Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

DeepQuantify: deep learning and quantification system of white blood cells in light microscopy images of injured skeletal muscles

Yang Jiao
Barbara St. Pierre Schneider
Emma Regentova
Mei Yang

DeepQuantify: deep learning and quantification system of white blood cells in light microscopy images of injured skeletal muscles

Yang Jiao,^{a,*} Barbara St. Pierre Schneider,^b Emma Regentova,^a and Mei Yang^a

^aUniversity of Nevada, Department of Electrical and Computer Engineering, Las Vegas, Nevada, United States

^bUniversity of Nevada, School of Nursing, Las Vegas, Nevada, United States

Abstract. White blood cells (WBCs) are the most diverse types of cells observed in the healing process of injured skeletal muscles. In the recovery process, WBCs exhibit a dynamic cellular response and undergo multiple changes of the protein expression. The progress of healing can be analyzed by the number of WBCs or by the number of specific proteins observed in light microscopy images obtained at different time points after injury. We propose a deep learning quantification and analysis system called DeepQuantify to analyze WBCs in light microscopy images of uninjured and injured muscles of female mice. The DeepQuantify system features in segmentation using the localized iterative Otsu's thresholding method, masking postprocessing, and classification of WBCs with a convolutional neural network (CNN) classifier to achieve a high accuracy and a low manpower cost. The proposed two-layer CNN classifier designed based on the optimization hypothesis is evaluated and compared with other CNN classifiers. The DeepQuantify system adopting these CNN classifiers is evaluated for quantifying CD68-positive macrophages and 7/4-positive neutrophils and compared with the state-of-the-art deep learning segmentation architectures. DeepQuantify achieves an accuracy of 90.64% and 89.31% for CD68-positive macrophages and 7/4-positive neutrophils, respectively. The DeepQuantify system employing the proposed two-layer CNN architecture achieves better performance than those deep segmentation architectures. The quantitative analysis of two protein dynamics during muscle recovery is also presented. © 2019 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.6.2.024006](https://doi.org/10.1117/1.JMI.6.2.024006)]

Keywords: deep learning; convolutional neural network; microscopy; quantification; white blood cell; macrophage; neutrophils.

Paper 18178RR received Aug. 20, 2018; accepted for publication Apr. 15, 2019; published online May 20, 2019.

1 Introduction

Skeletal muscle injuries can be caused by sports activities and natural disasters and even occur during combat operations. A muscle injury activates an inflammatory response, followed by muscle regeneration. As one part of this inflammatory response, monocytes and macrophages initiate healing and activate skeletal muscle regeneration.¹ During this process, white blood cells (WBCs), including neutrophils and monocytes/macrophages, travel from the bone marrow to the injured muscle via blood vessels. These WBCs exhibit a dynamic response by expressing a variety of proteins at different postinjury time points.² The dynamic response of WBCs can be analyzed and characterized by processing light microscopy images. Both uninjured muscles and injured muscles obtained in this study. At different time points after injury, from 4 to 192 h, injured muscles are processed and compared to understand the dynamics. The image analysis consists of cell detection and quantification of the amount of specific proteins.

Compared to manual analyses, which are laborious and prone to errors, computer vision techniques offer an efficient alternative. For automatic cell detection, image segmentation is a central problem that has been addressed in a great number of high-performance methods. For biomedical applications, these methods are included in commercial and publicly available software, such as ImageJ—an open source image processing software.³ For image analysis, such tools are used in a

semiautomatic mode, i.e., procedures and parameters are selected and fine-tuned manually based on the observed outcome. For example, Lejeune⁴ has demonstrated a semiautomated quantification of diverse, subcellular, immunohistochemical markers with Image-Pro Plus (Media Cybernetics, Silver Spring, Maryland). For microscopy images, segmentation using thresholding remains a popular approach because objects of interest exhibit distinct intensity levels as with light microscopy, wherein the stain makes WBCs distinct from the background tissue. Thresholding methods are less computationally extensive compared to the state-of-the-art segmentation methods, including those that use classifiers. Among thresholding methods, the Otsu's method is reported to be highly efficient.^{5,6} An adaptive segmentation method for WBCs was proposed by Liu et al.⁷ This method utilizes the outstanding edge density and low intensity of WBCs to discriminate them from red blood cells.⁷

Other than classic segmentation methods, classifiers are commonly used for direct detection of objects of interest. In recent years, advances in deep learning have made convolutional neural networks (CNNs) a major technology used for classification. Lecun et al.⁸ proposed LeNet-5, a CNN with two convolutional layers and two subsampling layers. The convolutional layer convolves image blocks with a preset filter and produces a special weight matrix. Then the subsampling layer selects the weighted feature from a local block and may abandon others to reduce the input size. After AlexNet⁹ gained success in an image

*Address all correspondence to Yang Jiao, E-mail: jiaoy1@unlv.nevada.edu

classification challenge in 2012, many successors followed.^{10–11} Later image segmentation benefited from deep networks, such as fully convolutional networks (FCN)¹² and SegNet.¹³ FCN adapted contemporary network classifications into FCN and transferred their learned representations by fine-tuning¹⁴ to the segmentation task. A skip architecture that allows deep layers to combine information from shallow layers is a significant contribution of FCN. Following FCN, SegNet has been presented as a better adapted deep learning segmentation architecture. The major part of the SegNet is an encoder network and a corresponding decoder network, followed by a pixel-wise classification layer. The encoder can be any typical deep learning classification architecture. The role of the decoder is to connect low-resolution encoder feature maps to full input resolution feature maps for pixel-wise classification. Instead of learning unreferenced functions, residual networks, or ResNets,¹⁵ explicitly reformulate the layers as learning residual functions with the reference to the layer inputs. While ResNets demonstrate solid efficiency in image recognition, DeepLabv3,¹⁶ which is based on atrous convolution ResNets, achieves superior segmentation accuracy.

Deep learning has found successful applications in biomedical image analysis. U-net¹⁷ employed CNN on neuronal structures in electron microscope stacks. Similar to the idea of FCN, U-net composes the feature maps of encoding layers with decoding layers to achieve a high-accuracy semantic segmentation. A fully convolutional residual network (FCRN)¹⁸ is an architecture proposed for specimen image segmentation. Combining the fully CNN and the residual network of a sand-clock shape improved the performance of FCRN. In 2017, fully convolutional dense-net (FCDN)¹⁹ demonstrated a state-of-the-art result on semantic segmentation due to the fact that the dense-net is more accurate and easier to train if each layer is directly connected to every other layer in a feed-forward fashion. Later, the idea of the dense-net has been employed to detect the scale of biomedical images.²⁰ In addition to 2-D deep learning architectures, 3-D architectures can be used in multi-dimensional biomedical data processing, including 3-D tissue segmentation²¹ and reconstruction.²² The multiscale input CNN architecture incorporates both local and contextual information aiming to improve segmentation results.²¹ Note deep learning segmentation requires pixel-wise labels of objects and thus needs a considerable amount of time for image labeling.

The goal of this study is to perform cell quantification in low-magnification light microscopy images. Most of the traditional and deep classification methods utilize high-magnification images and focus on WBC nuclei detection. The microscopy images in our study are of 100× magnification without cell nuclei stained. In addition, the light microscopy muscle images show textures and structures of connective tissues, blood vessels, nerve bundles, muscle spindles, and muscle fiber. All these factors impose significant challenges on the segmentation of WBCs. Furthermore, WBC markers are not always distinguishable from background structures; thus conventional segmentation methods could fail on the given type of imagery. Although deep segmentation architecture has potential to achieve a high degree of accuracy, this architecture requires a tremendous amount of time for marking objects of interest pixel by pixel. In addition, vague cell membrane and pseudo-color in microscopy image create a challenge for segmentation labeling. To attain a high-quantification accuracy under

acceptable computational and manual labeling costs, we propose a new segmentation and quantification system that combines a traditional segmentation method, such as thresholding, with deep learning classification. Benefits of the combination are twofold: (1) WBCs can be well discriminated from the background textures with the deep learning classifier and (2) in contrast to the above-mentioned deep segmentation networks, a considerable amount of time is saved because samples and ground truth labels are easier to obtain for deep classification.

In our previous work, we proposed a thresholding method called localized iterative Otsu (LIOtsu),²³ and a CNN²⁴ has been used to successfully detect CD68-positive WBCs. In this paper, we propose an extension of the CNN with an optimized architecture for other protein-expressed WBCs that have different appearance features. Proposed here, a system for the automatic quantification of WBCs in light microscopy muscle images called DeepQuantify includes the LIOtsu thresholding method and the CNN-based cell detection, followed by hybrid masking of segmentation results. Cells concentrated at muscle edges are excluded from consideration because of the possibility of false positives; thus, the muscle edge is to be recognized. The system automatically selects a region of interest (ROI) and quantifies WBCs pertaining to that ROI. The system efficiently handles various background textures and intensity variations within an image or across images/experiments. DeepQuantify is capable of adopting the state-of-the-art CNN architectures and detecting different WBC proteins.

The remainder of this paper is organized as follows. Section 2 presents the materials used in the research and CNN architectures. Section 3 describes the procedure of the DeepQuantify system. Section 4 presents the experimental and quantification results. Section 5 concludes this paper.

2 Materials and Methods

2.1 Dataset

All source images were acquired using the light microscope Eclipse E600 (Nikon, Melville, New York) with 10× objective, 100× magnification, and a color mosaic camera SPOT RT KE (Diagnostics Instruments, Sterling Heights, Michigan). The pixels are 24-bit values, and the image resolution is 1600 × 1200 pixels. The computer used for training and testing was built with Intel i7-7700K CPU, NVIDIA GTX 1080 GPU, 16G RAM, and HDD. The images were obtained from prepared slides of cross sections of crush-injured gastrocnemius muscle of female mice.² Injured and uninjured muscles were collected at 4, 8, 24, 32, 48, 96, and 192 h after the injury and then frozen. Each frozen muscle was sliced into 10-μm-thick sections. Serial sections, including both injured and uninjured muscles, were stained with an antibody that recognizes antigens present on target cells. Image exposure was balanced using Image-Pro Plus and Premier software (Media Cybernetics, Silver Spring, Maryland).

Table 1 presents the datasets used for the analysis. The left column of the table indicates that both uninjured and injured muscle slides were used. Note that for CD68-positive macrophages, the 4- and 8-h cases are not present because CD68-positive macrophages do not increase at 4- and 8-h postinjury.²⁵ Since 7/4 protein is not consistently observed in uninjured muscle,²⁵ no image was generated for the uninjured condition. Certain samples had been used in a previous study and,

Table 1 Light microscopy image dataset for uninjured and injured muscles.

Time	# of images for CD68 macrophages	# of images for 7/4 neutrophils
Uninjured	30	0
4-h after injury	0	5
8-h after injury	0	10
24-h after injury	5	5
32-h after injury	23	12
48-h after injury	17	11
96-h after injury	11	9
192-h after injury	10	10
Total	95	64

therefore, underwent a secondary analysis in the current study. Image size is 1600×1200 pixels.

2.2 CNN Classifier

To differentiate complex backgrounds encountered in light microscopy images, CNN classifier architectures need to be optimized, as explained below.

The convolutional layer contains neurons that can extract local elementary features, such as edges and corners. By overlaying the convolution layers, low-level features can be combined with complex features for successful classification. However, this type of overlay is computationally extensive if complete information of training samples is included. Thus, the filter size and the filter depth are to be adjusted.

At the subsampling layer, redundant or noninformative information can be discarded by averaging or maximizing a local weight matrix. Consequently, refining the feature map allows a deeper network, which yields better results. The rectifier transfer function is used because it has been proven to have a better performance compared to tanh and softplus activation functions.²⁶

Based on the above considerations, the designed CNN architecture contains a 16×16 input layer, two convolutional layers that employ 64 at 5×5 and 512 at 5×5 filters, two rectifiers after the convolutional layers, and two max-pooling layers with 2×2 filters and a stride size of 2. The architecture is shown in

Fig. 1 and summarized in Table 2, where C represents a convolutional layer and M is a maximal pooling layer. Here C1 (5, 64, 2) means that the first convolutional layer has a filter size of 5×5 , a filter depth of 64, and a zero-padding equal to 2. M1(2, 2) means that the first pooling layer has a filter size equal to 2×2 and a stride equal to 2.

All training samples were extracted from the source images manually. A total of 10,378 samples include 1816 positive and 8562 negative samples. The CNN has been tested on 1000 positive and 4156 negative samples.

The CNN is trained by stochastic gradient descent optimization with momentum.^{27–28} The iterative learning procedure is described by

$$\theta_{l+1} = \theta_l - \alpha \nabla E(\theta_l) + \gamma(\theta_l - \theta_{l-1}), \quad (1)$$

where l stands for the iteration number, $\alpha > 0$ is the learning rate, θ is the parameter vector, and $E(\theta)$ is the loss function. The stochastic gradient descent algorithm evaluates the gradient of the loss function $[\nabla E(\theta)]$ and updates the parameters using a subset of the training set. This subset is called a mini batch. γ determines the contribution of the previous gradient step to the current iteration. For training, the initial learning rate is set as 0.001, the mini batch size is 128, and the total training epochs is 50 until the validation accuracy is stable in $\pm 1\%$. For CNN classifiers, training, validation, and testing samples were selected from 96 images regions (400×400 pixels) across images of six time points, and 16×16 blocks of pixels were cropped and given to the network input. Uncertain samples are excluded. Training–validation–testing ratio is 4:1:1.

2.3 CNN Architecture Extension for Other Protein-Expression WBCs

To derive a CNN architecture for two proteins (CD68 and 7/4), features such as edges and corners are computed by three detectors: Harris corner detector,²⁹ MinEigen,³⁰ and Canny edge detector.³¹ Corner density and the average edge length will be used for defining the CNN architecture.

With the Harris detector, the corner response or corner quality is measured from the autocorrelation principal curvature space. In that space, (1) if two coordinates (α, β) are both low, the image intensity is constant, which means there is no edge or corner; (2) if one coordinate is high and the other one is low, an edge is contained; and (3) if both coordinates (α, β) are high, a corner is revealed. The quality or response R is measured by

$$R = \text{Det} - k \text{Tr}^2, \quad (2)$$

where

Fig. 1 The CNN architecture for CD68.

Table 2 The CNN architecture.

Protein	Name	Structure								
CD68	SS-CNN	Input	C1	Re	M1	C2	Re	M2	F1	F2
		16×16	(5, 64, 2)	Re	(2, 2)	(5, 512, 2)	Re	(2, 2)	8192	2

Table 3 CD68 and 7/4 features.

Harris detector				MinEigen detector				Canny detector	
Positive samples	CD68	7/4		CD68	7/4			Positive samples	
Corner/image (density)	2.9675	3.3637		4.12	4.56			Edge length (pixels)	
Average quality	1.2309×10^{-4}	4.3291×10^{-5}		5.5×10^{-3}	3.5×10^{-3}				

$Tr(M) = \alpha + \beta,$ (3)

$Det(M) = \alpha\beta,$ (4)

and α and β are the coordinates of the autocorrelation principal curvature space, k is a constant, and R is the quality or response.

The MinEigen corner detector uses the minimum eigenvalue algorithm to evaluate corners and edges. According to Ref. 28, the obtained symmetric 2×2 matrix Z must be both above the image noise level and well-conditioned. That means a window is accepted if the eigenvalues of Z are above a predefined threshold λ as

$(\lambda_1, \lambda_2) > \lambda.$ (5)

Corners that have a response above the threshold are selected for calculating the response R , and the corner density is measured per 16^2 pixels. The Canny edge detector returns a single pixel and connected edges that have a response above a threshold, where $0 < \text{threshold} < 1$. By setting the threshold to 0.99, only strong and connected edges are retained. Table 3 shows feature measurements for CD68 and 7/4 proteins.

An optimization hypothesis is derived based on object corner density and edge length. A high object corner density requires a smaller pooling stride to preserve the features. Longer edges

require a larger convolutional filter. Table 3 shows that CD68-positive macrophage samples contain, on average, 11.78% fewer Harris corners and 9.65% fewer MinEigen corners compared to those for 7/4-positive neutrophils. In contrast, CD68-positive macrophage samples display 15.61% stronger edges than 7/4-positive neutrophils. Following the optimization hypothesis, CD68-positive macrophages may require a relatively larger convolutional filter and a larger pooling stride, whereas 7/4-positive neutrophils, which exhibit a higher density feature, would require both a smaller convolutional filter and pooling stride. In Sec. 4.1, the experimental results of the best CNN architectures for CD68 and 7/4 are presented.

Figure 2 shows two sample images of CD68 and 7/4. The 7/4-positive neutrophils exhibit weaker edges and a lower edge density compared to the CD68-positive macrophages. Consequently, the architecture of CNN is defined for each type.

3 DeepQuantify System

Figure 3 shows the flowchart of the DeepQuantify system. It consists of preprocessing image segmentation using LIOTSu, CNN classification, muscle edge detection, and postprocessing with masks. Note that LIOTSu thresholding,²³ CNN classification, and muscle edge detection²³ can be implemented in parallel. The CNN classifier is designed as explained in Sec. 2.2.

Fig. 2 (a) 7/4 sample image and 7/4 WBC samples and (b) CD68 sample image and CD68 WBC samples.

Fig. 3 The DeepQuantify system.

In what follows, we describe the other steps of the overall detection and quantification scheme.

3.1 Preprocessing and LIOTSu-Based Segmentation

The preprocessing stage converts images into grayscale. Intensity variations within and across the images as well as contrast variations are common for the given type of imagery. Consequently, the input image is divided into 400×400 pixel blocks based on the results of the experiment. Next, the LIOTSu thresholding²³ is applied. Since the muscle fibers and the muscle edge can bias segmentation results, local thresholds are calculated in blocks.

Two parameters are preset: expected foreground/background ratio F and the object maximum count N . The value of F and N are obtained experimentally:

- If the local std ≥ 10 , $F < 0.5$ and $N < \text{std} \times 80$.
- If the local std < 10 , and $F < 0.25$, $N < \text{std} \times 40$.

The initial threshold value is obtained by maximizing the following variance:

$$\sigma_{\omega}^2 = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t), \quad (6)$$

where t is a threshold, ω is the probability of class separation by t , and σ^2 is the class variance. t is decided when σ_{ω}^2 is maximized.⁵ If the calculated foreground ratio and the maximum object area do not satisfy F and N , respectively, another iteration is invoked to acquire a new threshold value according to

$$t_i = St_{i-1}, \quad (7)$$

where i is an iteration number, t is a threshold, and S is a step size of each iteration. S satisfies the inequality $0.8 < S < 1$. For the i 'th iteration, if t_i satisfies

$$\frac{\text{area}_0(t_i)}{\text{area}_1(t_i)} < F \quad \text{AND} \quad N(t_i) < N, \quad (8)$$

then t_i is a determined threshold value.

Figure 4 illustrates the LIOTSu method and compares it to Otsu's method. Unlike its counterpart, the LIOTSu method works better for segmenting WBCs.

Fig. 4 Thresholding results for two sample images: (a) grayscale image, (b) Otsu results, (c) LIOTSu result, and (d) histogram and thresholds.

Fig. 5 The CNN postprocessing masks: (a) CNN classification result, (b) size mask, and (c) eccentricity mask.

3.2 Postprocessing Based on Size and Eccentricity Masks

For the muscle recovery study, the spatial distribution of WBCs is of interest. Size and eccentricity masks are created to determine features such as identity, size, connectivity, and eccentricity after the CNN completes the classification. The size mask defines the area of discrete or clustered cells. The eccentricity mask is derived from the width to length ratio (eccentricity) and provides useful information about the shape. These two masks are generated from the CNN classification results.

With an average cell size of 60 pixels, postprocessing is performed using a window of 8×8 pixels to ensure that every cell can be captured. For each sample image mask, normalization is applied. With defined lower and upper limits of the size (0 to Inf) and eccentricity (0 to 0.98 to exclude long connective tissue), we extract positive WBCs, as shown in Fig. 5. Details of postprocessing can be found in Ref. 24.

3.3 Muscle Edge Detection

As discussed earlier, WBCs located within 200 pixels of muscle edges are excluded from the subsequent analyses. Generally, the edge of the muscle can exhibit various visual properties such as continuity and intensity variance because of cell clustering. Thus, based on texture, it is more feasible to differentiate between void space out of the muscle edge and areas with muscle fiber or cells. For that purpose, we first enhance the contrast by linearly mapping the values of the input intensity to the full intensity range to enhance the muscle fiber. Next, the 3×3 range filter is applied. In areas with smooth texture or rough texture, the range of values in the neighborhood around a pixel will yield a smaller value or larger value, respectively. The output of the filter is then thresholded with a value between 0.05 and 0.1 as low variation is expected in the void areas. This procedure segments the image into textured and nontextured regions. The last step is the detection of an effective area adjacent to the image boundary based on the muscle edge location. Thereafter, objects at a distance of $d < 200$ pixels to the muscle edge are excluded.

3.4 Block-based Analysis in ROIs

The ROI is extracted such that only WBCs pertaining to the healing dynamics are quantified. Typically, ROIs shall not include the muscle edge and the surrounding tissues.

The ROI selection is based on the block unit. For each block, we use a score to determine if the block is a part of the ROI. The score is calculated based on the distance between each corner of the block and the muscle edge (< 200 pixels is effective; one point each); the number of effective objects (one point); and the percentage of the nonmuscle area ($> 50\%$ is effective; one point), which is an air gap in a block.²³ Effective ROI should contain at least two points. Figure 6 demonstrates the results of muscle edge detection and ROI extraction.

The ROI is divided into blocks of 200×200 pixels for the block-based analysis, which provides a fine-grain analysis of WBC dynamics in the local areas.

4 Experimental Results

In this section, we first report the experimental results of the best double convolutional layer architectures for different protein expressions based on the optimization hypothesis. The proposed architectures will be compared with the state-of-the-art classification architectures. Next, we analyze the segmentation results produced by LIOTsu, FCN, SegNet, U-Net, FCDN, DeepLabv3, and the proposed DeepQuantify system. Then WBC expressions of the CD68 and 7/4 proteins are quantified and analyzed using the proposed DeepQuantify system.

4.1 CNN Architectures for Two Proteins

4.1.1 Architectures selection

In Ref. 24, we have shown that a small stride of the pooling layer is not necessary for CD68-positive macrophages since a deep convolution filter can achieve the same performance at a faster rate. The experiments summarized in Table 4 indicate that the pooling strides of 1 and 2 nearly yield the same detection accuracy for CD68-positive macrophages. However, 7/4-positive neutrophils are most successfully classified with both a smaller convolutional filter and a smaller pooling stride. The experimental results confirm the optimization hypothesis. The depth of the convolutional filter is fixed to 64 because it is a maximal depth, considering limits of RAM.

Parameters of optimized CNN architectures and classifiers for two proteins are summarized in Table 5.

Fig. 6 Muscle edge detection and ROI selection: (a) original image, (b) thresholding result, (c) muscle edge detection, and (d) result of segmentation and an ROI.

Table 4 CNN architecture: pooling stride and convolutional filter depth.

CNN structure	M (2, 1)		M (2, 2)	
	Performance (%)	Time (s)	Performance (%)	Time (s)
CD68/performance				
C1 (3, 64, 1) to C2 (3, 64, 1)	99.2	344	98.7	42
C1 (5, 64, 2) to C2 (5, 64, 2)	99.2	355	99.2	45
C1 (7, 64, 3) to C2 (7, 64, 3)	99.3	362	99.2	52.5
C1 (5, 64, 2) to C2 (5, 512, 2) ¹⁷			99.3	193.5
7/4/performance				
C1 (3, 64, 1) to C2 (3, 64, 1)	97.9	635	96.3	77
C1 (5, 64, 2) to C2 (5, 64, 2)	97.3	666	96.6	83
C1 (7, 64, 3) to C2 (7, 64, 3)	96.8	675	96.4	95

4.1.2 Performance per validation of different architectures

Table 6 shows that deeper CNN classifiers can achieve the same or slightly better accuracy in classification than the proposed two-layer architectures. The result in Fig. 7 shows that the proposed CNNs achieve much faster convergence.

Table 6 CNN architecture: pooling stride and convolutional filter depth.

CNN structure	Performance (%)	Time (s)
CD68/performance		
Proposed CNN for CD68	99.3	193.5
VGG-16 ^{11,32}	99.7	1569
ResNet-18 ¹⁵	99.5	1175
ResNet-50 ¹⁵	99.1	9086
DenseNet-201 ³³	98.3	19,002
7/4/performance		
Proposed CNN for 7/4	97.9	635
VGG-16	97.1	1768
ResNet-18	97.9	1225
ResNet-50	97.9	20,302
DenseNet-201	96.1	41,746

4.2 Evaluation of DeepQuantify System

The system is evaluated by comparing the results with the ground truth provided by an expert in the field in the form of manually marked objects. Twenty images, which contain about 600 discrete cells overall, were evaluated. The system outputs are mapped on images and compared with the ground truth to evaluate quantification accuracy.

The ground truth was obtained as follows. The programmer provided the evaluated images (400 × 400 pixels). The expert,

Table 5 Optimized SS-CNN architectures.

Protein	Name	Structure						
		Input	C1	M1	C2	M2	F1	F2
CD68	SS-CNN	16 × 16	(5, 64, 2)	(2, 2)	(5, 512, 2)	(2, 2)	8192	2
7/4	SS-CNN	16 × 16	(3, 64, 1)	(2, 1)	(3, 64, 1)	(2, 1)	12,544	2

Fig. 7 Validation loss during CNN training.

as an observer, viewed each image using the microscope slide from which the image was captured. Using the 10× objective of a Nikon E600 bright field microscope, the observer examined the objects within this area. Infrequently, the observer used a 20× objective. CD68-positive and 7/4-positive cells were labeled manually based on several criteria: object color, size, shape, and edge texture. For example, after immunolabeling, these cells appear as shades of purple, ranging from light to dark. In contrast, debris or immunolabeling artifacts will often appear as white, black, or gray; exhibit rough edges; and be located above the section plane. Uncertain objects indicated by the observer are marked as “do not care” when evaluating.

The DeepQuantify system employing the proposed CNN architecture as well as those with other deeper CNN classifiers

are evaluated, including ResNet-18, ResNet-50, Vgg16, DenseNet-201. The system performance is also compared to the state-of-the-art deep segmentation and quantification architectures such as FCN,¹² SegNet,¹³ U-Net,¹⁷ FCDN,¹⁹ and DeepLabV3¹⁶ (ResNet). Specifically, since SegNet is composed of encoders and decoders, full-depth SegNet (Full-SegNet) and half-depth SegNet (Half-SegNet) are all considered. The time for sample preparation of deep segmentation architectures is 10 times higher compared to that of the DeepQuantify system, because the latter uses a sample-based classification, whereas the deep segmentation architecture performs a pixel-wise classification.

Tables 7 and 8 summarize precision, sensitivity, and accuracy computed by Eqs. (9)–(11). In terms of accuracy, first and second place are highlighted:

Table 7 Performance comparison for CD68-positive macrophages.

	False positives	False negatives	Total count	Precision (%)	Sensitivity (%)	Accuracy (%)
FCN-32s	102	4	422	75.83	98.77	75.12
FCN-16s	143	0	457	68.71	100.00	68.71
FCN-8s	70	3	389	75.59	98.76	81.38
Full-SegNet	22	39	323	93.19	88.53	83.15
Half-SegNet	22	40	329	93.31	88.47	83.20
U-Net	58	60	338	82.84	82.35	70.35
FCDN-103	93	19	412	77.43	94.38	74.01
DeepLabV3	32	30	334	90.42	90.96	82.97
DeepQuantify (proposed CNN)	25	7	335	93.43	97.81	90.64
DeepQuantify (ResNet-18)	42	3	359	88.30	99.06	87.57
DeepQuantify (ResNet-50)	28	9	333	91.59	97.13	89.18
DeepQuantify (Vgg16)	47	80	370	87.30	100	87.30
DeepQuantify (DenseNet-201)	70	0	392	82.14	100.00	82.14

Note: Bold items are the first and second place of the results.

Table 8 Performance comparison for 7/4-positive neutrophils.

	False positives	False negatives	Total count	Precision (%)	Sensitivity (%)	Accuracy (%)
FCN-32s	108	6	428	74.77	98.16	73.73
FCN-16s	28	32	296	90.54	89.33	81.71
FCN-8s	27	35	296	90.88	88.49	81.27
Full-SegNet	15	48	293	94.88	85.28	81.52
Half-SegNet	4	72	251	98.41	77.43	76.47
U-Net	28	46	337	91.69	87.04	80.68
FCDN-103	18	54	289	93.77	83.38	79.01
DeepLabV3	28	34	343	91.84	90.26	83.55
DeepQuantify (proposed CNN)	14	19	295	95.25	93.67	88.92
DeepQuantify (ResNet-18)	27	10	336	91.96	96.87	89.31
DeepQuantify(ResNet-50)	31	17	325	90.46	94.53	85.96
DeepQuantify (Vgg16)	10	42	259	96.14	85.57	82.72
DeepQuantify (DenseNet-201)	14	30	283	95.05	89.97	85.94

Note: Bold items are the first and second place of the results.

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (9)$$

$$\text{sensitivity} = \text{TP}/(\text{TP} + \text{FN}), \quad (10)$$

$$\text{accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}). \quad (11)$$

For all methods, there is a trade-off between the number of false positives and false negatives. For CD68-positive macrophages, full-SegNet and half-SegNet show higher accuracies and DeepLabv3 (ResNet-50) has balanced trade-offs than other segmentation networks. That has led to the conclusion that a high depth is not necessary in this case. The proposed DeepQuantify system with the proposed CNN achieves the best accuracy and then follows the DeepQuantify with ResNet-50.

Segmentation and quantification of 7/4-positive neutrophils are more challenging. The DeepQuantify with the ResNet-18 has the best performance in terms of the overall false, confirming the advantage of ResNets in information utilization particularly in more complex cases like 7/4. With only 0.39% lower in accuracy, the DeepQuantify system with the proposed CNN is ranked the second best. If we consider the down-sampling section (classifier) as an encoder and up-sampling section as a decoder, with the same encoder, the proposed DeepQuantify system achieves overall better quantification performance than those deep segmentation architectures. Figure 8 illustrates the major steps and segmentation results for CD68 macrophages and 7/4 neutrophils by the DeepQuantify system.

4.3 Quantification Results

To characterize the WBC dynamics, the following quantities are obtained:

- number of discrete cells,
- number of clusters,
- average size of a discrete cell,
- total area covered by clusters,
- number of discrete cells in clusters,

$$N_{\text{cic}} = \frac{\text{area}_{\text{cluster}}}{E(\text{size}_{\text{WBC}})} \quad (12)$$

- total number of cells,
- number of cells per block,
- frequency of occurrence of blocks with a number of cells within a certain range, i.e., a histogram.

The average values from all samples are presented in Table 9.

From Table 9, an average of 62.90 discrete CD68-positive macrophages and only 1.1 clusters are found in the uninjured muscles. At 24-h postinjury, the number of CD68-positive macrophages starts increasing in connective tissue. From 24- to 48-h postinjury, the increase of discrete macrophages slows. And the number of clusters increases at 32-h postinjury. The number of macrophage clusters is 2.27 times higher than that at 24 h. The number of CD68-positive macrophages and CD68-positive clusters reaches a peak at 96 h. Intensive clustering in the connective tissue can be observed at 96-h postinjury; there are 3.08 macrophages per 1000 μm^2 . After 96 h, both the CD68-positive discrete cell and cluster counts decrease. This result is confirmed by research in Ref. 2. The size of discrete CD68-positive cells slightly increases after regeneration begins.

Table 10 demonstrates the quantification result of 7/4-positive neutrophils. At 4-h postinjury, the 7/4 neutrophils show a high response to the injury. Both discrete and clustered cell

Fig. 8 The segmentation results by the DeepQuantify system for (1) and (2) 7/4 neutrophils and (3) and (4) CD68 macrophages: (a) original images, (b) segmentation results, (c) classification result, (d) application of masks, and (e) final results.

Table 9 Features extracted from muscle images for CD68-positive macrophages.

Number per image	Uninjured	24-h	32-h	48-h	96-h	192-h
Discrete cells, #	62.90	348.00	403.26	427.59	588.78	313.10
Clustered cells, #	1.10	27.60	62.57	100.00	116.11	30.60
Size of discrete cells (pixels)	43.75	57.11	58.39	59.82	58.510	56.44
Cells per 1000 μm^2 , #	0.07	0.52	1.04	1.36	3.08	0.55
Overall cell #	70.13	525.20	906.43	1328.12	3097.44	538.80

counts reach the peak at 4-h postinjury. From 8- to 192-h postinjury, the numbers of both discrete and clustered 7/4 neutrophils decrease steadily. After 48-h postinjury, 7/4 neutrophils appear as lighter and smaller objects, distributed sparsely. At 192-h postinjury, 7/4 neutrophils are not consistently present. Even at their peak presence, the number of 7/4-positive neutrophils is not as high as that of CD68-positive macrophages. On average, 0.96 cells per 1000 μm^2 are observed at 4-h following

injury. More clusters are present at 32 h than at 24 h. However, this difference could be due to the low number of 24-h samples. By comparing data for CD68 and 7/4 images, we have found that both discrete and clustered CD68 macrophages surpass 7/4 neutrophils at around 32-h after injury. It can be concluded that 7/4 neutrophils are active for the muscle injury recovery at early stages, whereas CD68 macrophages are more active at later stages. The dynamics of CD68 and 7/4 are presented in Fig. 9.

Table 10 Features extracted from muscle images for 7/4-positive neutrophils.

Per image	4-h	8-h	24-h	32-h	48-h	96-h	192-h
Discrete cells, #	427.60	377.70	406.60	353.55	308	80.67	4.00
Clustered cells, #	71.60	71.50	55.40	63.91	52.80	5.44	0.6
Size of discrete cells (pixels)	59.06	64.49	62.01	62.13	59.63	55.56	58.43
Cells per 1000 μm^2 , #	0.96	0.91	0.73	0.75	0.70	0.10	0.01
Overall cell #	971.80	882.20	745.20	755.55	690.00	106.56	9.60

4.4 Cell Density Analysis

The analysis of cell distribution in a local area is useful to study the dynamics across the injured area. The density distribution of CD68-positive cells is represented by the stacked view of the binned block histogram: the percentage of blocks containing 0 to 20, 21 to 40, ..., 141 to 160, and >160 cells, which are selected as representative bin sizes. Due to the variable size of ROIs, the distribution of blocks is normalized by ROI areas. The histogram of CD68-positive cell distribution is presented in Fig. 10. Figure 10(a) shows a scaled histogram for CD68. For uninjured muscles, over 99% blocks contain only 0 to 20 CD68 macrophages. After the injury, the number of blocks that contain more than 20 cells increases. Within 24-h after injury, although more blocks contain 21 to 40 cells, no significant clustering is observed. From 32- to 96-h postinjury, the number of CD68-positive macrophage clusters increases, and at 96-h postinjury, there is an intensive clustering.

Fig. 9 Number of CD68-positive and 7/4-positive cells at different time points: (a) discrete cells, (b) clusters, and (c) density.

Fig. 10 Block histograms: (a) CD68 and (b) 7/4.

For 7/4-positive neutrophils, as shown in Fig. 10(b), a steady decrease of density occurs between 4-and 48-h postinjury. Between 96 and 192 h, only 0 to 20 cells can be observed.

5 Conclusion

In this paper, we have developed an automatic quantification and analysis system called DeepQuantify to analyze WBCs in light microscopy muscle images. The system uses the LIOTSu thresholding method and a two-layer CNN classifier architecture that is optimized for each protein type. A hybrid masking approach was proposed and used on the segmentation result. The muscle edge detection and the ROI selection procedures exclude possibly false positive cells around the muscle edge. The DeepQuantify system avoids the segmentation labeling challenge in microscopy images and achieves better accuracy of WBC quantification than deep segmentation architectures. In DeepQuantify, several computation steps can be implemented in parallel. Thus it is also computationally efficient. In addition, DeepQuantify is fully automated, which makes it very suitable for processing large image sets. As the methodology proved to be efficient with high accuracy, it can be used for quantifying other protein expressions.

Disclosure

For this research, 106 female mice were involved. The animal procedures were approved by the UNLV IACUC and the Department of the Air Force AFMSA/SGEC Office of Research Oversight and Compliance. Mr. Jiao reports the support from UNLV Doctoral Graduate Research Assistantship during the conduct of the study. Dr. Schneider reports a grant from the Department of Defense Air Force Grant (No. FA-7014-10-2-0001) and the support from the UNLV Faculty Opportunity Award and UNLV Doctoral Graduate Research Assistantship during the conduct of the study. Dr. Regentova reports the support from the UNLV Faculty Opportunity Award during the conduct of the study. Dr. Yang reports the support from the UNLV Faculty Opportunity Award and UNLV Doctoral Graduate Research Assistantship during the conduct of the study.

Acknowledgments

This work was supported in part by the UNLV Faculty Opportunity Award, UNLV Doctoral Graduate Research Assistantship, and Department of Defense Air Force Grant (No. FA-7014-10-2-0001). Review of the material does not imply Department of the Air Force endorsement of factual accuracy or opinion.

References

1. S. Keeling et al., "Macrophage colony stimulating factor-induced macrophage differentiation influences myotube elongation," *Biol. Res. Nurs.* **15**(1), 62–70 (2012).
2. G. Dobek et al., "Mouse model of muscle crush injury of the legs," *Comp. Med.* **63**(3), 227–232 (2013).
3. C. Schneider, W. Rasband, and K. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nat. Methods* **9**, 671–675 (2012).
4. M. Lejeune et al., "Quantification of diverse subcellular immunohistochemical markers with clinicobiological relevancies: validation of a new computer-assisted image analysis procedure," *J. Anat.* **212**(6), 868–878 (2008).
5. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979).
6. S. Khobragade, D. Mor, and C. Patil, "Detection of leukemia in microscopic white blood cell images," in *Proc. Int. Conf. Inf. Process.*, pp. 435–440 (2015).
7. Y. Liu et al., "Segmentation of white blood cells image using adaptive location and iteration," *IEEE J. Biomed. Health Inf.* **21**(6), 1644–1655 (2017).
8. Y. Lecun et al., "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**(11), 2278–2324 (1998).
9. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, pp. 1097–1105 (2012).
10. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, pp. 1–14 (2015).
11. C. Szegedy et al., "Going deeper with convolutions," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1–9 (2015).
12. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015).
13. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **39**(12), 2481–2495 (2017).
14. J. Donahue et al., "DeCAF: a deep convolutional activation feature for genetic visual recognition," in *Proc. Int. Conf. Mach. Learn. (ICML)*, pp. 647–655 (2014).
15. K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, pp. 770–778 (2014).
16. L.C. Chen et al., "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2018).
17. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241 (2015).
18. Y. Li, L. Shen, and S. Yu, "HEP-2 specimen image segmentation and classification using very deep fully convolutional network," *IEEE Trans. Med. Imaging* **36**(7), 1561–1572 (2017).
19. S. Jegou et al., "The one hundred layers Tiramisu: fully convolutional DenseNets for semantic segmentation," arXiv:1611.09326v3 (2017).
20. S. Otalora et al., "Determining the scale of image patches using a deep learning approach," in *Proc. IEEE 15th Int. Symp. Biomed. Imaging (ISBI)*, pp. 843–846 (2018).
21. K. Kamnitsas et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.* **36**, 61–78 (2017).
22. R. Li et al., "Deep learning segmentation of optical microscopy images improves 3D neuron reconstruction," *IEEE Trans. Med. Imaging* **36**(7), 1533–1541 (2017).
23. Y. Jiao et al., "Automated quantification of white blood cells in light microscopy images of injured skeletal muscle," in *Proc. IEEE 8th Annu. Comput. and Commun. Workshop and Conf.*, pp. 1–7 (2018).
24. Y. Jiao et al., "Automated quantification of white blood cells in light microscopy muscle images: segmentation augmented by CNN," in *Proc. Int. Conf. Vision, Image and Signal Process.* (2018).
25. J. Nicholas et al., "Time course of chemokine expression and leukocyte infiltration after acute skeletal muscle injury in mice," *Innate Immun.* **21**(3), 266–274 (2014).
26. X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in *Proc. 14th Int. Conf. Artif. Intell. and Stat.*, pp. 315–323 (2011).
27. N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural networks* **12**(1), 145–151 (1999).
28. S. Ruder, "An overview of gradient descent optimization algorithms," arXiv:1609.04747 (2016).
29. C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conf.*, pp. 147–151 (1988).
30. J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 593–600 (1994).
31. J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**(6), 679–698 (1986).

32. S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit.*, pp. 730–734 (2015).
33. G. Huang et al., "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 2261–2269 (2017).

Yang Jiao received his BS and MS degrees in electrical and computer engineering. He is currently a PhD student at the Department of Electrical and Computer Engineering, University of Nevada, Las Vegas. Having more than 5 years' experience on image processing and 3 years' practice on bioimage processing, he focuses his research on cell and protein detection, counting, and tracking in microscopy and fluorescent images.

Barbara St. Pierre Schneider's work has amplified our nascent knowledge of the skeletal muscle healing process for more than 25 years. In many studies, she has used image analysis to quantify the antigen profile of leukocytes active during the healing process. Her research has been integral to the health care community with

the long-term goal of reducing issues, such as prolonged healing and disability, and ideally improving patients' quality of life.

Emma Regentova joined the University of Nevada, Las Vegas, in 2000. She is currently a professor in the Electrical and Computer Engineering Department. Her research interests comprise image processing, analysis, decision-making systems, digital design, and embedded DSP with applications to remote sensing, medical diagnostics, health sciences, intelligent transportation systems, and material identification for security and control systems.

Mei Yang received her PhD in computer science from the University of Texas at Dallas in August 2003. In August 2004, she joined in the Department of Electrical and Computer Engineering, University of Nevada, Las Vegas, where she has been a full professor since 2016. Her research interests include machine learning, computer architectures, networking, and embedded systems. She has published more than 150 peer-reviewed conference and journal papers in these areas.