Graph-Based Pancreatic Islet Segmentation for Early Type 2 Diabetes Mellitus on Histopathological Tissue

Xenofon Floros^{1,4,5}, Thomas J. Fuchs^{1,4,5}, Markus P. Rechsteiner^{2,5},
Giatgen Spinas^{3,5}, Holger Moch^{2,5}, and Joachim M. Buhmann^{1,5}

¹ Department of Computer Science, ETH Zurich, Switzerland
{`xenofon.floros,thomas.fuchs`}@inf.ethz.ch

² Institute of Pathology, University Hospital Zurich, University Zurich

³ Division of Endocrinology and Diabetes, University Hospital Zurich

⁴ Life Science Zurich PhD Program on Systems Biology of Complex Diseases

⁵ Competence Centre for Systems Physiology and Metabolic Diseases, Zurich

Abstract. It is estimated that in 2010 more than 220 million people will be affected by type 2 diabetes mellitus (T2DM). Early evidence indicates that specific markers for alpha and beta cells in pancreatic islets of Langerhans can be used for early T2DM diagnosis. Currently, the analysis of such histological tissues is manually performed by trained pathologists using a light microscope. To objectify classification results and to reduce the processing time of histological tissues, an automated computational pathology framework for segmentation of pancreatic islets from histopathological fluorescence images is proposed. Due to high variability in the staining intensities for alpha and beta cells, classical medical imaging approaches fail in this scenario.

The main contribution of this paper consists of a novel graph-based segmentation approach based on cell nuclei detection with randomized tree ensembles. The algorithm is trained via a cross validation scheme on a ground truth set of islet images manually segmented by 4 expert pathologists. Test errors obtained from the cross validation procedure demonstrate that the graph-based computational pathology analysis proposed is performing competitively to the expert pathologists while outperforming a baseline morphological approach.

1 Introduction

The computational pathology framework presented in this work aims at automated segmentation of type 2 diabetes mellitus (T2DM) islets. T2DM is a chronically progressive disease which is characterized by hyperglycaemia, insulin resistance, and insulin deficiency. Taken together, these factors lead to organ failure and the increased risk for cardiovascular diseases [1]. It is estimated that by 2010 there will be more than 220 million patients suffering from T2DM [2]. Thus, the search for diagnostic and treatment of this disease is pushing forward at tremendous speed not only in academia but in pharmaceutical industry as well.

Currently, the diagnosis of T2DM includes the measurement of Hemoglobin (Hb) A1c (normal: 4-6%, metabolic syndrome: 6-7%, T2DM: > 7%) which reflects the blood sugar levels [3]. As hyperglycaemia is a marker for the progressed disease status most of the patients are diagnosed as T2DM when the disease has already manifested. Therefore, the search for early T2DM and pre-diabetic markers is of urgent need. From mouse and rat models it is known that the pancreatic islets of Langerhans, which consists of beta cells for insulin production, increase in size to compensate the additional demand for insulin to maintain normoglycaemic blood levels [4]. The same situation seems to be true in humans as summarized in [5], hence indicating that pancreatic islet features could be used as early T2DM markers.

Motivation: New targets for T2DM prediction, prevention, and treatment generated by the available in vitro and in vivo animal models need to be quantitatively analyzed by high-throughput screening and later verified in human tissue. Therefore a computational pathology approach is necessary to be adopted. The aim of an automated analysis pipeline is first the detection of cell nuclei and second the segmentation of human pancreatic islets based on specific staining for α and β -cells. The robust segmentation of islets is the basis for further quantification of biomarkers regarding T2DM in human patients. Having correctly isolated the islets, it is possible to extract features (e.g. area of the islet) and test their ability to differentiate early T2DM patients and control cases.

Automated analysis of fluorescence images of human tissue poses two main difficulties which are going to be addressed in this work. (i) The 3D structure of the tissue leads to the problem that cell nuclei are not always perfectly cut in their maximum dimension producing numerous cutting artifacts. It has to be noted that this does not happen in applications with cell cultures or on blood smears for which the large majority of image processing tools in this field is developed. This problem is addressed in the presented work by training a robust classifier for object detection in contrary to using morphological or watershed based approaches. (ii) Variations in the production process of the histological slices can lead to areas of different thickness within one section. This preprocessing artifact produces not only blurred regions in the image but also illumination variation which are even worsened by variations in the fluorescent staining process. These problems are tackled by first using illumination invariant features for the classifier, second by employing clustering for the α and β -cell classification and third by facilitating a graph-based approach for the islet detection.

Tissue Preparation and Imaging: Human pancreatic tissue from either autopsies or biopsies were formalin fixed and paraffin embedded. Sections were cut at a thickness of 2 μm and stored at 4° Celsius till use.

For immunofluorescence, sections were deparaffinized and stained with antibodies specific for α and β -cells. Furthermore DAPI staining (DAKO, Carpinteria, CA) was used to label the cell nuclei. Fluorescence pictures were taken with a resolution of $1376 \times 1032 \times 3$ pixels and 20x magnification. Raw unedited material was used in the analysis (Figure 1).

Fig. 1. Human pancreatic tissue: Primary tissue is taken from either whole pancreas from autopsies (I: normal pancreas, II: adenocarcinoma, III: spleen)(a) or biopsy resectates. Sections from fixed and embedded tissue blocks are stained with specific antibodies for β -cells (b), α -cells (c) and DAPI (d). Pictures b-d were taken with a magnification of 20x.

Problem Formulation: From a computational viewpoint the input to the pipeline consists of three fluorescence images: (i) DAPI-Channel $\rightarrow I_D$ (staining specific for cell nuclei detection), (ii) Alpha-Channel $\rightarrow I_\alpha$ (staining specific for α -cells) and (iii) Beta-Channel $\rightarrow I_\beta$ (staining specific for β -cells).

The output of the algorithm is the segmented area that the pancreatic islet of interest occupies. Prior information from expert pathologists is incorporated in order to guide the search for a meaningful extraction of the islet area. The domain knowledge can be summarized in two main hypotheses:

$\mathcal{H}_1$: *The islets are defined as an area with high density of α and β -cells, with the α -cells being more specific in specifying the islet area.*

$\mathcal{H}_2$: *There is only one islet of interest per image.* In most of the images additional structures are observed, such as smaller islets, disrupted islets or outliers due to staining failures. The main goal is to extract only the dominant islet in each image while excluding artefacts.

The distinct steps of the computational pathology pipeline are described in detail in Section 2.

2 Methods

(2.1) Cell Nuclei Detection: Cell nuclei on DAPI stained images, I_D , are detected by following the approach in [6] which showed excellent results on histopathological tissue with immunohistochemical staining. To generate a set of positive and negative training patches of size 65×65 , a domain expert labeled two images from different patients. In addition to the selected cell nuclei their rotated and flipped counterparts were added to the positive training set. The negative class was down sampled to have a balanced training set. In total 1214 positive and 1214 negative samples were used for training.

For each of these samples a feature vector of length 281 was generated consisting of local binary patterns (LBP) [7] and a histogram of gray scale values. A great advantage of LBPs for this application is that they are illumination

invariant, i.e. invariant with respect to monotonic gray scale-changes and therefore no gray-scale normalization or histogram equalization is needed.

Based on these features a random forests classifier [8] was learned to differentiate between cell nuclei and background. A random forest classifier consists of a collection of tree-structured classifiers $\{h(\mathbf{x}, \Theta_k), k = 1, \dots\}$ where $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input $\mathbf{x}$. Random forests possess a number of advantages over classical boosting approaches to object detection as described in [8] and [6]. One of them is the internal out of bag (OOB) error which provides an unbiased estimate of the generalization error. For this application it is shown (supplement Figure 1) that the classifier converges to an OOB error of about 3% after 25 trees.

Finally, to detect the nuclei we classified each pixel of the DAPI test images to generate an accumulator map with a probability at each pixel for being a cell nucleus or not. After non maxima suppression the detections within a range of 20 pixels were clustered to one final hit which is approximately the size of an average nucleus. The output of this step consists of a list of the coordinates of all detected cell-nuclei, $\mathbf{x}_i \in \mathbf{R}^2$, $i = 1, \dots, N$.

(2.2) Cell Nuclei Classification: The two channels accounting for the staining of α and β cells, I_α and I_β respectively, are segmented into background and staining using k-means clustering (with $k = 2$ classes) on the intensity histograms. In order to classify each detected nucleus from step (2.1), a neighborhood of 10×10 pixels at the nucleus center is considered. The nucleus is classified based on a majority voting scheme of the segmented binary pixels in the patch of each channel I_α, I_β . If there is strong evidence provided from the segmented staining of channel I_α (I_β) then the nucleus is classified as α -cell (β -cell), otherwise we characterize it as “normal” cell. Thus, tuples of coordinates plus labels for all detected cells of the previous step are obtained : $(\mathbf{x}_i, y_i)$, $y_i = \{\alpha, \beta, n\}$. This approach mimics the workflow of the pathologists by first detecting all cell nuclei and then classifying them to their respective classes based on the intensity of the class-specific staining around each nuclei.

(2.3) Graph Construction: Based on the main hypothesis $\mathcal{H}_1$, a neighborhood graph on the identified α and β -cells is constructed, in such a way that clusters of cells correspond to connected components of the graph. Regions of the image with high cell density will be represented by a unique connected component in the graph. This construction is motivated in [9], where theoretical aspects of clustering with nearest-neighbor (NN) graphs are explored. In general this task can be solved either by constructing a knn graph or an ϵ -neighborhood graph. Empirical results showed that for this specific task of islet detection, the latter graph performed better, mainly because $\epsilon \in \mathbf{R}^+$ allows for more flexible structures.

Hence, given the set V of α and β -cells detected in the previous steps the ϵ -neighborhood graph $G = G_{eps}(V, \epsilon)$ is constructed, such that two nodes $\mathbf{x}_i, \mathbf{x}_j$ are connected with an edge iff $\|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq \epsilon$. The euclidean distance is an intuitive choice for this problem setting, because it captures the local structure of cell proximities in the images.

(2.4) Islet Selection and Segmentation: Given the constructed graph G_{eps} , clusters of α and β -cells are identified by isolating the connected components of the graph. Under the hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$, the largest cluster corresponds to the islet of interest. Therefore the largest connected component G^{islet} of graph G is extracted, as a first crude approximation of the islet area. This first approximation depends on the parameter ϵ of the graph construction, and can be viewed as the computational equivalent of an expert focusing in the densely stained regions of the image trying to get a first impression of the islet location. Furthermore, it acts on the nuclei level and not on the staining intensities, thus resulting in higher robustness.

Based on the crude estimation of the islet boundaries, an active contour scheme is employed, in order to refine the detected islet area. The basic idea in active contours, [10], is to evolve a curve, subject to constraints based on the given image, in order to detect objects in the image. In the proposed pipeline, we apply the model described in [11], which does not use an edge-detector to stop the evolving curve in the boundary, hence does not depend on the gradient of the image. Furthermore it is shown to be quite effective under the presence of noise and does not require any preprocessing (e.g. smoothing) of the initial image [11]. As motivated above, we initialize the curve on the convex hull of G^{islet} and apply it on the superposition of the two stained channels $I_\alpha + I_\beta$ in order to refine the boundary of the islet. The active contour model used, is governed by two parameters, (s, r) , with $s \in \mathbf{R}^+$ controlling the smoothness of the active contour and $r \in \mathbf{N}$ the number of iterations. The proposed initialization of the active contour is beneficial in two ways: (i) Active contours schemes are known to be sensitive in the curve initialization. A meaningful initialization is provided, tailored to the specific problem and (ii) starting close to the islet boundary also reduces the computation time needed.

The algorithm outputs a binary mask, I_{islet}^{seg} (of the same size as the input channels), which corresponds to the detected area of the human islet. The whole pipeline is governed by a tuple of parameters $\theta = (\epsilon, s, r)$. Based on this segmentation it is possible to automatically extract all biologically meaningful features that can be used as predictive markers for early T2DM, e.g. islet area, staining intensities, fractions of α and β -cells in the islet. Furthermore, the automatically extracted segmentation results are compared with manually segmented islets from expert pathologists in order to assess the algorithm performance against an objective ground truth.

Baseline method: According to our knowledge, there are no published approaches to the specific problem of pancreatic islet segmentation on histopathological tissue. The absence of a competing method was partially compensated by the construction of a baseline morphological approach which also exploits the prior knowledge on the islet segmentation, captured by hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$. The steps of the baseline method can be summarized as follows: (i) smooth the input staining $I_1 = I_\alpha + I_\beta$ using a gaussian filter, (ii) globally threshold I_1 to obtain a coarse segmentation, thus $I_2 = I_1 \geq t$, (iii) remove small holes by calculating the closing of $I_2 \rightarrow I_3$, (iv) extract the biggest contiguous region

from I_3 and return this as the detected islet I_{islet}^{seg} . An alternative version has an extra step (v) where in addition an active contour is initialized with I_{islet}^{seg} as in step (2.5) of the proposed pipeline. Similarly to the proposed approach the main parameters form a tuple $\theta = (t, s, r)$.

Statistical Evaluation: Given the binary mask of the algorithmic segmentation (I_{islet}^{seg}) and the manually segmented islet from the expert pathologist (I_{islet}^{man}) the cell nuclei agreement between the two masks is calculated. For example, $TP = \#cells \in \{I_{islet}^{seg} \wedge I_{islet}^{man}\}$, $FP = \#cells \in \{I_{islet}^{seg} \wedge \neg I_{islet}^{man}\}$ etc. From the error counts we extract common evaluation metrics, such as Precision ($P = \frac{TP}{TP+FP}$), Recall ($R = \frac{TP}{TP+FN}$) and F-measure ($F = 2 \times \frac{P \times R}{P+R}$).

3 Results

The training set consists of 18 triplets of images (three stained channels) corresponding to two patients with T2DM and one control case. Four expert pathologists independently segmented the islet of interest for each of the images in the training set. For each one of the 18 training cases we calculated the consensus over the 4 experts, thus obtaining a consensus ground truth. This enables us to compare the performance of the algorithm against a “consensus” expert, but also to estimate the intra-pathologist labeling agreement.

Fig. 2. (a) F-measure box plots (from left to right: pathologists, proposed algorithm, baseline with active contours (AC) and baseline without AC). The proposed pipeline performs comparably to the pathologists in terms of F-measure when compared to the expert consensus. Furthermore, the test error variance is low. Both baseline methods fail to achieve consistent segmentations as they perform well in some instances, but fail to segment properly a large number of cases in each cross validation fold. (b) ROC curves for parameter ϵ of the graph construction, keeping s fixed and equal to 0.2 and setting r equal to 100 (dashed line) and 500 iterations (continuous line).

A cross validation scheme was employed to compare the algorithmic approaches to the gold standard, i.e. the consensus of the pathologists. More specifically, a 3-fold cross validation is used, where in each fold the algorithm is trained on 12 cases (choosing the parameter values that minimize the error) and then the generalization performance is tested on the other 6 that are left aside as a validation set. For the pathologists, each one of them is compared against the consensus segmentations. It has to be noted that since the pathologists' segmentations are considered as the gold standard, the computational pathology approaches cannot perform better than the experts. The results are depicted in Figure 2(a).

Regarding the individual experts' annotations we observe that they are very close to the consensus ground truth (with an average F-measure of 0.97) and exhibit quite low variance. Such a high performance is expected since by construction the ground truth labels are computed by averaging the individual ones. The proposed algorithm performs comparably to the pathologists (with an average F-measure of 0.92 across folds) keeping also the variance in a reasonable range. On the other hand both baseline methods (with and without the active contour module) are outperformed by the graph-based segmentation in terms of the F-measure. Furthermore, we observe that the baseline segmentations exhibit high variance, which indicates also a tendency to generalize poorly to new data.

In Figure 2(b) a specific instance of a ROC curve is plotted to evaluate the performance of the proposed algorithm with respect to parameter ϵ which controls the graph construction and thus the initial key step of islet segmentation. More specifically parameter s , which controls the smoothness of the boundary, is kept fixed and for two values of parameter r , the number of iterations the active contour is updated ($r = 100, 500$), the true positive rate (TPR) is plotted against the false positive rate (FPR) over a wide range of parameter ϵ . At a first glance a complex behavior is observed for the large number of iterations in the active contour evolution ($r = 500$). For increasing values of ϵ , vertical ascents are observed in the plot where FPR stays the same and TPR increases. Furthermore, for some sequential increases of ϵ the FPR increases while TPR decreases, a behavior which is not usually observed in ROC curves. Both phenomena can be explained if we keep in mind that parameter ϵ does not directly affect the final segmentation, since the active contour based boundary refinement is applied in between. Increasing ϵ adds more nodes to the graph, thus increasing the initialization area. However if the active contour algorithm performs an adequate number of iterations it will dominate and converge to the islet, hence filtering out the false positive cells. A more balanced behavior is observed for a smaller number of iterations, where for increasing values of ϵ , in most of the times, both TPR and FPR are increased.

4 Conclusion

The computational pathology system presented in this work is able to, objectively and automatically, estimate the boundaries of human pancreatic islets.

The whole pipeline is transparent, modular and based on explicit hypotheses describing the domain knowledge. To the best of our knowledge this is the first framework that successfully tackles this specific segmentation problem. Cross validation results indicate that the algorithm performs competitively to human experts. Having a reliable pipeline to detect and isolate pancreatic islets from human histological tissue, enables researchers to test specific hypotheses regarding T2DM. We are convinced that the proposed framework can be the basis for further research regarding T2DM and that it can significantly assist the search for diagnostic and therapeutic markers.

Acknowledgments. This work has been partially supported by the Competence Center for Systems Physiology and Metabolic Diseases (CC-SPMD) of ETH Zurich.

References

1. Kasuga, M.: Insulin resistance and pancreatic beta cell failure. *J. Clin. Invest.* 116(7), 1756–1760 (2006)
2. Zimmet, P., Alberti, K.G., Shaw, J.: Global and societal implications of the diabetes epidemic. *Nature* 414(6865), 782–787 (2001)
3. Vijan, S., Stevens, D.L., Herman, W.H., Funnell, M.M., Standiford, C.J.: Screening, prevention, counseling, and treatment for the complications of type ii diabetes mellitus. Putting evidence into practice. *J. Gen. Intern. Med.* 12(9), 567–580 (1997)
4. Maedler, K., Schumann, D.M., Schulthess, F., Oberholzer, J., Bosco, D., Berney, T., Donath, M.Y.: Aging correlates with decreased beta-cell proliferative capacity and enhanced sensitivity to apoptosis: a potential role for fas and pancreatic duodenal homeobox-1. *Diabetes* 55(9), 2455–2462 (2006)
5. Bonner-Weir, S., O'Brien, T.D.: Islets in type 2 diabetes: in honor of dr. robert c. turner. *Diabetes* 57(11), 2899–2904 (2008)
6. Fuchs, T.J., Wild, P.J., Moch, H., Buhmann, J.M.: Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part II*. LNCS, vol. 5242, pp. 1–8. Springer, Heidelberg (2008)
7. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J(G.) (eds.) *ECCV 2004*. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
8. Breiman, L.: Random forests. *Machine Learning* 45, 5–32 (2001)
9. Maier, M., Hein, M., von Luxburg, U.: Cluster identification in nearest-neighbor graphs. In: Hutter, M., Servedio, R.A., Takimoto, E. (eds.) *ALT 2007*. LNCS (LNAI), vol. 4754, pp. 196–210. Springer, Heidelberg (2007)
10. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
11. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Transactions on Image Processing* 10(2), 266–277 (2001)