MIMO-NET: A MULTI-INPUT MULTI-OUTPUT CONVOLUTIONAL NEURAL NETWORK FOR CELL SEGMENTATION IN FLUORESCENCE MICROSCOPY IMAGES

Shan E Ahmed Raza[†] Linda Cheung [*] David Epstein [‡] Stella Pelengaris [*]
Michael Khan [*] Nasir M. Rajpoot [†]
[†] Department of Computer Science, University of Warwick, Coventry, UK
[*]School of Life Sciences, University of Warwick, Coventry, UK
[‡]Department of Mathematics, University of Warwick, Coventry, UK

ABSTRACT

We propose a novel multiple-input multiple-output convolution neural network (MIMO-Net) for cell segmentation in fluorescence microscopy images. The proposed network trains the network parameters using multiple resolutions of the input image, connects the intermediate layers for better localization and context and generates the output using multi-resolution deconvolution filters. The MIMO-Net allows us to deal with variable intensity cell boundaries and highly variable cell size in the mouse pancreatic tissue by adding extra convolutional layers which bypass the max-pooling operation. The results show that our method outperforms state-of-the-art deep learning based approaches for segmentation.

Index Terms**—** Cell Segmentation, Fluorescence Microscopy, Deep Learning

1. INTRODUCTION

Cell segmentation is an important task in biomedical image analysis involving cell level analysis [1]. In fluorescence microscopy images, this task is challenging for various reasons, for example the relatively large variation in the intensity of captured signal and the difficulty with separating neighbouring cells. It requires careful tuning of the algorithm to make it robust to intensity, shape, size and fusion of individual cellular regions. That process can require experimentation with a variety of features and can be time consuming. In contrast, deep learning based approaches have shown the great power of data-driven learning of features [2]. In this paper, we present a novel multi-input multi-output convolution neural network (MIMO-Net) to solve the problem of cell segmentation in fluorescence microscopy images.

A detailed review of cell segmentation methods has been presented by Meijering [3] for images from various modalities using membrane and cytoplasmic markers. We focus on segmentation of individual cells using fluorescent images of nuclear and membrane markers. Membrane markers such as E-cadherin (or Ecad) mark the boundary of individual cells, but the intensity of the membrane markers varies depending on the type and orientation of each cell.

Most of the existing approaches to cell segmentation employ thresholding, filtering, morphological operations, region accumulation, deformable model fitting [4], graph cut [5] and feature classification [6]. In this paper, however, we focus on deep learning based approaches using convolutional neural networks (CNNs) which have recently become popular with promising results for various image processing tasks such as segmentation [7, 8, 9, 10]. The fully convolutional network (FCN) for segmentation is considered to be a benchmark for segmentation tasks using deep learning [7]. The network performs pixel-wise classification to get the segmentation mask and consists of downsampling and upsampling paths. The downsampling path consists of convolution and max-pooling and the upsampling path consists of convolution and deconvolution (convolution transpose) layers. U-Net [9] is inspired by FCN but connects intermediate downsampling and upsampling paths to conserve the context information. DCAN [8] employs a modified FCN and trains the network for both object and contour features to perform segmentation. Another recently proposed multi-scale convolutional neural network [10] trains the network at different scales of the Laplacian pyramid and merges the network in the upsampling path to perform segmentation. In this paper, we propose a CNN which adds extra layers in the downsampling path which bypass the maxpooling operation to learn the parameters for segmentation of variable intensity cells. The network retains context information, interprets the output at multiple resolutions and trains the network at multiple input image resolutions in the downsampling path to learn the network parameters for variable cell sizes in the presence of variable intensities.

2. MATERIALS AND METHODS

2.1. Image Acquisition and Pre-processing

A multi-channel fluorescence microscope known as the Toponome Imaging System (TIS) [11], acquired images of tissue samples from mouse pancreata [12]. The TIS microscope is capable of capturing signals from multiple biomarkers, but for cell segmentation we only employ two channels corresponding to Ecad (membrane marker) and DAPI (nu-

978-1-5090-1172-8/17/$31.00 ©2017 IEEE

clear marker). After segmentation work is completed, the other channels are available to study individual cells, and to group similar cells together for statistical purposes. We performed alignment and normalization of the multi-channel images using the protocols designed for pre-processing of the TIS data [13, 14]. Next, ground truth for image segmentation, marked by an expert biologist, was used for training.

Sample images of exocrine cells and endocrine cells are shown in Fig. 1 as RGB composite images (enhanced for display), where membrane marker is shown in green, nuclear marker in blue and ground truth is overlaid in red with black boundaries. One can observe the variation in intensities of cell boundaries and that the nuclei are not always present and, if present, are not always positioned at the centre of the cell. In addition, endocrine cells are more tightly packed and are of smaller size compared to exocrine cells. These variations make it a challenging task to segment the cells in these images. In the next section, we propose MIMO-Net to segment these types of cells in this type of environment. The proposed approach can be readily extended to work with a variety of environments and a wider variety of cell types.

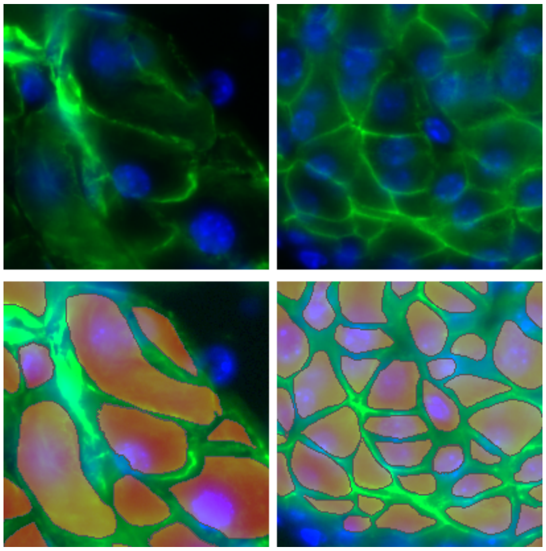

Fig. 1. Top row: Membrane marker is shown in green and nuclear marker in blue. Bottom row: ground truth is overlaid in red with black boundaries. Left: Exocrine Cells. Right: Endocrine Cells.

2.2. The Proposed Network

The architecture of the proposed MIMO-Net is shown in Fig. 2. The input to the network consists of two features, i.e., membrane and nuclear marker images. The network is divided into five groups and thirteen branches, the division depending on their function and the set of layers/filters.

The first group, which consists of four branches with output B1-B4, constructs the downsampling path. Each branch in Group 1 consists of convolution, max-pooling, resize and concatenation layers. The convolution and max-pooling layers perform standard operations as in conventional CNNs. We use $tanh$ activation after each convolution layer as our experiments showed that the network converges faster with $tanh$ activation. The resize layer resizes the image using bicubic interpolation so that the resized image dimension matches the corresponding dimension of the max-pooling output. We add the lower resolution input to retain the information from pixels that do not have the maximum response, because they are in the vicinity of a noisy neighbourhood. This is particulary useful when we are trying to detect cells with boundary markers having extreme intensities, even for individual cells as shown in Fig. 1. Thus learning the features in the presence of noise by bypassing the maxpooling operations. Another aspect of the resizing operation is to train the network on different sized cells as explained in Section 2.1. The output of branch 1 (B1) has feature depth of size 128 where the first half (64) of the features are the result of the max-pooling operation and the next half (64) are obtained by performing convolutions only on the resized image. The following branches in Group 1 double the feature depth of the previous branch but follow the same protocol in generating the branch output.

Group 2 consists only of branch 5 and performs convolution operations. Group 3 forms the upsampling path and consists of branches 6,7,8 & 9. Each of these branches takes two inputs, one from the previous branch and one from the branch with the closest feature dimension in the downsampling path. The output of each branch is double in height and width and half the depth of previous branch. The second input is added from the downsampling path for better localization and to capture the context information as in [9]. It also passes the convolution only features to the upsampling path, which helps to learn from the features which do not have maximum response in downsampling path. Compared to the U-Net [9], we add additional deconvolution layers instead of cropping the feature from the downsampling path. This allows us to produce a segmentation map of the same size as the input image and an overlap-tile strategy is not required. It also reduces the number of patches required to produce the desired segmentation output thus removing computational steps.

Group 4 & 5 generate the auxiliary and the main output and calculate the loss. Group 4 consists of three branches where each branch takes the output from one of B7-B9 and generates three auxiliary feature masks, which are fed into the main output branch. The output branch concatenates feature masks and performs convolution followed by softmax classification to get the segmentation output map $p_o(x)$ where x represents a pixel location. The output of branches B7-B9 are of different resolutions and so the deconvolution layer in each of the auxiliary branches is set to generate the output of the same size [8]. The deconvolution is followed by a convolution layer which produces the auxiliary feature mask. Each of the auxiliary feature masks is followed by a dropout layer (set to 50%) and the convolution layer followed by softmax classifi-

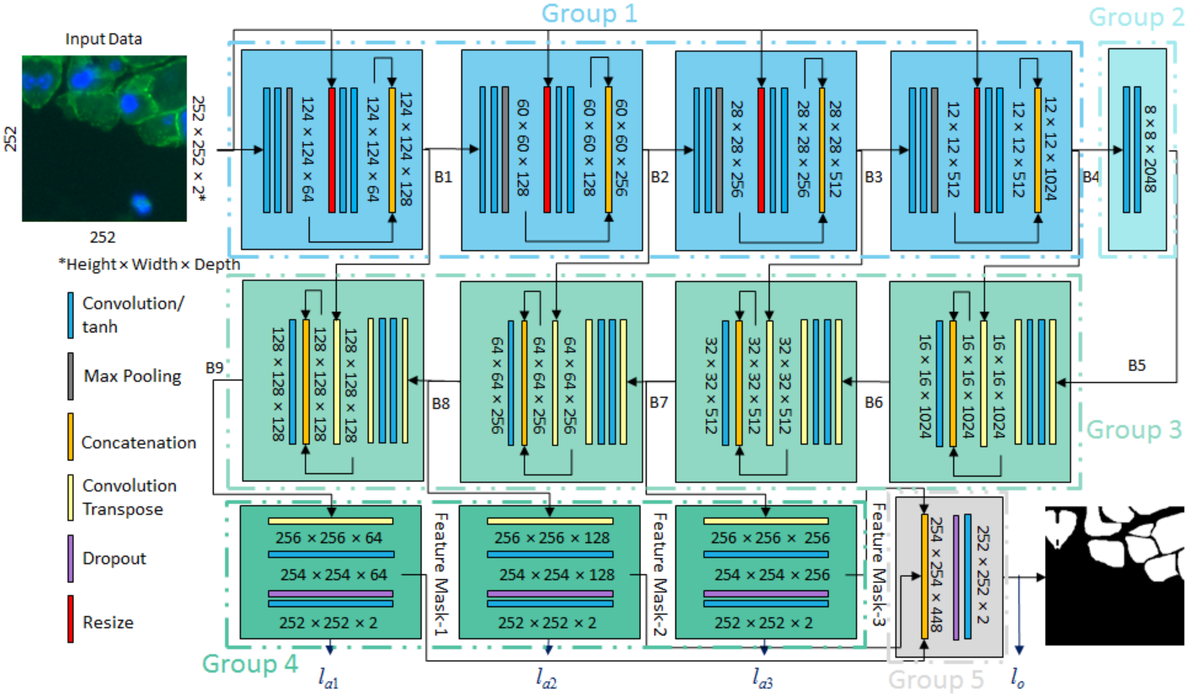

Fig. 2. The proposed MIMO-Net architecture.

cation to get the auxiliary outputs ($p_{a1}(x)$, $p_{a2}(x)$, $p_{a3}(x)$).

For training, we calculate weighted cross entropy loss for the main output (l_o) and the auxiliary outputs (l_{a1}, l_{a2}, l_{a3}) as

$$l_k = \sum_{x \in \Omega} w(x) \log(p_{k(x)}(x)) \tag{1}$$

where $k \in \{o, a1, a2, a3\}$ and Ω is the set of pixel locations in the input image. The weight function $w(x)$ gives higher weights to pixels which are at the merging cell boundaries, leading to a higher penalty [9]. The total loss(l) is calculated by combining auxiliary and main loss by using $l = l_o + (l_{a1} + l_{a2} + l_{a3})/epoch$ where $epoch > 0$ represents the number of training passes through the data. This strategy reduces the contribution of auxiliary losses for a higher $epoch$.

3. EXPERIMENTAL RESULTS

Our image data consists of 11,163 cells of which 6,641 (60%) are used for training and 4,522 (40%) for testing. We compare our results with the state-of-the-art FCN8 [7], DCAN [8] and U-Net [9] networks. The proposed network was implemented using TensorFlow v0.12 [15]. We start with a learning rate ($lr = 0.01$) and reduce it according to $lr = (epoch * 100)^{-1}$. To train the network, we perform data augmentation using Gaussian noise, lens distortion, flip and rotate. We used authors' implementation of FCN8 and trained it for our data, whereas DCAN and U-Net were implemented in TensorFlow.

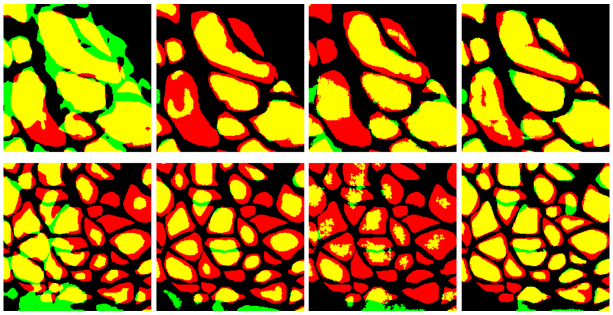

Fig. 3. Segmentation results, ground truth in red and output of the algorithm in green. Top row: Exocrine region. Bottom row: Endocrine region. Columns (left to right) are output from FCN8, U-Net, DCAN and MIMO-Net architectures.

The results (Fig. 3) show that FCN8 identified cellular regions but was not able to segment individual cells. DCAN is designed to learn the contour features and performed better segmentation of the cells in the exocrine region but performed poorly with smaller sized cells in the endocrine region. U-Net performed better than both FCN and DCAN but missed the cells with weaker boundaries. The proposed MIMO-Net method, performed better in the presence of variable intensities and variable size/shape of the cells. The output in Fig. 3 was post-processed for all the 5 algorithms using area opening (100 pixels) and hole filling operations to get the final output score in Table 1. For quantitative analysis, we used mea-

Table 1. Quantitative results for cell segmentation in terms of Dice coefficient, F1 score, Object Dice (OD), Pixel Accuracy (Acc) & Object Hausdorff (OH).

Network	Dice	F1	OD	PAcc	OH
FCN8 [7]	76.9%	8.2%	5.9%	73.8%	1350
FCN8W	71.4%	50.5%	50.9%	74.6%	91.8
DCAN [8]	76.0%	61.4%	63.8%	78.7%	42.3
UNet [9]	78.4%	66.4%	67.3%	80.3%	40.5
Proposed	**82.4%**	**71.8%**	**74.1%**	**83.5%**	**27.5**

sures which include Dice coefficient, F1 score, object Dice, pixel accuracy and object Hausdorff [2]. Hausdorff distance is lower and the rest of the measures are higher for better results. The quantitative results are shown in Table 1 which show that the proposed MIMO-Net method outperforms the state-of-the-art deep learning approaches with at least 3-4% margin in terms of average Dice, F1 score, object Dice, pixel accuracy and object Hausdorff. We modified the FCN8 algorithm (FCN8W) by introducing weighed loss [9] to improve segmentation of individual cells. FCN8W improved F1, object Dice, pixel accuracy and object Hausdorff but failed to increase the Dice coefficient. On average the network takes 2.50 sec for training and 0.39 sec for testing a batch of 5 images TitanX Maxwell on a Windows10 machine with Intel Xeon E5-2670 v2 CPU and 96 GB RAM.

4. CONCLUSIONS

Cell segmentation is an important step for cell-level analysis of biomedical images. With appropriate equipment and analysis, it enables us to compute protein profiles on the basis of individual cells, leading to cell phenotyping and detailed and soundly based cell level statistics. Images captured using fluorescence microscopy contain very weak and variable intensities which makes it difficult to segment the cells in these kind of images. The variable size of the cells makes it even more challenging for image processing algorithms to perform cell segmentation. We propose a MIMO-Net architecture to deal with both variable intensities and variable size and shape of cells. Intermediate connections between the layers allows the context and localization to be retained. The qualitative and quantitative results show that the MIMO-Net architecture outperforms state-of-the-art deep learning approaches. We plan to compare the architecture performance on histopathology images in future.

5. ACKNOWLEDGEMENTS

We are grateful to the BBSRC UK for supporting this study through project grant BB/K018868/1.

6. REFERENCES

[1] A. M. Khan et al., “Cell phenotyping in multi-tag fluorescent bioimages,” *Neurocomputing*, vol. 134, pp. 254–261, jun 2014.

[2] K. Sirinukunwattana et al., “Gland segmentation in colon histology images: The glas challenge contest,” *Medical Image Analysis*, vol. 35, pp. 489–502, 2016.

[3] E. Meijering, “Cell segmentation: 50 years down the road,” *Signal Processing Magazine, IEEE*, vol. 29, no. 5, pp. 140–145, 2012.

[4] J. Bergeest and K. Rohr, “Efficient globally optimal segmentation of cells in fluorescence microscopy images using level sets and convex energy functionals,” *Medical image analysis*, vol. 16, no. 7, pp. 1436–1444, 2012.

[5] S. Dimopoulos et al., “Accurate cell segmentation in microscopy images using membrane patterns,” *Bioinformatics*, vol. 30, no. 18, 2014.

[6] G. Li et al., “A novel multitarget tracking algorithm for myosin vi protein molecules on actin filaments in tirfm sequences,” *Journal of Microscopy*, vol. 260, no. 3, pp. 312–325, 2015.

[7] E. Shelhamer et al., “Fully convolutional networks for semantic segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , no. 99, pp. 1–1, 2016.

[8] H. Chen et al., “Dcan: Deep contour-aware networks for accurate gland segmentation,” *arXiv preprint arXiv:1604.02677*, 2016.

[9] O. Ronneberger et al., “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[10] Y. Song et al., “Accurate cervical cell segmentation from overlapping clumps in pap smear images,” *IEEE Transactions on Medical Imaging*, , no. 99, pp. 1–1, 2016.

[11] W. Schubert et al., “Analyzing proteome topology and function by automated multidimensional fluorescence microscopy,” *Nature biotechnology*, vol. 24, no. 10, pp. 1270–1278, 2006.

[12] S. Pelengaris, S. Abouna, et al., “Brief inactivation of c-myc is not sufficient for sustained regression of c-myc-induced tumours of pancreatic islets and skin epidermis,” *BMC Biology*, vol. 2, no. 1, pp. 26, 2004.

[13] S.E.A. Raza et al., “RAMTaB: robust alignment of multi-tag bioimages.,” *PLoS ONE*, vol. 7, no. 2, pp. e30894, jan 2012.

[14] S.E.A. Raza et al., “Robust normalization protocols for multiplexed fluorescence bioimage analysis,” *BioData Mining*, vol. 9, no. 1, pp. 11, 2016.

[15] M. Abadi, A. Agarwal, et al., “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, Software available from tensorflow.org.