Neuron Segmentation in Epifluorescence Microscopy Imaging with Deep Learning

Fernando González-Colín[a], Boris Escalante-Ramírez[b,d], Jimena Olveres[b,d], José Bargas[c], Miguel Serrano-Reyes[c]

[a]Posgrado en Ciencia e Ingeniería de la Computación, Universidad Nacional Autónoma de México, Mexico City, Mexico,[b]Facultad de Ingeniería, Universidad Nacional Autónoma de México, Mexico City, Mexico; [c]Instituo de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico City, Mexico; [d]Centro de Estudios en Computación Avanzada, Universidad Nacional Autónoma de México, Mexico City, Mexico.

ABSTRACT

Epifluorescence Microscopy Imaging is a technique used by neuroscientists for observation of hundreds of neurons at the same time, with single-cell resolution and low cost from living tissue. Recording, identifying and tracking neurons and their activity in those observations is a crucial step for researching. However, manual identification of neurons is a hard-working task as well as prone to errors. For this reason, automatized applications to process the recordings to identify functional neurons are required. Several proposals have emerged; they can be classified in four kinds of approaches: 1) matrix factorization, 2) clustering, 3) dictionary learning and 4) deep learning. Unfortunately, they have resulted inadequate to solve this problem. In fact, it remains as an open problem; two major reasons are: 1) lack of datasets duly labeled and 2) existing approaches do not consider the temporal dimension or just consider a tiny fraction of it, integrating all the frames in a single image is very common but inefficient because temporal dynamics are disregarded. We propose an application for automatic segmentation of neurons with a Deep Learning approach, considering temporal dimension through recurrent neural networks and using a dataset labeled by neuroscientists. Additional aspects considered in our proposal include motion correction and validation to ensure that segmentations correspond to truly functional neurons. Furthermore, we compare this application with a previous proposal which uses sophisticated digital image processing techniques on the same dataset.

Keywords: Digital image processing, deep learning, motion correction, neuroscience, neurons, epifluorescence microscopy imaging, segmentation

1. INTRODUCTION

Human brain has nearly 85 billion neurons [1], highly specialized cells whose main feature is generating and spreading electrical signals named action potentials. Sensitive cognitive functionalities emerge from the connection of these cells in very sophisticated networks. For this reason, it is valuable to understand the neuronal structure, the mechanisms which make it possible to generate and broadcast electric signals, network characteristics and the relation between those and neuronal activity patterns. Understanding the dynamics of neurons could lead to the understanding and treatment of neurological disorders like Parkinson's disease [2, 3]. In this work, neuronal activity is understood as the occurrence of action potentials in the cell. An action potential is an event which lasts approximately 1-2 ms and raises the voltage of the neuron up to 20 - 60 mv. Each neuron performs action potentials in different patterns, this is related to the physiology of the cell. Once an action potential is generated, it is spread across the neuron axon to reach another cell [4]. Understanding the nervous system requires understanding how neurons interact between them. Therefore, recording and analyzing neuronal activity are crucial steps in the search for that knowledge. A common technique to acquire recordings is calcium imaging. It allows recording neuronal activity across hundreds of neurons at the same time with single-cell resolution [2, 5, 6, 7, 8, 9]. Usually, the record is saved as a video, then neuroscientists analyze it meticulously looking for functional neurons. A functional neuron is one with fluorescence changes seen in videos as small blobs with intensity changes in their pixels across the time. Fluorescence changes are associated with action potentials. Thus, it is possible to find hundreds of small flashing blobs corresponding to hundreds of neurons to be identified. However, this task is

tedious, tiring and prone to human errors. Experienced neuroscientists need to review each frame for an average of 30 minutes, looking for slight intensity changes in neurons. In this scenario, it is necessary to develop tools to support researchers in this hard-working task [3]. The problem remains as an open one due to the challenges involved, as discussed later.

In this work we are presenting our proposal to tackle the described problem. As discussed in Section 2, several approaches have failed because they consider only spatial information while the temporal is compressed or disregarded. We are proposing a combination of Convolutional Neural Networks (CNNs) to analyze spatial information frame by frame and Recurrent Neural Networks (RNNs) to exploit temporal dynamics in order to achieve a deeper analysis. CNNs have been very successful in many segmentation tasks, specifically medical segmentation of cells and tissue [10]. At the same time, RNNs have proven a good performance identifying temporal patterns in natural language applications and other signal processing tasks [11]. They alone cannot offer a complete solution; preprocessing is needed to increase CNN performance and post-processing to refine and validate outputs, as we show in this work. The scope of this work is to validate the premise that considering RNNs to analyze temporal dynamics leads to better segmentations for the problem.

Section 2 reviews briefly related work in order to understand the challenges of the problem and the different proposed approaches so far. In Section 3 our proposal is explained. Experiments to test our proposal are described in Section 4 and results are presented in Section 5. Finally, in Section 6 our conclusions are drawn, and future work is proposed.

2. RELATED WORK

Even though several approaches have been proposed [3, 12, 13, 14, 15], there is a consensus that the problem remains open [5]. Most of the proposals can be gathered in four kinds of techniques: 1) matrix factorization, 2) clustering, 3) dictionary learning and 4) deep learning [15]. Traditional approaches, such as thresholding or mathematical morphology, have not been precise enough [3]. On the other hand, proposals grounded on deep learning have turned out prohibitive due to lack of properly labeled datasets [18]. A common practice is to identify, and segment observed cells in a cumulative image obtained from the maximum or the average value, pixel by pixel, across time. This leads to over-simplification and loss of temporal information; thus, resulting in suboptimal segmentation [5, 18]. The main challenges of the problem consist of video signal limitations such as noise, low contrast, fuzzy edges, uneven fluorescence emission, density and location variability, heterogeneous morphology, slight movement of the observed tissue and cells overlapping [3, 18, 8].

A major drawback of deep learning approaches is the lack of properly labeled datasets. There is not a standard protocol to identify and label functional neurons in this kind of videos. Even more, there is not an accepted benchmark. Probably, the most popular dataset is the Neurofinder Challenge [19, 5, 15], which consists of 19 calcium imaging videos with ground truth annotations for training and 9 without annotations for testing. However, many issues have been reported on this data, some of them dealing with the different conditions of the recorded data, different labeling techniques and several wrong annotated neurons [5]. Moreover, the videos are highly heterogeneous in appearance, present very variable statistical pixel values and have strongly unbalanced labels [15]. For these reasons Neurofinder Challenge has turned out unsuitable as a benchmark and the problem of lack of datasets remains unsolved.

3. METHOD

Our method consists of three main stages: 1) preprocessing, 2) segmentation and 3) post processing (Figure 1). Preprocessing stage's main function is to improve image quality. Frame by frame, local equalization and Gaussian blur are applied in order to get better contrast and suppress noise. Additionally, the Lucas-Kanade algorithm [16] is used to calculate optical flow between consecutive frames and thus compensate slight movements. Segmentation stage identifies and segments all functional cells observed in a video with two neural networks. The first one is a CNN U-Net [20] with an Atrous Spatial Pyramid Pooling (ASPP) [17] as bottleneck between encoder and decoder. The second network is a RNN Gated Recurrent Unit that performs convolution operations (CGRU) [11]. The U-Net is used to detect blobs in each frame, aiming at identifying all possible cells whether they are functional or not, then CGRU is fed with every segmented frame to preserve only those cells with fluorescence changes. A classifier is added aiming at building the output mask from the CGRU final state. Finally, post processing refines segmentation, removing structures too small to be considered regions of interest, applying mathematical morphology to close regions and validating that signal

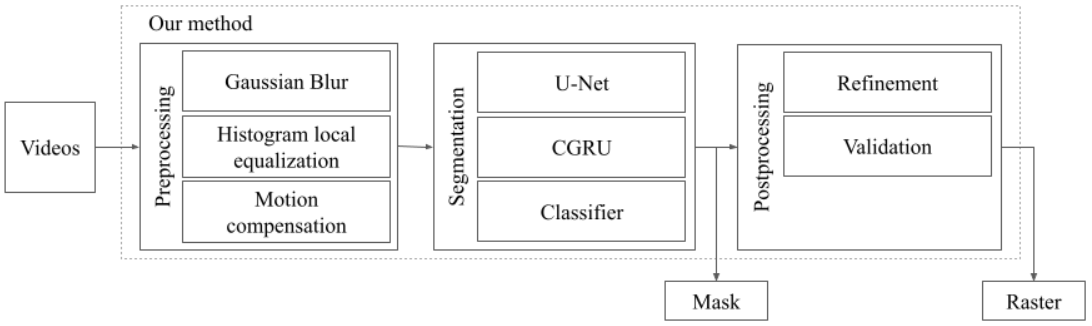

Figure 1. Global architecture of our proposed method is divided into three stages. The input consists of a video from an experiment and the output is the segmentation Mask and the activity matrix (Raster).

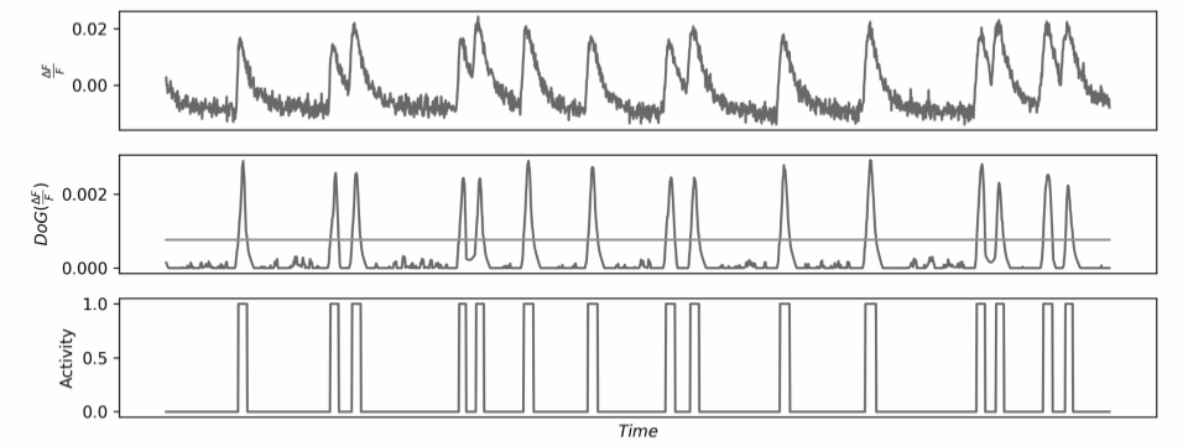

Figure 2. Example of activity extraction performance by Validation module. Upper plot shows the original calcium transient from a neuron. Medium plot displays the result of applying DoG to the previous, the horizontal line is the threshold to detect active neurons. Bottom plot shows the instances of time in which neurons show activity.

fluorescence (also known as calcium transients) of each cell corresponds to functional neurons. This is achieved using Differences of Gaussians in 1D data. Thresholding the analyzed signal retrieves a binary sequence of data that indicates instances in which cells are active (Figure 2). Activity sequences are arranged in a matrix named Raster where rows correspond to identified neurons and columns to time frames.

The combined use of CNNs and RNNs for spatio-temporal analysis presented in this work is a novel approach to the problem of functional neuron segmentation in epifluorescence microscopy imaging. The CNN selected is a U-Net [20], which has been very popular for image segmentation, trained end-to-end for pixel-wise prediction [21]. U-Net has an encoder-decoder architecture, whereas the encoder learns deep features, the decoder builds the segmentation based on the deep features learned. For this reason, the encoder is also named the analysis part and the decoder the synthesis part. Thanks to skip connections it is possible to propagate dense feature maps from the analysis layers to the synthesis layers, thus there is no information loss, and all deep features can be considered in the synthesis leading to more accurate outcomes.

Traditionally RNNs are meant to model relationships among sequential data [22], just as calcium imaging videos. However, long-term relationships are problematic to RNNs due to vanishing and exploding gradient problems. Fortunately, gated models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have been proposed to make up for gradient problems [11, 22]. The gated model's success relies on gating networks signals that use the current input and previous memory state to produce the next state. As gates have their own weights and parameters, they can be set through gradient descent in the learning phase [22]. Whereas LSTM model implements three different

gates, GRU does only two, however it has been proved that GRU presents a very similar performance with less parameters in comparison to LSTM [23]. Gated RNNs were designed to process 1D vectors, not images or tensors of higher dimensions. Vectorizing an image leads to two problems: 1) the resulting vector is very large, i.e., approximate length is about *width* * *height* * *channels* where width and height are given in pixels, 2) spatial relationships among pixels are ignored [23]. A feasible solution is a convolutional layer at the gates replacing product dot for convolutional operator. Thus, the objective is for filters to learn to convolve with the input image instead of learning weights for each pixel [23].

The problem we are tackling is binary segmentation, therefore a classifier is added in order to build the output mask from the CGRU's final hidden state and produces a segmentation map with the probabilities of each pixel belonging to a functional neuron. The classifier consists of a convolutional layer followed by batch normalization and ends with a sigmoid activation function.

4. EXPERIMENTS

The data gathered consists of 30 videos grouped in 7 experiments; each of them has been analyzed and labeled by an expert neuroscientist. All data was recorded at the Institute of Cellular Physiology of the National Autonomous University of Mexico. A confocal microscope 20X was used to observe spontaneous changes of fluorescence. The preparation is stimulated with light pulses of 48 nm and lasting of 15 to 50 ms plugged to the microscope with an optical fiber. The observed area has a size of 750 x 750 μm. In order to record and save observations a digital camera with cooling system is used. Frames are sampled at a rate of 100 to 250 ms/frame at intervals of 5 to 10 minutes. Videos are grayscale and have 3600 frames of 512 x 512 pixels.

As we aim at validating our hypothesis of the usefulness of combining CNNs with RNNs to tackle this problem, the following experiments aim at evaluating the advantages brought by this combination. First, using a previous work (SRNAME) [24] we generated a dataset formed by 1500 random frames from different experiments, these frames underwent our preprocessing module. Every frame was paired to a ground truth mask generated through SRANME's Blob Detection Module which is based on Difference of Gaussians. This first dataset was used to train a U-Net model (Figure 3). Then, all videos were processed frame by frame using the trained U-Net and the generated segmentations were used to train a CGRU model. In this case the ground truth was the output mask from the SRNAME system for each video. Since videos were large, up to 3600 frames, every video was split in k chunks without overlapping frames, then sequentially the CGRU model was trained with every chunk, performed backpropagation using video's ground truth and updated weights. The next chunk's initial hidden state is the last hidden state from the previous chunk; therefore, we could train the CGRU considering full videos, but gradients are reset between each chunk (Figure 4).

Our proposal was implemented in python, neural networks were implemented from scratch using Pytorch, training was executed in a Titan RTX GPU. The U-Net model was trained for 100 epochs, with a learning rate of 0.0001, the CGRU model was trained also for 100 epochs and a learning rate of 0.001. For both models, Adam was used as optimizer and binary cross entropy as loss function. Also, an early stop strategy was implemented to avoid overfitting. CGRU uses only one cell and the hidden state size equals 8. The U-Net was trained with 80% of the first data set samples and 10% for validation and testing. Samples in the testing set correspond to videos and experiments never seen in training nor validation sets. In the case of CGRU, 80% of all videos of five experiments were used for training and the remaining 20% for validation.

5. RESULTS

All models were evaluated using four metrics: 1) Sørensen-Dice coefficient (SDC), 2) Intersection over Union (IoU), 3) Precision and 4) Recall. Table 1 shows the results for the U-Net model and CGRU model with the testing set. Figure 5 shows some examples of the U-Net segmentation. Table 2 depicts the performance of our method analyzing two testing experiments, each of them consisting of three videos. For each video a mask was generated, afterwards all masks were combined taking the maximum value for every pixel.

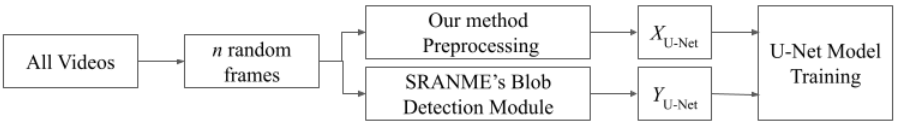

Figure 3. Training process for the U-Net model. $X_{U\text{-}NET}$ corresponds to the model input and $Y_{U\text{-}Net}$ ~~does~~ to the ground truth mask.

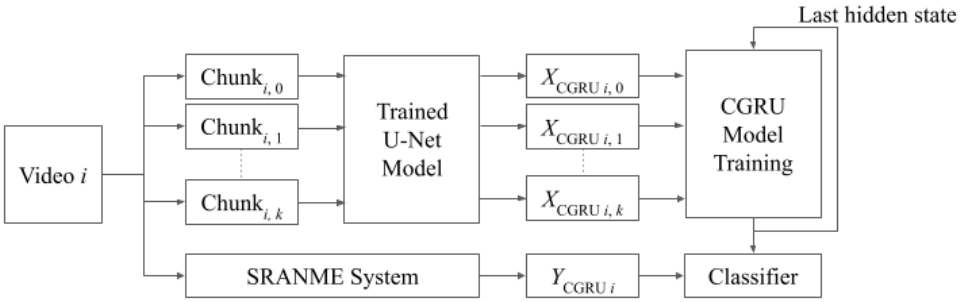

Figure 4. Training process for the CGRU model. Every video is 1) split into chunks and 2) processed with SRANME System to generate the ground truth mask $Y_{CGRU\ i}$. Each chunk is processed with the trained U-Net model to form $X_{CGRU\ i,j}$ input to feed the CGRU model. For each chunk, CGRU's weights are updated, and its final hidden state was used as the initial hidden state for the next chunk.

The U-Net model performs similarly to the SRANME's Blob Detector module, this generates a mask for each frame in a video, next they are combined to build a single mask [24] that corresponds to the common practice of synthesizing a single image discarding temporal information. In order to evaluate the advantages of using a RNN to analyze temporal dynamics, all segmentation masks generated by a U-Net are accumulated in two single images through maximum value and average value by pixel, respectively. Both synthesized masks are compared to masks generated by our method in Table 2. Figure 6 shows generated masks.

Table 1. Metrics at testing set by trained model.

Metrics at testing set				
Model	SDC	IoU	Precision	Recall
U-Net	0.814	0.686	0.817	0.811
CGRU	0.730	0.574	0.721	0.744

Table 2. Metrics by testing experiment and method.

Metrics for first experiment				
Method	SDC	IoU	Precision	Recall
Ours	**0.7585**	**0.6109**	**0.7452**	0.7722
Max.	0.4392	0.2813	0.2851	**0.9550**
Avg.	0.3631	0.2218	0.7128	0.2436
Metrics for second experiment				
Method	SDC	IoU	Precision	Recall
Ours	**0.7516**	**0.6021**	0.7061	0.8034
Max.	0.4529	0.2927	0.2987	**0.9356**
Avg.	0.4202	0.2660	**0.7456**	0.2925

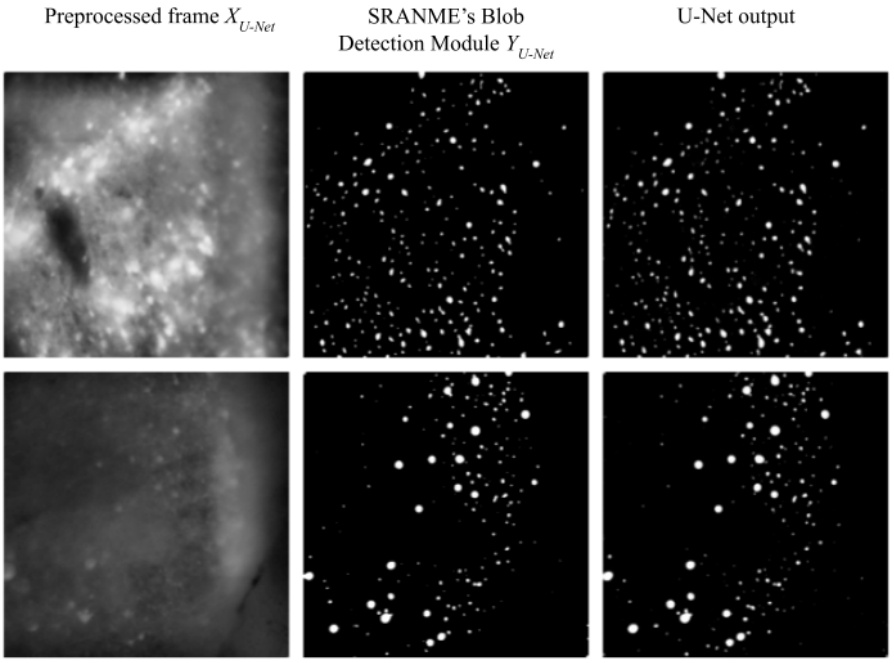

Figure 5. Two preprocessed frames at left column, their corresponding masks generated with SRANME's Blob Detection Module at central column and masks generated by the trained U-Net model at right.

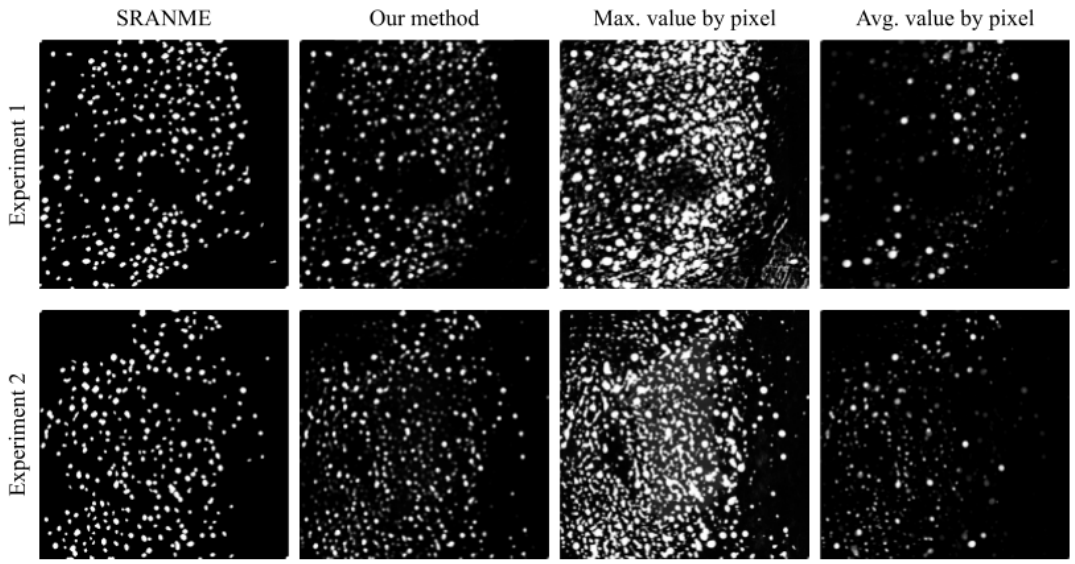

Figure 6. Comparison of masks obtained with different methods for two different experiments.

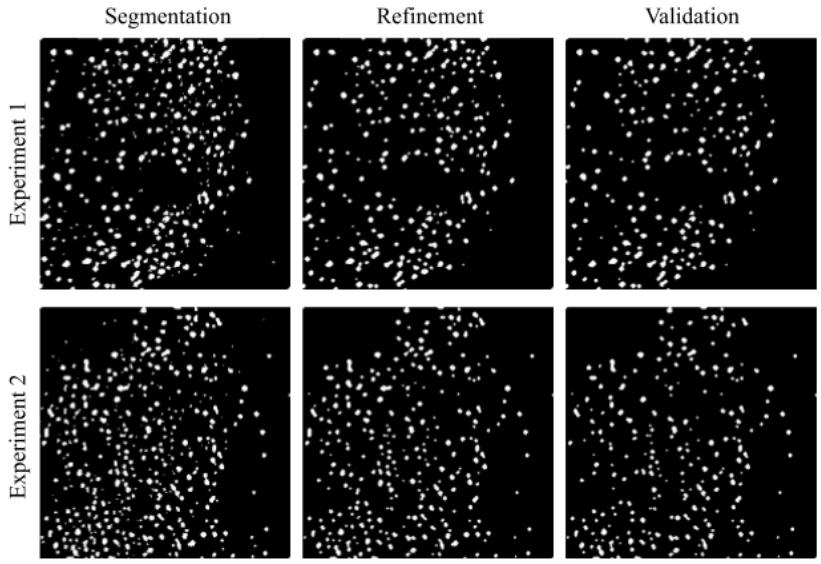

Figure 7. Outcomes from Refinement and Validation stages.

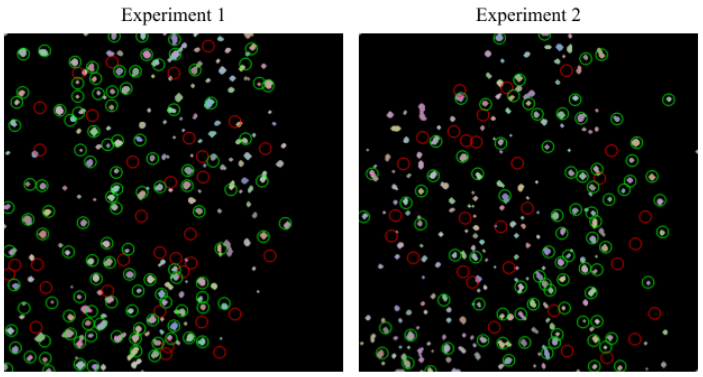

Figure 8. Comparison of our method's segmentation with manual labels. Correctly identified labels appear in green circle while missing cells are shown in red.

It is worth noticing that Max. Value by pixel mask exhibits over-segmentation, visually as well as objectively. A high recall value means a high number of positive pixels were correctly identified, but a low precision figure means that a small fraction of identified pixels are true positives. On the other hand, Average Value by Pixel exhibits sub-segmentation, showing low recall. Our method shows masks with high precision as well as recall values, so that an important part of positive pixels are identified, as SDC points out.

In Figure 7 refinement and validation outcomes are shown for both testing experiments and the comparative with manual labels are displayed in Figure 8. Recall rates of functional cells identified by the researcher and identified by our method are 0.7702 and 0.7413 for both experiments. Many functional cells have been identified and researchers recognize that since this task is to human errors of this task, it is feasible they have missed many of those cells or confuse others with functional ones. Thus, a deeper performance analysis is required.

6. CONCLUSIONS

Currently, neuroscientists count on very sophisticated imaging instrumentation to make observations of living brain tissue. Observations are stored in videos that are tough to analyze because it demands a lot of time and effort and it is a task prone to human error, therefore, digital applications are required for this task. However, developing such applications is a challenging task due to image quality limitations and issues related to expert abilities. Then emulating

and even outperforming researchers' abilities by means of computer vision is essential for this problem. Overcoming challenges related to image limitations such as noise or low contrast can be tackled with traditional image processing techniques like local equalization and filtering. Tackling the problem of identifying structures in the observed space could also be achieved with digital image processing as SRANME does with Differences of Gaussians. However, identifying patterns along time adds complexity to the task. Like the way humans do, it is necessary to compare different frames in order to identify some changes in pixel intensity and for that memory is required. Under that premise it has been proposed in this work to use two kinds of artificial neural networks, a convolutional one (U-Net) to identify spatial patterns and a recurrent one (CGRU) to identify temporal patterns.

This work also faced the shortage of reliable datasets. therefore, a previous proposal was used to build datasets to train and test our models. This allowed to validate our premise about the need to include a processing stage for temporal dynamics. The strategy of using a CGRU model together with a U-Net model leads to better segmentations than common strategies of synthesizing a single image using maximum or average value by pixel, which discards temporal information leading to over-segmentation and sub-segmentation respectively. It is worthy to stress the need of a final stage for refinement and validation; some constraints can be given to the neuronal networks as neuron size or closeness. This could be achieved in the future with a third network, or with other techniques such as mathematical morphology.

For future work, we propose to collect additional videos that allows to retrain our models with more diverse data. Moreover, training and testing should be done with manually validated labels and refined ground truth masks in order to outperform SRANME. Additionally, the Lucas-Kanade algorithm, used to compensate motion, is very sensitive to noise, which is a common limitation in microscopy imaging, so improved optical flow estimation should be considered.

ACKNOWLEDGEMENTS

Authors acknowledge the graduate scholarship from CONACyT and thank UNAM PAPIIT grants TA101121 and IV100420.

REFERENCES

[1] Shepherd, G. M., "The Synaptic Organization of the Brain," 5th Edition. New York: Oxford University Press, (2004).

[2] Pérez, J., Duhne, M., Lara, E., Plata, V., Gasca, D., Galarraga, E., Hernandez, A. and Bargas, J., "Pathophysiological signatures of functional connectomics in parkinsonian and dyskinetic striatal microcircuit," Neurobiology of Disease, 91, 347-61 (2016).

[3] Baglietto, S., Kepiro, I. E., Hilgen, G., Sernagor, E., Murino, V. and Sona, D., "Automatic Segmentation of Neurons from Fluorescent Microscopy Imaging," BIOESTEC, CCIS 881, 121-133 (2018).

[4] Dayan, P. and Abbott, L. F., "Theorical Neuroscience," 1st Edition. London: The MIT Press (2005).

[5] Kirschbaum, E., Bailoni, A. and Hamprecht, A. H., "DISCo: Deep Learning, Instance Segmentation, and Correlations for Cell Segmentation in Calcium Imaging," ICCAI 2020, LNCS 12265, 151–162 (2020).

[6] Badea, T., Goldberg, J., Mao, B. and Yuste, R., "Calcium Imaging of Epileptiform Events with Single-Cell Resolution", Journal of Neurobiology 48, 215-227, (2001).

[7] Cossart, R., Ikegaya, Y. and Yuste, R., "Calcium imaging of cortical networks dynamics". Cell Calcium 37, 451–457, (2005).

[8] Ferran, D., Reichinnek, S., Both, M. and Hamprecht, F. A., "Automated identification of neuronal activity from calcium imaging by sparse dictionary learning," IEEE 10th International Symposium on Biomedical Imaging (2013).

[9] Serrano-Reyes, M., García-Vilchis, B., Reyes-Chapero, R., Cáceres-Chávez, V. A., Tapia, D., Galarraga, E. and Bargas, J., "Spontaneous Activity of Neuronal Ensembles in Mouse Motor Cortex: Changes after GABAergic Blockade", Neuroscience, (446): 304-322 (2020).

[10] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B. and Sánchez, C. I., "A survey on deep learning in medical image analysis," Medical image analysis (MedIA), vol. 42, 60–88 (2017).
[11] Siam, M., Valipour, S., Jagersand, M. and Ray, N., "Convolutional Gated Recurrent Networks for Video Segmentation," arXiv.org (2016).
[12] Lee, Y., Xie, J., Lee, E., Sudarsanan, S., Lin, D-T., Chen, R. and Bhattacharyya, S., "Real-Time Neuron Detection and Neural Signal Extraction Platform for Miniature Calcium Imaging," Frontiers in Computational Neuroscience Vol 14 (2020).
[13] Apthorpe, N., Riordan, A., Aguilar, R., Homann, J., Gu, Y., Tank, D. and Seung, H., "Automatic Neuron Detection in Calcium Imaging Data Using Convolutional Networks," 30th Conference on Neural Information Processing Systems (2016).
[14] Giovannucci, A., Friedrich, J., Gunn, P., Kalfon, J., Brown, B. L., Koay, S. A., Taxidis, J., Najafi, F., Gauthier, J. L., Zhou, P., Khakh, B. S., Tank, D. W., Chklovskii, D. B. and Pnevmatikakis, E. A., "CaImAn an open source tool for scalable calcium imaging data analysis," eLife Neuroscience (2019).
[15] Klibisz, A., Rose, D., Eicholtz, M., Blundon, J. and Zakharenko, S. "Fast, Simple Calcium Imaging Segmentation with Fully Convolutional Networks," LMIA/ML-CDS 2017, LNCS 10553, 285–293 (2017).
[16] Lucas, B. and Kanade, T., "An iterative image registration technique with an application to stereo vision," In Proceedings of the International Joint Conference on Artificial Intelligence, 674–679 (1981).
[17] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," arXiv.org (2017).
[18] Soltanian-Zadeh, S., Sahingut, K., Blau, S., Gong, Y. and Farsiu, S., "Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning," PNAS Vol. 116, 8554-8563 (2019).
[19] (February 2022). *Neurofinder*. Retrieved from http://neurofinder.codeneuro.org/.
[20] Ronneberger, O., Fischer, P. and Brox, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," arVix.org (2015).
[21] Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P. and Johansen, H. D., "DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation," arXiv.org (2020).
[22] Day, R. and Salem F. M., "Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks," IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), 1597-1600 (2017).
[23] Chung, J., Gulcehre, C., Cho, K. and Bengio, Y., "Empirical evaluation of gated recurrent neural networks on sequence modeling," arXiv.org (2014).
[24] González, F., "Sistema de reconocimiento automático de neuronas en microscopía de epifluorescencia," Thesis, Universidad Nacional Autónoma de México, Mexico City (2019).