

# Deep Learning Tissue Segmentation in Cardiac Histopathology Images

# 8

Jeffrey J. Nirschl\*, Andrew Janowczyk<sup>†</sup>, Eliot G. Peyster\*, Renee Frank\*,  
Kenneth B. Margulies\*, Michael D. Feldman\*, Anant Madabhushi<sup>†</sup>

University of Pennsylvania, Philadelphia, PA, United States\* Case Western Reserve University,  
Cleveland, OH, United States<sup>†</sup>

## CHAPTER OUTLINE

<b>8.1</b>	<b>Introduction</b>	180
<b>8.2</b>	<b>Experimental Design and Implementation</b>	183
8.2.1	Data Set Description	183
8.2.2	Manual Ground Truth Annotations	183
8.2.3	Implementation	183
8.2.3.1	Network Architecture	183
8.2.3.2	Building a Training Database	183
8.2.3.3	Training a Deep Learning Model	185
8.2.4	Training a Model Using Engineered Features	185
8.2.5	Experiments	186
8.2.5.1	Experiment 1: Comparison of Deep Learning and Random Forest Segmentation	186
8.2.5.2	Experiment 2: Evaluating the Sensitivity of Deep Learning to Training Data	187
8.2.6	Testing and Performance Evaluation	188
<b>8.3</b>	<b>Results and Discussion</b>	188
8.3.1	Experiment 1: Comparison of Deep Learning and Random Forest Segmentation	188
8.3.2	Experiment 2: Evaluating the Sensitivity of Deep Learning to Training Data	188
<b>8.4</b>	<b>Concluding Remarks</b>	191
	<b>Notes</b>	191
	<b>Disclosure Statement</b>	191
	<b>Funding</b>	192
	<b>References</b>	192

**CHAPTER POINTS**

---

- Deep learning is a representation learning approach that learns informative features directly from data. It is domain independent and a generic workflow can be applied for multiple uses in digital pathology
- Deep learning is an efficient and robust method for myocyte and stroma segmentation in cardiac histopathology images with significant improvement over traditional handcrafted feature approaches

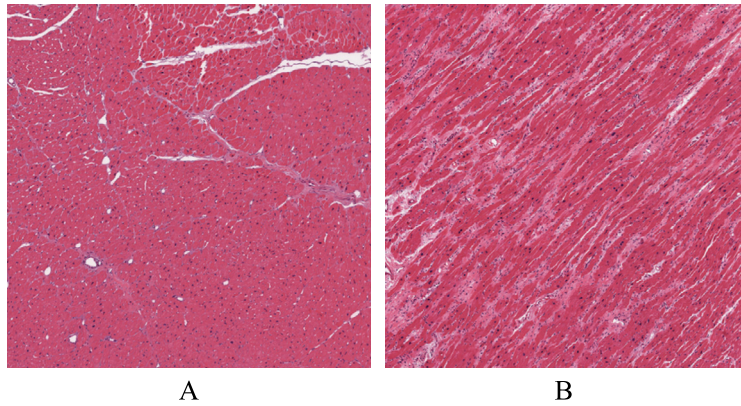
---

**8.1 INTRODUCTION**

The advent of whole-slide imaging (WSI) has increased the digitization of pathology slides, which is generating large, information-rich data sets where modern computer vision and machine learning methods thrive [1–3]. Traditional approaches for image analysis in digital pathology have involved manually engineering image features to use in a machine learning classifier [3]. These features generally involve pixel intensity statistics, texture descriptors, and image decompositions, as well as task-specific features. However, representation learning methods, such as deep convolutional neural networks or “deep learning”, have made significant advances in computer vision and are becoming increasingly common in medical image analysis [3,4]. Deep learning has many potential applications to digital pathology because it excels at tasks with large and complex training data sets, such as WSI. In addition, since deep learning models do not carry domain or tissue-specific information, a general workflow can be rapidly applied to problems across different domains. In this chapter, we provide a brief introduction to digital pathology, deep learning, and cardiac histopathology. We then present a workflow for segmentation of myocytes versus stroma in cardiac histopathology images. The long-term motivation for developing this segmentation workflow is to allow the exploration of whether computer-extracted histomorphometric features of tissue architecture from cardiac biopsy images are predictive of cardiac failure and/or rejection in the context of heart transplant patients.

Manual pathological examination takes advantage of the human visual system, which excels at automatically and efficiently partitioning an image into distinct components. The pathologist uses domain knowledge to provide biological context to the image and interpret structures into their appropriate tissue or cellular subtypes. This is followed by the identification of image features that can be used to label tissue as normal or abnormal. The final diagnosis is made through a combination of diagnostic criteria, histopathological image features, and prior experience. Computational analysis of digital pathology slides takes a similar process; image segmentation generally precedes the extraction of tissue-specific and disease-relevant features, which are ultimately used for diagnosis.

Image segmentation in digital pathology involves the partitioning the image into histologic primitives, or the biologically-relevant tissue, cellular, or sub-cellular structures (e.g. stromal tissue, nuclei, etc.) [5–8]. This allows the extraction and quantification of tissue or cell-specific features that can be used for diagnosis or prognosis

**FIGURE 8.1**

Example normal and abnormal cardiac tissue. (A) Cardiac biopsy from a patient without heart failure shows stroma limited to perivascular regions as well as regular, dense arrays of cardiomyocytes. (B) A biopsy from a patient with clinical heart failure shows an expansion of stromal tissue that disrupts the cardiomyocyte arrays. In addition, myocytes and their nuclei can be enlarged relative to normal tissue. Segmentation of myocytes and stroma is a first step toward quantifying these tissue abnormalities. Images are  $2000\text{ }\mu\text{m} \times 2000\text{ }\mu\text{m}$  at  $5\times$  magnification.

[9–14]. The process of applying computerized image analysis to quantify histologic features is known as quantitative histomorphometry. One important reason to segment histologic primitives prior to quantitative histomorphometry is to account for features that may have different predictive power depending on their context within the tissue. For example, normal cardiac tissue has stroma but it is limited to perivascular regions. However, when stroma is embedded between myocytes it is abnormal and can represent disease or fibrosis, as shown in Fig. 8.1. Although segmentation is an important step in digital pathology, manual segmentation is neither efficient nor routinely incorporated into clinical practice. Thus, developing automated algorithms for robust tissue segmentation is an important prerequisite for quantitative histomorphometry and predictive modeling.

Representation learning methods, such as deep convolutional neural networks or “deep learning”, have made significant advances in computer vision and are becoming increasingly common in medical image analysis [3,4,15]. Deep learning has many potential applications to digital pathology because it excels at tasks with large and complex training data sets, such as WSI. In addition, since deep learning models do not carry domain or tissue specific information, a general workflow can be applied to multiple cases. Deep learning differs from conventional approaches in that important features are not specified and designed *a priori*, but are rather learned directly from the data. A representation learning approach to image segmentation in digital

pathology reduces the time spent engineering features and ensures that the biological diversity and technical variance of the data set is captured.

In general, deep neural networks work by arranging nodes or artificial “neurons” in successive, convolutional, max-pooling, and fully connected layers as described elsewhere [16,17]. Parameters that specify neuron weighting are iteratively adjusted during training to approximate functions from data in a way that minimizes a loss function, which is typically correlated with the task at hand (e.g. class separation). The net result is the extraction of hierarchical features that map the input image into abstract feature vectors that capture complex and nonlinear relationships in the data.

Deep learning has recently been successfully applied to a number of digital pathology problems including localizing invasive ductal carcinoma [18], mitosis detection [5,6], colon gland segmentation [19], epithelial and stromal segmentation [8], and histopathological diagnostic support [14]. In this work, we explore the use of a deep learning framework for tissue segmentation in cardiac histopathology.

The two most prominent tissue classes in cardiac tissue are the myocytes, responsible for heart contraction, and the stroma, consisting of cellular and acellular support tissue. The stromal tissue is generally minimal in normal heart tissue. However, in the context of heart failure or cardiac rejection there is an expansion of the stromal tissue (see Fig. 8.1).

Cardiac biopsy is performed in the setting of new-onset heart failure, or for post-transplant rejection surveillance in adult heart transplant recipients [20,21]. Histopathological examination of these cases shows an expansion of the cellular and acellular stromal tissue, fibrosis, and morphological changes within myocytes, among other features. Manual review of biopsies is time consuming and intra- and inter-observer variability is high, even using revised diagnostic guidelines [22].

In cardiac biopsies, some features based on diagnostic criteria and domain knowledge include the percent tissue area that is stroma (fibrosis), lymphocyte presence and number in each tissue compartment, as well as morphological features of myocytes and their nuclei. Together, these features may allow automatic and quantitative prediction of the risk of heart failure or cardiac rejection, which would significantly improve efficiency, reliability, and quality of care for the treatment of heart failure and heart transplant patients.

In the remainder of this chapter, we focus on developing a deep learning framework for myocyte and stroma segmentation in cardiac histopathology images. Section 8.2 describes the design and implementation of two models for tissue segmentation, a deep learning approach or a random forest classifier with handcrafted features. We present our results for these experiments in Section 8.3. The first set of experiments compares the segmentation performance of a deep learning model to a traditional feature-engineered approach and supervised machine learning classifier. The second highlights the sensitivity of deep learning to the fidelity of the training annotations as well as patch representation within the training set. We conclude in Section 8.4 with potential downstream applications of myocyte and stroma segmentation in heart failure and rejection.

---

## 8.2 EXPERIMENTAL DESIGN AND IMPLEMENTATION

### 8.2.1 DATA SET DESCRIPTION

The data set includes 103 whole-slide images of H&E stained, full-thickness cardiac biopsies obtained at the time of death for non-heart-failure controls (52 patients) or heart transplant for patients with end-stage heart failure (51 patients). Images were scanned at  $20\times$  magnification using an Aperio ScanScope slide scanner and down-sampled to  $5\times$  magnification for image processing and analysis. An equal number of patients from each disease class (10 failing or non-failing) were randomly allocated to a held-out validation set and the remaining patients were used for training, such that 83 patient WSI were used for training models and 20 patients were kept as a held-out validation cohort.

### 8.2.2 MANUAL GROUND TRUTH ANNOTATIONS

For each patient in the training data, at least  $1 \times 10^4 \mu\text{m}^2$  tissue area of the WSI, per class, was manually annotated as myocyte or stroma. Training patches were selected from these annotated regions with hypersampling of regions at the interface between two tissue classes (myocyte or stroma), which we refer to as edge-hypersampling. See Fig. 8.2 for example training annotations used for patch selection.

Manual ground truth annotations from the 20 patients in the held-out validation set were used to assess segmentation performance. These validation annotations were performed on one ROI of  $500 \mu\text{m} \times 500 \mu\text{m}$   $5\times$  magnification image for each patient. In each ROI, the entire field was annotated, in detail, to allow a pixel-level comparison of the manual and automated segmentation methods. Example ground truth validation annotations are shown in Section 8.4.

### 8.2.3 IMPLEMENTATION

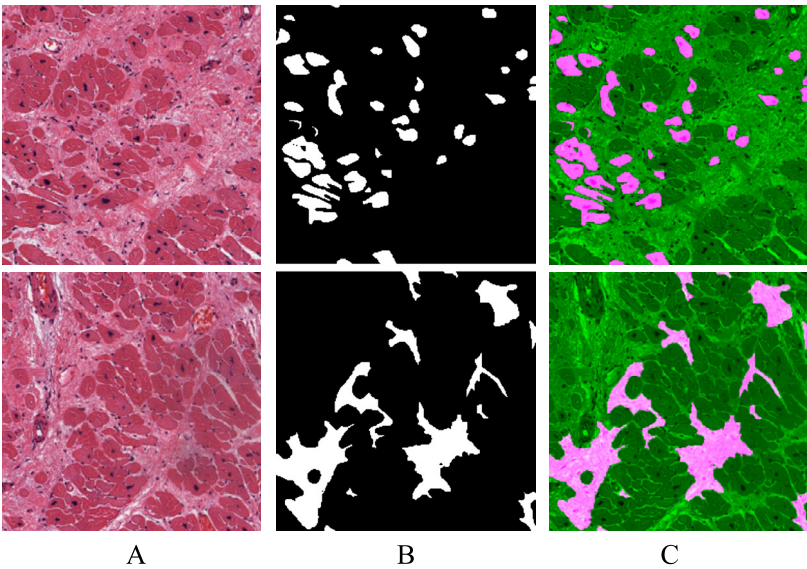
There are a number of excellent and well-maintained platforms for implementing deep learning. Apart from language, speed, and ease of implementation, an actively maintained package with a strong development community is important. We provide an abbreviated list of platforms in Table 8.1. In this tutorial, we use the open-source framework Caffe [23]. We direct the reader to [23] and the Caffe online documentation for additional information.

#### 8.2.3.1 Network Architecture

Many software platforms have existing networks that can be modified for new specific uses cases. In this example, we use a modified version of the AlexNet architecture [24,25], which was used in the  $32 \times 32$  CIFAR-10 challenge [26].

#### 8.2.3.2 Building a Training Database

The CIFAR-10 network requires training patches to be  $32 \times 32$  pixels. The first step, then, is to determine the appropriate magnification for training patches. In general, we



**FIGURE 8.2** Ground truth for myocytes or stroma. (A) Original images. (B) Binary ground truth mask for myocytes (top) or stroma (bottom) for pixel-level training annotations. White pixels indicate class membership. (C) Overlay image of binary mask in B on original image in A. Images are 500  $\mu\text{m} \times 500 \mu\text{m}$  at 5 $\times$  magnification.

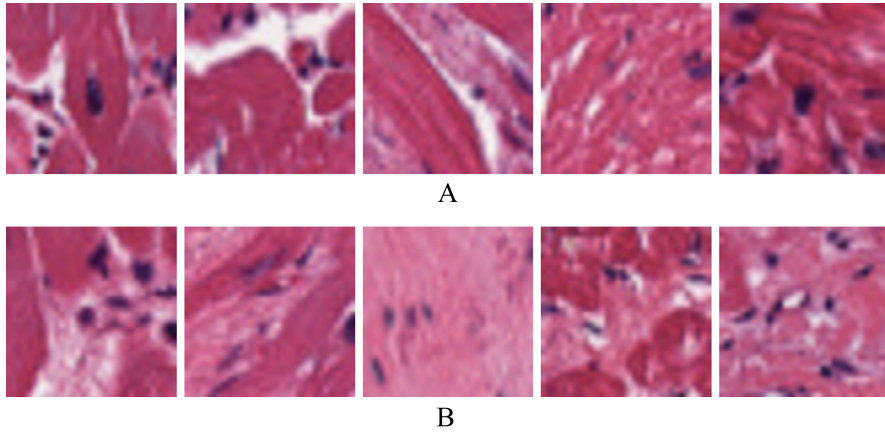
**Table 8.1** Abbreviated list of deep learning software packages

Software	Language	URL
Caffe	C++	<a href="http://caffe.berkeleyvision.org/">http://caffe.berkeleyvision.org/</a>
Torch	LUA, C	<a href="http://torch.ch/">http://torch.ch/</a>
Theano	Python, C	<a href="http://deeplearning.net/software/theano/">http://deeplearning.net/software/theano/</a>
TensorFlow	Python, C++	<a href="https://www.tensorflow.org/">https://www.tensorflow.org/</a>

recommend the highest magnification that allows the histologic primitive of interest to fit within the patch window, while still containing contextual information outside of the histologic primitive. The patch should also have enough resolution and context for an expert to correctly classify the center pixel of image. For cardiac histopathology, we downsample the 20 $\times$  magnification WSI to 5 $\times$  apparent magnification, which allows a typical myocyte in cross-section to occupy approximately 50% of the patch window. Example training patches are shown in Fig. 8.3.

As mentioned previously, deep learning identifies discriminative features directly from training data. Thus, the next important consideration is the choice and representation of training patches. We find that training is more efficient when the transition zones, or the interfaces between one or more classes, are over-represented in the data



**FIGURE 8.3**

Example training patches. Example  $32 \times 32$  training patches used for both the deep learning and RF approach, where the center pixel of the patch determines the class label. (A) Myocyte training patches. (B) Stroma training patches. Each patch is  $64 \mu\text{m} \times 64 \mu\text{m}$  at  $5\times$  magnification.

set, as we later show in Experiment 2. To this end, we hypersample patches from the edges of the annotated training masks such that edge patches are over-represented in the training data set based on their distance to the mask edge.

### 8.2.3.3 Training a Deep Learning Model

We train the deep learning classifier using  $5 \times 10^3$  patches per patient, and augment the training set by rotating each patch by  $90^\circ$ ,  $180^\circ$ , or  $270^\circ$  for  $1.66 \times 10^6$  patches per class and a total of  $3.32 \times 10^6$  training patches. These patches were split into three training and testing-folds at the patient level with balanced myocyte and stroma patches. Each fold was trained for  $6 \times 10^5$  iterations on a Tesla K20c GPU with CUDA 7.0 using the cuDNN setting, the AdaGrad solver [27] built into Caffe, and a fixed batch size. The network parameters for the AlexNet architecture are shown in Table 8.2.

## 8.2.4 TRAINING A MODEL USING ENGINEERED FEATURES

We compare our deep learning approach to a traditional segmentation workflow using engineered features and a Random Forest classifier. Random forests were chosen as a classifier because they are among the top performing supervised machine learning algorithms [28,29], they perform well on biomedical image data sets [30], and require fewer adjustments to obtain stable results.

The Random Forest classifier for myocyte and stroma segmentation was trained using 1000 trees and 333 engineered-features based on intensity and texture. Ta-

**Table 8.2** Deep learning hyperparameters

Variable	Setting
Batch size	128
Initial learning rate	0.001
Learning rate schedule	AdaGrad
Rotations	[0, 90, 180, 270]
Number of iterations	600,000
Number of epochs	23
Weight decay	0.004
Random mirror	Enabled
Transformations	Mean-centered

**Table 8.3** Description of features used for the Random Forest stroma classifier

Category	Length	Features
Intensity	13 × 3 images	Energy, entropy, min, max, standard deviation, variance, mean, median, median absolute deviation, range, root mean squared, skewness, and kurtosis of patch intensity
Texture	98 × 3 images	13 Haralick texture features [33], 25 Laws texture features [34], 1 Local Binary Pattern [36], 21 XY gradient features, 6 Gabor features [32], 26 CoLIAGe features [31], 6 CORF features [35]

*Intensity features were calculated on a 32 × 32 sliding window with the center pixel assigned the computed value for each patch. The three grayscale images of each patch that were used include a grayscale conversion of the RGB image and a color-deconvolved hematoxylin or eosin grayscale image.*

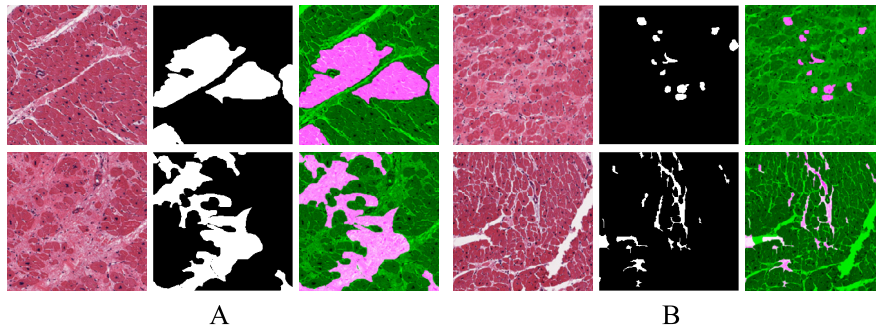
ble 8.3 lists the features used here. The intensity features capture the differences in H&E staining between myocytes and stroma by considering intensity magnitude, variation, and higher-order moments. Texture features such as Co-occurrence of Local Anisotropic Gradient Orientations (CoLIAGe) features [31], Gabor features [32], and Haralick features [33], among other texture features [34–36], were designed to detect the edges of class boundaries as well as the heterogeneity and disorder of the stroma relative to myocytes. These features were computed for each 32 × 32 patch of the grayscale converted RGB image as well as the color-deconvolved hematoxylin or eosin grayscale image [37]. The feature vector and label for the center pixel patch was used for training. All of the training patches were used to construct a single RF model.

## 8.2.5 EXPERIMENTS

### 8.2.5.1 Experiment 1: Comparison of Deep Learning and Random Forest Segmentation

Experiment 1 compares the segmentation performance of deep learning model to a random forest model using engineered features in cardiac histopathology images. Each approach was trained using the same set of image patches and the output from



**FIGURE 8.4**

Coarse versus fine training annotations. This figure shows two different approaches for generating training annotations for deep learning. (A) Coarse annotations maximize the tissue area that is annotated, but may lack precise class boundaries. (B) Fine training annotations capture less tissue area and may require more effort, but generate delineate class boundaries with pixel-level precision. Experiment 2 shows the effect of training annotation and patch representation on the deep learning output. Myocytes are annotated in the top row and stromal tissue in the bottom row. Images are  $500\ \mu\text{m} \times 500\ \mu\text{m}$ .

each model is an image where each pixel is assigned a probability of class membership, between 0 and 1. Thresholding the probability output at 0.5 should maximally separate the classes, which is why a fixed threshold of 0.5 was used to create binary masks.

#### **8.2.5.2 Experiment 2: Evaluating the Sensitivity of Deep Learning to Training Data**

Routine tissue annotation in digital pathology is generally performed at intermediate magnifications and produces masks that lack delineation of precise, pixel-level tissue boundaries. While this approach can rapidly annotate large regions of the WSI, it may have consequences for the data that ultimately go into the network. Since deep learning techniques learn models from the data presented, it is important to consider the training annotation method as well as the best procedure for sampling patches from within these annotations. Fig. 8.4 shows an example of coarse annotations (A) versus fine annotations with pixel-level boundaries (B).

Uniform random sampling of patches from training annotations assumes that all patches provide equally valuable information for classification. This is not true in general, though, as WSI often contain large regions of tissue that are qualitatively very different and easily classified using simple approaches. However, the transition zones between tissue classes are more challenging, and we can direct the deep learning network to learn these boundaries better by modifying the patch representation within the training data. We modify patch representation by hypersampling patches near the edge of the training annotation, or edge-hypersampling, an approach also

used in [38]. In this experiment we compare uniform and edge-hypersampling approaches for both coarse and fine, pixel-level training annotations.

### 8.2.6 TESTING AND PERFORMANCE EVALUATION

The fully annotated ROIs from the held-out patients were used to assess segmentation performance for both sets of experiments. We evaluate the segmentation performance for each experiment using multiple criteria including: the Area Under the Curve (AUC), F1 score, True Positive Rate (TPR), True Negative Rate (TNR), Positive Predictive Value (PPV), Modified Hausdorff Distance (MHD) [39,40].

---

## 8.3 RESULTS AND DISCUSSION

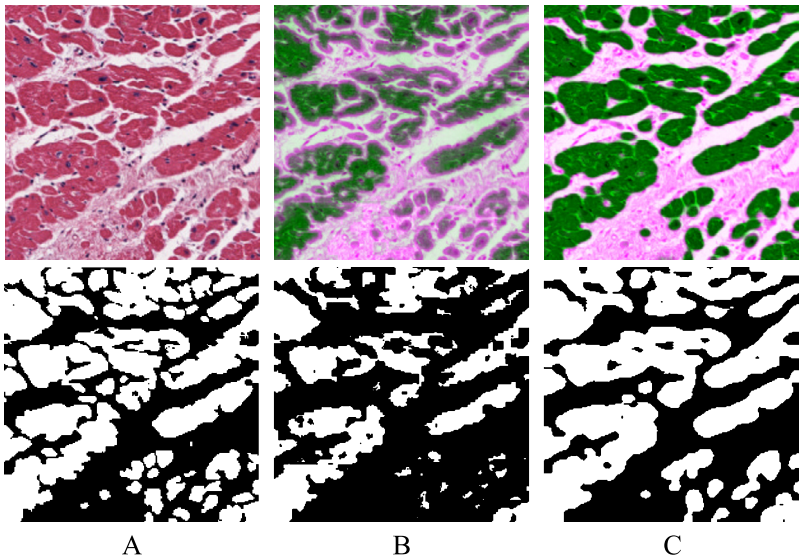
### 8.3.1 EXPERIMENT 1: COMPARISON OF DEEP LEARNING AND RANDOM FOREST SEGMENTATION

An overlay of the probability output on the original image for each classifier is shown in Fig. 8.5. Table 8.4 shows the quantitative evaluation of each classifier on 20 held-out patients using the performance metrics listed in the previous section (AUC, F1 score, TPR, TNR, PPV, MHD). Qualitatively, the deep learning model produces much more confident predictions of class membership, as noted by the green or magenta intensity of the probability outputs. It is also clear that both the deep learning and random forest models perform well on regions far from the interface with the opposing class, as expected. Since large regions of the image are easier to classify both models achieve a high F1 score and TPR. However, only the deep learning model is able to achieve both a high TPR (sensitivity) and TNR (specificity).

What is also evident from Fig. 8.5 is that deep learning more accurately delineates the class boundaries. This is quantified in Table 8.4 using the MHD, which is a robust measurement for matching objects based on their edge points [40]. The MHD shows that deep learning more accurately delineates tissue boundaries compared to the random forest model. Finally, related to the previous point, deep learning excels at detecting fine features such as small myocytes embedded in stromal tissue. These are often assigned low probabilities in the random forest model, which are lost with thresholding. Together, these data suggest that deep learning should be used when precise delineation of tissue boundaries are desired. However, in the next experiment we examine the sensitivity of deep learning and highlight an important caveat with deep learning models.

### 8.3.2 EXPERIMENT 2: EVALUATING THE SENSITIVITY OF DEEP LEARNING TO TRAINING DATA

Fig. 8.6 shows the effect of coarse versus fine training annotations and edge hypersampling on deep learning segmentation. The first result is that coarse annotations

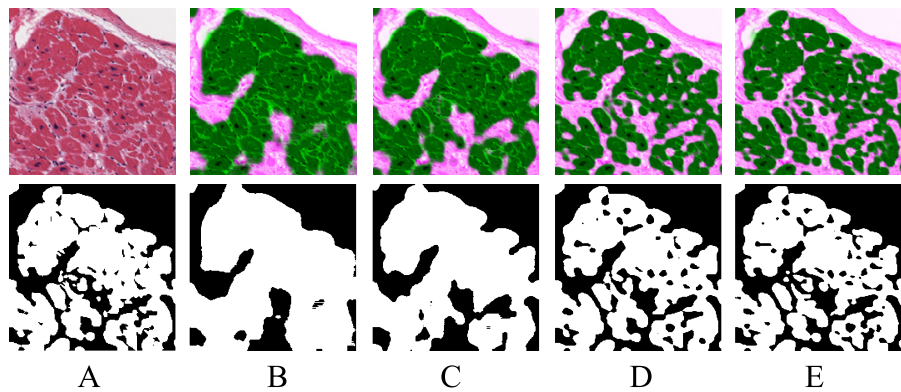


**FIGURE 8.5** Qualitative comparison of segmentation performance. (A) The original H&E image and ground truth annotation. (B) Random forest probability output overlaid on the original image and binarized segmentation mask. (C) Deep learning probability output overlaid on the original image. The green or magenta channel intensity in the probability output probability of the pixel belonging to the myocyte or stroma class, respectively. White pixels in the binary masks represent the myocyte class. Images are  $500\ \mu\text{m} \times 500\ \mu\text{m}$ .

**Table 8.4** Mean  $\pm$  SD segmentation performance on validation set for Experiment 1

Metric	RF	DL Fold 1	DL Fold 2	DL Fold 3	DL Mean
AUC	$0.80 \pm 0.12$	$0.95 \pm 0.02$	$0.94 \pm 0.03$	$0.95 \pm 0.02$	$0.95 \pm 0.02$
F1 score	$0.91 \pm 0.05$	$0.96 \pm 0.03$	$0.96 \pm 0.04$	$0.96 \pm 0.03$	$0.96 \pm 0.03$
TPR	$0.92 \pm 0.09$	$0.96 \pm 0.03$	$0.96 \pm 0.05$	$0.96 \pm 0.04$	$0.96 \pm 0.04$
TNR	$0.67 \pm 0.29$	$0.94 \pm 0.03$	$0.93 \pm 0.05$	$0.94 \pm 0.03$	$0.94 \pm 0.04$
PPV	$0.90 \pm 0.05$	$0.96 \pm 0.03$	$0.97 \pm 0.02$	$0.97 \pm 0.03$	$0.97 \pm 0.03$
MHD	$122 \pm 68$	$50 \pm 11$	$54 \pm 12$	$49 \pm 10$	$51 \pm 11$

produce coarse results. While it is not entirely surprising, the visual results are striking. Edge hypersampling can partially compensate for coarse annotations, but the best results are obtained after re-annotation with precise pixel-level boundaries. The effect of edge-hypersampling is less clear in the specific case of fine myocyte and stroma annotations. This is likely due to the fact that a majority of the annotations are already at the tissue interface and already considered “edges”. The addition of edge-



**FIGURE 8.6**

Effect of training annotations fidelity on DL performance. (A) Original image (top) and ground truth annotation. (B) DL network trained with uniform random patch sampling from coarse masks. (C) Network trained with edge-hypersampling of patches from coarse masks. (D) Network with uniform random patch sampling of fine training masks. (E) Network trained with edge-hypersampling of fine training masks. Probability outputs overlaid on the original images (top row) and binary masks (bottom). Images are 500  $\mu\text{m}$   $\times$  500  $\mu\text{m}$ .

**Table 8.5** Mean  $\pm$  SD segmentation performance on validation set for Experiment 2

Metric	Coarse-Unif	Coarse-Edge	Fine-Unif	Fine-Edge
AUC	0.80 $\pm$ 0.10	0.84 $\pm$ 0.08	0.95 $\pm$ 0.02	0.95 $\pm$ 0.02
F1 score	0.90 $\pm$ 0.06	0.92 $\pm$ 0.06	0.96 $\pm$ 0.05	0.96 $\pm$ 0.03
TPR	0.94 $\pm$ 0.08	0.96 $\pm$ 0.05	0.96 $\pm$ 0.03	0.96 $\pm$ 0.03
TNR	0.67 $\pm$ 0.24	0.71 $\pm$ 0.19	0.95 $\pm$ 0.03	0.94 $\pm$ 0.03
PPV	0.88 $\pm$ 0.08	0.89 $\pm$ 0.07	0.96 $\pm$ 0.03	0.97 $\pm$ 0.03
MHD	250 $\pm$ 148	226 $\pm$ 146	50 $\pm$ 11	49 $\pm$ 10

hypersampling to fine annotations has little effect, but importantly does not harm performance (Table 8.5).

Although, we manually annotate myocytes and stroma, due to their irregular boundaries, different histologic primitives may benefit from less time-consuming approaches. For example, reliable nuclei annotations can be obtained using an approach similar to Janowczyk et al. [38]. In this case, high fidelity “edge” annotations for the non-nuclei class were generated by dilating the nuclei mask and then subtracting the original mask. In any case, it is important for digital pathology users to match the granularity of their training annotations with the quality of the desired output.

---

## 8.4 CONCLUDING REMARKS

In this chapter we described the use of deep learning for the specific case of segmenting myocytes and stroma in cardiac histopathology images. We developed two models to segment myocytes and stroma, a deep learning approach and the use of a random forest classifier in conjunction with handcrafted features. A deep learning approach was used because this method has shown excellent performance in computer vision tasks and it excels with large data sets.

We evaluated the performance of our two models using a held-out data set of 20 patient images with expert annotated tissue boundaries, where deep learning was superior in all metrics. Most notably, deep learning segmentation had both a high sensitivity and specificity as well as more accurate tissue boundaries, compared to a random forest model. In addition, our results suggested that deep learning is highly sensitive to the fidelity of the training annotations and that edge-hypersampling can improve segmentation performance.

In summary, our deep learning framework required no feature engineering and showed improved segmentation performance compared to a classifier trained using handcrafted features. This will enable the identification and quantification of tissue-specific features predictive of heart failure or cardiac transplant rejection. Future work using these segmentation results will allow the development of features predictive of cardiac disease, and may eventually lead to pipelines for image-based predictors of outcome based off cardiac histopathology. This has the potential to improve cardiac biopsy screening, provide real-time feedback for clinicians, and serve as an objective second reader for pathologists.

---

## NOTES

For additional examples of deep learning in digital pathology, please see: <http://www.andrewjanowczyk.com/category/deep-learning>.

---

## DISCLOSURE STATEMENT

Dr. Madabhushi is the co-founder and stake holder in Ibris Inc., a cancer diagnostics company. Drs. Madabhushi and Feldman are equity holders and have technology licensed to both Elucid Bioimaging and Inspirata Inc. Drs. Madabhushi and Feldman are scientific advisory consultants for Inspirata Inc. and sit on its scientific advisory board. Dr. Feldman is also a consultant for Phillips Healthcare, XFIN, and Virbio. Dr. Margulies hold research grants from Thoratec Corporation and Merck and serves as a scientific consultant/ advisory board member for Janssen, Merck, Pfizer, Ridgetop Research, Glaxo-Smith-Kline, NovoNordisk.

---

## FUNDING

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award numbers R21CA179327-01, R21CA195152-01, U24CA199374-01, the National Institute of Diabetes and Digestive and Kidney Diseases under award number R01DK098503-02, the DOD Prostate Cancer Synergistic Idea Development Award (PC120857), the DOD Lung Cancer Idea Development New Investigator Award (LC130463), the DOD Prostate Cancer Idea Development Award, the Case Comprehensive Cancer Center Pilot Grant, the VelaSano Grant from the Cleveland Clinic, the Wallace H. Coulter Foundation Program in the Department of Biomedical Engineering at Case Western Reserve University, the I-Corps@Ohio Program. JJN was supported by NINDS F30NS092227.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

---

## REFERENCES

1. Anant Madabhushi, Digital pathology image analysis: opportunities and challenges, *Imaging Med.* 1 (1) (2009) 7–10, <http://dx.doi.org/10.2217/iim.09.9>, URL <http://www.futuremedicine.com/doi/abs/10.2217/iim.09.9>.
2. Farzad Ghaznavi, Andrew Evans, Anant Madabhushi, Michael Feldman, Digital imaging in pathology: whole-slide imaging and beyond, *Annu. Rev. Phytopathol.* 8 (2013) 331–359, <http://dx.doi.org/10.1146/annurev-pathol-011811-120902>, URL <http://www.ncbi.nlm.nih.gov/pubmed/23157334>.
3. Rohit Bhargava, Anant Madabhushi, Emerging themes in image informatics and molecular analysis for digital pathology, *Annu. Rev. Biomed. Eng.* 18 (March) (2016) 387–412, <http://dx.doi.org/10.1146/annurev-bioeng-112415-114722>.
4. Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444, <http://dx.doi.org/10.1038/nature14539>, URL <https://www.ncbi.nlm.nih.gov/pubmed/26017442>.
5. Dan Cireşan, Ueli Meier, Juergen Schmidhuber, Multi-column deep neural networks for image classification, *Proc. Int. Conf. Pattern Recognit.* 25 (February) (2012) 3642–3649, <http://dx.doi.org/10.1109/CVPR.2012.6248110>.
6. Haibo Wang, Angel Cruz-Roa, Ajay Basavanahally, Hannah Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio Gonzalez, Anant Madabhushi, Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features, *J. Med. Imaging* 1 (3) (2014) 034003, <http://dx.doi.org/10.1117/1.JMI.1.3.034003>.
7. J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, A. Madabhushi, Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images, *IEEE Trans. Med. Imaging* (2015), <http://dx.doi.org/10.1109/TMI.2015.2458702>.
8. Jun Xu, Xiaofei Luo, Guanhao Wang, Hannah Gilmore, Anant Madabhushi, A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images, *Neurocomputing* (2016) 1–10, <http://dx.doi.org/10.1016/j.neucom.2016.01.034>, URL <http://linkinghub.elsevier.com/retrieve/pii/S0925231216001004>.

9. A.H. Beck, A.R. Sangoi, S. Leung, R.J. Marinelli, T.O. Nielsen, M.J. van de Vijver, R.B. West, M. van de Rijn, D. Koller, Systematic analysis of breast cancer morphology uncovers stromal features associated with survival, *Sci. Transl. Med.* 3 (108) (2011) 108ra113, <http://dx.doi.org/10.1126/scitranslmed.3002564>, URL <http://stm.sciencemag.org/content/scitransmed/3/108/108ra113.full.html>.
10. Ajay Basavanahally, Michael Feldman, Natalie Shih, Carolyn Mies, John Tomaszewski, Shridar Ganesan, Anant Madabhushi, Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: comparison to oncotype DX, *J. Pathol. Inform.* 2 (2) (2011) S1, <http://dx.doi.org/10.4103/2153-3539.92027>, URL <http://www.ncbi.nlm.nih.gov/pubmed/22811953>.
11. Anant Madabhushi, Shannon Agner, Ajay Basavanahally, Scott Doyle, George Lee, Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data, *Comput. Med. Imaging Graph.* 35 (7–8) (2011) 506–514, <http://dx.doi.org/10.1016/j.compmedimag.2011.01.008>.
12. Scott Doyle, Michael Feldman, John Tomaszewski, Anant Madabhushi, A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies, *IEEE Trans. Biomed. Eng.* 59 (5) (2012) 1205–1218, <http://dx.doi.org/10.1109/TBME.2010.2053540>.
13. James S. Lewis, Sahirzeeshan Ali, Jingqin Luo, Wade L. Thorstad, Anant Madabhushi, A quantitative histomorphometric classifier (QuHbIC) oropharyngeal squamous cell carcinoma, *Am. J. Surg. Pathol.* 38 (1) (2014) 128–137, <http://dx.doi.org/10.1097/PAS.0000000000000086>.
14. Geert Litjens, Clara I. Sánchez, Nadya Timofeeva, Meyke Hermesen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-van de Kaa, Peter Bult, Bram van Ginneken, Jeroen van der Laak, Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis, *Sci. Rep.* 6 (April) (2016) 2628, <http://dx.doi.org/10.1038/srep26286>, URL <http://www.nature.com/articles/srep26286>.
15. Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8) (2013) 1798–1828, <http://dx.doi.org/10.1109/TPAMI.2013.50>.
16. Yoshua Bengio, Learning deep architectures for AI, *Found. Trends Mach. Learn.* 2 (1) (2009) 1–127, <http://dx.doi.org/10.1561/22000000006>.
17. Jürgen Schmidhuber, Deep learning in neural networks: an overview, *Neural Netw.* 61 (2015) 85–117, <http://dx.doi.org/10.1016/j.neunet.2014.09.003>.
18. Angel Cruz-Roa, Ajay Basavanahally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, Anant Madabhushi, Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks, *Proc. SPIE* 9041 (216) (2014) 904103, <http://dx.doi.org/10.1117/12.2043872>.
19. Philipp Kainz, Michael Pfeiffer, Martin Urschler, Semantic segmentation of colon glands with deep convolutional neural networks and total variation segmentation, *arXiv:1511.06919*, 2015.
20. L.T. Cooper, K.L. Baughman, a.M. Feldman, a. Frustaci, M. Jessup, U. Kuhl, G.N. Levine, J. Narula, R.C. Starling, J. Towbin, R. Virmani, The role of endomyocardial biopsy in the management of cardiovascular disease: a scientific statement from the American Heart Association, the American College of Cardiology, and the European Society of Cardiology endorsed by the Heart Failure Society of America and the Heart Failure Association of the European Society of Cardiology, *Eur. Heart J.* 28 (24)



- (2007) 3076–3093, <http://dx.doi.org/10.1093/eurheartj/ehm456>, URL <http://eurheartj.oxfordjournals.org/cgi/doi/10.1093/eurheartj/ehm456>.
21. Maria Rosa Costanzo, Anne Dipchand, Randall Starling, Allen Anderson, Michael Chan, Shashank Desai, Savitri Fedson, Patrick Fisher, Gonzalo Gonzales-Stawinski, Luigi Martinelli, David McGiffin, Francesco Parisi, Jon Smith, David Taylor, Bruno Meiser, Steven Webber, David Baran, Michael Carboni, Thomas Dengler, David Feldman, Maria Frigerio, Abdallah Kfoury, Daniel Kim, Jon Kobashigawa, Michael Shullo, Josef Stehlik, Jeffrey Teuteberg, Patricia Uber, Andreas Zuckermann, Sharon Hunt, Michael Burch, Geetha Bhat, Charles Canter, Richard Chinnock, Marisa Crespo-Leiro, Reynolds Delgado, Fabienne Dobbels, Kathleen Grady, Walter Kao, Jaqueline Lamour, Gareth Parry, Jignesh Patel, Daniela Pini, Sean Pinney, Jeffrey Towbin, Gene Wolfel, Diego Delgado, Howard Eisen, Lee Goldberg, Jeff Hosenpud, Maryl Johnson, Anne Keogh, Clive Lewis, John O’Connell, Joseph Rogers, Heather Ross, Stuart Russell, Johan Vanhaecke, The International Society of Heart and Lung Transplantation guidelines for the care of heart transplant recipients, *J. Heart Lung Transplant.* 29 (8) (2010) 914–956, <http://dx.doi.org/10.1016/j.healun.2010.05.034>.
  22. Annalisa Angelini, Claus Boegelund Andersen, Giovanni Bartoloni, Fiona Black, Paul Bishop, Helen Doran, Marny Fedrigo, Jochen W.U. Fries, Martin Goddard, Heike Goebel, Desley Neil, Ornella Leone, Andrea Marzullo, Monika Ortmann, Francois Paraf, Samuel Rotman, Nesrin Turhan, Patrick Bruneval, Anna Chiara Frigo, Francesco Grigoletto, Alessio Gasparetto, Roberto Mencarelli, Gaetano Thiene, Margaret Burke, A web-based pilot study of inter-pathologist reproducibility using the ISHLT 2004 working formulation for biopsy diagnosis of cardiac allograft rejection: the European experience, *J. Heart Lung Transplant.* 30 (11) (2011) 1214–1220, <http://dx.doi.org/10.1016/j.healun.2011.05.011>.
  23. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, Trevor Darrell, Caffe: convolutional architecture for fast feature embedding, in: *Proceedings of the ACM International Conference on Multimedia*, 2014, pp. 675–678, arXiv:1408.5093, 2015.
  24. Alex Krizhevsky, G. Hinton, Convolutional deep belief networks on CIFAR-10, Unpublished manuscript, pp. 1–9, URL <http://www.cs.utoronto.ca/~kriz/conv-cifar10-aug2010.pdf>, 2010.
  25. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* (2012) 1–9, <http://dx.doi.org/10.1016/j.protcy.2014.09.007>.
  26. Alex Krizhevsky, Learning Multiple Layers of Features from Tiny Images, Science Department, University of Toronto, 2009, pp. 1–60, URL <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>.
  27. John Duchi, Elad Hazan, Yoram Singer, Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.* 12 (2011) 2121–2159, <http://dx.doi.org/10.1109/CDC.2012.6426698>, URL <http://jmlr.org/papers/v12/duchi11a.html>.
  28. Rich Caruana, Alexandru Niculescu-Mizil, An empirical comparison of supervised learning algorithms, in: *Proceedings of the 23rd International Conference on Machine Learning*, vol. C(1), 2006, pp. 161–168, URL <http://portal.acm.org/citation.cfm?doid=1143844.1143865>.
  29. Rich Caruana, Nikos Karampatziakis, Ainur Yessenalina, An empirical evaluation of supervised learning in high dimensions, in: *Proceedings of the 25th International Conference*

- on Machine Learning, 2008, pp. 96–103, URL <http://portal.acm.org/citation.cfm?doid=1390156.1390169>.
30. Chintan Parmar, Patrick Grossmann, Johan Bussink, Philippe Lambin, Hugo J.W.L. Aerts, Machine learning methods for quantitative radiomic biomarkers (Supplement), *Sci. Rep.* 5 (2015) 13087, <http://dx.doi.org/10.1038/srep13087>.
  31. Prateek Prasanna, Pallavi Tiwari, Anant Madabhushi, Co-occurrence of local anisotropic gradient orientations (CoLIAGe): distinguishing tumor confounders and molecular subtypes on MRI, *Lect. Notes Comput. Sci.* 8675 (Part 3) (2014) 73–80, [http://dx.doi.org/10.1007/978-3-319-10443-0\\_10](http://dx.doi.org/10.1007/978-3-319-10443-0_10).
  32. Dennis Gabor, Theory of communication. Part 1: The analysis of information, *J. Inst. Electr. Eng., Part 3, Radio Commun. Eng.* 93 (26) (1946) 429–441, <http://dx.doi.org/10.1049/ji-3-2.1946.0074>, URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5298517](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5298517).
  33. Robert Haralick, K. Shanmugan, I. Dinstein, Textural features for image classification, URL <http://dceanalysis.bigr.nl/Haralick73-Texturalfeaturesforimageclassification.pdf>, 1973.
  34. K.I. Laws, Textured Image Segmentation, PhD thesis, Univ. of Southern California, 1980.
  35. George Azzopardi, Nicolai Petkov, A CORF computational model of a simple cell that relies on LGN input outperforms the Gabor function model, *Biol. Cybern.* 106 (3) (2012) 177–189, <http://dx.doi.org/10.1007/s00422-012-0486-6>.
  36. L. Wang, Texture unit, textural spectrum and texture analysis, *IEEE Trans. Geosci. Remote Sens.* 28 (4) (1990) 509, <http://dx.doi.org/10.1109/TGRS.1990.572934>.
  37. Arnout C. Ruifrok, Ruth L. Katz, Dennis A. Johnston, Comparison of quantification of histochemical staining by hue–saturation–intensity (HSI) transformation and color-deconvolution, *Appl. Immunohistochem. Mol. Morphol.* 11 (1) (2003) 85–91, <http://dx.doi.org/10.1097/00129039-200303000-00014>.
  38. Andrew Janowczyk, Scott Doyle, Hannah Gilmore, Anant Madabhushi, A resolution adaptive deep hierarchical (RADHicaL) learning scheme applied to nuclear segmentation of digital pathology images, *Comput. Methods Biomech. Biomed. Eng.* 1163 (April) (2016) 1–7, <http://dx.doi.org/10.1080/21681163.2016.1141063>, URL <http://www.tandfonline.com/doi/full/10.1080/21681163.2016.1141063>.
  39. D.M.W. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation, *J. Mach. Learn. Technol.* 2 (1) (2011) 37–63.
  40. Marie-Pierre Dubuisson, Anil K. Jain, A modified Hausdorff distance for object matching, *Proc. Int. Conf. Pattern Recognit.* (1994) 566–568, <http://dx.doi.org/10.1109/ICPR.1994.576361>.