AUTOMATED 3D MUSCLE SEGMENTATION FROM MRI DATA USING CONVOLUTIONAL NEURAL NETWORK

Shrimanti Ghosh[1], Pierre Boulanger[1], Scott T. Acton[2,3], Silvia S. Blemker[3], Nilanjan Ray[1]

[1]Dept. of Computing Science, University of Alberta, Edmonton, Alberta, Canada
[2]Dept. of Electrical & Computer Engineering, University of Virginia, USA
[3]Dept. of Biomedical Engineering, University of Virginia, USA

ABSTRACT

In this paper, we propose an automated segmentation algorithm for human leg muscles from 3D MRI data using a deep convolutional neural network (CNN). Using a generalized cylinder model, a 3D human leg muscle is represented by two smooth 2D images. The CNN predicts these two images from raw 3D voxels. For our base CNN, we use a pre-trained AlexNet that is coupled with a principle components analysis (PCA)-head. The AlexNet predicts a compressed vector, which is then back-projected by the PCA-head into two 2D images representing a 3D leg muscle boundary. This structured-output CNN architecture is fine-tuned in an end-to-end fashion. Our proposed CNN out-performs the conventional model-based approach, the active appearance model (AAM) image segmentation algorithm. The average Dice score between the ground truth segmentation and the obtained segmentation image is 0.85 using our CNN model, whereas the AAM yields a Dice score of 0.60.

Index Terms— Magnetic resonance imaging (MRI), leg muscle segmentation, 3D modeling, principal component analysis (PCA), convolutional neural networks (CNN).

1. INTRODUCTION

In the anatomical imaging community, automated segmentation of the leg and its constituent muscles is regarded as an important *open problem*. The 3D segmentation of tissues and structures from MRI is often performed in conjunction with pre-and post- surgical planning and diagnosis. Skeletal muscle is a very important tissue in the body as it serves various physiological functions such as the generation of movement and support. Muscle volume is also an important determinant of muscle functional capacity [1]. Segmenting multiple muscles in MRI is challenging as each muscle groups have very similar intensity values in the MRI [2, 3]. Current methods of muscle segmentation in MRI images require a high-level of user input where clinician must segment various muscle boundaries by hand using their anatomical knowledge and keen observation. Manual segmentation is usually accurate but is impractical for large datasets because it is tedious and time consuming. For this reason, we propose an automated muscle segmentation algorithm to reduce human effort and the amount of manual segmentation time.

Recently, automated segmentation of various portions of human leg muscles has been proposed in [4]. However, the problem of segmenting the entire leg muscle architecture automatically is still an open challenge. Such an automation has the potential to provide an objective measure of effectiveness of pre- and post-surgery treatments.

Artificial neural networks (ANN) have been developed for image segmentation, registration and object detection [6]. Many advances have been achieved in recent past years in the area of medical image segmentation from MRI or CT (computed tomography) images [7-9]. A fully automatic brain tumor segmentation approach using convolutional neural networks (CNN) from MRI images was proposed and 93% accuracy was achieved compared to human segmentation [8].

To attain the goal of automated muscle segmentation, we use a generalized cylinder model similar to [5] to represent the geometry of a three-dimensional (3D) structure of leg muscles contours. Using this representation, one segment of a 3D leg muscle is converted to two 2D parametric images $x(z,t)$ and $y(z,t)$ where z is the slice position in the z direction and t is a parametric position along the contour. *Then we use the MRI images of 3D leg muscle as input to the CNN and the corresponding segmentation (i.e., 2D parametric image pair) of leg muscle contours as labels in the network.*

The performance of our method improved remarkably when we impose a structure on the output by principal component analysis (PCA) to reduce the high dimensional predicted vector to a low-dimensional one. Also, we compared our result with a classical image segmentation model-based algorithm, the active appearance model (AAM). One of the most significant advantages of the proposed method is that *no initialization* is needed, in contrast to AAM.

2. METHODOLOGY

2.1. Datasets

Our database contains fully segmented images of leg muscles from athletes [5], and as an initial step, we concentrate, in this paper, on a specific leg muscle the "*adductor magnus*". The adductor magnus is a large triangular muscle, situated on the medial side of the thigh [10]. It is a muscle of the thigh made up of three cords and is in the inner part of the thigh. In total, we had 600 3D datasets. The complete database is divided into *75% for training and 25% for testing*.

2.2. 3D to 2D Parametric Mapping

We used a generalized cylinder to model 3D human leg muscles [5]. These cylinders are a series of closed contours which are stacked in the z-direction as shown in Fig. 1. One can mathematically represent the 3D structure of a muscle by using two 2D parametric images [5] that correspond to the trajectory of the contour at every slice z. To do so, a 3D-to-2D mapping is applied here. The two 2D parametric images are: $x(z, t)$ and $y(z, t)$, so that the muscle boundary has the coordinates: $(x(z, t), y(z, t), z)$, where $t \in [0, 1]$ is the normalized distance that parameterizes the contour.

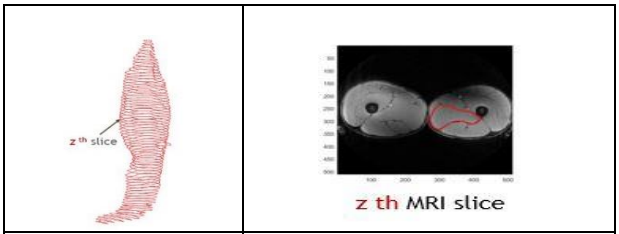


Figure 1: **(a)** A human leg muscle represented by stacked closed contours (generalized cylinder model) [5] & **(b)** 2D MRI slice and the corresponding segmentation (red contour) over-laid on the slice

On the 2D images we imposed a smoothness constraint which is referred as reparameterization operation. This can be achieved with the following minimization [5, 11]:

$$\min_{s_z}[\sum_{z,t} | x(z,t+s_z) - x(z-1,t+s_z) | + | y(z,t+s_z) - y(z-1,t+s_z) |].$$

The variable s_z represents re-parameterizations of t, i.e., s_z is the amount of circular shift of the z^{th} row of x and y matrices. Following re-parametrization, a circular shift smoothing is performed on the 2D parametric images. The smooth parametric images are shown in Fig. 2.

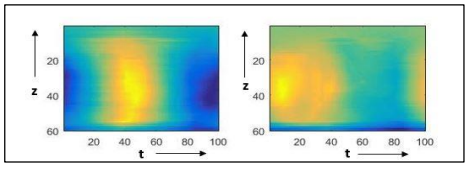


Figure 2: Segmentation of the 3D MRI stack (Left image is for $x(z, t)$ and right image is for $y(z, t)$), t is the number of parameters (set as 100) and z is the number MRI of slices

2.3. Proposed Deep Neural Network Architecture

Recently, convolutional neural networks have achieved success at large-scale image classification or recognition problems such as ImageNet [12, 13]. Here we applied deep CNN architecture of the classification/regression network AlexNet [14] to our problem.

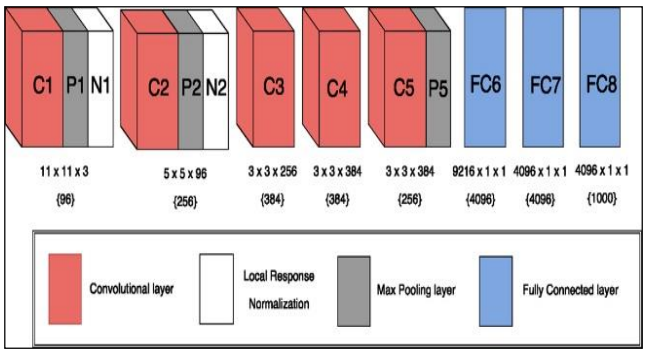


Figure 3: Conventional AlexNet Architecture [17]

The conventional architecture of AlexNet shown in Fig 3 consists of 8 weight layers including 5 convolutional layers, 3 fully connected layers and the normalization layers. In total, it has 21 layers and the number of parameters is about 61 million. We use AlexNet pre-trained for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and fine-tune it to our MRI dataset.

The first convolutional layer and the last fully connected layers are modified by changing the initial weights according to our problem. Then an L2 normalization layer is added at the end which is considered as the objective function or the loss function.

We pre-process the data to generate more training examples for the CNN. Each MRI image slice is of size 512 × 512 (pixels) in grayscale and we resized all images to size to 256 × 256 (pixels) in grayscale to reduce the dimension – this reduction in spatial dimension allowed us to increase batch size of 3D image stacks during training. Each 3D MRI image stack is divided into several overlapping stacks, each consisting of 10 image slices. Corresponding to each 10-slice stack, we have a pair of 2D parametric images, $x(z, t)$ and $y(z, t)$ indicating hand segmented 3D volume for a leg muscle.

In our case these labeled parametric images x and y are 10-by-100. So, concatenating x and y and forming a vector we have a 2000-by-1 response/label vector for each 10-slice 3D MRI stack. Our dataset consists of 600 3D stacks (256-by-256-by-10) along with response vector of dimension 2000-by-1 for each such 3D stack. Thus, the input to our CNN is a 256-by-256-10 gray scale image and the structured output or response vector is 2000-by-1. To accommodate this structured output within a deep architecture, we utilize PCA as described next.

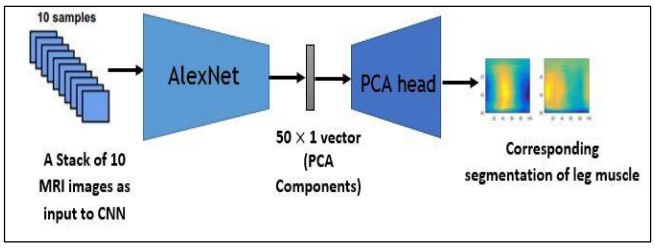


Figure 4: Our proposed CNN architecture.

Given a set of training output or response vectors $\boldsymbol{r_i}$, the principal component analysis (PCA) framework is as follows. p output vectors are combined into a matrix $[\boldsymbol{r_1}\ \boldsymbol{r_2}\ \ldots\ \boldsymbol{r_p}]$ of size m-by-p (here, m= 2000). Then the matrix is centered $\boldsymbol{R} = [\boldsymbol{r_1} - \bar{\boldsymbol{r}}\ \boldsymbol{r_2} - \bar{\boldsymbol{r}}\ \ldots\ \boldsymbol{r_p} - \bar{\boldsymbol{r}}]$, where $\bar{\boldsymbol{r}} = \left(1/p\right) \sum \boldsymbol{r_i}$. Using singular value decomposition (SVD) [15], we obtain, $\boldsymbol{R_{m\times p}} = \boldsymbol{U_{m\times p}}\ \boldsymbol{D_{p\times p}}\ \left(\boldsymbol{V_{p\times p}}\right)^T$. To achieve dimensionality reduction, we take the first 50 columns of $\boldsymbol{U}$, so that a 2000-by-1 vector $\boldsymbol{r}$ is reduced to $\hat{\boldsymbol{r}} = (\boldsymbol{U_{2000\times 50}})^T\ (\boldsymbol{r} - \bar{\boldsymbol{r}})$.

Instead of using AlexNet to predict a 2000-by-1 response vector $\boldsymbol{r}$, we now ask the net to predict a 50-by-1 response vector $\hat{\boldsymbol{r}}$. Next, we can reconstruct the 2000-by-1 as follows:

$$\boldsymbol{r} \approx \boldsymbol{U_{2000\times 50}}\,\hat{\boldsymbol{r}} + \bar{\boldsymbol{r}} \quad (1)$$

Moreover, being a linear function (i.e. in the form of $\boldsymbol{Ax} + \boldsymbol{B}$, where $\boldsymbol{A} = \boldsymbol{U_{2000\times 50}}$ and $\boldsymbol{B} = \bar{\boldsymbol{r}}$ are known), this reconstruction in equation (1) can be performed within the deep architecture. This is shown in Fig. 4, where AlexNet predicts $\hat{\boldsymbol{r}}$ and PCA-head reconstructs $\boldsymbol{r}$. This is our deep CNN architecture for structured output.

2.4. Comparison with Active Appearance Model

The AAM model is trained from manually drawn contours and finds the main variations in the training data using PCA, which enables the model to automatically recognize if a contour is a possible/good object contour. The method is very useful for automatic segmentation and recognition of biomedical objects [16]. The disadvantage of this model is, it needs the initial position of segmentation contour for prediction and in our case the position of the contours on MRI image slices are different. Thus, to automate this process, we first binarize an image slice by Otsu's thresholding method [18]. Next, we take the two largest connected components from the binarized slice. Because, we have confined our study to the left leg muscle, we initialize the contour at the centroid of the connected component for the left leg on the MRI image. Note that this initialized shape is the mean shape $\bar{r}$ from the training set.

3. EVALUATION METHOD

For evaluation, we use Dice similarity coefficient (DSC). The DSC is used for comparing the *similarity* and diversity of sample sets. The DSC measures the spatial overlap between two segmentations, **A** (ground truth) and **B** (segmentation obtained by proposed method) and is defined as **DSC (A, B) = 2(|A∩B|)/(|A|+|B|)** where ∩ is the intersection and | . | implies cardinality of a set. A higher DSC value is associated with a better segmentation result.

4. RESULTS AND DISCUSSIONS

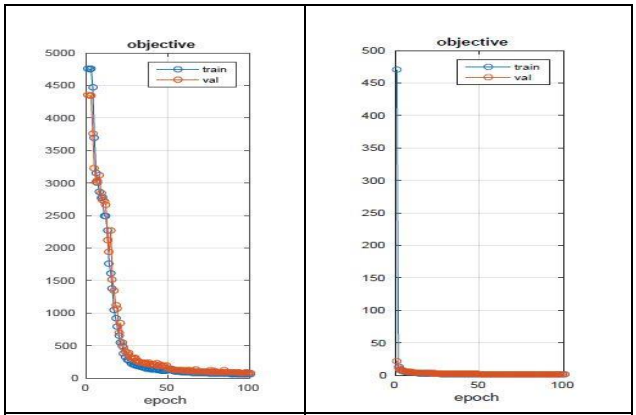
Two different approaches for this problem are discussed here: one where the parameters, $U_{2000\times 50}$ and $\bar{r}$ of the PCA-head were not modified during training and the other where these parameters were modified by backpropagation.

4.1. Predicting 50 PCA Components with CNN

We inputted a 256-by-256-by-10 stack of grayscale MRI images into the AlexNet to predict a 50-by-1 response vector. The training was performed with 100 epochs and with a batch size of 6 and learning rate 10^{-6}. Then we reconstruct the 3D cylinder (i.e. two parametric 2D images of size 10 × 100 each) from the 50 PCA components by PCA-head. From this method, the obtained highest Dice score is 0.9487 and DSC range is (0.70 – 0.92) for 90% of the test data. The L2 error curve with different epochs is shown in Fig 5 (a) and the error bar along with mean and SD is shown in Fig 6.

4.2. End-to-end Training with PCA-head

As a second approach, we initialize the weights and the bias term of the PCA-head with $U_{2000\times 50}$ and $\bar{r}$ respectively. Then, we allow backpropagation to modify these parameters in the PCA-head. From this method, the obtained highest DSC is 0.9088 and DSC range is (0.60 – 0.90) for 90% of the test data. The error bar for this method is shown in Fig 5. We believe that, the principal reason of obtaining a lower Dice score than the previous method described in Section 4.1 is likely due to over-fitting for this small dataset. The L2 error curve with different epochs is shown in Fig 5 (b) and the error bar along with mean and SD is shown in Fig 6.



(a) (b)

Figure 5: L2-error vs Epoch curve **(a)** CNN (predicting 50 PCA components) and **(b)** End-to-End CNN

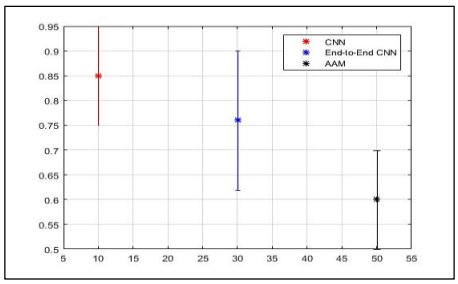


Figure 6: Error-bar for 3 methods with Mean and SD.

We compare our deep learning method with a conventional image segmentation method AAM. Comparison between ground truth and automated segmentation from CNN and AAM methods is shown in Fig 7 and Fig 8 below with different DSC (Highest, 75th, 50th, 25th percentile).

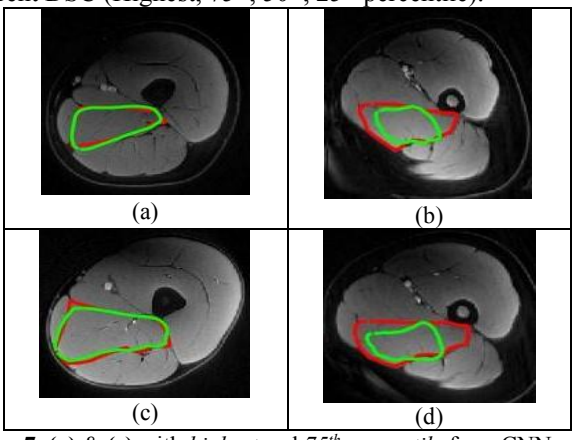


Figure 7. (a) & **(c)** with *highest* and *75th percentile* from CNN and **(b)** & **(d)** with *highest* and *75th percentile* from AAM respectively (manual ground truth segmentation as red contour & obtained segmentation as green contour)

In Fig 7, the segmentation obtained from CNN are (a) with highest dice score 0.9487 and (c) with 75th percentile 0.9044. The segmentation from AAM are (b) with highest dice score 0.7004 and (d) with 75th percentile 0.6682.

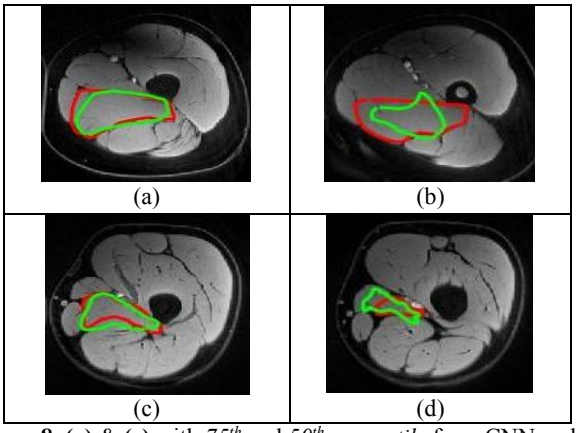


Figure 8. (a) & **(c)** with *75th* and *50th percentile* from CNN and **(b)** & **(d)** with *75th* and *50th percentile* from AAM respectively (manual ground truth segmentation as red contour & obtained segmentation as green contour)

In Fig 8, the segmentation obtained from CNN are (a) with 50th percentile 0.8612 and (c) with 25th percentile 0.7918. The segmentation from AAM are (b) with 50th percentile 0.6201 and (d) with 25th percentile 0.5575. The visual results achieved from CNN are remarkably superior to those obtained from AAM.

Table. 1.
Dice scores between original segmentation & segmentation obtained by CNN and AAM in test dataset

CNN with 50 PCA Components	End-to-End CNN Training with PCA head	Active Appearance Model
0.85 ± 0.10	0.76 ± 0.14	0.60 ± 0.10

In Table 1, the mean and the standard deviation for the same test dataset for 3 different approaches are given in a tabular form. CNN performs more accurately when it predicts 50 PCA components instead of predicting 2000 pixel values of the corresponding segmentation.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel approach to automated human leg muscle segmentation using deep convolutional neural networks. By applying PCA, we have imposed a strong structured regression approach to our model. But when we used the basis and mean obtained from PCA, it over-fits the training data and the performance on the test dataset is reduced. The average achieved Dice score is 0.85 which is better than one of the conventional methods – the AAM. This high accuracy is achieved by applying PCA to reduce the high dimensional output vector of the neural network. One of the biggest advantages of deep learning based method is that it does not need initialization of contours unlike conventional methods such as the AAM.

As an initial starting point, we have focused on one specific kind of leg muscle. But for future work, we would like to apply our method for segmentation of other leg muscle automatically. To prevent the over-fitting that occurred when the PCA basis and mean are used while training the neural net, we can lower the learning rate for this layer so that it limits the co-adapting of the model. Also, we can apply other network architectures such as the VGGNet for larger datasets.

6. REFERENCES

[1] Geoffrey G. Handsfield, Craig H. Meyer, Joseph M. Hart, Mark F. Abel, Silvia S. Blemker, "Relationships of 35 lower limb muscles to height and body mass quantified using MRI", Journal of Biomechanics, Volume 47, pages 631-638, 2014.

[2] Katherine R.S. Holzbaur, Wendy M. Murray, Garry E. Gold, Scott L. Delp, "Upper limb muscle volumes in adult subjects", Journal of Biomechanics, Volume 40, pages 742-749, 2007.

[3] Yan Geng, Sebastian Ullrich, Oliver Grottke, Rolf Rossaint, Torsten Kuhlen, Thomas M. Deserno, “Scene-based segmentation of multiple muscles from MRI in MITK”, Bildverarbeitung für die Medizin 2009.

[4] Salma Essafi, G. Langs, J-F. Deux, A. Rahmouni, G. Bassez4, N. Paragios, “Wavelet-driven knowledge-based MRI calf muscle segmentation”, IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009.

[5] Nilanjan Ray, Satarupa Mukherjee, Krishna Kanth Nakka, Scott T. Acton, Silvia S. Blanker, “3D-to-2D mapping for user interactive segmentation of human leg muscles from MRI data”, Signal and Information Processing (GlobalSIP), IEEE Global Conference, 2014.

[6] Mostafa Jabarouti, Moghaddam and Hamid Soltanian-Zadeh, “Medical Image Segmentation Using Artificial Neural Networks”, Artificial Neural Networks - Methodological Advances and Biomedical Applications.

[7] H. P. Ng, S. H. Ong, J. Liu, S. Huang, K. W. C. Foong, P. S. Goh and W. L. Nowinski, “3D Segmentation and Quantification of a Masticatory Muscle from MR Data Using Patient-Specific Models and Matching Distributions”, Journal of Digital Imaging, Vol 22, No 5, pp 449-462, October, 2009.

[8] Raunar Rewari, “Automatic Tumor Segmentation from MRI scans”, Stanford University.

[9] Erwan Jolivet, Elisabeth Dion, Philippe Rouch, Guillaume Dubois, Remi Charrier, Christine Payan, and Wafa Skalli, “Skeletal muscle segmentation from MRI dataset using a model-based approach”, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization Vol. 2, Iss. 3,2014.

[10] Anatomy of the muscular system, (chapter 10), Link http://www.coursewareobjects.com/objects/evolve/E2/book_pages/thibodeau/pdfs/03470394_A03718_10.pdf .

[11] Seyed Mehdi Moghadas Tabatabaei Zavareh, “A 3D Framework for the Musculoskeletal Segmentation of Magnetic Resonance Images”, University of Ottawa, March, 2015.

[12] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. “ImageNet: A large-scale hierarchical image database”. In Proc. CVPR, 2009.

[13] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. “ImageNet large scale visual recognition challenge”. CoRR, abs/1409.0575, 2014.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “ImageNet classification with deep convolutional neural networks”. In NIPS 2012.

[15] Gene H. Golab and Charles F. Van Loan, “Matrix Computations”, Third Edition, The Johns Hopkins University Press, London.

[16] T.F. Cootes, G.J Edwards, and C,J. Taylor "Active Appearance Models", IEEE Transactions on Pattern Analysis and Machine Intelligence 2001.

[17] Srinivas S, Sarvadevabhatla RK, Mopuri KR, Prabhu N, Kruthiventi SSS and Babu RV (2016) “A Taxonomy of Deep Convolutional Neural Nets for Computer Vision”. Front. Robot. AI 2:36. doi: 10.3389/frobt.2015.00036.

[18] N. Otsu, "A threshold selection methods from grey-level histograms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 9, pp. 62-66, 1979.