Detection of overlapping cells in microscopic scans of GFP brain tissues (TPA 11)

Saha Arrman Anicket, *EE16B148* and Subhankar Chakraborty, *EE17B031*

Abstract

There is a need in modern neuroscience for accurate and automated image processing techniques for analyzing the large volume of neuroanatomical imaging data. Even at light microscopic levels, imaging mouse brains produces individual data volumes in the TerraByte range. A fundamental task involves the detection and quantification of objects of a given type, e.g. neuronal nuclei or somata, in brain scan dataset. Traditionally this quantification has been performed by human visual inspection with high accuracy, that is not scalable. When modern automated CNN and SVM based methods are used to solve this classification problem, they achieve accuracy levels that range between 85 to 92%. However, higher rates of precision and recall that are close to that of human performance are necessary. In this report, we implement the stage-I of an unsupervised, iterative algorithm from [1], which provides a high performance for a specific problem of detecting Green Fluorescent Protein labeled nuclei in 2D scans of mouse brains. We also experiment an optimization based approach to improve the detections obtained from stage-I.

I. Introduction

UNDERSTANDING the cytoarchitecture of the brain is an important goal in the field of neuroscience. In particular, the identification of locations and quantification of neurons across the brain is a vital task. To achieve this goal, modern imaging methods have allowed users to produce images at very high resolution for further analysis. Entire mouse brains can now be imaged using light microscopy at sub-micron resolution from entire mouse brains (around $1cm^3$ volume), with individual data volumes in the terrabyte range. Optical microscopic techniques applied at such high resolutions, help to study the brain at a single neuron resolution but the large data volumes make it a significant computational and data analytics challenge. In addition to the advancements in imaging techniques, advanced methods such as the expression of the Green Fluorescent Protein (GFP) have ensured that specific types of neurons can be tagged and studied in great detail. Manually studying the distribution of these GFP expressing neurons across the entire brain is often time consuming and is susceptible to human errors over large datasets. To overcome this, an automated process of detection and quantification of neurons must be designed based on well grounded image processing (or computer vision) algorithms. However, the level of sophistication in the analysis techniques has not yet caught up with the significant improvements in microscopic imaging methods. We take inspiration from [1] and try different changes. This report is a summary of all those observations and results obtained in the process.

II. ALGORITHMIC DESCRIPTION

A. Brief description of the algorithm used

The proposed method in [1] consists of two sub-processes. Phase-I sub-process consists of steps, which enhance the foreground areas (FGR) followed by feature extraction, using the Distance Transform (DT) from edge map of FGR and ridge lines from the

modulated DT (mDT) map. Peaks of mDT map and features on ridge lines correspond to candidate centers of detected objects of interest. Phase-II sub-process fills the cell areas (detected centers with edges) detected in Phase-I with the background color on FGR to obtain a residual map (RFM). Significant areas of the RFM are then iteratively filled using Hough Transforms on convex arc segments that were identified on the outer boundary edges of the objects in the RFM. The output of the cell center locations provided by the two phases are eventually merged using a union operation (spatial locations of very nearby cell centers are averaged) to create the final set of detected cell centers. The images we were provided were 12 bit images with their green channels already preporcessed as mentioned in [1]. Most of our efforts were concentrated at replicating parts of stage-I using those preprocessed images and experimenting with that. We also extracted the RFM after 1 iteration using the centers from stage-I. The overall work can be summarized as

- Perform edge-detection on the binary image of the FGR (refer [1]) map.
- Distance Transform (DT) on edge map
- Compute mDT as 0.1*FGR + 0.9*DT
- Get the local maxima
- Perform Vessel detection using Frangi filter
- Get bifurcation points (BF)
- Fit ellipses to the image using the detected points as initial guesses
- Getting the RFM after 1 iteration

For a lot of our code, we used functions from the *skimage, OpenCV,* and *SciPy* libraries. Brief descriptions are provided for the sections we coded ourselves.

B. Detection of Bifurcation Points

The basic idea is that a bifurcation point splits the small region around it into more than two parts. We took the output from the Frangi filter and skeletonized it appropritately so that we have 1 pixel thick vessels. On that skeletonized image, we used a sliding window of size $k \times k$ and checked if the skeleton within the window divided that region into more than two components.

C. Fitting Ellipses to the Image using the detected points

In this section, we experiment with an optimization based method to improve the cell center detections from stage-I. We take a region or size $m \times m$ around the centers from stage-I and improve it by fitting an ellipse using differential evolution and the MSE criteria. The ellipses are of the form

$$\frac{(xcos\theta - ysin\theta)^2}{a^2} + \frac{(ycos\theta + xsin\theta)^2}{a^2(1-e^2)} = 1$$

where a is the length of the major axis of the ellipse, e is the eccentricity, and θ is the slant of the ellipse. The centers of the fit ellipse are considered as the improved centers. The algorithm is defined in Algorithm 1.

Algorithm 1: Fitting Ellipses

Input: Centers from stage-I for rectification, window size, Image

Output: Improved cell centers

```
n = Number_of_centers_from_stage-I
X,Y = size_of_Image
new_centers = []
for i in 1:n do
    center = Centers_from_stage-I_for_rectification[i]
    x, y = coordinates of center
    x_1 = max(x-window size, 0)
    y_1 = max(y-window size, 0)
    x_2 = min(x+window size, X)
    y_2 = min(y+window size, Y)
    image patch = Image[x_1:x_2, y_1:y_2]
    ellipse = best_fit_ellipse_on_image_patch_using_differential_evolution_and_MSE_criteria
    x,y = center_of_ellipse
    push (x,y) into new_centers
return new_centers
```

III. OUTPUT

Some of our outputs can be seen in Figure 1 and 2. Figure 1 contains comparisions of our outputs to those from the paper. Figure 2 is an illustration of our proposed ellipse fitting method on the Bifurcation Points from stage 1. The method has been run only on the Bifurcation Points and not the merger of mDT local maximas with Bifurcation Points to prevent clutter and see a clear comparison. Also, please note the fact that the RFM produced in Figure 1l has been generated taking centers directly as the local maximas of the mDT and Bifurcation Points to get a fair comparision with the RFM from [1].

A. Observation

We made the following observations during the experiments.

- The DT, and hence the mDT are extremely sensitive to the threshold changes. Hence, choosing an appropriate threshold is of paramount importance to get a good guess of maximas for stage-I.
- To get single pixel wide vessels after the Frangi filter, we tried various methods like skeletonize, medial axis and thinning. Skeletonize gave markedly the best results for us.
- As it can be seen in Figure 1, the DT, mDT and RFM generated by us are very close to those from the paper. The results of the Frangi filter need more fine-tuning.
- The proposed ellipse fitting method manages to better center some of the bifurcation points as it can be seen in Figure 2a and Figure 2b.

IV. Conclusion and Future Work

From the outputs we can see that we have quite successfully implemented stage-I. The RFM shows that we have something promising to start implementing stage-II on. The improvement of cell centers by fitting ellipses gave promising results on a few detections too. A future course of work would involve trying it on centers detected by stage-II once that is implemented and see if it can cause any improvements in that too. Another course of work might be to try other preprocesing steps to reduce the dependency on the iterative stage-II. We did try a few of those but the experiments were not very methodically done and hence have not been included in the report.

References

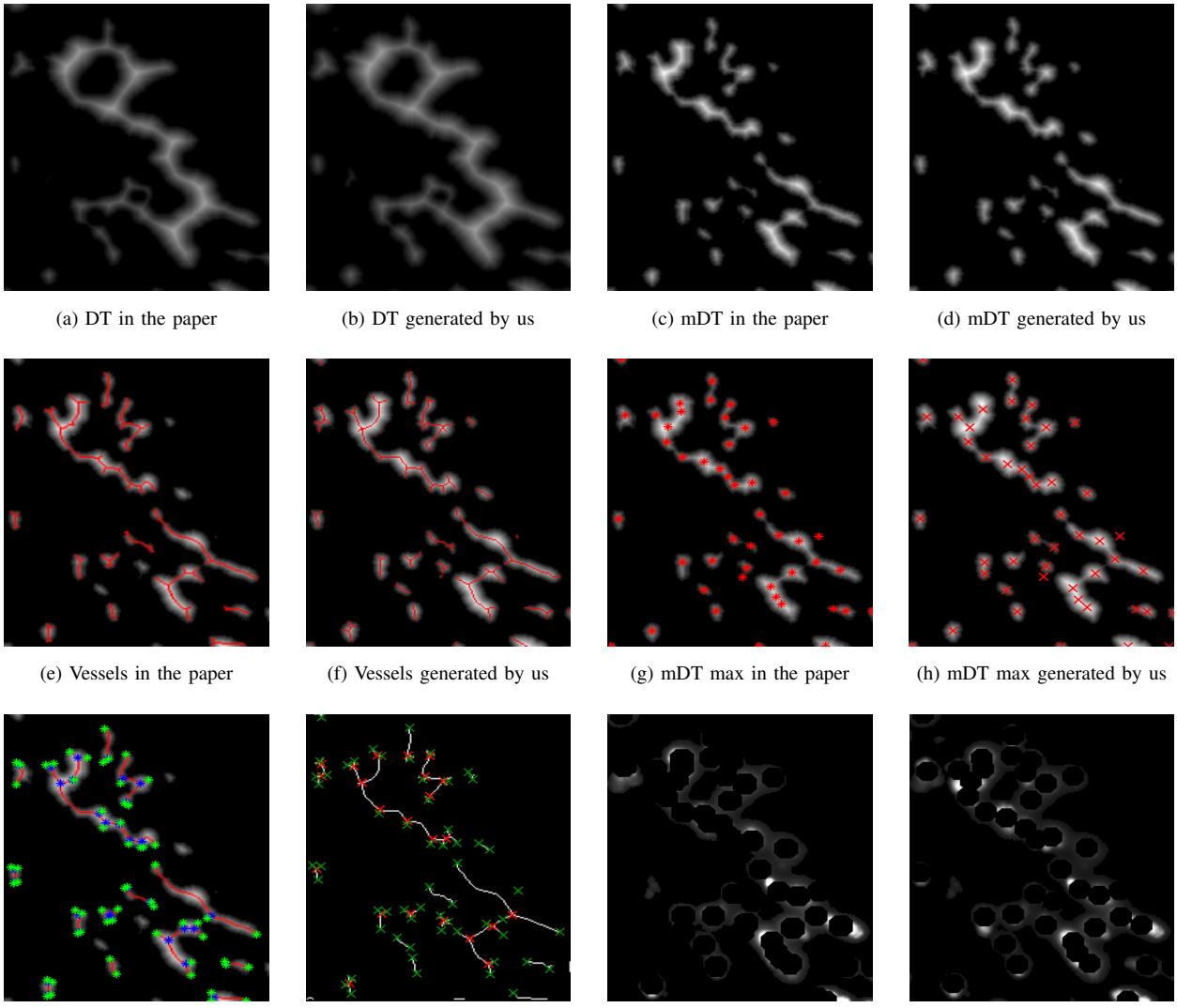
[1] S. Das, J. Jayakumar, S. Banerjee, J. Ramaswamy, V. Vangala, K. Ram, P. Mitra *High precision automated detection of labeled nuclei in Gigapixel resolution image data of Mouse Brain.* BioRxiv 2019 Jan 1:252247

[2] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever *Multiscale vessel enhancement filtering.* InInternational conference on medical image computing and computer-assisted intervention 1998 Oct 11 (pp. 130-137). Springer, Berlin, Heidelberg.

V. APPENDIX

The following is the range of different parameters used for the experiments.

Parameter	**Description**	**Value**
th	Threshold used for binarizing FGR	0.2
dc	Minimal distance for Merging Centers of Cells	6 pixels
k	Window size to detect Bifurcation Points	4 pixels
m	Window size for fitting ellipses	40 pixels
a	Length of major axis for fitting ellipses	6-12 pixels
e	Eccentricity of the ellipses	0-0.85
θ	Slant of the ellipses	$(\frac{-\pi}{2}, \frac{\pi}{2})$ radians

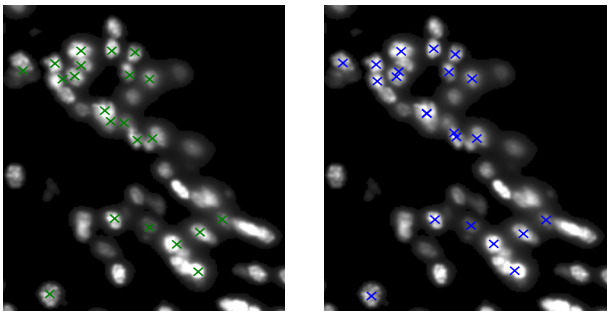


(a) DT in the paper (b) DT generated by us (c) mDT in the paper (d) mDT generated by us

(e) Vessels in the paper (f) Vessels generated by us (g) mDT max in the paper (h) mDT max generated by us

(i) Bf and RE in the paper (j) BF and RE generated by us (k) RFM from the paper (l) RFM generated by us

Fig. 1: Comparision of results generated by us to those in the paper



(a) BF points obtained directly (b) BF points after improvement

Fig. 2: Bifurcation points from stage-I before and after improvement using ellipses.