AUTOMATIC MUSCLE PERIMYSIUM ANNOTATION USING DEEP CONVOLUTIONAL NEURAL NETWORK

Manish Sapkota[1,2] *Fuyong Xing*[1,2] *Hai Su*[2] *Lin Yang*[2]

[1]Department of Electrical and Computer Engineering, University of Florida
[2]J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida

ABSTRACT

Diseased skeletal muscle expresses mononuclear cell infiltration in the regions of perimysium. Accurate annotation or segmentation of perimysium can help biologists and clinicians to determine individualized patient treatment and allow for reasonable prognostication. However, manual perimysium annotation is time consuming and prone to inter-observer variations. Meanwhile, the presence of ambiguous patterns in muscle images significantly challenge many traditional automatic annotation algorithms. In this paper, we propose an automatic perimysium annotation algorithm based on deep convolutional neural network (CNN). We formulate the automatic annotation of perimysium in muscle images as a pixel-wise classification problem, and the CNN is trained to label each image pixel with raw RGB values of the patch centered at the pixel. The algorithm is applied to 82 diseased skeletal muscle images. We have achieved an average precision of 94% on the test dataset.

***Index Terms*—** Perimysium annotation, muscle, convolutional neural network

1. INTRODUCTION

Recently histopathological study has shown growing evidence that skeletal muscle extracellular matrix (ECM) affects the normal function of muscle [1]. ECM is very important in the maintenance, transmission and repair of the muscle fibre force. Idiopathic Inflammatory Myopathies (IIMs), a rare form of muscle inflammatory disease that causes muscle weakening and pain, exhibits clinical manifestation in the regions of perimysium [2]. Figure 1 shows typical mononuclear cell infiltration in the perimysium region in one sample Hematosin & Eosin (H&E) stained diseased skeletal muscle image. Accurate delineation of the perimysium region can provide support for infiltration characterization, which is helpful for effective diagnosis and prognosis of the muscle disease. However, manual annotation in a large number of digitized muscle specimens is time consuming, laborious and subjective.

This research is funded, in part, by NIH 1R01AR065479-01A1.

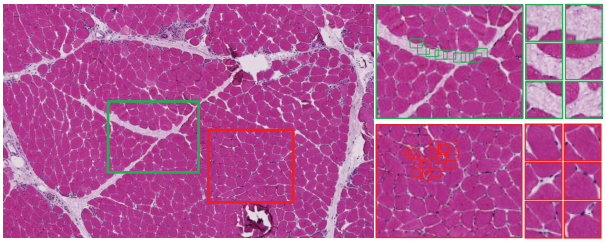

Fig. 1. An example of H&E stained skeletal muscle image. **Left:** The cross sectional area of the skeletal muscle scan cropped at 4x magnification. The green/red box indicates a muscle region with/without perimysium, respectively. **Middle:** The zoomed-in regions of sample regions shown in **Left**. **Right:** Several zoomed-in small image patches displayed in **Middle**. These patches are used as training samples for our learning model. Green boxes indicate positive samples and red boxes represent negative samples.

Computer-aided algorithms provide a promising strategy for automated annotation on histopathology images. Xu *et al.* [3] have proposed a context-constrained multiple instance learning (MIL) method to achieve pixel-wise segmentation/anotation on colon histopathology images. Due to the high variability of the patterns shown in histopathology images, it is difficult to design an effective feature descriptor for automatic image analysis. In recent years, there is an encouraging evidence that learned representation of biomedical images might perform better than the handcrafted features [4, 5, 6]. Cruz-Roa *et al.* [7] have proposed a deep neural network for automated basal cell carcinoma cancer detection, and a unified deep representation learning model is reported [8] for automatic prostate MR image segmentation. Recently, a deep convolutional neural network [9] has been successfully applied to mitosis detection in breast cancer histopathology images. However, none of these methods deal with digitized muscle specimens, which are significantly different from other types of histopathology images. Since some of the perimysium regions are very similar to other ECMs, it is difficult to achieve automatic perimysium annotation.

In this paper, we present an automated perimysium annotation approach on skeletal muscle images, which is based on a deep convolutional neural network (CNN), as shown in

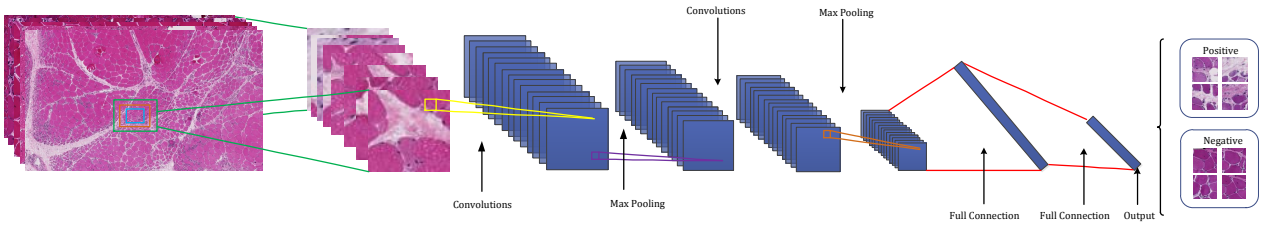

Fig. 2. The architecture of our proposed CNN model.

Figure 2. The problem is formulated into a pixel-wise classification framework, where a CNN model is trained with raw RGB values of image data and automatically learns a set of hierarchical features for classification. In order to introduce scale invariance, we feed the CNN model with multi-scale training image inputs. In the testing stage, the learned CNN model is applied to the images in a sliding window, differentiating pixels in the perimysium region from others to achieve automatic annotation. To the best of our knowledge, this is the first attempt to automate the analysis of muscle pathology of perimysium. This approach provides effective perimysium annotation results, which can serve as a basis for further image analysis of skeletal muscle disease.

2. METHODS

Given a set of training RGB image patches $I_i \in R^{r\times c\times 3}, i = 1, ..., N$ with dimensionality $r \times c$ for each of the 3 channels, we propose to learn a CNN-based mapping function to predict the class labels. The patches with center pixels located in the perimysium regions are labeled as positive, otherwise negative (see Figure 1).

2.1. CNN Architecture

Convolutional neural network (CNN) is a feed-forward network which has alternating layers of convolution and max-pooling, followed by some fully connected layers [10]. It can provide progressively abstract representation of the input with the increment of the number of layers. The CNN structure used in our implementation is summarized in Table 1. The convolutional layer calculates a set of output feature maps by performing multiple 2D filters on input images. Formally, define M_j^l as the j-th output feature map of the l-th layer, we have the following equation

$$M_j^l = f(\sum_i M_i^{l-1} * K_{ij}^l + b_j^l), \qquad (1)$$

where K_{ij}^l and b_j^l represent convolutional kernel and bias corresponding to the i-th input feature map and the j-th output feature map, respectively. The $f(x)$ is a nonlinear activation function, referred to as rectified linear units (ReLUs) [11]$f(x) = \max(0, x)$. It enables fast model training and potentially improves the classification performance. We chose a kernel size of 5×5 for convolution layers based on the input image size of 32×32. A larger kernel size would decrease the discriminative power of the network, and too small would give ambiguous feature representation.

Table 1. The structure of the CNN used in our algorithm.

Layer No.	Layer Type	Feature Map	Kernel Size
1	Input	$32 \times 32 \times 3$	-
2	Convolutional	$28 \times 28 \times 6$	5×5
3	Max-pooling	$14 \times 14 \times 6$	2×2
4	Convolutional	$12 \times 12 \times 12$	3×3
5	Max-pooling	$6 \times 6 \times 12$	2×2
6	Fully-connected	64×1	-
7	Output	2×1	-

Max pooling layer is used to preform dimension reduction by keeping the most promising value in the given subregion. It also introduces local shift and translation invariance, and corresponds to a kernel of size 2×2 without overlapping in our design. Fully-connected layer consists of ReLUs aiming to learn global feature representation. The last (Output) layer is a fully-connected layer with a softmax function, which is used for final classification.

2.2. CNN Model Training and Testing

In our implementation, each pixel is represented by a patch centered at this specific pixel. Therefore, the patch size plays a significant role in the automatic perimysium annotation. Learning hierarchical features with multi-scale input images have shown to improve the classification performance [12]. In order to incorporate scale invariance into the classifier, we train the model with a multi-scale version of the input images. Specifically, we crop image patches from the whole slides with different window sizes at the same pixel location: $28 \times 28, 32 \times 32$, and 64×64, and upsample or downsample these patches to have a unified size of 32×32.

The model is trained using backpropagation with stochastic gradient descent [13], which locally minimizes the negative log-likelihood objective function. In order to achieve fast convergence in training, all the image patches are normalized to have zero mean and unit variance. The learning rate is an important parameter in our model. It is initialized as 0.1 and decayed by a factor of $(1 + d \times t)$ within each epoch, where d is equal to 10^{-3} and t is the epoch index, until the vali-

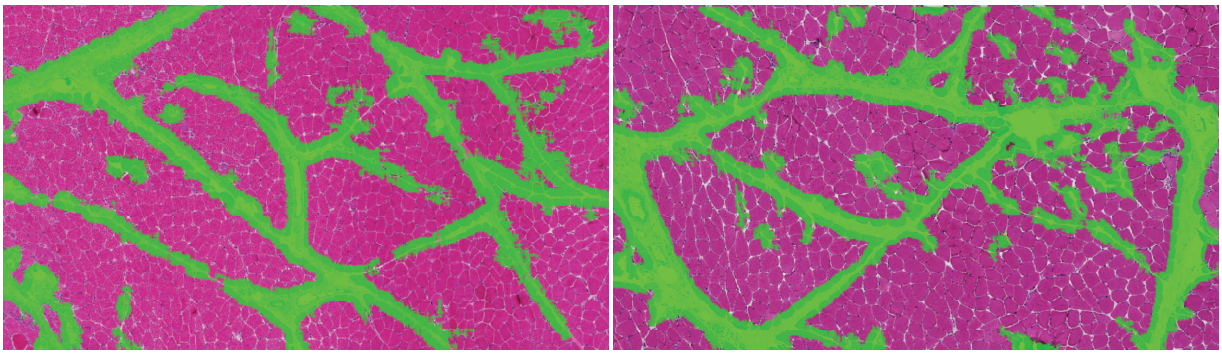

Fig. 3. Automatic perimysium annotation results using our CNN based method.

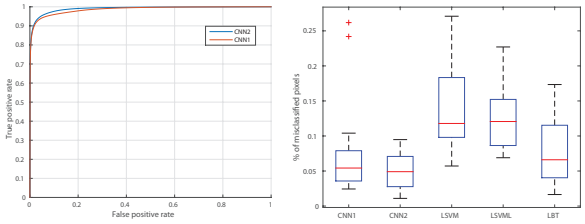

Fig. 4. Comparison between the proposed CNN (CNN1) and its variation (CNN2). **Left:** the ROC curves; **Right:** box plot for PFP.

dation error stops improving with the current learning rate. This early stopping strategy is an important step to avoid overfitting [14]. Batch size and the momentum are kept fixed during the training, as 100 and 0.5, respectively.

In the testing stage, automatic annotation is achieved by applying the CNN model to new images using a sliding window of 32×32. The patches are normalized in a similar way as training. The patches partially outside the image boundaries are ignored. The softmax layer outputs the probabilities that each pixel is located in the perimysium or other regions. We predict patch labels by choosing the category associated with a higher probability.

3. EXPERIMENTAL RESULTS

The proposed method is evaluated both quantitatively and qualitatively using 82 skeletal muscle images (roughly 1100×700), which are cropped at 4x magnification from 39 Hematosin & Eosin stained whole slides cross section biopsy scans. These slides represent two types of muscle diseases: Dermatomyositis (DM) and Polymyositis (PM), which exhibit different ranges of perimysium infiltration. The perimysium regions in the images are manually annotated as ground truth. Approximately 75% of the images are randomly selected for training and cross validation, and the remaining 25% is used for testing. From the training images, in total, 312000 square patches are generated for training and 78000 for validation. Figure 3 shows the automatic annotation on two sample images, where the perimysium regions are accurately annotated using green color in the muscle images.

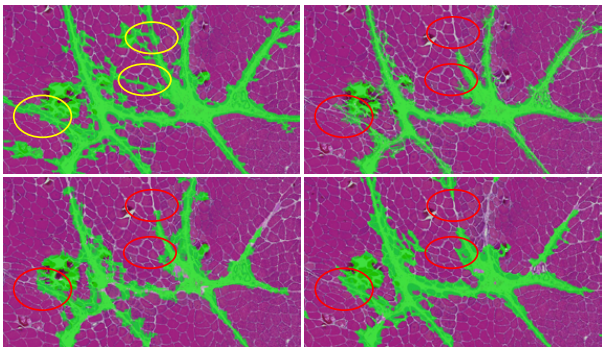

Fig. 5. The automatic annotation results using different algorithm. **Top-left:** The proposed CNN; **Top-right:** LSVM [15]; **Bottom-left:** LSVML [16]; **Bottom-right:** LBT [17]. The yellow ellipses overlaid on the image indicate that the narrow perimysium regions can be successfully annotated by the proposed CNN, however, other methods failed to detect these thin perimysium regions which are are marked with red ellipses.

In the first set of experiments, we evaluate different structures of CNN. For comparison, we have trained another deep convolutional neural network (CNN2) by removing the first fully-connected layer of the proposed framework. We evaluate the pixel-wise classification for quantitative analysis, and the receiver operating characteristic (ROC) curves for the proposed CNN (CNN1) and its variation (CNN2). The quantitative experimental results are displayed in Figure 4. Area under the curves (AUCs) observed for CNN1 and CNN2 are 0.98 and 0.99, respectively. In addition, Figure 4 also shows the percentage of falsely classified pixels: $PFP = \frac{(N_{FP}+N_{FN})}{N_{total}}$ where N_{FP}, N_{FN}, and N_{total} represent the number of false positive, false negative, and total pixels, respectively. We can see that CNN2 performs slightly better than CNN1 which has a more deeper architecture, and this might be due to the limited training dataset.

In addition to the comparison of different CNN structures, we also compare the CNN based methods with two state of

Table 2. Summary of the evaluation compared with ground truth on CNN and other methods.

Methods	Precision				Recall				F_1-score			
	$Mean \pm std$	Max	Min	80%	$Mean \pm std$	Max	Min	80%	$Mean \pm std$	Max	Min	80%
CNN1	0.94 ± 0.04	0.99	0.85	0.99	0.87 ± 0.10	0.99	0.67	0.97	0.90 ± 0.05	0.97	0.80	0.96
CNN2	0.93 ± 0.07	0.99	0.76	0.99	0.92 ± 0.07	1.0	0.73	0.98	0.92 ± 0.04	0.98	0.80	0.96
LSVM	0.92 ± 0.06	0.99	0.80	0.97	0.76 ± 0.12	0.91	0.49	0.89	0.82 ± 0.08	0.94	0.63	0.90
LSVML	0.78 ± 0.15	0.97	0.41	0.90	0.87 ± 0.09	1.0	0.67	0.96	0.81 ± 0.08	0.91	0.59	0.87
LBT	0.79 ± 0.13	0.97	0.58	0.91	0.96 ± 0.04	1.0	0.84	1.0	0.86 ± 0.09	0.98	0.69	0.94

the arts: 1) A large-scale SVM [15] based classifier using raw pixel intensities as feature vectors (LSVM), and locality binary pattern [16] (LSVML). 2) A logistic boosting classifier using texton features (LBT) [17]. As one can tell, the deep learning based models provide lower PFP errors than those shallow learning methods. For quantitative comparison, we calculate precision (P), recall (R), and F_1-score as

$$P = \frac{N_{TP}}{(N_{TP} + N_{FP})}, \; R = \frac{N_{TP}}{(N_{TP} + N_{FN})}, \; F_1 = \frac{2PR}{(P + R)}, \tag{2}$$

where N_{TP} denotes the number of true positive pixels. Figure 5 shows the qualitative automatic annotation results using different methods. As one can tell, our proposed CNN based method can handle narrow perimysium regions which present some challenges for other learning algorithm using hand-crafted features. Table 2 shows the quantitative comparison among the CNN based methods and other state of the arts. It is clear that our method and its variation consistently provide the best classification results. This is attributed to the fact that the proposed CNN models are an end-to-end learning method that can automatically learn hierarchical features that are best suitable for automatic annotations.

4. CONCLUSION

We have presented an automated perimysium annotation approach in skeletal muscle images using convolution neural network. In order to handle scale variations, multi-scale versions of input images are used for model training, and automatic annotation is achieved by performing pixel-wise classification with a sliding window on testing images. The comparative experiments demonstrate the effectiveness of its superior performance. Our method is a general learning framework, which can be applied to other automatic image annotation for microscopic image analysis.

5. REFERENCES

[1] A. R. Gillies and R. L. Lieber, "Structure and function of the skeletal muscle extracellular matrix," *Muscle & nerve*, vol. 44, no. 3, pp. 318–331, 2011.

[2] C. Malm and J. Yu, "Exercise-induced muscle damage and inflammation: re-evaluation by proteomics," *Histochemistry and cell biology*, vol. 138, no. 1, pp. 89–99, 2012.

[3] Y. Xu, J. Zhang, E. Chang, M. Lai, and Z. Tu, "Context-constrained multiple instance learning for histopathology image segmentation," in *MICCAI*, 2012, vol. 7512, pp. 623–630.

[4] G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao, and D. Shen, "Unsupervised deep feature learning for deformable registration of mr brain images," in *MICCAI*, 2013, vol. 8150, pp. 649–656.

[5] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *MICCAI*, 2013, vol. 8150, pp. 246–253.

[6] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations," in *MICCAI*, 2014, pp. 520–527.

[7] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *MICCAI*, 2013, pp. 403–410.

[8] S. Liao, Y. Gao, A. Oto, and D. Shen, "Representation learning: A unified deep learning framework for automatic prostate mr segmentation," in *MICCAI*, 2013, vol. 8150, pp. 254–261.

[9] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *MICCAI*, 2013, pp. 411–418.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[11] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *ICML*, 2010, pp. 807–814.

[12] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *TPAMI*, vol. 35, no. 8, pp. 1915–1929, 2013.

[13] Y. A. LeCun, L. Bottou, G. B. Orr, and K. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*, 1998, vol. 1524, pp. 9–50.

[14] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[15] N. Djuric, L. Lan, S. Vucetic, and Z. Wang, "Budgetedsvm: A toolbox for scalable svm approximations," *JMLR*, vol. 14, pp. 3813–3817, 2013.

[16] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, pp. 51–59, 1996.

[17] D.J. Foran, L. Yang, O. Tuzel, W. Chen, J. Hu, T.M. Kurc, R. Ferreira, and J.H. Saltz, "A cagrid-enabled, learning based image segmentation method for histopathology specimens," in *ISBI*, 2009, pp. 1306–1309.