. Changes in MR images also correlate with clinical outcome measures potentially serving as imaging biomarkers for clinical research [4]. Current analysis approaches invariably include hand-tracing of individual compartments that is time consuming and less than ideal for clinical trials. Automated segmentation of calf muscle is therefore essential for developing quantitative biomarkers of muscular disease progression and can contribute to its prediction.

A plethora of methods has been developed to separate calf muscle region, subcutaneous adipose tissue (SAT), and intermuscular adipose tissue (IMAT) from the lower leg tissue region, based on which a muscle fat percentage can be obtained as a measurement of fatty degeneration/infiltration that has been shown to have correlation with disease progression [5]. For example, Valentinitsch *et al.* [6] applied multi-stage K-means clustering [7] to segment calf muscle, SAT and IMAT. Amer *et al.* [8] first used a fully convolutional network (FCN) to segment the whole muscle mask and then classified healthy muscle and IMAT from the segmented mask by deep convolutional auto-encoder. These whole muscle segmentation approaches rely on the intra-object homogeneity to separate muscle and adipose tissues. Afterward, an overall muscle fat percentage can be obtained. However, specific muscle compartments may be more affected in different neuromuscular diseases (e.g., the posterior compartment shows initial changes in this population [9]). Assessing the changes in individual muscle compartments may be more sensitive than measuring change in the whole calf. In order to study the disease progression in individual muscle compartments [10, 11, 12], segmentation of individual muscle compartments is critical.

Different from the entire muscle-region segmentation, automated segmentation of individual muscle compartments is a challenging task due to the unique characteristics of MR muscle images as shown in Fig. 1. All non-diseased muscle compartments have similar appearances while the MR bias field exists across the whole image region. Besides, muscular dystrophy can introduce substantial appearance changes to a part or the whole compartment (Fig. 1d–e). Therefore, identifying individual compartments using only local characteristics is unrealistic. On the other hand, shape model based approaches are unsuitable due to large shape variations and deformations caused by the disease as well as by patient's positioning in the scanner.

Attempts to segment individual muscle compartments on MR images are rare. Essafi *et al.* [13] used a landmark-based approach with diffusion wavelets to represent shape variations for 3D segmentation of TA and PL and achieved a mean DICE coefficient of 0.55. Wang *et al.* [14] encoded shape prior by a point distribution model in a higher-order Markov Random Field framework to segment the medial Gas compartment on the same dataset used in [13] and obtained an averaged landmark error around 7 mm. Commean *et al.* [12] presented a semi-automated method to segment five individual muscle compartments by thresholding

and edge detection to study MR imaging measurement reproducibility. Troter *et al.* [15] used a multi-atlas registration approach for individual muscle segmentation in quadriceps femoris and achieved an averaged DICE of 0.87 ± 0.11 on an MRI dataset of healthy young adults. Rodrigues *et al.* [16] adopted a two-stage mechanism to segment individual muscle compartments by first identifying all muscle voxels using Adaboost classifier and then registering the muscle mask to a reference atlas for muscle compartment labeling. However, the individual muscle compartment segmentation results were not visually promising and the accuracies were not reported.

Compared with traditional techniques, FCN has a large model capacity to learn complex representations and enables pixel-to-pixel training, which makes it suitable for this application. To overcome difficulties mentioned above, an FCN with strengthened neighborhood relationship is desired. Many methods have been reported that imposed high-level neighborhood-aware or edge-aware relationships to either refine FCN outputs or directly change FCN internal architectures. Bauer *et al.* [17] adopted a Conditional Random Field strategy to regularize classification results for brain tumor segmentation. Similarly, Guo *et al.* [18] applied a topology-wise graph to refine FCN output for pancreatic tumor segmentation. Both Chen *et al.* [19] and Shen *et al.* [20] used a multi-task FCN to predict object region and edge maps simultaneously for histological object segmentation and brain tumor segmentation respectively, with each task regularized by a cross entropy loss term. Recently, Kampffmeyer *et al.* [21] proposed a single-task FCN to predict pixel level connectivity maps based on n-neighborhood (n=4, 8) relationships, and reverted the prediction to segmentation masks using pixel-pair connectivity agreement.

In this paper, we propose a novel neighborhood relationship-aware FCN based on a variant of 3D UNet [22], called FilterNet, for automated segmentation of all five calf muscle compartments. We enhance neighborhood relationships in two ways: efficiently enlarge convolution receptive field and explicitly derive object boundaries directly from object prediction maps in an end-to-end training optimization framework. Specifically, by enlarging the convolution receptive field, information in a larger neighborhood is taken into consideration when generating the prediction for each central voxel, increasing the model robustness. Additionally, we use kernel-based edge detectors on the prediction maps to regularize the voxel-level probability dissimilarity inside a neighborhood region defined by the kernel size. Motivations behind such kernel edge detector-based constraints are 3-fold. First, the sizes of medical datasets are often small, which makes it not favorable to learn edge regularization from scratch. Edge detectors assess pre-defined neighbor relations (often using derivative formulas) and can be regarded as high-level initialization of the edge regularization module. Second, kernel edge detectors played an important role in medical image segmentation approaches over the past several decades, and they are compatible with CNN end-to-end training. Third, this mechanism provides flexibility to further fine-tune hyper-parameters of the kernel, by making the parameters of interest trainable.

Compared with previously reported approaches, our work has several contributions: a) we report a fully automated approach for 3D segmentation of five calf muscle compartments simultaneously; b) by considering the similar textures shared by individual muscles, we are

able to efficiently impose edge constrains in an end-to-end training manner; c) our method is robust in MR images from both healthy subjects and patients with DM1.

To the best of our knowledge, this is the first automated approach for five calf muscle compartments segmentation. Methodologically, our work is the first attempt to regularize neighborhood relationships in the form of kernel-based edge detection on prediction maps that allows direct back-propagation.

2. Methods

2.1. Pre-processing

In the pre-processing step, a bias field correction method described in [23] is first applied to reduce image intensity inhomogeneity by estimating bias fields as Gaussian distributions and maximizing the high-frequency content of the estimated unbiased image. Each image is further normalized to zero mean and unit variance to reduce inter-subject variations.

Afterwards, to remove large portions of background and reduce model complexity, we utilize Otsu's thresholding [24] and K-means clustering (k=2) to localize and separate left and right leg-areas. All right legs are mirrored to conform to left legs.

The workflow and dimensional change of images are shown in Fig. 2. Note that the pre-processing step is completely unsupervised.

2.2. FilterNet

A typical FCN often consists of an encoder and a decoder. In the encoding phase, multiple levels of features are extracted from the raw input by down-sampling operations to obtain deep and compact representations. The deep representations are then up-sampled to the full resolution in the decoding phase. During the down-sampling and up-sampling steps, a significant problem is the loss of resolution and the associated loss of fine details. UNet [25, 22] and RetinaNet [26] include long skip connections to concatenate encoding features to decoding features at the same feature scale to restore information lost during down-sampling. FC-ResNet [27] further uses residual blocks with identity mapping introduced by ResNet [28] to allow direct gradient back-propagation to earlier layers.

Considering the numerous successes that UNet based neural networks achieved in medical image segmentation problems [18, 8], we based our FilterNet on a UNet-like architecture (still called UNet for simplicity). The network details of both the base UNet we used and our FilterNet are shown in Fig. 3 and Fig. 4. There are mainly two differences between the two networks. First, in order to preserve fine details and enlarge receptive field during the encoding phase, block B is used in the FilterNet. Block B in the FilterNet utilizes short skip connections to enable gradient identity mapping and therefore preserve fine details. It also provides a portal to increase convolution kernel size that allows the network to take account of a broadened view and utilize contextual information from a large neighborhood. However, a large kernel size significantly increases the number of parameters. To avoid this, FilterNet reduces feature channels in block B at down-sampled scales by a factor of 2 compared with block A used in UNet. Second, the FilterNet further employs an edge gate to extract

localized neighborhood relationships in the form of edge detection that allows gradient back-propagation. Different from other works that predict object edges as an extra task to be fused with predictions of object regions [19, 20], FilterNet derives object edge information directly from the object region probability maps through the edge gate and regularizes the derived edge relationship using true edges. In this way, the neighborhood relationship is directly encoded into the region-based probability maps within a single-task FCN, without introducing a number of extra training parameters. In addition, FilterNet implements edge constraints in an end-to-end training manner, it avoids designing and fine-tuning an additional framework for post-processing purpose as [17] and [18] did. Details of block A, block B, edge gate, and loss function are described as follows.

2.2.1.   Backbone blocks—Both blocks A and B work as backbone blocks for the FilterNet, as shown in Fig. 4. Block A consists of double layers of convolution (*Conv*) with kernel size $k = 3 \times 3 \times 3$ (padding $p = (k - 1)/2$, stride $s = 1 \times 1 \times 1$), batch normalization (BN), and rectified linear unit (ReLU) activation function. Block B adds one more *Conv* layer with undecided kernel size ζ at the beginning, and a short skip connection of addition to achieve identity mapping. During the encoding phase, the kernel sizes in block B are set as 7, 5, and 3 for three scales of features to enlarge convolution receptive field. As a result [29], FilterNet increases receptive filed size by 22 voxels along each dimension. Note that due to padding and size restriction, the actual increased receptive field size along z axis is less than 22 voxels.

2.2.2.   Edge gate—Instead of stopping at FCN output of object pixel-wise probability maps as UNet does, an extra step is used in the FilterNet for the edge gate to directly and dynamically derive the true and predicted edge information, respectively, from the ground truth map and the output probability map and to impose constraints on the predicted edges. For the edge gate to support end-to-end training, kernel-based edge detections are used as convolutions with designed kernels inside the network. In this study, Laplacian of Gaussian (LoG) [30] is used. Suppose an input image patch is denoted as I, the function of LoG is defined as

$$F_{LoG}(I) = \kappa_G * \kappa_L * I, \tag{1}$$

where κ_L and κ_G represent the Laplacian kernel and the Gaussian smoothing kernel and $*$ is the convolution operation. Since F_{LoG} finds double edges, while the boundary of muscle compartments in this application is not sufficiently clear to define inside and outside edges, we add an activation function to remove the negative edges, and obtain a new function $F_{L\rho G}$. To increase the flexibility of our edge gate and reduce the dependency on handcrafted parameters, the standard deviation σ of the Gaussian smoothing function is made trainable. Thus, under the condition of a σ, the edge gate respond function can be written as

$$F_{L\rho G}(I \mid \sigma) = \kappa_{G_{(\sigma)}} * \rho(\kappa_L * I), \tag{2}$$

where ρ is a variant of the non-linear hard tanh function that restricts the input value into range [0, 1] such that

$$\rho(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}.$$ (3)

The MR images used in this study have in-plane (on x-y plane) resolution of 0.7 mm and slice thickness (along z direction) of 7 mm. Therefore, convolution kernels κ_L and $\kappa_{G_{(\sigma)}}$ are defined as below and applied to x-y slices.

$$\kappa_L = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \in R^{3 \times 3},$$ (4)

$$\kappa_{G_{(\sigma)}} = \tau [a_{i,j}]_{i,j} \in R^{5 \times 5},$$ (5)

where τ is a constant adjusting factor under a given σ that ensures $\sum_{i=1}^{5} \sum_{j=1}^{5} a_{i,j} = 1$. Element $a_{i,j}$ in $\kappa_{G_{(\sigma)}}$ is defined as,

$$a_{i,j} = \frac{1}{2\pi\sigma^2} e^{-\frac{(i-\bar{i})^2 + (j-\bar{j})^2}{2\sigma^2}}.$$ (6)

In Eq. 6, $\bar{i} = \bar{j} = 3$.

The gradient back-propagation keeps the fixed kernel κ_L unchanged and updates kernel κ_G with trainable σ. For the edge gate, an explicit neighborhood relationship is derived from the probability maps.

2.2.3. Loss function—For an input image $I \in R^{W \times H \times D}$, the FilterNet predicted output map $\hat{Y}$ can be denoted as,

$$\hat{Y} = D_A(E_B(I)) \in R^{N \times W \times H \times D},$$ (7)

where E_B is the encoder that uses block B's and D_A is the decoder that consists of block A's.

The loss function of FilterNet consists of two terms, L_c for region learning amd L_e for edge constraining, balanced by a weighting factor λ such that

$$L = (1 - \lambda)L_c + \lambda L_e,$$ (8)

where L_c is a multi-class cross-entropy loss defined as

$$L_c = \sum_{n=1}^{N} -Y_n \cdot log(\hat{Y}_n),$$ (9)

where $N = 6$, Y_n is the one-hot encoded region label for class n, and $\widehat{Y}_n \in \widehat{Y}$ is the corresponding predicted map for class n. L_e represents the least absolute error (L1-norm) loss between the true edge maps and the derived edge maps. $W = \{w_n | n = 1, …, 6\}$ is a weighting array.

$$L_e = \sum_{n=1}^{N} w_n \left\| F_{L\rho G}(Y_n \mid \sigma) - F_{L\rho G}(\widehat{Y}_n \mid \sigma) \right\|. \tag{10}$$

During gradient back-propagation, the partial derivative of loss L in terms of σ is,

$$\frac{\partial L}{\partial \sigma} = \lambda \frac{\partial L_e}{\partial \sigma}. \tag{11}$$

According to the derivative chain rule,

$$\frac{\partial L_e}{\partial \sigma} = \frac{\partial L_e}{\partial \kappa_{G_{(\sigma)}}} \cdot \frac{\partial \kappa_{G_{(\sigma)}}}{\partial \sigma} = \frac{\partial L_e}{\partial \kappa_{G_{(\sigma)}}} \cdot \left[\frac{\partial a_{i,j}}{\partial \sigma} \right]_{i,j}, \tag{12}$$

where $\frac{\partial L_e}{\partial \kappa_{G_{(\sigma)}}}$ can be obtained from the differentiation of the Gaussian smoothing *Conv* layer, and $\left[\frac{\partial a_{i,j}}{\partial \sigma} \right]_{i,j}$ is calculated according to the definition of $a_{i,j}$ in Eq. 6.

The final multi-class classification output can be defined as the indices of the maximum values on probability maps $\widehat{Y}$ along the channel dimension.

FilterNet is optimized by stochastic gradient descent [31]. The initial learning rate is 10^{-3}, which is divided by 5 every 10 epochs. In order to increase the robustness and generalization of the network, the input training patches are sub-regions sized $120 \times 120 \times 28$, cropped from the localized leg-areas (described in Fig. 2) with a step size of 20 voxels along x and y directions, which results in 9 times as large the number of the training patches as that of the leg-areas. The batch size is set as 2. We train the network with 30 epochs. Data augmentation is performed, where 3 more patches are generated for each training patch. Namely, a rotation value is randomly chosen between -10° to 10°, two scaling factors are randomly chosen between 0.8 and 1.2 along x and y directions. The initial value of λ is 0.001, multiplied by 10 every 10 epochs. W is [0,0.2,0.2,0.15,0.15,0.3] for the 6 classes of background, TA, TP, Sol, Gas, and PL, respectively. The initial value of σ is 1.

3. Experiments and Results

3.1. Experimental Setting

40 lower leg T1-weighted MR images of 40 subjects (10 were healthy, 30 with DM1) were included in this work. The original image size is 512×512×30 and the voxel size is 0.7×0.7×7 mm. The acquisition of these images used the first echo of a 3-point Dixon gradient echo sequence with repetition time (TR) 150 ms, echo time (TE) 3.5 ms, field of

view (FOV) 36 cm, bandwidth 224 Hz/pixel, and scan time 156 s. Expert-traced muscle compartment segmentations served as the independent standard.

Besides FilterNet, we also designed several other neural networks for performance evaluation and comparison purposes. In Section 3.2, based on the differences between UNet and FilterNet, an ablation study was conducted to show the effectiveness of block B and edge gate. Then in Section 3.3, thorough performance comparisons were presented among the UNet, a multi-task FCN that aggregates region and edge predictions, and FilterNet to demonstrate the superiority and efficiency of the FilterNet. All neural networks were implemented using the PyTorch platform [32] and applied to the same dataset. The training parameters were identical for these methods. The models were trained on Nvidia GeForce GTX 1070 GPU with 8 GB of memory.

Given a limited-size dataset, 4-fold cross-validation was used to evaluate the performance of each method. The dataset that included both legs was divided in the 4 fold-groups at the subject level so that data from the same subject were never simultaneously used for both training and testing. In 4-fold cross-validation, the 40 subjects were evenly and randomly divided to 4 groups. Each time, one group was taken as the test set, and the remaining three groups were used as the training set. The process was repeated 4 times so every group served as a test set exactly once. As a result, each subject was used for testing just once. DICE Similarity Coefficient (DSC) and absolute surface-to-surface distance (ASSD, in mm) between the automated surface and the manual surface were used as evaluation metrics. For each subject, the performance was averaged for left and right legs.

3.2. Ablation Study

In order to reveal the performance improvement introduced by block B and edge gate respectively, ablation experiments that included UNet, UNet with block B, UNet with edge gate, and FilterNet were conducted. Therefore, in addition to FilterNet, the other three were described as follows.

- *UNet:* The details of UNet architecture used in this application are displayed in Fig. 3(a). UNet output $\hat{Y}$ is

$$\hat{Y} = D_A(E_A(I)), \tag{13}$$

  where E_A is the encoder consisting of block A's. The loss function is a multi-class cross-entropy loss the same as in Eq. 9.

- *BUNet:* BUNet utilizes block B's in the encoding path of UNet to enlarge convolution receptive field. BUNet output $\hat{Y}$ is the same as FilterNet output in Eq. 7. However, compared with FilterNet, BUNet does not have edge gate. Therefore, λ in Eq. 8 is set as 0, such that the loss function here is

$$L = L_c \tag{14}$$

- *UNet-F:* Similarly, UNet-*F* and UNet share the same network architecture. The prediction output $\hat{Y}$ of UNet-*F* is the same as in Eq. 13. However, different from

UNet, edge gate is added onto $\hat{Y}$ in UNet-*F*. Thus the loss function of UNet-*F* is the same as in Eq. 8.

Fig. 5 shows mean values of DICE and ASSD for five individual muscle compartments obtained from the aforementioned approaches. Overall, FilterNet achieved the best accuracies in terms of DICE and ASSD among the four methods. Both BUNet and UNet-*F* outperformed UNet. However, UNet-*F* had better surface positioning accuracy than BUNet.

3.3. Performance Comparison

In this section, the performance of FilterNet was thoroughly compared with the performance of UNet and a multi-task FCN, called Boundary-Aware FCN.

- *Boundary-Aware FCN:* Boundary-Aware FCN follows the basic idea of the kind of FCN proposed in [20], where the network integrates the predictions for region and edge maps explicitly. The schematic of Boundary-Aware FCN used in this application is shown in Fig. 6. The input patches, encoder and decoders are the same as those in UNet. One decoder D_A^r attempts to predict region maps while the other decoder D_A^e learns the corresponding edge maps. Then the predicted region and edge maps are concatenated and fed into several *Conv*, *BN*, and *ReLU* layers (denoted as Φ) to get the final region prediction $\hat{Y}_c$. Thus, $\hat{Y}_c$ can be described as,

$$\hat{Y}_c = \Phi\big(\hat{Y}_r \odot \hat{Y}_e\big), \tag{15}$$

where $\odot$ is the concatenation operation, $\hat{Y}_r = D_A^r(E_A(I))$ is the predicted region map, and $\hat{Y}_e = D_A^e(E_A(I))$ is the predicted edge map.

The loss function L includes three cross-entropy loss terms, region loss L_r, edge loss L_e, and a final combined loss L_c.

$$L = \sum_{i = e, r, c} L_i = \sum_{i = e, r, c} - Y_i log\big(\hat{Y}_i\big), \tag{16}$$

where each Y_i is the corresponding label map. Note that our implementation of Boundary-Aware FCN uses the same encoder and decoder as described in our baseline UNet design, which is different from the original description in [20].

As a result, Table 1 summarizes DSC and ASSD between the automated segmentations and the independent standard on five calf muscle compartments for the UNet, Boundary-Aware FCN and FilterNet. Compared with UNet, FilterNet generated significantly better results for each compartment in terms of both DICE and ASSD. FilterNet was also shown to have significant differences from Boundary-Aware FCN in DICE for TP and Sol and in ASSD for each muscle compartment.

From top to bottom, Fig. 7 displays four 2D segmentation examples from images of four patients, with each representing a unique situation. The first example shows a normal subject

with calf muscle surrounded by a thick SAT layer. In this case, Gas segmented by UNet had leakage into Sol and SAT, while Gas and Sol segmented by Boundary-Aware FCN also leaked into the SAT layer. The second example is from an MR image of a patient with severe DM1, where TA segmentation obtained from UNet spread into TP and PL, as well as holes existing in Sol. TP segmentation from Boundary-Aware FCN had false positives in true PL, and a hole appeared in Gas. The third example shows notable intensity inhomogeneity around the boundary of TP and Sol. Both UNet and Boundary-Aware FCN were sensitive to the inhomogeneity. The last example presents a 2D cross-sectional slice that is near to one end of the lower leg, where the muscle compartment boundaries are complicated and tough to identify. UNet and Boundary-Aware FCN generated strangely shaped Sol and some voxels inside the true TA were misclassified as TP. In contrast to the situations happened to UNet and Boundary-Aware FCN, though the segmentations from FilterNet were not always perfect, FilterNet was able to relief these problems and appeared to be more robust to image inhomogeneity and object shape maintenance on these examples.

Fig. 8 lists the 3D shapes of the muscle segmentations from the same leg of a patient. From the x-y view, FilterNet generated smoother and topologically superior segmentations. When observing the individual muscle compartment models, UNet and Boundary-Aware FCN showed obvious region leakages of TA, TP, and PL, while FilterNet segmentations were free from such leakages. Table 2 compared the number of model parameters, memory usage and averaged training time per epoch. FilterNet has the lowest number of parameters and memory usage, and UNet ran the fastest during training.

4. Discussion

4.1. Ablation Study

The superiority of BUNet against UNet as shown in Fig. 5 indicates the effectiveness of block B. However, the calculated ASSD index for Gas segmentation by BUNet became worse compared with UNet. Gas has a long and flexible shape on the x-y plane and also shares similar appearance patterns with nearby objects, when neighborhood was broadened while proper edge-aware regulations were lacking, false positives might have been increased across the whole scope and reflected in worsened surface positioning accuracy.

When edge gate was added, UNet-F generated segmentations with apparent improvement in ASSD over UNet. The loss term associated with edge gate in Eq. 10 implies that a) when a pixel is not close to an object boundary, either inside or outside of the object, the probability values of all voxels located inside the neighborhood that belongs to the same class should be similar while b) when a pixel is close to an object boundary, probability value differences between such voxels may show dissimilarity. Therefore, with edge gate, UNet-F was able to impose neighborhood-wise constraints to the predicted output and achieved more topologically correct results.

Compared with UNet, FilterNet reached a new level of accuracy. As we mentioned earlier, when neighborhood is broadened, edge-aware regulations are needed to reduce false positives and strengthen boundary information. Likewise, when edge-aware regulations are presented, broadened neighborhood helps reduce sensitivity to local noise and results in

improved segmentation accuracy. Therefore, the two mechanisms can benefit from each other and lead to the best performance of FilterNet in the ablation study.

4.2. FilterNet vs. UNet, Boundary-Aware FCN

UNet is a single-task FCN that includes only one decoder to learn object region maps, while Boundary-Aware FCN is a multi-task FCN with two decoders to learn region and edge maps respectively and extra layers to fuse the output of the two tasks, with the cost of substantially increased numbers of parameters and computational time, as shown in Table 2. With emphasized boundary information and integration of region and edge learning, Boundary-Aware FCN achieved superior performance in terms of DICE and ASSD compared with UNet (Table 1).

Both FilterNet and Boundary-Aware FCN intended to take advantage of object boundary to improve object segmentation accuracy. Boundary-Aware FCN learns object boundary from scratch and regards predicted boundary maps as extra feature channels into region learning. While FilterNet derives object boundary directly from the predicted region maps and encodes the neighborhood-wise relationship by designed edge detectors with little cost. Instead of learning from scratch, FilterNet avoided introducing a lot of extra parameters to learn the desired edge pattern. The edge detectors turned out to be more efficient than an edge-learning decoder, from the numerical and visual comparisons in Table 1, Fig. 7, and Fig. 8. There are two potential reasons behind such phenomena. First, given a small training dataset, a model with a huge amount of parameters may result in over-fitting, and thus the performance may be compromised. Second, calf muscle compartment boundaries are hard to learn, due to the fact that all muscle compartments share similar appearance patterns. Besides, edge-like appearances caused by intramuscular nerves and vessels [33, 34] can be misleading.

Fig. 7 and Fig. 8 further demonstrated that when multiple objects are close to each other and share very similar textures, optimization that is based on pure pixel-level classification loss may cause disjoint regions or holes inside the true object. When neighborhood-wise dissimilarity penalty were added for voxels away from the boundary, as edge gate did in FilterNet, the situations of disjoint regions and holes were mitigated.

4.3. Approach and Future Work

The application of calf muscle compartment segmentation represents a category of multi-class segmentation problems, where nearby objects are next to each other and have very similar textures. In order to segment each object accurately and at the same time maintain object shape topology, we proposed FilterNet as a neighborhood relationship enhanced FCN that has broadened convolution receptive field and an edge detector based gate to apply constraints directly to the probability maps in an end-to-end training manner.

Besides increasing convolution kernel size as we did with block B, there are other ways to broaden the receptive field or integrate context information from a larger scope. For example, dilated convolutions [35] can be applied to enlarged ranges without increasing filter sizes. Attention [36] has the potential to take into account of multi-scale features

simultaneously by trainable weights. Exploration of more efficient ways to enlarge feature neighborhood will remain as future work.

For the edge gate used in our FilterNet, Laplacian edge detector and Gaussian smoothing kernel with trainable σ were applied to derive object edges. After optimization, we obtained a σ of 0.89, 0.85, 0.92 and 0.90 pixel for each fold, respectively. We have also explored a sole use of the Laplacian edge detector in the edge gate, $F_{L\rho G}$ with fixed σ ($\sigma = 1$) while leaving largest connected component for each label of the results to be generated by $F_{L\rho G}(\sigma = 1)$ as a post processing step. Note that only keeping the largest connected component is only feasible when the desired object is known to be a single region while our method does not have this restriction. It turns out that though the performance differences in terms of DSC are small, the performance differences in terms of ASSD are notable. As the ASSD values shown in Table 3, our current FilterNet outperforms the other tested methods. This means that the usage of Gaussian smoothing makes the edge gate more robust to noise and the neural network has the ability to fine-tune hyper-parameters like σ to further improve the performance. Finding advanced convolution-based detectors with trainable hyper-parameters is also worth exploring in the future. In addition to reducing the sensitivity to noise, Gaussian smoothing also increases edge response area sizes, given extremely small portion of edge voxels in a volume. Due to the sparsity of edge response areas and the optimization efficiency of L1-norm in this case [37], the L1-norm was used in Eq. 10 instead of the L2-norm.

Use of edge constraints in a light-weight manner was previously considered, e.g., by Ronneberger *et al.* [25] who computed a weight map that highlighted the edge areas and calculated the weighted cross-entropy loss. This approach increases the importance of edge areas, but the weight map calculation is based on ground truth only. Our edge gate is applied to prediction maps during each back-propagation, thus has a high penalty in holes and undesired disjoint objects that may otherwise appear during prediction. In [38], deep features are used to directly learn shape parameters that reflect shape differences from a mean shape representation for cervical vertebrae on X-ray images. In this way, disjoint regions are eliminated and the edge is smooth. Cervical vertebrae are rigid with stable (position-independent) shapes, while muscle compartments consist of soft tissues that may change their shapes dramatically in response to changes in position, large deviations from the mean shape representation result. A boundary loss term based on the summation of nearest distances from segmentation pixels to the true boundary in level set representation was proposed in [39] with pixels given penalties according to the distance from the true boundary. However, since the calculation is only carried out for pixels segmented as the target, regions corresponding to holes inside the prediction do not have any penalty associated with them. As an highly relevant improvement, our method penalizes both holes and regions away from the true edges while being computationally efficient. In this study, we applied edge constraints on the axial plane only, due to the extreme anisotropic resolution of the dataset. The clinical reality of MR imaging protocols that are used when imaging calf muscles unfortunately results in such anisotropic acquisition parameters. If more isometric data become available, we plan to use 3D edge detectors in the future and expect to see additional performance gains.

Different from muscle segmentation by atlas-based methods applied to MR images from homogeneous populations [40], the dataset of 40 images we used in this study was quite diverse and included unaffected patients (10/40) and patients with DM1 (30/40). For patients with DM1, the posterior compartment muscles, gastrocnemius and soleus, are usually the earliest and eventually most severely affected muscles. The changes are not limited to these muscles. However, for the purpose of measuring changes over time for eventual clinical trials, being able to track muscles that are less severely affected may be equal or more important than those that are affected earliest and most severely in DM1. In this work, we have shown the feasibility and robustness of our approach in segmenting individual calf muscle compartments in normals and when DM1 was presented. In the future, for each muscle compartment segmentation, disease-affected regions can be easily clustered from the healthy muscle structure to develop compartment-based biomarkers guiding a disease progression study in the clinic.

5. Conclusion

A neighborhood relationship enhanced FCN was reported and applied to individual calf muscle compartment segmentation on T1-weighted MR images. With an increased convolution receptive field, resolution-preserving skip connections, and explicitly edge-aware regulations by a kernel-base edge gate to constrain pixel-level probability values inside a neighborhood, our FilterNet considered the specialty of adjacent multi-class muscle segmentation and delivered a striking performance improvement (DSC>0.88, ASSD<1.66 mm) over previously-reported methods (DSC<0.55, ASSD~7 mm), suggesting clinical-use feasibility of automated calf-compartment segmentation.

Acknowledgment

References

[1]. Yaman A, Ozturk C, Huijing PA, Yucesoy CA, Magnetic resonance imaging assessment of mechanical interactions between human lower leg muscles in vivo, Journal of Biomechanical Engineering 135 (9) (2013) 091003.

[2]. Wattjes MP, Kley RA, Fischer D, Neuromuscular imaging in inherited muscle diseases, European Radiology 20 (10) (2010) 2447–2460. [PubMed: 20422195]

[3]. Stramare R, Beltrame V, Dal Borgo R, Gallimberti L, Frigo A, Pegoraro E, Angelini C, Rubaltelli L, Feltrin G, MRI in the assessment of muscular pathology: a comparison between limb-girdle muscular dystrophies, hyaline body myopathies and myotonic dystrophies, La Radiologia Medica 115 (4) (2010) 585–599. [PubMed: 20177980]

[4]. Willcocks RJ, Rooney WD, Triplett WT, Forbes SC, Lott DJ, Senesac CR, Daniels MJ, Wang DJ, Harrington AT, Tennekoon GI, Russman BS, Finanger EL, Byrne BJ, Finkel RS, Walter GA, Sweeney HL, Vandenborne K, Multicenter prospective longitudinal study of magnetic resonance biomarkers in a large duchenne muscular dystrophy cohort, Ann Neurol 79 (4) (2016) 535–547. [PubMed: 26891991]

[5]. Wren TA, Bluml S, Tseng-Ong L, Gilsanz V, Three-point technique of fat quantification of muscle tissue as a marker of disease progression in duchenne muscular dystrophy: preliminary study, American Journal of Roentgenology 190 (1) (2008) W8–W12. [PubMed: 18094282]

[6]. Valentinitsch A, Karampinos DC, Alizai H, Subburaj K, Kumar D, Link TM, Majumdar S, Automated unsupervised multi-parametric classification of adipose tissue depots in skeletal muscle, Journal of Magnetic Resonance Imaging 37 (4) (2013) 917–927. [PubMed: 23097409]

[7]. MacQueen J, et al., Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Oakland, CA, USA, 1967, pp. 281–297.

[8]. Amer R, Nassar J, Bendahan D, Greenspan H, Ben-Eliezer N, Automatic segmentation of muscle tissue and inter-muscular fat in thigh and calf MRI images, in: MICCAI, Springer, 2019, pp. 219–227.

[9]. Heskamp L, van Nimwegen M, Ploegmakers MJ, et al., Lower extremity muscle pathology in myotonic dystrophy type 1 assessed by quantitative MRI, Neurology 92 (24) e2803–e2814.

[10]. Gourgiotis S, Villias C, Germanos S, Foukas A, Ridolfini MP, Acute limb compartment syndrome: a review, Journal of Surgical Education 64 (3) (2007) 178–186. [PubMed: 17574182]

[11]. Alizai H, Nardo L, Karampinos DC, Joseph GB, Yap SP, Baum T, Krug R, Majumdar S, Link TM, Comparison of clinical semi-quantitative assessment of muscle fat infiltration with quantitative assessment using chemical shift-based water/fat separation in MR studies of the calf of post-menopausal women, European radiology 22 (7) (2012) 1592–1600. [PubMed: 22411305]

[12]. Commean PK, Tuttle LJ, Hastings MK, Strube MJ, Mueller MJ, Magnetic resonance imaging measurement reproducibility for calf muscle and adipose tissue volume, J Magn Reson Imaging 34 (6) (2011) 1285–1294. [PubMed: 21964677]

[13]. Essafi Salma and Langs Georg and Jean-François Deux and Alain Rahmouni and Guillaume Bassez and Nikos Paragios, Wavelet-driven knowledge-based MRI calf muscle segmentation, in: ISBI, IEEE, 2009, pp. 225–228.

[14]. Wang C, Teboul O, Michel F, Essafi S, Paragios N, 3D knowledge-based segmentation using pose-invariant higher-order graphs, in: MICCAI, Springer, 2010, pp. 189–196.

[15]. Le Troter A, Fouré A, Guye M, Confort-Gouny S, Mattei J-P, Gondin J, Salort-Campana E, Bendahan D, Volume measurements of individual muscles in human quadriceps femoris using atlas-based segmentation approaches, Magnetic Resonance Materials in Physics, Biology and Medicine 29 (2) (2016) 245–257.

[16]. Rodrigues R, Pinheiro AM, Segmentation of skeletal muscle in thigh Dixon MRI based on texture analysis, arXiv preprint arXiv:1904.04747

[17]. Bauer S, Nolte L-P, Reyes M, Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization, in: MICCAI, Springer, 2011, pp. 354–361.

[18]. Guo Z, Zhang L, Lu L, Bagheri M, Summers RM, Sonka M, Yao J, Deep LOGISMOS: deep learning graph-based 3D segmentation of pancreatic tumors on CT scans, in: ISBI, IEEE, 2018, pp. 1230–1233.

[19]. Chen H, Qi X, Yu L, Dou Q, Qin J, Heng P-A, DCAN: Deep contouraware networks for object instance segmentation from histology images, Medical Image Analysis 36 (2017) 135–146. [PubMed: 27898306]

[20]. Shen H, Wang R, Zhang J, McKenna SJ, Boundary-aware fully convolutional network for brain tumor segmentation, in: MICCAI, Springer, 2017, pp. 433–441.

[21]. Kampffmeyer M, Dong N, Liang X, Zhang Y, Xing EP, ConnNet: A long-range relation-aware pixel-connectivity network for salient segmentation, TIP 28 (5) (2019) 2518–2529.

[22]. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International conference on medical image computing and computer-assisted intervention, Springer, 2016, pp. 424–432.

[23]. Tustison N, Gee J, N4ITK: Nicks N3 ITK implementation for MRI bias field correction, Insight Journal 9.

[24]. Otsu N, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics 9 (1) (1979) 62–66.

[25]. Ronneberger O, Fischer P, Brox T, U-net: Convolutional networks for biomedical image segmentation, in: MICCAI, Springer, 2015, pp. 234–241.

[26]. Lin T-Y, Goyal P, Girshick R, He K, Dollár P, Focal loss for dense object detection, in: ICCV, IEEE, 2017, pp. 2980–2988.

[27]. Drozdzal M, Chartrand G, Vorontsov E, Shakeri M, Di Jorio L, Tang A, Romero A, Bengio Y, Pal C, Kadoury S, Learning normalized inputs for iterative estimation in medical image segmentation, Medical Image Analysis 44 (2018) 1–13. [PubMed: 29169029]

[28]. He K, Zhang X, Ren S, Sun J, Identity mappings in deep residual networks, in: ECCV, Springer, 2016, pp. 630–645.

[29]. Dumoulin V, Visin F, A guide to convolution arithmetic for deep learning, arXiv preprint arXiv:1603.07285

[30]. Sonka M, Hlavac V, Boyle R, Image processing, analysis, and machine vision, Cengage Learning, 2014.

[31]. Bottou L, Large-scale machine learning with stochastic gradient descent, in: Proceedings of COMPSTAT'2010, Springer, 2010, pp. 177–186.

[32]. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A, Automatic differentiation in PyTorch, in: NIPS Autodiff Workshop, 2017.

[33]. FUXE K, SEDVALL G, The distribution of adrenergic nerve fibres to the blood vessels in skeletal muscle, Acta physiologica Scandinavica 64 (1–2) (1965) 75–86. [PubMed: 14347281]

[34]. Sinha S, Sinha U, Edgerton VR, In vivo diffusion tensor imaging of the human calf muscle, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 24 (1) (2006) 182–190.

[35]. Yu F, Koltun V, Multi-scale context aggregation by dilated convolutions, arXiv preprint arXiv:1511.07122

[36]. Chen L-C, Yang Y, Wang J, Xu W, Yuille AL, Attention to scale: Scale-aware semantic image segmentation, in: CVPR, IEEE, 2016, pp. 3640–3649.

[37]. Melkumova L, Shatskikh SY, Comparing ridge and lasso estimators for data analysis, Procedia engineering 201 (2017) 746–755.

[38]. Al Arif SMR, Knapp K, Slabaugh G, Spnet: Shape prediction using a fully convolutional neural network, in: MICCAI, Springer, 2018, pp. 430–439.

[39]. Kervadec H, Bouchtiba J, Desrosiers C, Granger E, Dolz J, Ayed IB, Boundary loss for highly unbalanced segmentation, in: MIDL, 2019, pp. 285–296.

[40]. Nguyen H-T, Croisille P, Viallon M, Leclerc S, Grange S, Grange R, Bernard O, Grenier T, Robustly segmenting quadriceps muscles of ultra-endurance athletes with weakly supervised U-Net.
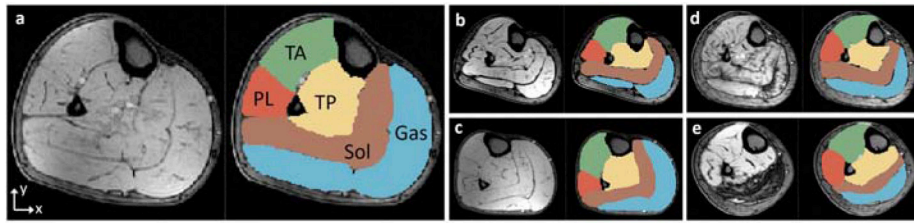
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1:
Examples of T1-weighted MR images of calf muscle cross sections. Each panel shows the original image and the corresponding expert segmentations for TA, TP, Sol, Gas and PL. (a–c) Normal subjects. (d–e) Patients with severe DM1. Best viewed in color.
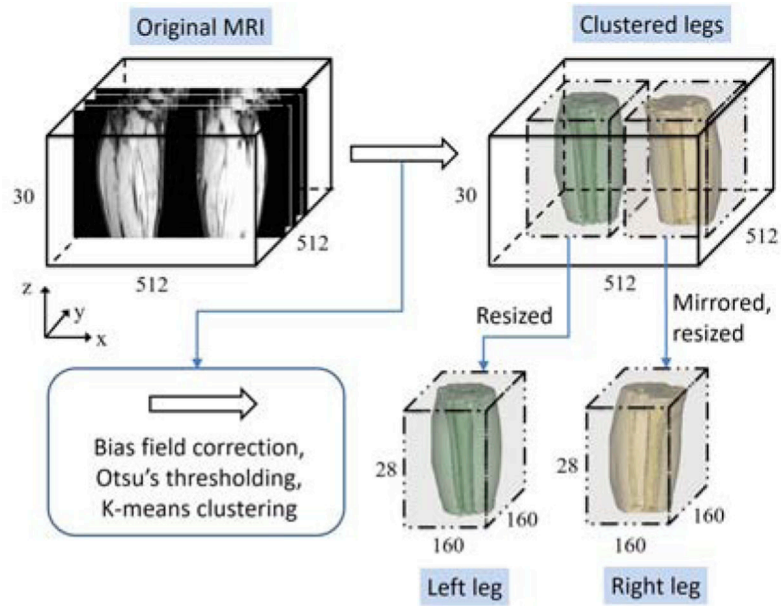
Figure 2:
Workflow of the pre-processing step. After Bias field correction, Otsu's thresholding, and K-means clustering, each leg-area is localized on the original MR images. The leg-areas are then extracted and resized to get a uniform dimensionality. Right leg is mirrored to left.
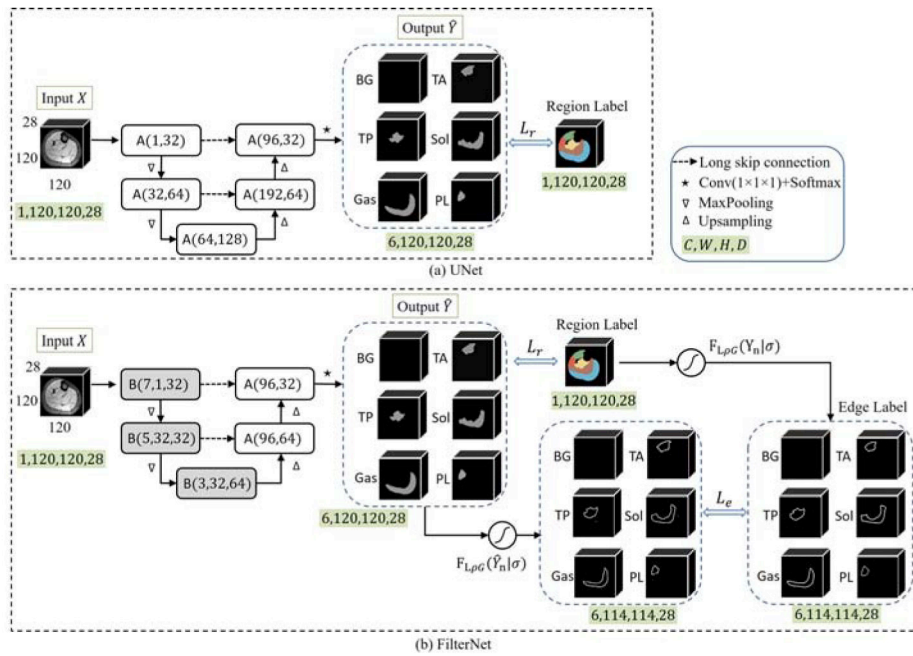
Figure 3:
UNet (a) vs. FilterNet (b) for multi-class calf muscle segmentation. The two networks have the same input and output. [C, W, H, D]: channel, width, height, depth. "BG" represents background. Best viewed in color.
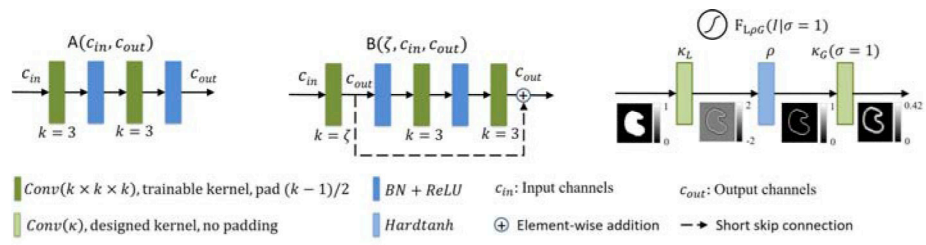
Figure 4:
Details of block A, block B, and edge gate in FilterNet. An example is shown along with edge gate flow chart. Best viewed in color.
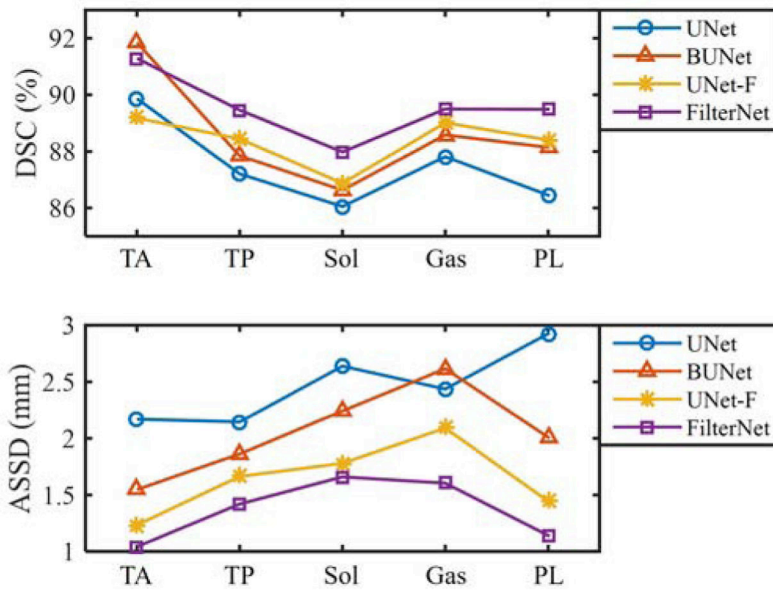
Figure 5:
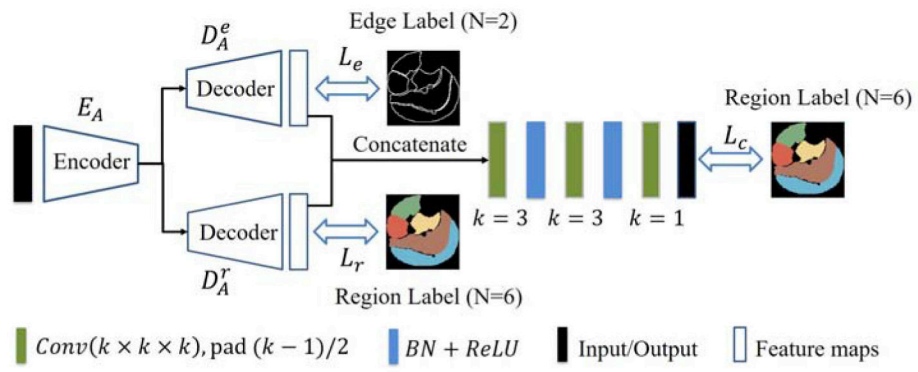Mean values of DICE and ASSD (in mm) from UNet, BUNet, UNet-*F* and FilterNet.

Figure 6:
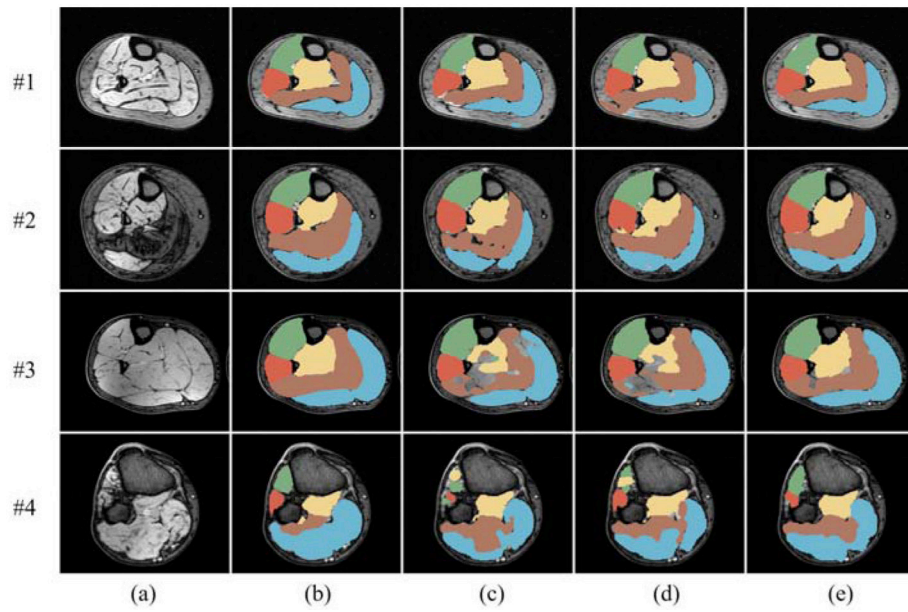The schematic of Boundary-Aware FCN. Best viewed in color.

Figure 7:
Cross-sectional segmentation examples overlaid with MR images. (a) Original scan. (b) Ground truth. (c) UNet. (d) Boundary-Aware FCN. (e) FilterNet. Each row represents a 2D cross-sectional example from a different image. Best viewed in color.
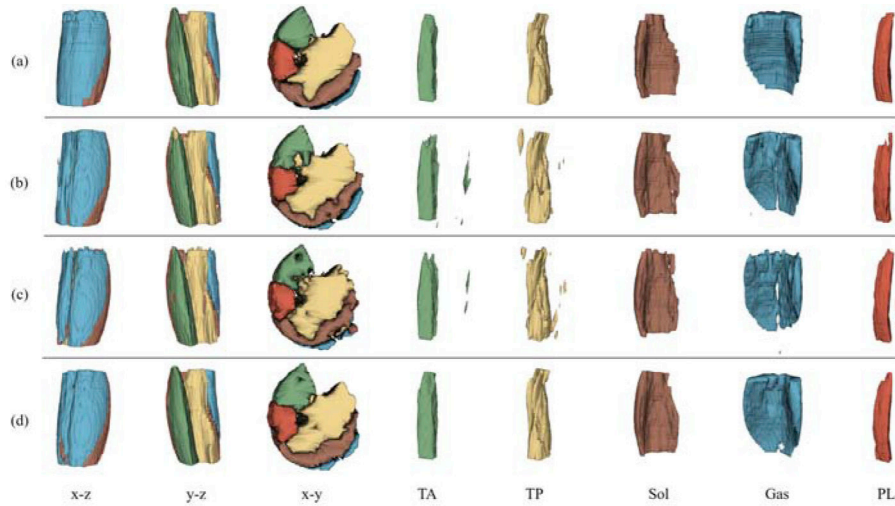
Figure 8:

3D demonstration of muscle compartment segmentations from the same leg. Each row represents a method while each column represents a view category. (a) Ground truth. (b) UNet. (c) Boundary-Aware FCN. (d) FilterNet. The first three columns are three orthogonal views of the five muscle compartment segmentations, followed by five columns showing individual segmentations. Note the extra objects generated by UNet and Boundary-Aware FCN for TA, TP, Gas and PL. Best viewed in color.

Table 1:

DSC and ASSD (mean±std) for five calf muscle compartments from UNet, Boundary-Aware FCN and FilterNet. The unit for ASSD is mm. Statistical significance in bold.

Methods		UNet		Boundary-Aware FCN		FilterNet	
		Mean±STD	p value*	Mean±STD	p value*	Mean±STD	p value*
TA	DSC	89.86±11.07	**0.033**	90.51±13.22	0.333	91.29±10.11	/
	ASSD	2.17±2.00	**≪0.001**	1.80±1.97	**≪0.001**	1.04±0.81	/
TP	DSC	87.20±6.89	**0.007**	88.11±6.16	**0.043**	89.46±4.19	/
	ASSD	2.15±1.39	**≪0.001**	2.10±1.56	**0.002**	1.42±0.66	/
Sol	DSC	86.05±8.94	**≪0.001**	87.09±8.50	**0.041**	88.00±7.92	/
	ASSD	2.64±1.87	**≪0.001**	2.35±1.87	**0.001**	1.66±<.82	/
Gas	DSC	87.81±10.19	**0.017**	88.33±9.84	0.075	89.50±7.99	/
	ASSD	2.44±1.99	**≪0.001**	2.38±2.17	**0.001**	1.60±1.29	/
PL	DSC	86.45±12.36	**≪0.001**	89.22±11.12	0.568	89.49±12.44	/
	ASSD	2.93±3.37	**≪0.001**	1.64±1.60	**0.007**	1.14±1.02	/

*Paired t-test with FilterNet (significance level $p < 0.05$)

Table 2:

Number of model parameters, memory usage, and averaged training time per epoch for UNet, Boundary-Aware FCN and FilterNet. Best performance in bold.

Methods	# parameters (Millions)	Memory (GB)	Training time (mins)
UNet	1.97	4.67	**11**
Boundary-Aware FCN	2.39	6.01	25
FilterNet	**1.44**	**3.76**	16

Averaged ASSD for each muscle compartment. F_L: FilterNet with only Laplacian kernel used in the edge gate. $F_{L\rho G}(\sigma = 1)$: FilterNet with fixed σ=1 in the Gaussian kernel. $F_{L\rho G}(\sigma = 1)$ & LLC: leaving largest connected component for each label is applied to the results of $F_{L\rho G}(\sigma = 1)$. $F_{L\rho G}(\cdot/\sigma)$: FilterNet with trainable σ. Best performance in bold.

Method	TA	TP	Sol	Gas	PL
F_L	1.57	1.85	2.28	2.24	2.02
$F_{L\rho G}(\sigma = 1)$	1.48	1.84	2.11	2.15	1.85
$F_{L\rho G}(\sigma = 1)$ & LLC	**1.01**	1.56	1.69	1.65	1.21
$F_{L\rho G}(\cdot/\sigma)$	1.04	**1.42**	**1.66**	**1.60**	**1.14**