DEEP LEARNING FOR AUTOMATIC CELL DETECTION IN WIDE-FIELD MICROSCOPY ZEBRAFISH IMAGES

Bo Dong⋆† Ling Shao⊙ Marc Da Costa‡ Oliver Bandmann‡ Alejandro F Frangi⋆†

⋆ Centre of Computational Imaging & Simulation Technologies in Biomedicine (CISTIB)
† Department of Electronic & Electrical Engineering, The University of Sheffield
‡ Department of Neuroscience, The University of Sheffield
⊙ Department of Computer Science and Digital Technologies, Northumbria University

ABSTRACT

The zebrafish has become a popular experimental model organism for biomedical research. In this paper, a unique framework is proposed for automatically detecting Tyrosine Hydroxylase-containing (TH-labeled) cells in larval zebrafish brain z-stack images recorded through the wide-field microscope. In this framework, a supervised max-pooling Convolutional Neural Network (CNN) is trained to detect cell pixels in regions that are preselected by a Support Vector Machine (SVM) classifier. The results show that the proposed deep-learned method outperforms hand-crafted techniques and demonstrate its potential for automatic cell detection in wide-field microscopy z-stack zebrafish images.

1. INTRODUCTION

Zebrafish is an excellent system model for neurodegenerative diseases such as Parkinson's Disease (PD). It is easy to visualise the dopaminergic neurons in the zebrafish brain by using a process called whole-mount in situ hybridisation (WISH) with a probe against messenger ribonucleic acid (mRNA) for tyrosine hydroxylase (TH). TH is rate-limiting enzyme that synthesises L-dihydroxyphenylalanine (L-DOPA), the precursor of dopamine. Some of the genetic zebrafish models of PD show a reduction of around 25% of their dopaminergic neurons (assessed by WISH for TH), as early as 3 days post fertilisation [1]. Once WISH has been performed on the larvae, they are mounted onto microscope slides and a series of images are taken of each larva in slices (called z-stack images). When these z-stack images are combined, they can be rendered to give a three-dimensional (3D) image of the dopaminergic neurons in the larva’s head. Using this 3D image, it is possible to detect or count the individual neurons. At the moment, this is a manual, time consuming, subjective, and error-prone process. The expectation is that automatic algorithms would free neuroscientists from having to do this detection or counting process manually.

Detecting TH-labeled neuron in wide-field microscopy images is difficult. Different neurons have highly variable appearances, and the TH-labeled cells are often clustered together, which makes detecting individual cells harder. In addition, different imaging conditions, such as the exposure time, the light source intensity and the transparency of the specimen, can cause variations of the image intensity in each of the three RGB channels resulting in cells appearing with different colours in different specimens. Furthermore, when imaging z-stacks of the zebrafish larva using wide-field microscopy, there is slight out-of-focus light that appears in the recorded z-stack image. In this paper, deblurring or deconvolution [2] is not applied on the z-stack images on purpose to make the task more challenging and the proposed method applicable to realistic scenarios where the deconvolution parameters are difficult or impossible to obtain. Fig. 1 shows that the further away from the central frame, the blurrier the cell.

Deep learning is a set of methods in machine learning that attempts to represent high-level features in data with a multiple-layer architecture [3]. CNN, as one of the most effective deep learning techniques, was introduced in [7]. Then, it was further developed by other researchers [3]. Instead of adopting hand-crafted features, such as Histogram
Fig. 2: The work flow of the TH-labeled cell detection framework. The off-line training process and the detection process are shown in red and green dotted rectangles, respectively.

Fig. 3: Colour normalisation through IIS.

of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT), deep learning can learn high-level features from training data automatically. Furthermore, CNN is able to reduce learn-able parameters significantly by sharing the same basis function across different image locations [3]. Pooling methods are used for building translational invariant features. We use max-pooling in this paper.

In this paper, we propose an automatic deep learning based cell detection framework for 3D z-stack microscopy zebrafish images. In the framework, to improve the efficiency and accuracy of training a CNN from the large-size images, an SVM classifier is applied to detect cell regions for collecting the CNN training set. Major contributions of this paper include that we apply the deep learning technique, a supervised max-pooling CNN, to a new and significant area, and the results show potential capability for automatic cell detection in z-stack images. Moreover, we make accessible a wide field zebrafish microscopy database with TH-labeled neurons and manual annotations by experts on neuron localisation that can be used for evaluation and benchmarking purposes by other competing methods.

2. RELATED WORKS

Cell detection in microscopy images is a significant step for further cell based experiments. Several researchers focus on cell analysis in zebrafish images, for example, segmentation and characterisation of zebrafish retinal horizontal neurons [8], counting of zebrafish neuron somata [9], and high-throughput analysis for region detection and quantification of zebrafish embryo images [10]. There are also many cell detection algorithms published in computer vision and medical image analysis areas [11, 12]. However, most of the automated 3D cell detection methods are not a suitable replacement for manual cell detection [12].

In the past few decades, two kinds of cell detection algorithms have emerged. The first kind is segmentation or thresholding based approaches [13] and various software implementations emerged including some plugins for 'ImageJ' [11] and the 'FARSIGHT' toolkit [14]. The second class is model or feature based methods [15, 16]. With the development of the machine learning techniques, learning based cell detection has potential capability for automatic cell detection. Some learning based cell detection methods [17, 18] are proposed for two-dimensional (2D) immunohistochemistry images. However, there is no particular automatic TH-labeled cell detection method for light microscopy z-stack zebrafish images. In this paper, we propose an automatic TH-labeled cell detection framework with deep learning for wide-field microscopy z-stack zebrafish images.

3. DATASET, EXPERIMENTS AND RESULTS

3.1. Zebrafish Dataset

The dataset contains 35 stacks scanned under 20X magnification through a wide-field microscope. 25 of these stacks are selected for the off-line training and 10 stacks for testing. In total, there are approximately 1000 TH-labeled cells. The volume (in voxels) of each stack is \(1024 \times 1344 \times z\), and different stacks have different values of \(z\). The spatial resolution is \(3\, \mu m\) and the axial resolution is \(1.5\, \mu m\). The magnification of the objective is \(20 X\), and the numerical aperture (\(NA\)) is 0.7. We make this dataset publicly available for other researchers to develop and evaluate their TH-labeled cell detection or cell counting algorithms. For each stack in this dataset, a professional observer labelled all centres of TH-labeled cells as the ground-truth using 'Point Picker' [19], a plug-in of the open-source JAVA program 'ImageJ' [11].

3.2. Pre-processing: Colour Normalisation

The zebrafish larvae are recorded in several sessions spanning a number of days for completing the whole dataset. The expos-
sure time is not guaranteed to be the same for each session of recording through the light microscope, so the colour of each stack may be different. We apply Image Intensity Standardisation (IIS), which was first introduced in [20] for intensity normalisation of 2D grey scale images. More recently, Bogunović et al. [13] modified the IIS algorithm for normalising the intensity of the 3D grey scale Rotational Angiography. In this paper, we apply the original IIS as a colour normalisation method for 3D light microscopy images.

We calculate three histograms of the three channels of the whole RGB stack first. Then, the stack histogram of each channel is aligned to the corresponding reference based on the non-linear registration method described in [20]. An example of this process is shown in Fig. 3.

### 3.3. Cell Region Detection

We determine the cell regions \( R \) to discard the irrelevant background regions. Selecting background training patches is also important for training a CNN. Therefore, we detect cell regions efficiently and roughly using an SVM classifier, and then cell and background training patches are collected from \( R \) instead of the whole stack.

When recording the zebrafish larvae, all the TH-labeled cells are located in the centre of the image. We notice that all TH-labeled cells have distinctive colours from the background, and the largest part of the image area is covered by background pixels. To collect supportive training patches for training the CNN detector, we need to collect patches near the TH-labeled cells. All the stacks have a similar RGB histogram after IIS colour normalisation, so the colour histogram is the most useful and reliable feature to distinguish the cell and non-cell region. The binary SVM classifier based on RGB histogram features (SVM-RGB Histogram) is used as a rough and fast cell region detector in this paper. Using the SVM-RGB Histogram detector also guarantees that all TH-labeled cells are detected in cell regions. Another reason why we detect cell region first is that although CNN is a powerful tool, the training set need to be carefully selected.

Based on the size of the TH-labeled cell, we extract patches of \( (9\mu m) \) and an axial radius \( r_a = 2 \) pixels \((3\mu m)\) along \( z \) direction. All pixels in the cube are considered as cell-sample centres \( C_p \). All the remaining pixels in stacks belong to background-sample centres \( C_n \). We extract all cell patches from \( C_p \), in 20 training stacks, and extract the same number of background patches randomly from \( C_n \). In total, approximately 0.1 million cell patches and the same number of background patches are extracted to train the SVM-RGB Histogram detector.

The SVM detector is used to detect the cell region for collecting CNN training patches, which can remove most of the background pixels. This stage is more like a feature selection pre-process. In this case, the CNN could also be more accurate to detect cell samples in the cell region. Similarly, in the test stage, for accuracy, we also first apply the SVM detector to detect those regions. We will compare the proposed method with training the CNN without this cell region detection stage. For training the conventional CNN, the cell samples are the same, but the non-cell samples are randomly selected from the background.

### 3.4. Training Set Pre-processing

After the cell region \( R \) is detected by the SVM-RGB Histogram detector, we extract cell and background patches in region \( R \) from all test stacks for training CNN with the same size of patches and neighbourhood setting described in 3.3. Pixels in the cell region \( R \) have similar colours, so colour feature in the cell region is not reliable for distinguishing cell and background patches. To reduce the training time, we transfer all RGB patches into the YUV colour space, and only the Y-channel patches are needed. For each Y-channel cell patch, we rotate 0, 90, 180, 270 degrees to make the detector rotation invariant and increase the number of cell samples as well. The cell and background patches can have overlapping pixels, which is beneficial for increasing the probability of detecting true cells. About 0.5 million cell patches are extracted from all training stacks, and about the same number of background patches from the cell region \( R \).

### 3.5. Cell Pixel Detection

After the max-pooling CNN is trained, we test it on the test stacks. Firstly, the cell region \( R \) is detected by the SVM-RGB Histogram detector in every frame of each stack in the test dataset. Then, the pre-trained CNN is used for detecting cells by scanning every pixel in \( R \), and each pixel is given a probability value \( P_c \). It requires less than 300 seconds to process a slice. We will get a 3D probability map \( M_p \) for a stack, and a Gaussian filter is used for smoothing this probability map \( M_p \). Finally, all 3D local maxima can be found in smoothed \( M_p \) using method described in [22], which requires a self-turning threshold.

![Fig. 4: Cell region detection using SVM-RGB Histogram detector.](a) Original Frame  (b) Cell region \( R \) with size \( 41 \times 41 \) \((123\mu m \times 123\mu m)\). To increase the number of cell samples, for each labelled ground-truth central-pixel \( C(x, y, z) \), we draw a cube with a spatial radius \( r_s = 3 \) pixels.
3.6. CNN Training

We train the CNN network demonstrated in Fig. 5. Training this CNN network needs 3 days of computation using a MATLAB R2011b implementation with C++ library [23] in a personal computer with an i5-3470 CPU clocked at 3.4GHz, 14GB memory and 64-bit operating system. Less than 15 epochs are required to reach the minimum error (less than 10%) with a total of 1 million training patches.

3.7. Results

We evaluate the detection results according to Human Observer’s labelled cells on the test set. Detected cells that are closer than 10 pixels (30μm) in 2D slice, and 5 pixels (7.5μm) in vertical direction from a ground-truth centroid are defined as True Positives. Besides the number of True Positives (N_TP), we also count the number of False Positives (N_FP) and False Negatives (N_FN). Then, performance measures including recall (recall = N_TP/(N_TP + N_FN)), precision (precision = N_TP/(N_TP + N_FP)), and F1 (F1 = 2PR/(P + R)) score are calculated. We named our proposed method ‘Refined CNN’, and we compare it with CNN without the cell region detection stage (CNN), and the SVM method with different features, including RGB histogram features (RGBHist), RGB Colour & Scale-Invariant Feature Transform features (RGB&SIFT), and RGB&SIFT combined with Histogram of Oriented Gradients (RGB&SIFT+HOG). Table 1 shows the proposed method performs significantly better than the compared techniques. Additionally, it can be seen from the example result in Fig. 6 that the detection accuracy is good if cells are not clustered together and the proposed method need to be improved in the cell cluttering areas.

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>Precision</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refined CNN (Proposed)</td>
<td>0.7692</td>
<td>0.6452</td>
<td>0.7018</td>
</tr>
<tr>
<td>CNN</td>
<td>0.6071</td>
<td>0.3542</td>
<td>0.4474</td>
</tr>
<tr>
<td>RGBSift+HOG</td>
<td>0.5600</td>
<td>0.2593</td>
<td>0.3544</td>
</tr>
<tr>
<td>RGBHist</td>
<td>0.4839</td>
<td>0.1136</td>
<td>0.1840</td>
</tr>
<tr>
<td>RGBSift</td>
<td>0.3704</td>
<td>0.0962</td>
<td>0.1527</td>
</tr>
</tbody>
</table>

4. CONCLUSION AND FUTURE WORK

In this paper, an automatic deep learning based cell detection approach for light microscopy zebrafish z-stack images has been introduced. The results show that the proposed framework has potential for 3D cell detection and outperforms techniques based on hand-crafted features. Future work will aim at validating our approach on larger datasets with two observers’ label information and extending the 2D CNN model to a 3D RGB CNN model and making it more straight-forward for cell detection, with the ultimate goal of gradually bringing automated 3D cell detection into practice. Furthermore, other feature learning methods [24, 25, 26] will be investigated as well.

5. REFERENCES


field microscopy with sparse priors: Application to zebrafish imagery,” in ICPR 2014.


