Automatic Liver Tissue Section Image Characterization

Fei (Frank) Yang[1] and Ayooluwakunmi Jeje[2]

***Abstract*— Analyzing liver cross-section images is a manual and laborious task, slowing down critical research toward finding alternative cures for patients with end-stage liver disease. In this paper, we present methods to automatically count hepatocytes cells, count nuclei and classify liver vessel types, using input images of cell boundary and cell nuclei based on a dataset of 21 images. Compared to a trained researcher, the methods are able count cells, segment overlapping nuclei, and classify vessel types, including portal vein, central vein, and bile duct with reasonable precision and accuracy. Future work includes improving classification accuracy and detecting other cell types.**

***Index Terms*— microscopy image processing, cell counting, nuclei segmentation, vessel classification, SVM, Canny edge detection, Hough transform**

I. INTRODUCTION

Liver transplantation is currently the only cure for patients with end-stage liver disease, yet currently over 14,000 patients are on the liver transplant waitlist awaiting a donor [1]. Alternative treatment options are critically needed for these patients.

Physicians and scientists contemplating new treatment methods face the same major obstacle: understanding the molecular mechanisms by which the proliferation of an important type of liver cell called hepatocytes is regulated. The Nusse Lab of the Stanford Institute of Stem Cell Biology and Regenerative Medicine studies this class of specialized cells and recently published work in identifying a population of hepatocyte stem cells in the liver [2]. A major obstacle in scaling the research is the laborious task of manual cell counting from tissue section images.

This project leverages image processing and machine learning techniques to tackle the unique challenges presented by tissue section images. First, compared to tissue culture images, cells in a tissue section image is highly heterogenous with complex structures and ambiguous boundaries, making cell counting a more difficult task. Second, tissue section images may contain different types of cells, including hepatocytes, bile duct epithelial cells, endothelial cells, and immune cells. In this project, we focus exclusively on the hepatocytes cells which is the focus of the research. Third, in contrast to most other cells in our body, hepatocytes can have more than one nucleus. A final challenge is that in addition to cells, the liver section images also show important features of a liver, including central veins, portal veins, and bile ducts.

There are three distinct goals in the project: accurately count the number of hepatocyte cells irrespective to their stain, accurately segmented clustered nuclei, and accurately classify liver features, including central vein, portal vein and bile duct. This work focuses on developing algorithms to accomplish each goal and analyzing their performances. The algorithm is applied to a dataset provided by the Nusse Lab, consisting of 21 liver tissue section images, which are manually labeled by a trained researcher.

This paper is structured as the following: first we will present the implementation pipelines for cell boundary segmentation, cell nuclei segmentation, and liver vessel type classification. Next, the performance of each task using various techniques is analyzed and compared.

II. BACKGROUND

Microscopy has been the tool used by biomedical scientists to advance our fundamental understanding of life at the cellular level. Advances in modern microscopic techniques, such as the use of fluorescent markers and microscopic systems with sub-light diffraction limits, lowered the cost of acquiring images and enhanced the level of details of these images [3]. On the flip side, the abundance, heterogeneity, dimensionality, and complexity makes manual image analysis a laborious and expensive task. Consequently, automating the analysis of these images can help accelerate the progress in cell biology and other related biomedical fields.

There are numerous benefits associated with potentially automating these tasks, based on a survey of existing literature. Automating these tasks using image processing and machine learning may result in significant time savings as well as reduce measurement variability due to operator-dependent and parameter-sensitive conditions [3]. Additionally, automation has the potential of quantifying numerous cell topology characteristics that are difficult or expensive to do so manually [4].

There have been many papers published on developing automated methods for segmenting and counting cells in microscopy images. Common approaches range from classical image processing techniques, such intensity thresholding and morphological operations, to modern approaches such as tensor voting schemes, neural networks, and Markov random fields [3]. Due to the astounding variety of the different types of cell analysis problems, there are just as many cell segmentation methods that combines existing techniques and apply it to a new problem, and our project is no exception.

III. METHOD

The high-level image processing pipeline is illustrated in figure 1. Each step in the pipeline is described in detail in the following sections.

[1]F. Yang: Department of Electrical Engineering, Stanford University
[2]A. Jeje: Department of Computer Science, Stanford University

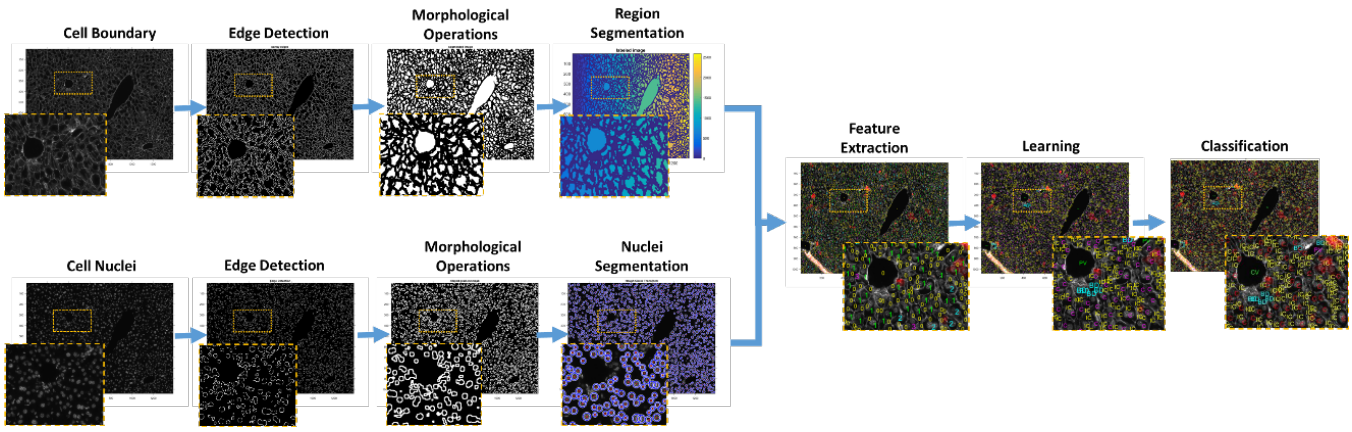

Fig. 1: Image processing pipeline for cell boundary segmentation, cell nuclei segmentation, and liver features classifications.

A. Cell Boundary Segmentation

a) Edge Detection: The cell boundaries of a gray-scale cell boundary input image are detected using the Canny edge gradient. The Canny edge detector is chosen over other edge detection methods such as Prewitt and Sobel edge detection because the Canny edge detector calculates gradients based on derivative of a Gaussian filter and therefore is less susceptible to errors due to ambiguous cell boundaries. A sample image resulting from Canny Edge detection is shown in figure 2 .

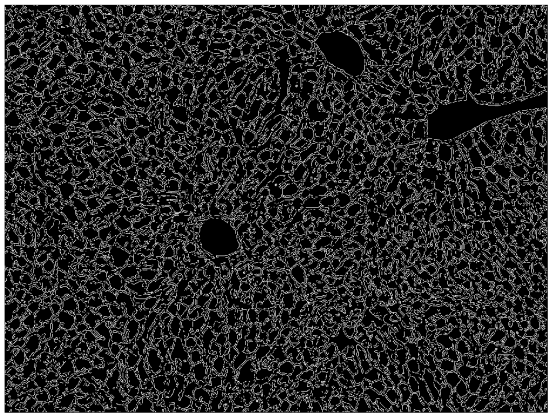

Fig. 2: A sample cell boundary output image using Canny edge detection.

b) Morphological Operations: A few morphological operations are performed to the resulting edge image to enhance the salient cell boundaries so that distinct cell regions can be identified. First, the image is dilated using a line structure of 2 pixel length in both the horizontal and vertical direction. Second, the image is negated to so that only the cell boundary information is present. Next, small regions are flood filled to reduce noise. Finally, the cell regions are smoothed using opening and closing operations with a 1 pixel diamond structuring element. Sample output images from each of the morphological operations is shown in figure 4.

c) Region Segmentation: Following edge detection and morphological operation that enhance the boundary of cell structures, regions are segmented by getting 8-connected neighborhoods. While there are many techniques in the technical literature for extracting regions in microscopy images, including the principal component analysis approach [5], [6] or the graph mining process [7], segmenting regions based on connectivity matches the identification pattern of a researcher thereby providing satisfactory results. A sample image of segmented regions, including potential cells, non-cellular regions, and veins, are distinctively colored in a gradient in figure 3.

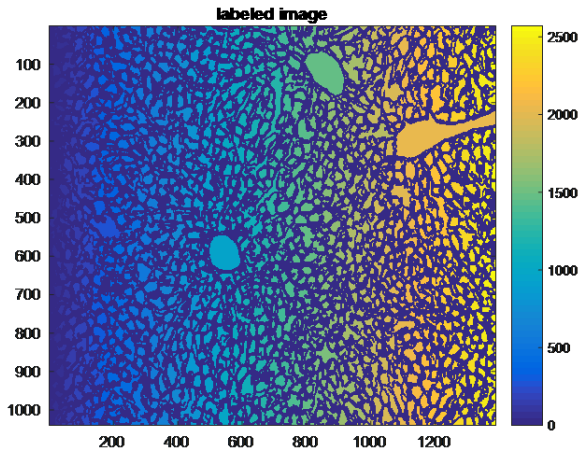

Fig. 3: A sample image of segmented regions, including potential cells, non-cellular regions, and veins, each distinctively colored in a gradient.

B. Cell Nuclei Segmentation

a) Pre-processing: An input gray-scale image of the nuclei is preprocessed using global image thresholding via Otsu's method that is adjusted by a small tuning parameter to reject ambiguous regions in the image.

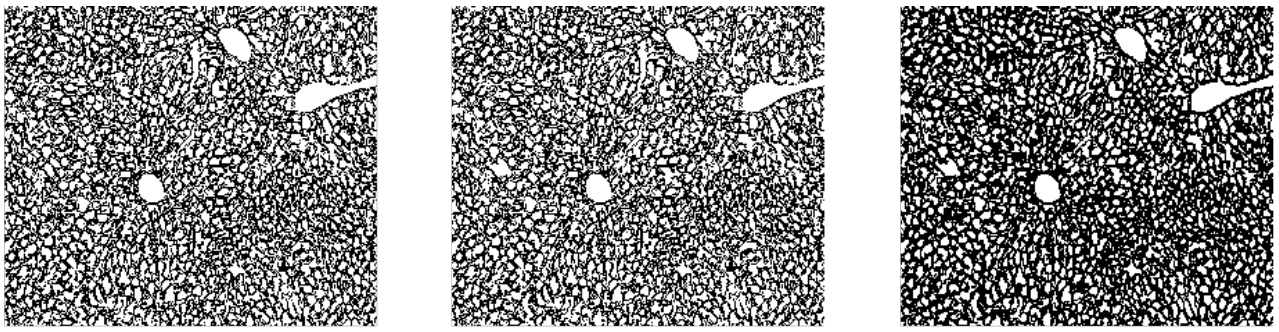

Fig. 4: A sample image illustrating the following morphological operations: a) dilated gradient mask, b) binary image with small holes filled in, c) smoothed cell boundaries using opening and closing.

b) Edge Detection: Following the binarization of the image, Canny edge detection is used to detect nuclei edges in the image. Canny edge detection also works well for nuclei segmentation due to its robustness to noise.

c) Morphological Operation: Following edge detection, the nuclei boundaries are enhanced by dilating the image using a line structuring element of 3 pixel length in the horizontal and vertical directions.

d) Hough Transform: To detect the circular shaped hepatocyte cell nuclei, circular Hough Transforms emerged as the best option among other techniques attempted which include SIFT keypoint, SURF keypoint, and morphological operation. The performances of each operation are compared in the results section. MATLAB implements the circular Hough Transform via the function imfindcircles, which takes in the following parameters:

- Radius: Circle radius is the approximate radius of the circular objects to detect, specified as a scalar of any numeric type.
- Sensitivity: Sensitivity factor is the sensitivity for the circular Hough transform accumulator array. Increasing the sensitivity factor leads to detection more circular objects and a potential increase in false detections.
- EdgeThreshold: Edge gradient threshold sets the gradient threshold for determining edge pixels in the image. Decreasing the threshold leads to detection of more circular objects with both weak and strong edges.

Based on testing with training images the circular Hough Transform is robust to the EdgeThreshold parameters, because the Hough Transform is applied on an edge map instead of a raw input image. Also, there is little variance in the optimal sensitivity parameter among images, since the microscopy images are taken in uniform and controlled lighting conditions. However, the algorithm is highly sensitive to the radius range of the objects that it searches over.

Because the nuclei image potentially contains nuclei of many types of cells, which have different nuclei size and shapes, the algorithm is adjusted to detect only the hepatocyte cell nuclei that have large, circular shapes, borrowing ideas from the existing literature [8], [9]. To do so, the equivalent diameter of the connected regions are estimated using the dilated edge map. The minimum radius is selected to be the 25th percentile of the range limits to avoid false positive detections of non-hepatocyte cells, while the maximum radius is selected to be the 99th percentile to avoid missing large cells. In summary, a sample image of the four steps in nuclei segmentation are shown in figure 5.

C. Liver Vessel Type Classification

a) Feature extraction: From the results of nuclei and cell boundary segmentation, relevant features are extracted from each image in the dataset, including geometric properties, boundary properties, cell properties, neighboring region properties. While all the geometric properties and a couple of boundary properties are obtainable via the MATLAB function RegionProps, the interesting features including SIFT keypoint density, distance from bile duct, and average cell size vs. relative distance are intended to quantify the domain knowledge used by an expert researcher classifying vessels present in a tissue as portal vein or central vein.

The implementations of domain-specific features are described in detail below:

- SURF key point density: The number of SURF keypoints that appear divided by area around a region. This is the Speeded-Up Robust Features keypoint detector that captures corners and edges in images. This parameter is also intended to capture the roughness or smoothness around the boundary of a region.
- Count of nuclei: The number of nuclei that is present in a cell. This parameter is computed by checking if the nuclei centroid detected via nuclei segmentation lie in any of the nearest regions detected via region segmentation.
- Size of nuclei: The total area of nuclei that are present in a cell. This number is a proxy for the amount of DNA matter in a cell. This parameter is computed by summing the total area of all nuclei detected in a cell region.
- Average cell size vs. relative distance from vessel: The average cell size relative to its distance away from a cell. This set of features is computed by averaging the size of cells at different distance away from the

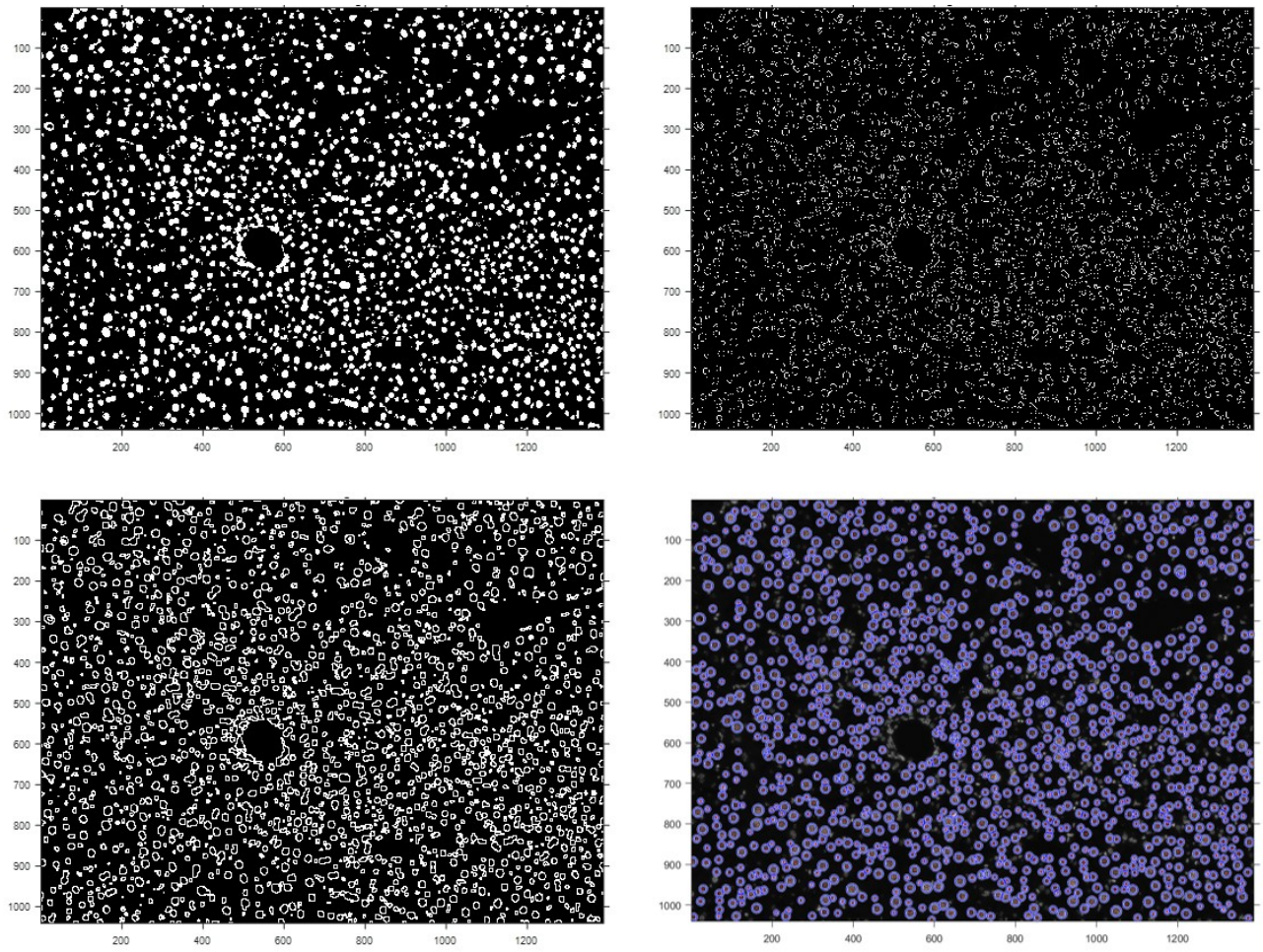

Fig. 5: A set of sample output images illustrating the steps of nuclei segmentation: a) binarized image using global thresholding, b) edge image using Canny edge detection, c) enhanced edge image using dilation, d) detected nuclei using Hough Transform

boundary of a vessel. They are intended to capture the domain knowledge that cells tends to be more compressed around a portal vein than a central vein, which can be observed the colored live section images in the Appendix(figures 14 and 15). For example, the following figure 6 demonstrates the phenomena that average cell size nearby a portal vein is smaller than that nearby a central vein.

- Distance to Bile Duct: Distance to the nearest bile duct computed as the centroid of the cell to the nearest pixel of the region. This feature intends to capture the domain-knowledge that portal veins tend to have adjacent bile ducts, which can be observed in the two sample liver section images provided in the Appendix(figures 14 and 15).

b) Learning and Classification: The features extracted above are used to create feature vectors for bile duct, central vein and portal vein training examples in 12 images. These training examples are passed into the following classifiers and the models are tested using the remaining 9 images:

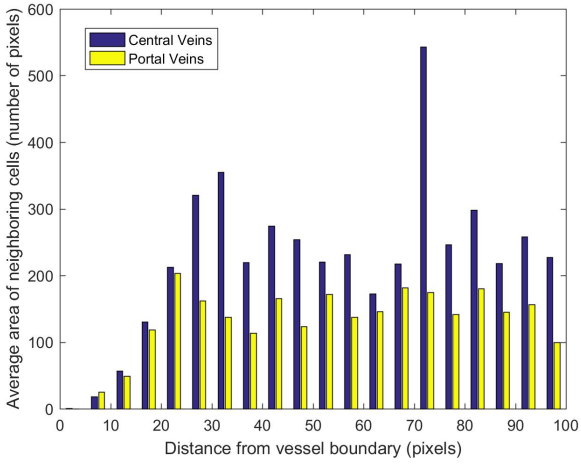

Fig. 6: Average cell size nearby a portal vein vs a central vein.

- Support Vector Machine: A multiclass SVM model is fitted using an error-correcting codes multiclass model with a one-vs-one SVM for each class and a gaussian kernel.
- K-Nearest Neighbors: The majority class of the three nearest neighbors of a feature vector is used for prediction.
- Multinomial Logistic Regression: A multinomial logistic regression model is fitted for the 3 classes with the training set.

IV. Results and Analysis

A. Cell Counting

The performance of the cell counting method based on the cell boundary and nuclei segmentation pipelines is analyzed in this section. An overlay image of cell boundary and nuclei shows stained hepatocyte cells, which may have more than 1 nucleus. The positions of regions segmented via cell boundary are registered with the positions of segmented nuclei, and a region is counted as a cell if it contains at least one nucleus. A sample image showing regions labeled with the number of nuclei present in that region is shown in figure 7.

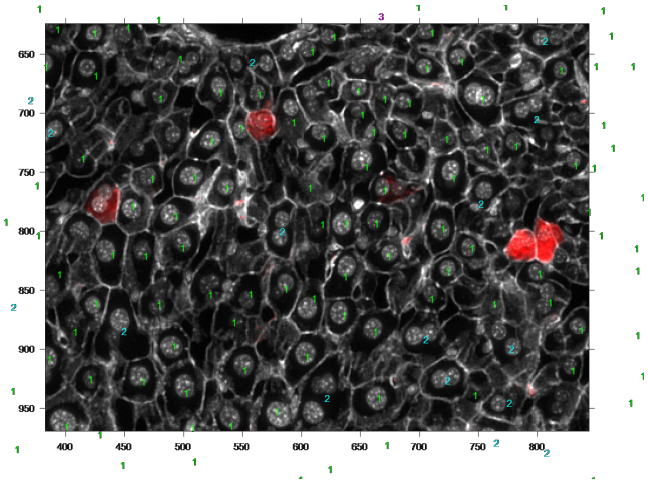

Fig. 7: A sample image showing cells labeled with the number of nuclei present

Eight images are used for error analysis. The comparison between the number of cells manually counted by a trained researcher and these automatically counted are show in table I.

Discrepancies between human counting vs. machine counting may arise in many ways. For example, when a cell boundary is ambiguous, a cell could be interpreted as one or two cells. Also, nuclei counts within detected cells could be erroneous. These discrepancies are quantified by an accuracy metric, which is the difference between automatic cell counts and manual cell counts over the total manual cell counts, as shown in figure 8.

The automatic cell counting method tends to overcount the number of 1 nuclei cells as the 1 nuclei cells have lower

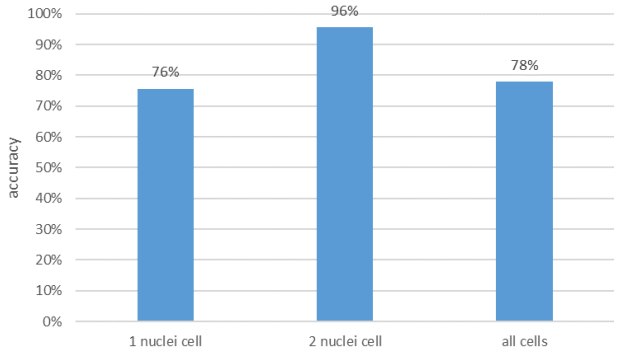

Fig. 8: A graph showing the comparison between the number of cells manually counted by a trained researcher and these automatically counted

accuracy. On the other hand, the automatic method is fairly accurate counting the number of 2 nuclei cells. Overall, the automatic algorithm achieves a respectable accuracy of 78% for all cells.

B. Nuclei Segmentation

The performance of the nuclei segmentation method is analyzed in this section. The output of the nuclei segmentation pipeline is the count of hepatocyte cell nuclei that have a characteristic large, circular shape. A sample output of each of the four methods is shown in the figure 9, where a detected nucleus is traced by a circle outline.

The Hough transform not only detects nuclei of different size but also distinguishes between two overlapping nuclei. SIFT keypoint detection fails to positively identify many nuclei perhaps due to the uniform nature of nuclei that lack corners for SIFT to positively identify as keypoints. SURF keypoint detection does a much better job than SIFT keypoint detection, however SURF still creates many false positives, i.e. nuclei identified where none exists. Finally, nuclei counting based on connected regions post morphological erosion and dilation is not robust enough because it cannot be finetuned to work across a range of images, and it detects many false positives.

The performance of each methods is quantified by comparing with the number of nuclei counted by a trained researcher. The number of false positives and false negatives based on test images is shown in figure 10. Because nuclei counts are cross-referenced with cell boundary segmentations by eliminating any nuclei not belonging in a cell, recall is more important than precision.

C. Vessel Type Classification

The classification results shown in figures 11, 12, 13 validates the choice of features used, especially those extracted based on domain knowledge. The SVM model unsurprisingly performs the best out of the three learning models tested. The features extracted capture the characteristics of bile ducts as all three methods gave great prediction for them. However, the lack of very clear distinctive characteristics between

	Manual Labeling Cell Counts				Automatic Detection Cell Counts			
	1 nuclei cell	2 nuclei cell	3 nuclei cell	2 nuclei cell	1 nuclei cell	2 nuclei cell	3 nuclei cell	2 nuclei cell
img1	34	6	0	0	37	10	3	0
img2	28	0	0	0	26	6	1	1
img3	115	22	1	0	146	21	6	0
img4	41	7	0	0	48	8	0	1
img5	79	19	1	0	90	20	3	0
img6	51	19	0	0	67	12	1	0
img7	18	7	0	0	34	5	1	0
img8	24	7	0	0	37	5	1	0

TABLE I: A table showing the comparison between the number of cells manually counted by a trained researcher and these automatically counted.

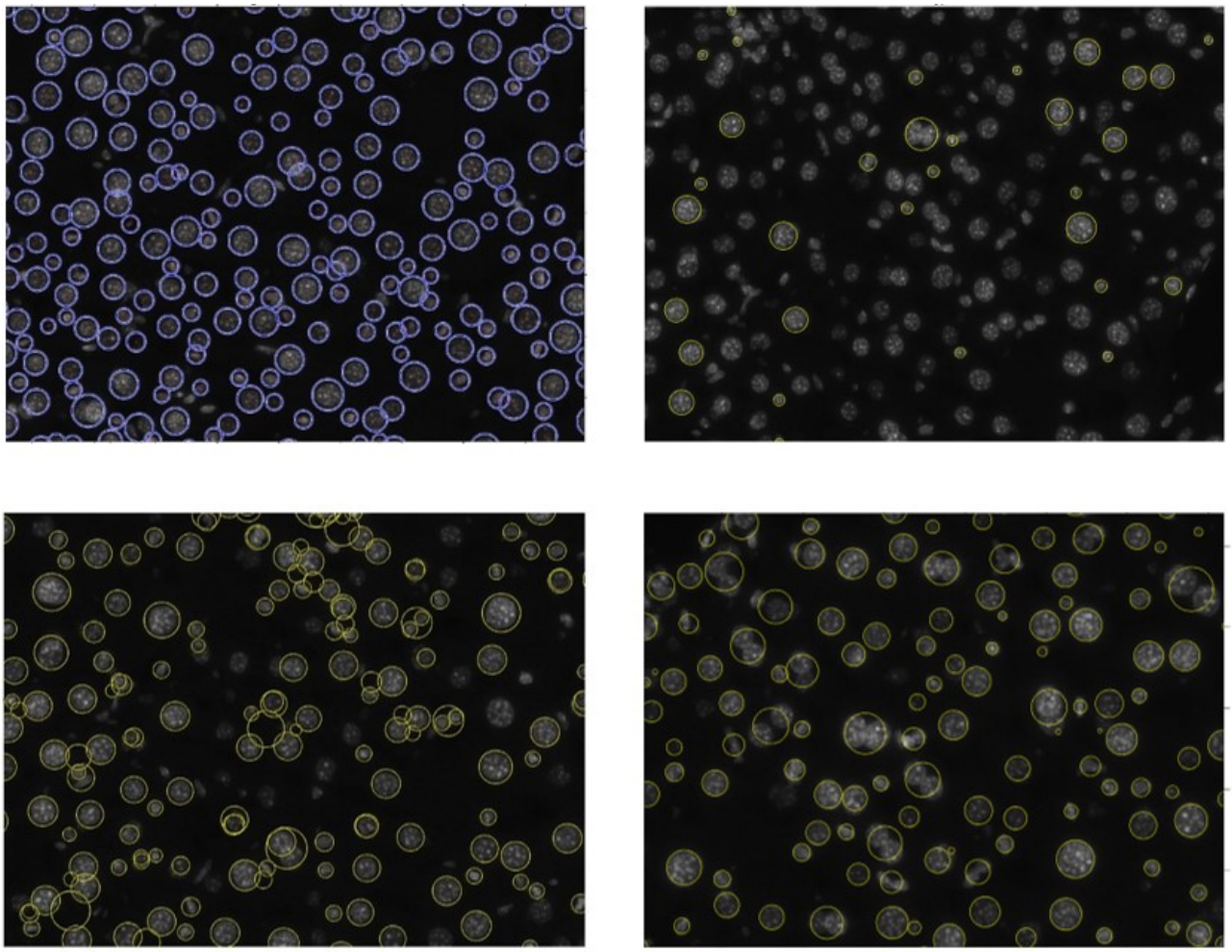

Fig. 9: Sample nuclei segmentation output using the following techniques: a) Hough Transform, b) SIFT keypoint, c) SURF keypoint, and d) morphological connected neighbors

central veins and portal vein image regions leads to some misclassifications.

V. Conclusion

The goals of this projects are threefold: 1) accurately detect cell counts, 2) correctly classify liver vessel types, and 3) accurately segment overlapping nuclei. Given the limited timeframe and resources available, this project successfully developed methods that automatically count cells with about 80% accuracy and segment overlapping nuclei with about 60% precision and 100% recall. Also portal vein, central vein, and bile duct in our very limited dataset were classified with 91% precision.

There were many challenges encountered that were resolved to varying degrees in the project. Ambiguous cell boundary is the biggest factor that contributes to over-counting 1-nuclei cells, because the 2-nuclei cells with

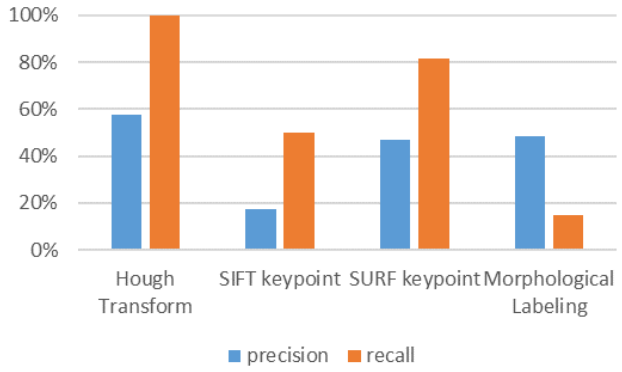

Fig. 10: A sample image showing cells labeled with the number of nuclei present

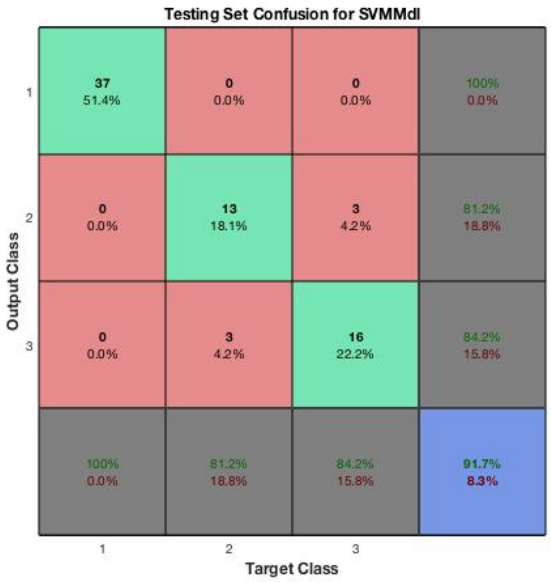

Fig. 11: Multiclass SVM model test confusion matrix

ambiguous boundaries may detected as two 1-nuclei cells. Another major challenge is extracting the relevant features for classification of vessel types based on domain knowledge. For example, average cell size relative to distance from vessel boundary is computed based on the observation that cells around portal veins tend to be more compressed than these around central veins. A final major challenge is the expensive nature of generating labeled datasets, limiting the scope of supervised learning algorithms that are feasible.

VI. Future Work

Future works could focus on improving the classification accuracy of the vessel types, by building up a datasets with thousands of labeled images. This could present the opportunity of using autoencoders to extract classification features for the models used here and perhaps a convolutional neural network as a classification model. Based on a robust

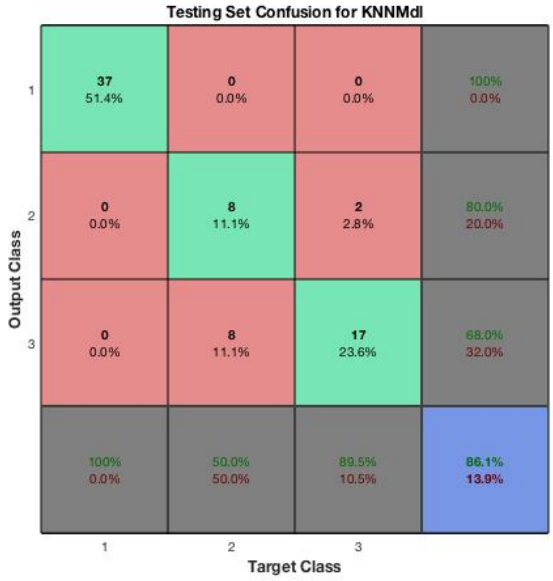

Fig. 12: K-nearest neighbors model test confusion matrix

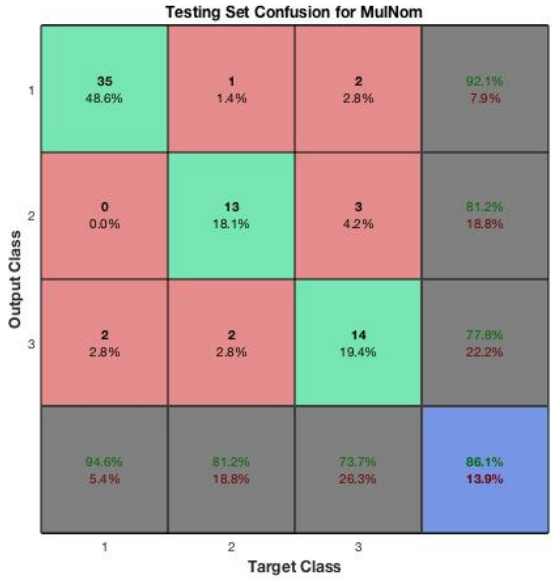

Fig. 13: Multinomial logistic regression model test confusion matrix

feature classification algorithm, certain expensive research tasks can also be automated, such as measuring the number of cells closer to certain types of vessel. Finally, the project scope can be expanded to include detecting other cell types, such as endothethial cells and bile duct epithelial cells.

VII. Acknowledgment

We would like to thank Prof. Gordon Wetzstein, the course staff Hershed Tilak and Jean-Baptiste Boin, and MD/PhD

student Dani Zhao of the Nusse Lab for their help and suggestions throughout the course of this class. Their support and mentorship were invaluable to the successful completion of this project.

REFERENCES

[1] *Data from United Network for Organ Sharing (UNOS), http://www.unos.org*, 2014.

[2] W. B, Z. L, F. M, L. CY, and N. R, "Self-renewing diploid axin2+ cells fuel homeostatic renewal of the liver.," *Nature*, 2015.

[3] E. Meijering, "Cell segmentation: 50 years down the road," *IEEE Signal Processing Magazine*, vol. 29, no. 5, September 2012, pp. 140145.

[4] I. Grishagin, "Automatic cell counting with imagej," *Analytical Biochemistry*, 2015.

[5] C. G. Loukas, G. D. Wilson, B. Vojnovic, and A. Linney, "An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections," *Cytometry Part A*, vol. 55A, no. 1, pp. 30–42, 2003.

[6] P. Wuttisarnwattana, M. Gargesha, W. van't Hof, K. R. Cooke, and D. L. Wilson, "Automatic stem cell detection in microscopic whole mouse cryo-imaging," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 819–829, March 2016.

[7] G. M. Faustino, M. Gattass, S. Rehen, and C. J. P. de Lucena, "Automatic embryonic stem cells detection and counting method in fluorescence microscopy images," *IEEE*, 2009.

[8] J. Cheng and J. C. R. ^{ast}, "Segmentation of clustered nuclei with shape markers and marking function," *IEEE Transactions on Biomedical Engineering*, vol. 56, pp. 741–748, March 2009.

[9] F. Maruhashi, S. Murakami, and K. Baba, "Automated monitoring of cell concentration and viability using an image analysis system," *Cytotechnology*, vol. 15, no. 1, pp. 281–289, 1994.

VIII. APPENDIX

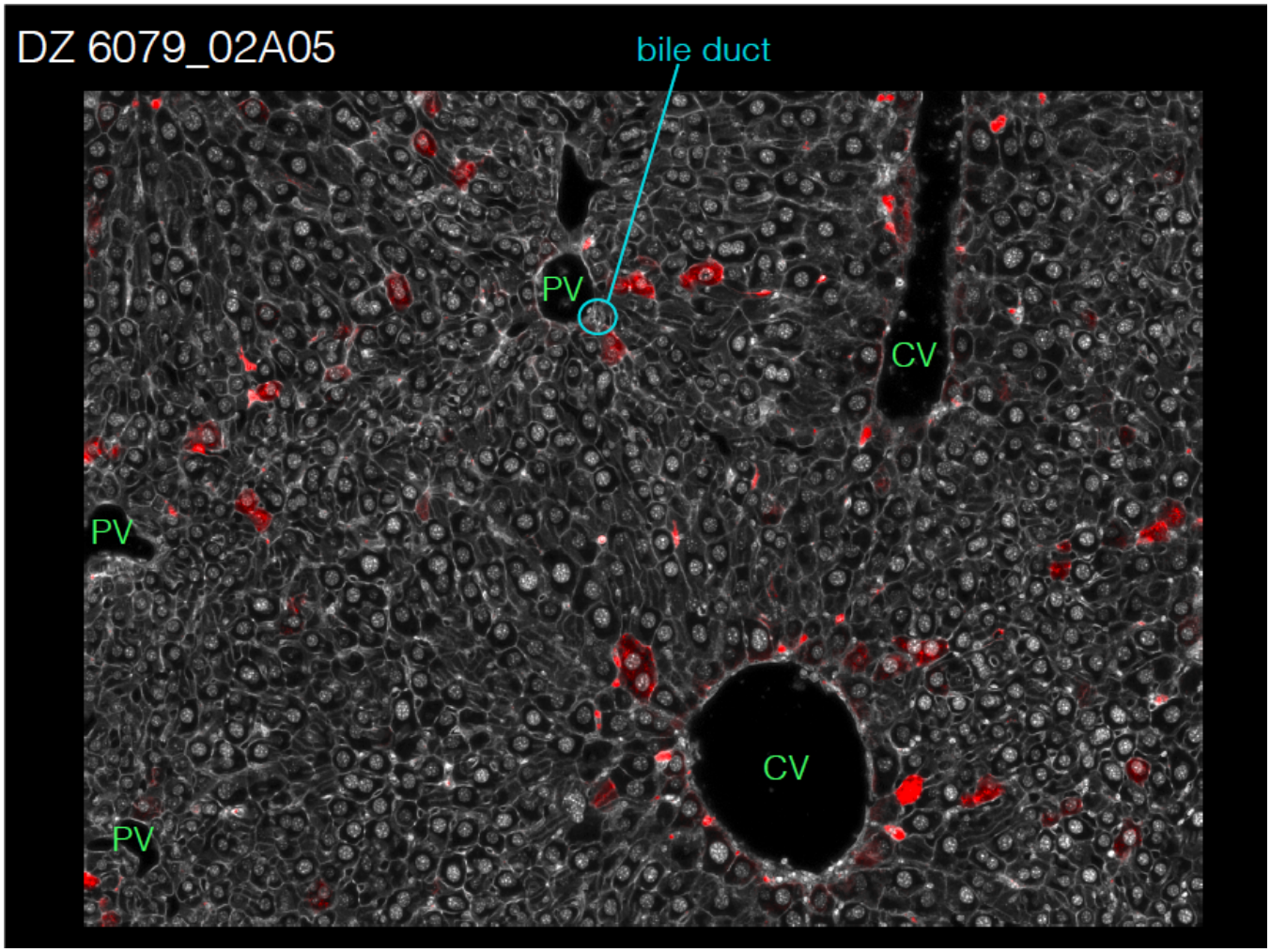

Fig. 14: Sample liver tissue section image with vessel types labeled by a trained researcher, showing portal vein (PV), central vein (CV), and bile duct (BD)

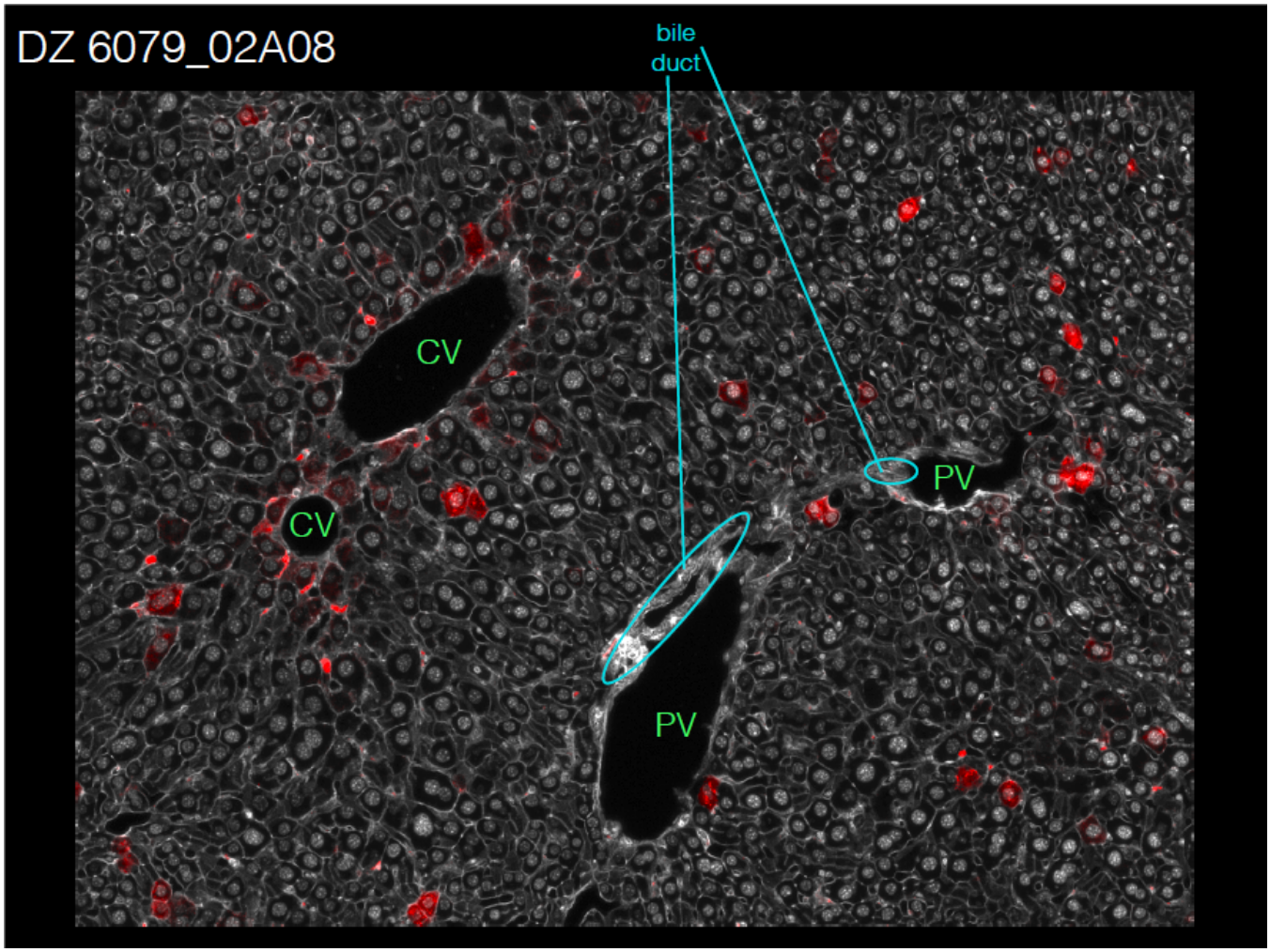

Fig. 15: Sample liver tissue section image with vessel types labeled by a trained researcher, showing portal vein (PV), central vein (CV), and bile duct (BD)

Feature Type	Feature Name	Feature Description
Geometric	Area	The number of pixels in the region
	Eccentricity	The eccentricity of the ellipse that has the same second moments as the region
	Perimeter	The distance around the boundary of the region
	Equivalent Diameter	The diameter of a circle with the same area as the region
	Major Axis Length	The length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region
	Minor Axis Length	The length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region
	Orientation	The angle between the x-axis and the major axis of the ellipse that has the same second-moments as the region. The value is in degrees, ranging from -90 to 90 degrees.
Boundary	Surf key point density	The number of SURF keypoints that appear divided by area around a region. This is the Speeded-Up Robust Features keypoint detector that captures corners and edges in images
	Bounding Box	The smallest rectangle containing the region
	Convex Image	Binary image (logical) that specifies the convex hull, with all pixels within the hull filled in (set to on)
	Pixel List	A p-by-Q matrix that specifies the locations of pixels in the region.
Cell Properties	Count of nuclei	The number of nuclei that is present in a cell
	Size of nuclei	The total area of nuclei that is present in a cell
Neighboring Regions	Average cell size vs. relative distance	The average cell size relative to its distance away from a cell
	Distance to Bile Duct	The distance to the nearest bile duct computed as the centroid of the cell to the nearest pixel of the region

TABLE II: Descriptions of features used in learning and classification