Automatic paraspinal muscle segmentation in patients with lumbar pathology using deep convolutional neural network

Wenyao Xia[1], Maryse Fortin[2], Joshua Ahn[3], Hassan Rivaz[2,4], Michele C. Battié[5], Terry M. Peters[1], and Yiming Xiao[1]

[1] Robarts Research Institute, Western University, London, Canada
[2] PERFORM Centre, Concordia University, Montreal, Canada
[3] Health, Kinesiology & Applied Physiology, Western University, London, Canada
[4] Electrical and Computer Engineering, Concordia University, Montreal, Canada
[5] School of Physical Therapy and Western's Bone and Joint Institute, Western University, London, Canada

Abstract. Recent evidence suggests an association between low back pain (LBP) and changes in lumbar paraspinal muscle morphology and composition (i.e., fatty infiltration). Quantitative measurements of muscle cross-sectional areas (CSAs) from MRI scans are commonly used to examine the relationship between paraspinal muscle characters and different lumbar conditions. Current investigation primarily uses manual segmentation that is time-consuming, laborious, and can be inconsistent. However, no automatic MRI segmentation algorithms exist for pathological data, likely due to the complex paraspinal muscle anatomy and high variability in muscle composition among patients. We employed deep convolutional neural networks using U-Net+CRF-RNN with multi-data training to automatically segment paraspinal muscles from T2-weigthed MRI axial slices at the L4-L5 and L5-S1 spinal levels and achieved averaged Dice score of 93.9% and mean boundary distance of 1mm. We also demonstrate the application using the segmentation results to reveal tissue characteristics of the muscles in relation to age and sex.

Keywords: segmentation, deep learning, lumbar pathologies, MRI

1 Introduction

Low back pain (LBP) is the most common musculoskeletal disorder in adults, with a lifetime prevalence of up to 84% [1]. To better understand the underlying pathology and facilitate treatment and rehabilitation, there is an increasing interest to study the association between LBP and changes in morphology and composition (e.g., fat vs. muscle ratio) of lumbar paraspinal muscles. Most commonly, paraspinal muscle cross-sectional area (CSA) measurements are obtained from axial magnetic resonance imaging (MRI) scans, with published studies only relying on time-consuming and expertise-intensive manual segmentation for their analyses. So far, only a few methods [2-4] have been proposed for automatic

paraspinal muscles segmentation, mostly based on computed tomography (CT) images and healthy subjects. However, much higher morphometric and composition variations exist in patients with LBP. In this paper, we use deep learning to automatically identify paraspinal muscles, including the multifidus and erector spinae muscles, as well as the intervertebral disc, spinal bone (i.e., spinous and transverse processes), and psoas muscles for computer-assisted analysis of LBP. More specifically, we leverage the strength of U-Net [5] and conditional random fields (CRF)-based probabilistic graphical modelling, which is reformulated as a Recurrent Neural Network (RNN) [6] to incorporate spatial information in tissue labeling. Furthermore, we also demonstrate the application of the proposed technique to help characterize muscle tissue properties in relation to the factors of sex and age, among a small cohort of patients with LBP.

2 Materials and Methods

2.1 Subjects, imaging, and preprocessing

Lumbosacral T2-weighted (T2w) MR images of 112 patients (59 male, age=30∼59y) were selected from the European research consortium project, Genodisc, on commonly diagnosed lumbar pathologies (physiol.ox.ac.uk/genodisc). For each sex, the subjects' ages are roughly uniformly distributed for the included range. Axial MRI slices of the L4-L5 and L5-S1 spinal levels at mid-disc were acquired for analysis. The multifidus (MF), erector spinae (ES) and psoas muscles, as well as the disc and spinal bone (9 labels) were manually segmented for all subjects at the L4-L5 and L5-S1 levels, using the software ITK-SNAP (itksnap.org). All cross-sectional MR images were first processed with N4 inhomogeneity correction [7] to remove field non-uniformity in the image (see Fig. 1). Then, the processed images were linearly transformed to the space of population-averaged paraspinal muscle atlases [8] at L4-L5 and L5-S1 spinal levels, and resampled to a standard image size of 256×256 with the resolution of 1x1 mm^2. This mitigates the large individual body size variation and images resolution differences for efficient feature learning. The manual segmentations were also transformed with the associated transformation and resampled to the same image size and resolution with nearest-neighborhood interpolation.

2.2 Deep convolutional neural networks

With deep learning, many biomedical image segmentations can be performed with close-to-human accuracy. Instead of hand-crafting features to identify the target object, the discriminative image features can be learned automatically from examples via convolutional neural networks. In 2015, Ronneberger et al. [5], developed U-Net, which contains feature map concatenations and multiple de-convolution layers with learnable weight filters. In this study, we use U-Net to obtain the probability of each label given input MR image,$P(l_i|I)$ and we train a standard U-Net to directly predict all 9 labels from the input MRI image

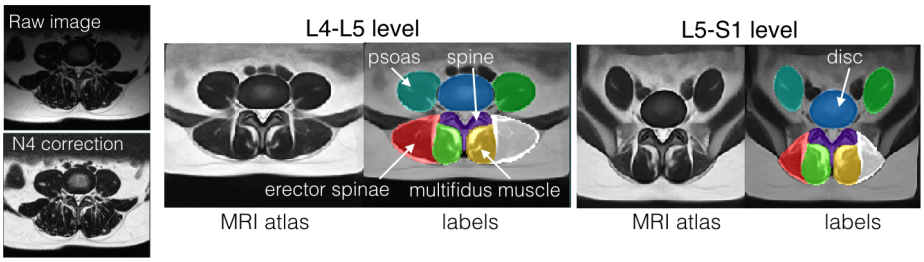

Fig. 1. Cross-sectional MRIs of paraspinal muscles at the L4-L5 and L5-S1 levels with different tissue classes using population-averaged MRI atlases.

simultaneously. As described in [5], the standard U-Net consists of a contracting and expanding path, each with 4 resolution steps with a total of 23 convolution layers. We train our network using a combination of Dice coefficient and cross-entropy loss [9]: $L_{total} = L_{Dice} + L_{CE}$, where the Dice loss is computed for all samples individually and averaged over the batch.

2.3 Conditional random fields

Each muscle or bone component is highly localized and has smooth boundaries. However, due to image noise and local minima in training, the final label map tends to have small scattered mis-classifications. To resolve this problem, we combine conditional random fields [6] as recurrent neural networks with our U-Net to allow spatial constraints between labels. This can improve the delineation and integrity of our segmentations. We formulate our final labels as the inference from CRFs, given the probability map from U-Net and the input image. We model our CRFs to minimize the energy function: $E(x) = \sum_i \phi_u(l_i) + \sum_{i<j} \phi_p(l_i, l_j)$, where $\phi_u(l_i) = -log(P(l_i|I))$ is the unary energy measuring the inverse likelihood of pixel i assigned as label l_i, and $\phi_p(l_i, l_j)$ is the pairwise energy measuring likelihood of the neighboring pixel pair i and j assigned as label l_i and l_j, respectively. The pairwise energy term provides an image dependent smoothness cost, which encourages the neighboring pixels to share similar labels. The label l_i that minimizes above cost function, is chosen as our final label and we use the meanfield approximation algorithm [6] to minimize the cost function with 6 iterations (Fig.2(b)).

2.4 Multi-data training with gradient magnitude map

We also propose multi-data training to improve the efficiency of learning from examples with large morphological and intensity variations. Aside from the input training MRI, we compute the Laplacian gradient magnitude map of the training MRI and use them as additional samples to expand the training dataset. Instead

of predicting labels based on input MRIs, the same network also has to learn how to identify the muscles & bones based on the gradient map, which contains no intensity information. The purpose of such augmentation training is to force the network to extract shared features between the MRI and its gradient magnitude for label classification. As a result, the network favors structural and spatial information from the MRI, and becomes less reliant on its intensity. The added gradient map is used solely for training and only the MRI is used for testing. The benefit of this approach is to increase the robustness of the trained model by reinforcing gradient feature learning (Fig.2 (a)).

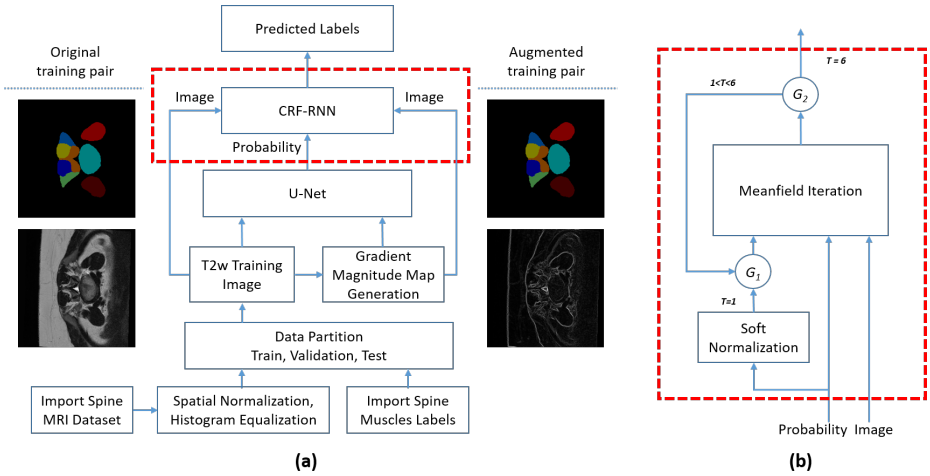

Fig. 2. (a) Overview of the proposed MRI segmentation pipeline. (b) Implementation of CRF-RNN, where G represents gating function and T represents iteration number.

2.5 Implementation and training

Our neural networks were implemented in TensorFlow using the NiftyNet framework [10] and trained on a NIVIDIA Titan XP GPU. We used Adam optimizer to train the network at a learning rate of 0.001. Of 112 patients, 75 were used for training, 10 for validation, and 27 for testing. Both training and testing datasets contained approximately equal splits in terms of gender and sex to fairly include muscle variations due to these factors. For training, the maximum iteration number was set at 15000 with a batch size of 20 and early stopping to prevent over-fitting.

2.6 Muscle-fat separation and analysis

To demonstrate the application of the proposed segmentation method, we employed the segmentation results of the test dataset (27 patients) to obtain the CSAs of the multifidus and erector spinae muscles, as well as their fat percentages. We used k-means to separate the fat and muscle tissues while using the segmented muscle labels to constrain the region of interest. Spearman partial correlation between each of these metrics and the factor of sex or age was computed to reveal the physiological differences between different subgroups, as this relationship is being actively investigated in the LBP research.

3 Results

For segmentation, we first use Dice coefficient, sensitivity (recall), and positive predictive value (PPV) to quantitatively assess the performance of our trained neural networks. Experimental comparisons with standard U-Net and U-Net+CRF-RNN trained only with MRIs are also made to demonstrate the performance improvement using proposed training framework. Since Dice could be biased for imbalanced classes, we also compute the mean boundary distance (MBD) between the segmentations and the groundtruth to evaluate the performances of trained models. The Dice scores and MBDs are shown in Table 1. From our testing, we see that U-Net alone can produce highly accurate segmentation results with mean Dice of 92.4%, mean recall of 92.7%, mean PPV of 93.0%, and MBD of 1.93 mm. After CRF-RNN is incorporated, Dice (92.6%), PPV (93.4%), and MBD (1.48mm) are improved, but there are no significant improvements for recall (92.7%). In comparison, the proposed multi-data training yields the best segmentation results for all muscle and bone components with overall average Dice of 93.8%, recall of 93.9%, PPV of 94.2%, and overall MBD of 1.00 mm. Comparing to U-Net+CRF-RNN trained only with MRIs, introducing multi-data training can improve the segmentation accuracies for all muscles and bones, especially for left & right psoas muscles with 3-4% improvements in terms of Dice and 1-2 mm improvements in terms of MBD. This is further reflected through the significant decrease in standard deviations for both Dice and MBD as shown in Table 1, indicating improvements in terms of robustness. We also performed paired-sample t-test using averaged Dice and MBD over all the labels. Our segmentations with multi-data training shows significant ($p<0.05$) and consistent improvement over U-Net+CRF-RNN with only MRI training both in terms of Dice ($p=0.001$) and MBD ($p= 0.043$). As for U-Net vs. U-Net+CRF only trained with MRIs, there is no significant improvement in Dice ($p=0.15$), but there is significant improvement in terms of MBD ($p = 0.004$).

To further illustrate the performance of our network, segmentation results for the two spine levels from our trained network using proposed framework are shown in Fig. 3. As seen from Fig. 3(c), U-Net+CRF-RNN reduces the small scattered false positives and smoothes the segmentation boundary. From Fig. 3(d), our network yields highly accurate segmentations for all muscle and bone components compared to the ground truth despite large muscle variations.

Table 1. Segmentation performance in different setups measured in Dice and mean boundary distance as mean ± standard deviation

Dice (%)	spine	disc	left psoas	right psoas	left MF	right MF	left ES	right ES	overall
U-Net	91.8 ±2.9	96.9 ±1.9	90.5 ±15.1	91.5 ±15.5	94.4 ±2.6	93.9 ±3.0	89.7 ±6.5	90.5 ±5.3	92.4 ±3.9
U-Net+CRF	91.3 ±5.3	97.0 ±2.4	90.9 ±15.1	91.9 ±15.5	94.4 ±2.4	94.2 ±2.3	89.9 ±6.4	90.8 ±5.1	92.6 ±3.9
Multi-data training	92.7 ±2.9	97.4 ±1.1	94.6 ±4.3	94.8 ±4.0	95.0 ±2.3	94.5 ±1.9	90.6 ±5.8	91.3 ±4.6	93.9 ±0.2
Mean Boundary Distance (mm)	**spine**	**disc**	**left psoas**	**right psoas**	**left MF**	**right MF**	**left ES**	**right ES**	**overall**
U-Net	1.26 ±1.4	1.09 ±1.8	2.99 ±4.6	4.04 ±13.6	1.10 ±0.7	1.30 ±1.0	1.89 ±1.1	1.80 ±1.3	1.93 ±1.9
U-Net+CRF	0.97 ±1.0	0.84 ±0.8	2.10 ±2.8	2.93 ±13.5	0.96 ±0.4	1.05 ±0.5	1.61 ±0.9	1.42 ±0.8	1.48 ±1.8
Multi-data training	0.68 ±0.2	0.57 ±0.3	1.18 ±1.0	0.93 ±0.6	0.87 ±0.4	0.94 ±0.4	1.46 ±0.8	1.34 ±0.6	1.00 ±0.3

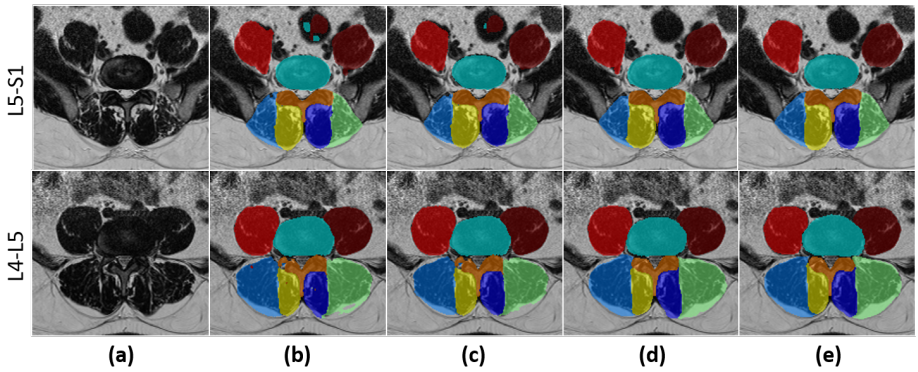

Fig. 3. Segmentation examples with MRIs at the L4-L5 and L5-S1 levels shown on the first and second rows, respectively. Columns: (a) Input T2w MRI (b) U-Net segmentation (c) U-Net+CRF-RNN segmentation (d) Our segmentation (e) Ground truth.

The results of the k-means algorithm to separate the muscle and fat tissue are demonstrated in Fig. 4. From the muscle morphometric analysis in the atlas reference space, at the L4-L5 level, the female group is significantly correlated ($p < 0.05$) with higher fat percentage within the multifidus muscles bilaterally

(r=0.46 for both sides) while controlling for age. At the same time, this metric within the left erector spinae muscles is positively correlated with increasing age (r=0.40). At the L5-S1 level, the multifidus muscle fat content is correlated with sex (female) on both sides (r=0.53 and 0.47 for the left and right). Our analysis with CSAs did not yield any significant correlations.

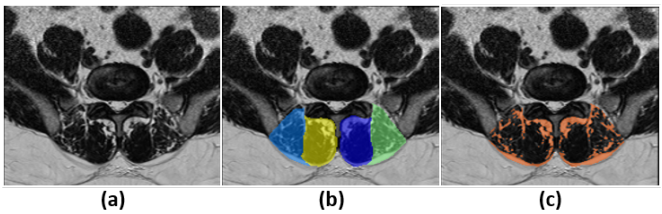

Fig. 4. Fat separation using k-means. (a) Input T2w MRI (b) Segmentations of multifidus and spinae muscles (c) Separated fat within the muscles (in orange color).

4 Discussion and future work

Our results demonstrate that adding gradient magnitude images as additional training set can significantly improve the segmentation accuracy without increase network complexity. Hence, it is easy to implement, computationally efficient, and can be generalized to other MR or CT images. Although this technique is unconventional as we normally desire the network to learn the gradient features automatically, it can be very helpful to improve learning efficiency in applications with small datasets. Leveraging the automatic segmentation results, the muscle morphometric analysis demonstrates that female sex and aging is correlated with increased fatty infiltration in multifidus and erector spinae muscles among patients with lumbar pathologies. This trend is consistent with previous reports [11]. While the trend of fatty infiltration is similar on both sides of the body, only the left erector spinae muscle showed significant correlation with age. This is likely due to the side that is commonly affected by disease.

CRF-RNN is able to reduce issues of island labels when only using U-Net. Although many directly employ CRFs models as a post-processing step for U-Net, the RNN implementation allows end-to-end training. However, with high variations of muscle tissue properties, separation between muscles, especially between multifidus and erector spinae muscles, can be challenging. Future work will include improved definition of muscle groups with multi-contrast MRI, and we will also seek to incorporate models to enhance our network with the implied shape constraints to further improve the boundary smoothness of the segmentation. In the future, we will also include a wider age range of patients in the training set to further improve our network.

5 Conclusion

We have proposed the first technique with deep convolutional neural networks to automatically segment the paraspinal muscles from MRIs in patients with LBP, and further demonstrated its application to facilitate computer-assisted analysis of muscle characteristics. With CRF-RNN to help add spatial constraints to the tissue labeling, we proposed to incorporate additional gradient magnitude maps of the original images in training to enhance the performance without adding network complexity. We expect the resulting technique to greatly benefit the investigation of paraspinal muscle and common lumbar disorders.

6 Acknowledgment

This work was supported by CIHR, CFI, NSERC and BrainsCan, as well as the Seventh Framework Programme (Health-2007-2013, grant agreement NO: 201626: GENODISC) and Canada Reearch Chairs program. We acknowledge the support of NVIDIA Corporation and thank Dr. Yingli Lu for his help.

References

1. Balagu, F., Mannion, A. F., Pellis, F., Cedraschi, C.: Non-specific low back pain. The lancet 379(9814), 482–491 (2012)
2. Kamiya, N., Li, J., Kume, M., et al.: Fully automatic segmentation of paraspinal muscles from 3D torso CT images via multi-scale iterative random forest classifications. Int J Comput Assist Radiol Surg 13(11), 1697–1706 (2018)
3. Engstrom, C. M., Fripp, J., Jurcak, V.,et al.: Segmentation of the quadratus lumborum muscle using statistical shape modeling. Journal of Magnetic Resonance Imaging 33(6), 1422-1429 (2011)
4. Wei, Y., Xu, B., Tao, X., Qu, J.: Paraspinal muscle segmentation in CT images using a single atlas. In: Proc. PIC. pp. 211–215. IEEE (2015)
5. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI, vol. 9351, pp.234–241. Springer (2015)
6. Zheng, S., Jayasumana, S., Romera-Paredes, et al.: Conditional random fields as recurrent neural networks. ICCV, pp.1529–1537. (2015)
7. Tustison, N. J., Avants, B. B., Cook, P. A., et al.: N4ITK: improved N3 bias correction. IEEE transactions on medical imaging 29(6), 1310 (2010)
8. Xiao, Y., Fortin, M., Batti, M. C., Rivaz, H.: Population-averaged MRI atlases for automated image processing and assessments of lumbar paraspinal muscles. European Spine Journal 27(10), 2442–2448 (2018)
9. Isensee, F., Petersen, J., Klein, A., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486. (2018)
10. Gibson, E., Li, W., Sudre, C., et al.: NiftyNet: a deep-learning platform for medical imaging. Computer methods and programs in biomedicine 158, 113-122 (2018)
11. Urrutia, J., Besa, P., Lobos, D., et al.: Lumbar paraspinal muscle fat infiltration is independently associated with sex, age, and inter-vertebral disc degeneration in symptomatic patients. Skeletal radiology 47(7), 955–961 (2018)