

Bayesian Polytrees with Learned Deep Features for Multi-Class Cell Segmentation

R.Navin Kumar MCA., M.Phil.* , Sarath Kumar P **

* Assistant Professor, Department of MCA, Nandha Engineering College (Autonomous), Erode, Tamilnadu, India.
Email: navinsoccer07@gmail.com

** Final MCA, Department of MCA, Nandha Engineering College, Erode, Tamilnadu, India.
Email: sarathkumarpsk08@gmail.com

Abstract:

This project studies the application of machine learning in the analysis and diagnosis of Multi-class cell segmentation. The algorithm is evaluated on simulated data and on two publicly available fluorescence microscopy datasets, outperforming directed trees and three state-of-the-art convolutional neural networks, namely SegNet, DeepLab and PSPNet. And then, four machine learning algorithms including Bayesian Polytree, linear regression, support vector machine and logistic regression have been employed to the data sets. The performance comparisons of accuracy and recall rate among different algorithms indicate that the Bayesian Polytree algorithm has the optimal performance over the other two in both data sets. Moreover, the comparisons have been carried out in the cases with and without deviation standardization for each algorithm, and the results demonstrate that the deviation standardization has a certain effect on the accuracy improvement.

Keywords — Cell Segmentation, Microscopy, Bayesian polytree, SegNet, Machine Learning, Neural Networks.

I. INTRODUCTION

Accurate and efficient image segmentation of complex spatial object arrangements composed of multiple constituting structures (or classes) is challenging yet paramount for biological discoveries underpinned by quantitative imaging. For example, the identification of different cells within tissue or organelles within cells, The sub-cellular localization of proteins, the interactions of different cell types in organ development, or the immune response during infection, as just a few examples of relevant problems in biology. To assess the morphological and behavioral characteristics of these cells (some having unknown

causes quantitative metrics are devised, which require image segmentation as an unavoidable first step. Additionally, histology images are increasingly used for disease diagnosis and grading. Quantitative analysis of these images through the developed metrics (e.g. for abnormal nuclei as a potential indicator of cancer) helps pathologists by providing a supporting diagnosis and disease progress evaluation. Still, at a finer resolution, the biology of cell nucleus, i.e. the organization of the genome and the proteins, has a functional relevance with the biological cell processes, and their mis-localization (hence segmentation) can be a valuable indicator for many pathologies.

2. LITERATURE REVIEW.

With a massive influx of multimodality data, the role of data analytics in health informatics has grown rapidly in the last decade. This has also prompted increasing interests in the generation of analytical, data driven models based on machine learning in health informatics. Deep learning, a technique with its foundation in artificial neural networks, is emerging in recent years as a powerful tool for machine learning, promising to reshape the future of artificial intelligence. Rapid improvements in computational power, fast data storage, and parallelization have also contributed to the rapid uptake of the technology in addition to its predictive power and ability to generate automatically optimized high-level features and semantic interpretation from the input data. Deep learning has gained a central position in recent years in machine learning and pattern recognition. we have outlined how deep learning has enabled the development of more data-driven solutions in health informatics by allowing automatic generation of features that reduce the amount of human intervention in this process.

3. MACHINE LEARNING

Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a subset of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory

data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

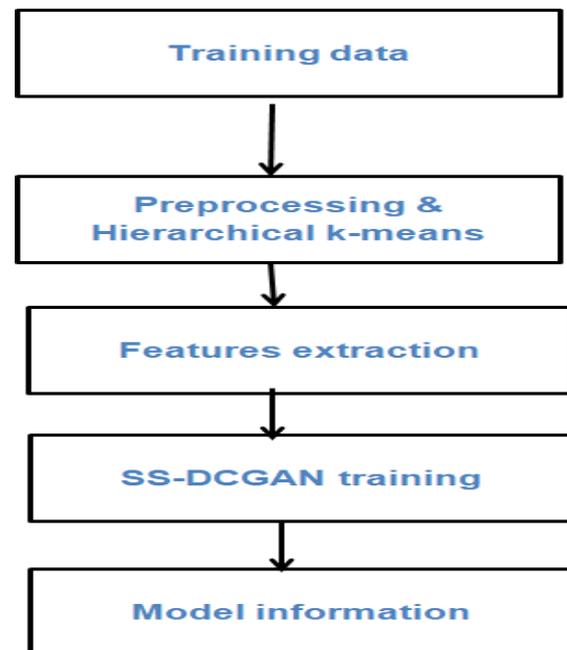
4. MODULES

A. PREPROCESSING

Image pre-processing is the name for operations on cell images at the lowest level of abstraction whose aim is an improvement of the Cell data that suppress undesired distortions or enhances some image features important for further processing. ... Its methods use the considerable redundancy in cell images Here we implemented HOG based pre-processing technique

B. FEATURE EXTRACTION

Feature extraction involves reducing the number of resources required to describe a large set of data. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the cell data with sufficient accuracy. Here we implemented K-means Cluster technique



C.SEGMENTATION

In digital image processing and computer vision, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images.

D.CLASSIFICATION

SVM classification is a process to analyze the number of data sets and extracts the meaning of data. Support vector machine provides methods and techniques for transformation of the data into useful information for decision making. **With accuracy of 98%.**These techniques can make process fast and take less time to predict the Glaucoma with more accuracy. The healthcare sector assembles enormous quantity of healthcare data which cannot be mined to uncover hidden information for effectual decision making. It becomes more influential in case of Glaucoma that is considered as the predominant reason behind death all over the world. In medical field, Data Mining provides various techniques and has been widely used in clinical decision support systems that are useful for predicting and diagnosis of various diseases.

4. EXISTING SYSTEM

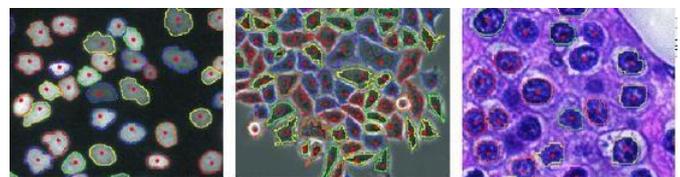
As a consequence, it is nearly impossible to obtain a large number of samples to satisfy the data quantity requirement of deep learning. Taking this reason into account, researches and clinical applications based on deep learning methods have been hindered. On the contrary, the traditional machine learning algorithms can obtain high accurate results even only based on small-scale data sets through manually selecting appropriate features. Consequently, traditional machine learning algorithms are widely adopted in the field of the traditional Chinese medicine (TCM) diagnosis and treatment researches. These studies have played a

significant role in the exploration of medical classification rules. Second, AI-related researches on facial and head cell data have been rarely reported, especially on the Microscopically cell datas.

5. PROPOSED SYSTEM

Considering a large number of cell examination items, only parts of these items are included. Specifically, two types of cell examination data are involved. F-MNCS and ABR are selected on account of the lower data dimension but a larger amount. The traditional machine learning algorithm using cell data based on small-scale data set has been adopted to carry out the related researches on the clinical application. Meanwhile, two data sets are established after data cleaning. Furthermore, detailed comparisons and discussions are conducted on processed results of four algorithms, including the effect comparison in the cases with and without data standardization. The traditional machine learning algorithm is better than the DL method in the cell data. The dimension would have a great impact on the model with small-scale data set. It can be predicted that with the increase of the data set size, the model accuracy without standardization would be promoted. It also provide a significant reference for the diagnosis based on clinical data and the improvement of medical efficiency.

6. RESULTS



The bag of local Bayesian classifier approach is implemented in a scale space. We compute local histograms with several different window sizes (for example, based on the cell sizes

in our phase contrast image datasets, we choose $w = \{10, 20, 30\}$). To save computational cost, we use integral histogram technique [9] for histogram computation and Nystrom method for spectral clustering. It costs 50 seconds to classify 1.4M pixels on a common desktop workstation, and it can be faster by parallel computing due to the pixel-wise classification. show a microscopy image to be segmented and its corresponding cell posterior probability after ensembling 9 local Bayesian classifiers (three clusters on each of three window scales). After the pixel-level MAP classification, we group the connected cell pixels into cell candidate blobs and remove small blobs probably due to non-cell pixels. The classification achieves good cell segmentation results compared to manual-labeled ground truth, as shown in Fig.4(c). We quantitatively evaluate the segmentation on the object level. In the ground truth, a cell object is detected (true positive, TP) if most (e.g. 90%) of its component pixels are correctly classified, otherwise it is missed (false negative, FN). A cell candidate blob by classification is a false positive (FP) if most of its pixels do not match the ground truth. Therefore, we define precision as $P = |TP|/(|TP|+|FP|)$, recall as $R = |TP|/(|TP| + |FN|)$, and F-measure as the Harmonic mean of precision and recall. Our approach is validated on four types of cells of different appearances captured by different imaging modalities and device settings: (1) C2C12 muscle stem cells imaged by Zeiss Axiovert 135TV microscope at 5X magnification; (2) bovine aortic endothelial cells imaged by Olympus IX71 microscope at 10X; (3) bovine vascular endothelial cells imaged at 10X; (4) Central Nervous System (CNS) stem cells imaged by Zeiss Axiovert 135TV microscope at 40X. We use intensity feature on the first three phase contrast microscopy sequences and joint intensity-gradient features on the fourth DIC microscopy sequence. Table 1 shows the complete evaluation results with 92.5% average accuracy (F-measure). Fig.5 shows some samples of the

microscopy images and corresponding cell posterior probabilities. In the end, we perform parameter sensitivity analysis on K over one microscopy sequence. As shown in Fig.6, the segmentation result is poor with only one local Bayesian classifier. After $K \geq 2$, the results are comparable to each other (e.g. we choose $K = 2^5$ in our system).

7. CONCLUSION

We propose a bag of local Bayesian classifier approach for cell segmentation in microscopy imagery. Local Bayesian classifiers (experts) are learned from clustered training image patches. Any new pixel to be classified is assigned a posterior probability about how likely it is a cell or background pixel based on the mixture-of-experts model. The binary segmentation results are obtained by MAP classification. We evaluate our approach quantitatively on four different types of microscopy images with 92.5% average accuracy.

REFERENCES.

- [1] C. Fowlkes, et al., "Spectral Grouping Using the Nystrom Method," IEEE Transaction on Pattern Analysis and Machine Intelligence, 26(2): p214-225, 2004.
- [2] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," International Conference on Machine Learning (ICML), 1996.
- [3] D. House et al., "Tracking of Cell Populations to Understand their Spatio-Temporal Behavior in Response to Physical stimuli," IEEE CVPR Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), 2009.
- [4] K. Li, M. Chen and T. Kanade, "Cell Population Tracking and Lineage Construction with Spatiotemporal Context," Intl. Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2007.
- [5] K. Li and T. Kanade, "Nonnegative Mixed-Norm Preconditioning for Microscopy Image Segmentation," International Conference on Information Processing in Medical Imaging (IPMI), 2009.

[6] N. Otsu, "A Threshold Selection Method from GrayLevel Histograms," IEEE Transactions on Systems, Man, and Cybernetics, 9(1): 62-66, 1979.