NOISE-TOLERANT DEEP LEARNING FOR HISTOPATHOLOGICAL IMAGE SEGMENTATION

A Thesis

by

WEIZHI LI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Jim Ji
Committee Members,	Xiaoning Qian
	Yoonsuck Choe
	Nick Duffield
Head of Department,	Miroslav Begovic

December 2017

Major Subject: Electrical Engineering

ABSTRACT

Devising a promising algorithm based on the handcrafted feature for histological images (histo-images) segmentation is demanding due to the complexity of histo-images. Deep network models have achieved dominating performances for being capable of capturing high-level features. However, a major hurdle hindering the application of deep learning in histo-image segmentation is to obtain ground-truth for training. Taking the segmentations from simple off-the-shelf algorithm as training data will be a new way to solve the problem of scarce of clean ground truth. These off-the-shelf segmentations are considered to be noisy data requiring new learning scheme for deep learning segmentation. The majority of noisy label deep learning research is for image classification. Motivated by the lack of research in noisy label deep learning for image segmentation and the realistic necessity in histo-image segmentation, we study whether and how integrating imperfect or noisy ground-truth from simple off-the-shelf segmentation algorithms may help achieve better performance so that the deep learning can be applied in histo-image segmentation with the manageable effort.

Two noise-tolerant deep learning architectures are proposed in this thesis based on the Noisy at Random (NAR) Model and the Noisy Not at Random (NNAR) Model with the largest difference that NNAR based architecture assume the label noise is dependent on features of the image. Unlike the most works only researching in the single type of noisy data, we also study how to integrate multiple types of noisy data into one specific model which has extensive application in the circumstance when segmentations from multiple off-the-shelf algorithms are available. The implementation of the NNAR based architecture demonstrates its effectiveness and superiority over off-the-shelf and other existing deep-learning-based image segmentation algorithms.

DEDICATION

This thesis work is dedicated to my family, friends and instructors for their support and help throughout the study.

ACKNOWLEDGMENTS

I would like to thank my parents who work diligently to financially support my study so that I have the opportunity to accept high-quality education in Texas A&M University. In past two years during my master study, I was unable to stay with them due to studying abroad. It was their understanding and encouragement that give me the most comfort in the road of research and inspire me to conquer any difficulties.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This thesis work was mainly completed in conjunction with Dr. Jim Ji and Dr. Xiaoning Qian. Duo to the academical freedom in

The data analyzed for Chapter X was provided by Professor XXXX. The analyses depicted in Chapter X were conducted in part by Rebecca Jones of the Department of Biostatistics and were published in (year) in an article listed in the Biographical Sketch.

All other work conducted for the thesis (or) dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University and a dissertation research fellowship from XXX Foundation.

NOMENCLATURE

OGAPS	Office of Graduate and Professional Studies at Texas A&M University
B/CS	Bryan and College Station
TAMU	Texas A&M University
SDCC	San Diego Comic-Con
EVIL	Every Villain is Lemons
EPCC	Educator Preparation and Certification Center at Texas A&M University - San Antonio
FFT	Fast Fourier Transform
ARIMA	Autoregressive Integrated Moving Average
SSD	Solid State Drive
HDD	Hard Disk Drive
O&M	Eller Oceanography and Meteorology Building
DOS	Disk Operating System
HDMI	High Definition Multimedia Interface
L^1	Space of absolutely Lebesgue integrable functions; i.e., $\int \lvert f \rvert < \infty$
L^2	Space of square-Lebesgue-integrable functions, i.e., $\int \lvert f \rvert^2 < \infty$
$PC(S)$	Space of piecewise-continuous functions on S
GNU	GNU is Not Unix
GUI	Graphical User Interface

PID	Principal Integral Domain
MIP	Mixed Integer Program
LP	Linear Program

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION AND LITERATURE REVIEW

Histopathological image (histo-image) is considered to be the "gold standard" in clinical diagnosis for the reason that the histo-image includes comprehensive information of the disease by retaining most of the intricate structure in preparation. Given great advances on the database of digitized histological tissue, the histo-image has not only been used for the diagnosis of disease but also for the biomarker discovery which help detecting the risk of potential disease. To relieve doctors or clinicians from time-consuming work on the analysis of histo-images, researchers expect to design specific computer algorithm for the automatic histo-image analysis which is challengeable due to the high complexity of histo-images. Being specific in the histo-images affected by Duchenne muscular dystrophy (DMD), clinician are always interested in the proportion of fibrosis (stained to blue), muscle (stained to red) and the rest stuff (mostly stained to white) to diagnose the seriousness of disease. Some of these histo-images are shown in Figure 1.1. The inhomogeneity and variability of color spectrum distributed over the DMD affected histo-images are the main obstacle to develop the automatic algorithm for histo-image analysis.

Numerous traditional machine learning based methods has been proposed since the early 1990s [1]: for the histo-image analysis but most of these methods require a huge effort on the extraction of handcrafted features regarding to the specific properties of the

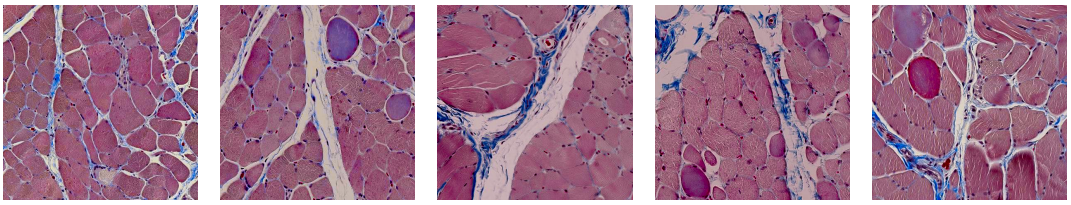


Figure 1.1: Histo-images affected by DMD

context within histo-images to achieve the promising results. With the recent dominant success of deep learning in computer vision field, researchers begin to apply or construct the specific deep models for the histo-image analysis. Though the deep learning based methods often can achieve better results than traditional machine learning based methods in histo-image analysis, the challenge of the scarce of ground truth hinders the widespread application of deep learning in the histo-image analysis, especially for histo-image segmentation where the fine resolution makes manual annotation by experts extremely time-consuming.

The ultimate goal of our research in this thesis is to apply the deep learning in histo-image segmentation with manageble efforts meanwhile generating promising results in the circumstance when ground truth is unavailable. We take the segmentations from the simple off-the-shelf algorithms as the ground truth and expect to recover clean segmentation from these noisy ground truth by deep learning. Such trick is similar to crowd-sourcing but the difference is that obtaining the data from the off-the-shelf algorithm is less resource-demanding than those for the crowd-sourcing. The essence of our research is actually the noisy label learning and many related works have been proposed recently. However, most of these works in computer vision are mainly for the image classification and also do not consider a model for multiple types of noisy label. In this thesis, we propose the noise-tolerant network (NTN) and adaptive noise-tolerant network (ANTN) for noisy label deep learning in image segmentation. The ANTN outperforms the competing state of the art with two novel aspects: (1) multiple noisy labels can be integrated into one deep learning model; (2) noisy segmentation modelling, including probabilistic parameters, is adaptive, depending on the given testing image appearance. Implementation of our model on both the synthetic data and histo-images demonstrates its effectiveness and superiority over off-the-shelf and other existing deep-learning-based image segmentation algorithms.

1.1 Histopathological Image Analysis

1.1.1 Machine Learning Method

Most of the traditional machine learning methods for histo-image analysis require designing handcrafted features based on the complete domain knowledge to generate promising results. By using linear discriminant analysis (LDA) and the Forward/Backward Search methods, Petushi *et al*. [2] first select the distribution of dispersed chromatin cell nuclei and the distribution of tubular cross sections as the highly correlated features with the breast cancer and then graded the breast cancer with the these features by quadratic classifier. Sertal *et al*. [3] extract the intermediate features from the cytological components in histo-images and combined them with the low level color texture feature for the follicular lymphoma grading. They reduce the dimensionality of the feature space by implementing principal analysis (PCA) and the linear discriminant analysis (LDA) then classify the follicular lymphoma by the Bayesian classifier. Nguyen *et al*. [4] aggregate 19 features based on the nuclei, cytoplasm, and lumen shape to detect the prostate cancer using the support vector machine (SVM). The features in the above works are mainly comprising of first order statistical information such as mean, standard deviation and median generated based on the relative cytological characteristics. There are works incorporating higher order statistical features for histo-image analysis. For example, Demir *et al*. [5] innovatively represent the low magnification tissue histo-image of the breast cancer by constructing a augmented cell graph in which node weight represent the size of cell cluster and the edge weight is defined as the Euclidean distance between cell clusters. With the the augmented cell graph, higher statistical order features are constructed with the set of the eigenvalues generated by decomposing the graph.

1.1.2 Deep Learning Method

The traditional machine learning method takes a huge effort in the extraction of features which usually requires being familiar with the cytological characteristics within histo-images to generate promising results. In contrast to that, the recent prosperous deep learning function based on the convolutional neural network (CNN) to extracting high-level features and gain enormous successes in computer vision. Due to the advantage of more effective feature extraction process with less demanding requirement such as domain knowledge, many researchers turn to deep learning method for automatic histo-image analysis. Cireşan *et al*. [6] implement a feed-forward deep neural network taking the patches of histo-images as input to detect the mitosis. The patch input is a square window of RGB values from the histo-image being mapped to the class of the central pixel as mitosis or non-mitosis. Chen *et al*. [7] revise the u-net [8] to the deep contour-aware network (DCAN) for gland segmentation. The DCAN first combined multi-level contextual information with auxiliary supervision in each of two branches for object segmentation and contour segmentation, and then fuse results from the two branches to generate more detail aware gland segmentations. Unlike the [6, 7] which train the convolutional filters from random initialization or use the pretrained parameters from other networks, Cruz-Roa *et al*. [9] first apply auto-encoding technique to learn the feature representation of the histo-image patches and then take these learned feature weight as the convolutional filters to construct CNN. They also incorporate a additional layer for visualization of the feature pattern about the cancer region to enhance the interpretability of CNN.

1.2 Noisy Label Learning

Despite the tremendous potential of deep networks, the supervised nature hinders their wider application in histo-image analysis, especially for histo-image segmentation since the manual pixel-wise annotation of high-resolution histo-images is time-consuming and

labor-intensive. To solve the problem of scarce of data, Albarqouni *et al*. [10] implement the crowdsourcing technique to obtain a large scale annotation from non-experts, and incorporated the process of crowdsourcing to CNN for mitosis detection. The novelty of their deep model for crowdsourcing data is to have an additional aggregation layer which aggregate the the ground-truth from crowdvotes matrix to refine the model based on the sensitivity and specificity of each annotator. Though crowdsourcing provides a large scale of annotated data, the quality of data is not guarantee and the process of recruiting non-experts is still resource demanding in practice. **Instead of relying on crowd-sourcing, we resort to existing image segmentation algorithms to obtain noisy segmentation labels and design a new deep learning model to recover clean segmentations by integrating these noisy labels from different segmentation algorithms.**

There are several existing noisy label deep learning models [11, 12, 13, 14, 15, 16] that address the problem when the labels of the training datasets are "noisy". As discussed in [17], most of these methods focus on image classification or patch-labeling applications.

Hinton *et al*. [11] have pioneered to use the deep network to incorporate the label-flip noise in aerial image labeling. They assume the label-flip noise is only dependent on the true label and adopt an EM algorithm to train network model parameters iteratively, considering the true labels as latent variables. Benoît *et al*. [17] point out it is more realistic that mislabeling is dependent on input features and Xiao *et al*. [12] take such assumption into consideration and integrate three types of label noise transition probabilities given the same true label for clothing classification. Instead of modeling true labels as latent variables in [11, 12], Veit *et al*. [13] have introduced a multi-task label cleaning architecture for image classification, in which an image classifier is supervised by the output of the label cleaning network trained using the mixture of clean and noisy labels, which is effective in learning large-scale noisy data in conjunction with a small subset of clean data.

All the aforementioned methods [11, 12, 13] require a small clean dataset to assist

model inference. Sukhbaatar *et al*. [14] propose a noisy label image classification model that is capable of learning network parameters from noisy labels solely by diffusing label noise transition probability matrix from the initial identity matrix with a weight decay trick; but such model inference can be unstable. Reed *et al*. [15] introduce "prediction consistency" in training a feed-forward autoencoder with noisy data, requiring that the same label prediction should be made given similar image features. Similar to the idea of avoid overfitting, Kakar *et al*. [16] add a regularization term on the coefficients of hidden units during training to obtain stable results.

1.3 Thesis Contribution

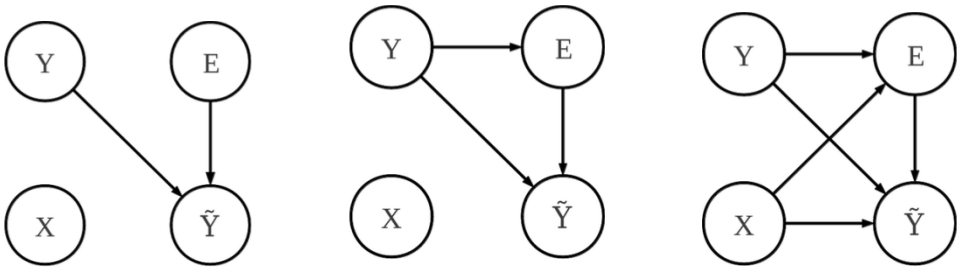
Most of the existing noisy label deep network models [11, 12, 13, 14, 15, 16] are on image classification or patch-labeling. Besides, there is still no existing method to flexibly integrate multiple types of noisy data in the literature to the best of our knowledge. In this thesis, we propose two deep architectures for noise learning problem in image segmentation. Implementation of our model on both the synthetic data and histo-images demonstrates its effectiveness and superiority over off-the-shelf and other existing deep-learning-based image segmentation algorithms.

2. Model Formulation of Noisy Label Learning for Image Segmentation

Being different from the feature noise which affects the observed values of the feature, label noise pollutes the observed labels of instances by altering the label class. The works [18, 19, 17] state the label noise poses more harm than feature noisy in learning problem for the two reasons: 1) multiple features determine how the instance is classified in learning whereas only one label is assigned to one instance and 2) the importance of features is varied whereas labels assigned to instances always affect learning to a large extent. There are a lot of similarities between dealing with label noise and outliers detection. Actually, mislabelled instances are considered to be outliers if the mislabelling occur in the vicinity of instance in sample space with low probability and the instances of these outliers often looks anomalous regarding to the class corresponding to the incorrect label. It it such similarity makes many noisy label learning works very close to outlier detection [17]. In additional to the outlier alike label noise, there are mislabelling occurring in specific condition with high probability (*e.g.*The boundary region where all the classes are equiprobable always comes with the mislabelling error) and the instances of these mislabel does not look anomalous as outliers. In this thesis, we mainly research on such label noise coming from the off-the-shelf algorithms or the appearance of images for histo-image segmentation. In this chapter, we will first discuss the categorization of label noise, and then move to the u-net [8] which is the fundamental architecture we implement in our experiment for noisy label deep learning. Following that, two noise-tolerant deep learning models for image segmentation will be discussed.

2.1 Categorization of Label Noise

According to [17], three types of label noise models exist: the noisy completely at random model (NCAR), noisy at random model (NAR) and the noisy not at random (NNAR).



(a) NCAR (b) NAR (c) NNAR

Figure 2.1: Graphical probabilistic model of label noise process

These graphical probabilistic models of label noise are shown in Figure 2.1. We represent the features of data by X, true class by Y and the observed class by $\tilde{Y}$. To model the label noise process, the binary variable E indicating whether the instance is mislabelled or not is also introduced. We now explain these label noise models following the [17].

1) **Noisy Completely at Random (NCAR) Model:** The NCAR model depicting the label noise independent of features of data X and true class Y is shown in Figure 2.1(a). The binary indicator E for mislabelling indicates the true class Y is altered to other class with a certain probability as observed $\tilde{Y}$ if E is one and vice versa. In the case of binary cases, it is certain to be symmetric for NCAR noise in both classes which means both classes are mislabelled with the same percentage in the process of data generation. Obviously these data carry no useful information with a probability $Pr(E = 1)$ to be 0.5. In contrast, the true label is altered uniform randomly in the case of multiclass when $E = 1$. Such NCAR noise at multiclass case is analogous to flipping a biased coin first to decide whether mislabelling occurs or not and then a $|\mathcal{Y}| - 1$ faces fair dice label is tossed to decided which class the true class is altered to be if the mislabelling occurs. The NCAR model is uncommon in real practice for its oversimple assumption of label noise process.

2) **Noisy at Random (NAR) Model:** The NAR model has a broader application than NCAR for it considering the influence of true class on label noise. The probabilistic model

of NAR is shown in Figure 2.1(b) and the arrow from the Y to E illustrates probability of mislabelling is affected by the true class. Having assumed a direct effect between true class Y and mislabelling indicator E, the NAR is capable of modelling asymmetric label noise, *i.e.*certain classes are more likely to be altered than other classes. For example, the existed objects such as buildings, roads or alleys disappear in the aerial images due to the incompleteness of the maps and this is called omission noise in aerial image learning [11]. Such omission noise in aerial image learning can actually be modelled as the NAR noise and [11] has proposed using the neural network to label aerial images from noisy data based on the NAR model. Another case where NAR noise occurs is control subjects in medical case-control studies. For the reason that the test used to label control studies may be too invasive or expensive, the control studies are replaced by suboptimal diagnostic test so that the control subjects is prone to be mislabelling [20].

3) **Noisy Not at Random (NNAR) Model:** The NNAR model shown in Figure 2.1(c) considers a type of more complete and general label noise process where the mislabelling is determined by both features of data and true class, *i.e.*, incorrect labelling are more likely to occur for certain class and in certain regions of the sample space for X. For example, the instances distributed in the classification boundary or low density region of sample space are prone to be mislabelling and such label noise are considered to be NNAR. This situation occurs in real practice such as speech recognition challengeable for phonetic similarity between the recognized words and correct one [21]. Therefore, the features of words are supposed to be involved in the impact on the mislabelling. In addition to speech recognition, another domain NNAR model applies is image classification/segmentation. Xiao *et al.* [12] modelled the relationship between noisy data and clean based on NNAR model and proposed an end to end convolutional neural network (CNN) for image classification of clothing applied in the scenario when the large scale of well-labelled data is hard to obtain. Our thesis mainly research on the case of histo-image segmentation where

the labelling of data is not from clinical expert but the off-the-shelf algorithms thus the assumption of NNAR model is reasonably appropriate.

2.2 End to End Convolutional Neural Network (CNN) : U-net

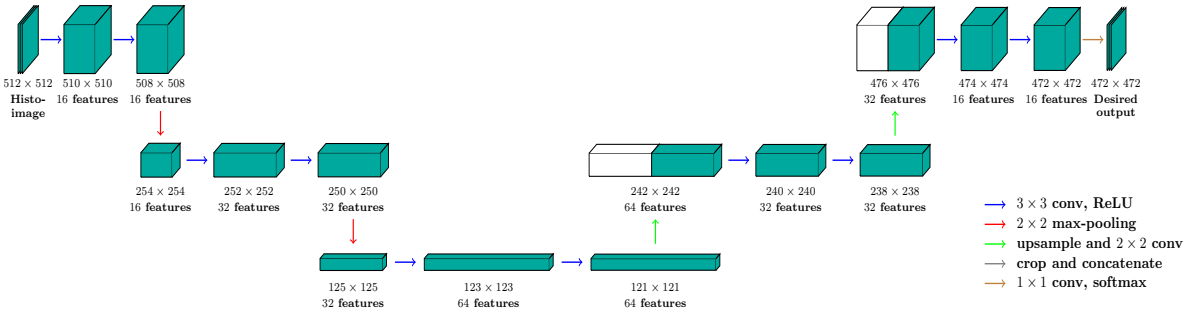


Figure 2.2: Schematic illustrations of u-net

U-net is an image-to-image deep learning framework shown to be effective in biomedical image segmentation [8, 7]. Unlike CNN-based deep learning with only contracting layers for image classification and annotation, u-net adds an expanding module to enable pixel-wise labeling (Figure 2.2). In our implementation, 3×3 multi-scale convolutional filters followed by rectified linear units (ReLU) are applied in three levels of the contracting layers. Between every two layers, 2×2 max-pooling is applied to derive more abstract non-linear features. For expanding layers, the derived feature maps are up-sampled twice and concatenated with the convolutional feature maps at the corresponding scale of the contracting layers. Another two convolutional layers and a final softmax output layer are then applied to derive the final pixel-wise labeling for histo-image segmentation. Such a u-net implementation has a 15-layer network architecture as shown in Figure 2.2.

U-net is a supervised deep learning framework, requiring accurate segmentation labels for training. However, for the histo-image segmentation, usually the manually annotated

histo-images are not available. In order to enable u-net histo-image segmentation, one work-around is to apply traditional image segmentation algorithms, such as K-Means, and use the resulting segmentations with reasonably high accuracy to train u-net. However, there is no guarantee that these segmentation results have good enough quality, especially due to large histo-image appearance variation.

2.3 Noise-Tolerant Network (NTN)

To alleviate the requirement of accurately segmented histo-images for u-net training, we propose to adjust the original u-net to be noise-tolerant following the NNAR model so that the performance will be robust to potentially noisy training segmentations. Given T training images $X = \{X_1, \ldots, X_T\}$, we can construct the probabilistic relationship between hidden clean segmentation Y and the corresponding noisy segmentation $\hat{Y}$ as

$$\begin{aligned} Pr(\hat{y}_n = j|X) &= \sum_{i=1}^{3} Pr(\hat{y}_n = j|y_n) Pr(y_n = i|X) \\ &= \sum_{i=1}^{3} q_{ij} Pr(y_n = i|X) \end{aligned} \tag{2.1}$$

where y_n is the pixel label indexed by n, i and j are label, and N is the total pixel. Following (2.1), the negative log likelihood L can be constructed as

$$L = -\frac{1}{N} \sum_{n=1}^{N} log[\sum_{i=1}^{3} q_{ij} Pr(y_n = i|X)] \tag{2.2}$$

The framework of such noise-tolerant network (NTN) is shown in Figure 2.3. The main difference of the NTN from the original u-net is an additional noise-tolerant layer incorporating parameters of transition probability q_{ij}'s after the clean label prediction network. The additional noise-tolerant layer is shown in Figure 2.4. The parameters of the

added noise-tolerant layer can be represented by a 3×3 transition matrix $Q = \left(q_{ij}\right)_{3\times 3}$ with the constraints: $0 \leq q_{ij} \leq 1$ and $\sum_j q_{ij} = 1, \forall i$. The NAR based deep learning model is motivated by the "label flip noise model" in a recent noise-tolerant AlexNet-based image classification framework [14] that addresses a similar noisy label problem. The difference is that our NTN is for pixel-wise labeling in histo-image segmentation but the method in [14] is for the whole image classification.

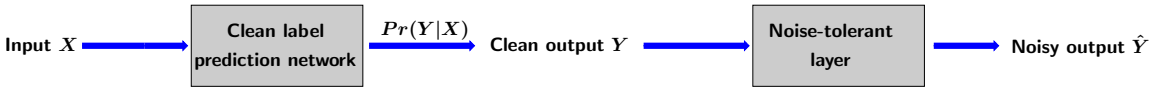


Figure 2.3: Schematic illustrations of noise-tolerant u-net

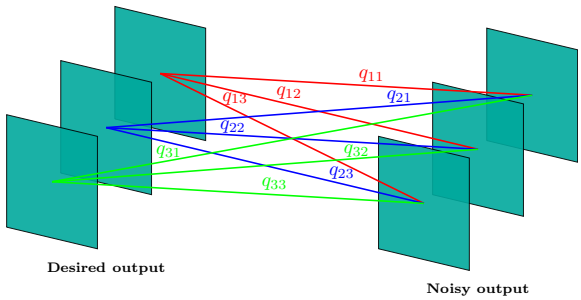


Figure 2.4: Illustration of the "noise-tolerant" layer

By the total probability theorem, it is clear that (2.2) is equivalent to the maximum likelihood estimates of involved parameters in the modified u-net with noisy segmentation $\hat{Y}$ as $L = -\frac{1}{N}\sum_{n=1}^{N} log[Pr(\hat{y}_n|X)]$. The training of the other layers simply follows the back-propagation procedure for the original u-net. More importantly, we can rewrite

$$L = -\frac{1}{N}\sum_{i=1}^{3}\sum_{n\in\mathcal{S}_i} log\big[Pr^i(\hat{y}_n = j|X;Q)\big],$$

where $\mathcal{S}_i$ is the set of pixels that have the true label i, and $Pr^i(\hat{y}_n = j|X;Q)$ denotes the full model prediction probability for pixel n in $\mathcal{S}_i$. Asymptotically when $N \to \infty$, $L \to -\sum_{i=1}^{3}\sum_{j=1}^{3} q_{ij}^* log[Pr^i(\hat{y} = j|X;Q)] \geq -\sum_{i=1}^{3}\sum_{j=1}^{3} q_{ij}^* log(q_{ij}^*)$, achieving the minimum when $Pr^i(\hat{y} = j|X;Q) \to q_{ij}^*$ which is actual flip transition probability. Denote the confusion matrices for clean and noisy segmentations by $C_s = \left(c_{ij}^s\right)$ and $C_r = \left(c_{ij}^r\right)$ respectively, where $c_{ij}^s = \frac{1}{|\mathcal{S}_i|}\sum_{n\in\mathcal{S}_i} Pr^i(y_n = j|X)$ and $c_{ij}^r = \frac{1}{|\mathcal{S}_i|}\sum_{n\in\mathcal{S}_i} Pr^i(\hat{y}_n = j|X;Q)$. It is clear $C_r = C_sQ$. If we know the actual label flip transition matrix $Q = Q^*$, minimizing L will asymptotically force $c_{ij}^r = Pr^i(\hat{y} = j|X;Q) \to q_{ij}^*$ hence $C_r = C_sQ^* \to Q^*$ forcing C_s converging to identity. Therefore, training the noise-tolerant u-net using noisy segmentations with actual transition matrix Q^* directly forces the clean label network to predict the true labels. In practice, minimizing L does not guarantee Q converging to Q^* [14]. In order to derive well-behaved solutions, either a trace norm or a ridge regularization term for Q can be added to the objective function when training the noise-tolerant layer. Based on the reasoning in [14], we use the ridge regularization and fix the corresponding weight decay parameter to 10^{-4} in our experiments.

2.4 Adaptive Noise-Tolerant Network (ANTN)

The NAR based characteristic of the NTN model may limit the performance of noisy label learning for image segmentation since it assumes that the label noise is only dependent on the label. Also, the NTN model does not consider the case of multiple noisy segmentations. To overcome the above shortcomings, we propose an Adaptive Noise-Tolerant Network (ANTN) which is a NNAR based model assuming the label noise is dependent on both appearance of image and the label, and multiple noisy segmentations can be incorporated in training as well. In ANTN, probabilistic dependency between the input image pixels, the ground-truth segmentation, and "noisy" segmentation labels from off-

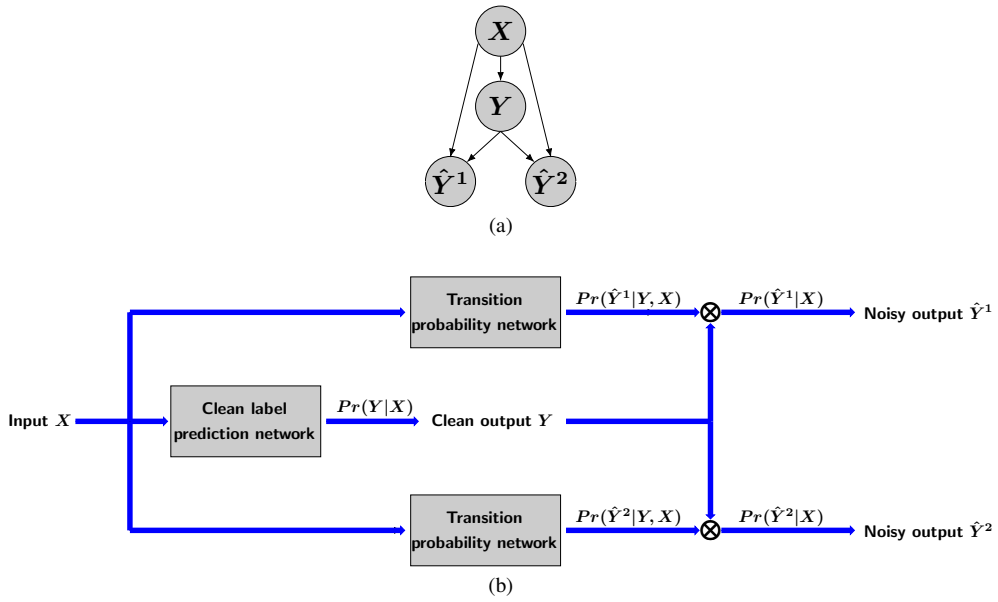


Figure 2.5: (a) Graphical probabilistic model and (b) architecture of the Adaptive Noise-Tolerant Network (ANTN). X represents the input image, Y represents the ground-truth segmentation, $\hat{Y}^1$ and $\hat{Y}^2$ represent noisy segmentations.

the-shelf image segmentation algorithms can be modelled explicitly. By adaptively modeling image-dependent label-flip noise from different segmentation algorithms, ANTN can borrow signal strengths from *multiple* noisy labels to achieve better segmentation results. The graphical probabilistic model and architecture of the network are shown in Figure 2.5.

Given a set of training images $X = \{X_1, X_2, \ldots, X_T\}$, which could be sub-images or patches, we can apply S selected off-the-shelf segmentation algorithms to obtain noisy or imperfect segmentations $\hat{Y}^1 = \{\hat{Y}_1^1, \hat{Y}_2^1, \ldots, \hat{Y}_T^1\}$, $\hat{Y}^2$, ..., $\hat{Y}^S$. To clearly convey the idea, we focus on the settings with $S = 2$ in the thesis. We can model the relationships between input images and noisy segmentations based on the following general probabilistic model:

$$Pr\big(\hat{Y}^1, \hat{Y}^2|X\big) = \sum_{Y \in C^{|\mathcal{I}|}} Pr\big(\hat{Y}^1, \hat{Y}^2, Y|X\big) = \sum_{Y \in C^{|\mathcal{I}|}} Pr\big(\hat{Y}^1|Y, X\big) Pr\big(\hat{Y}^2|Y, X\big) Pr\big(Y|X\big), \tag{2.3}$$

in which C is the total number of label classes for segmentation; Y denotes the clean or perfect segmentations; and $|\mathcal{I}|$ represents the total number of pixels in X indexed by the pixel set $\mathcal{I}$. We note that in [11, 14] and the NTN model, (1) the clean pixel-wise labels indexed by n: y_n's, are conditionally independent given X: $Pr\big(Y|X\big) = \Pi_{n \in \mathcal{I}} Pr(y_n|X)$; and (2) the noisy pixel-wise labels $\hat{y}_n$'s are conditionally independent with X given Y and the pixel-wise label transition probabilities are identical: $Pr\big(\hat{Y}|Y, X\big) = \Pi_{n \in \mathcal{I}} Pr(\hat{y}_n|y_n)$. Hence, the log-likelihood with one set of noisy labels for X can be written as:

$$L = \log Pr\big(\hat{Y}|X\big) = \sum_{n \in \mathcal{I}} \log \Big[\sum_{y_n=1}^{C} Pr\big(\hat{y}_n|y_n\big) Pr\big(y_n|X\big) \Big], \tag{2.4}$$

with which the Noise-Tolerant Network (NTN) in section 3 is proposed to recover clean segmentations.

In the proposed ANTN model (2.2), we relax the second assumption in NTN when integrating multiple types of noisy labels. For different pixels, the transition probability of the noisy label given the clean label will be dependent on X since segmentation results from different algorithms can be dependent on both images and segmentation algorithms. Let $Pr_n(\hat{y}_n^s|y_n, X)$ denote the new transition probabilities, where $s = 1, 2$ for different noisy segmentations. Following the dynamic filter network models in [22], we can rewrite the probabilistic model (2.2):

$$Pr\big(\hat{Y}^1, \hat{Y}^2|X\big) = \prod_{n \in \mathcal{I}} \sum_{y_n=1}^{C} Pr_n(\hat{y}_n^1|y_n, X) Pr_n(\hat{y}_n^2|y_n, X) Pr(y_n|X). \tag{2.5}$$

With this model, we can construct respective deep learning models for all the involved probability distribution functions, including the clean label probability $Pr(Y|X)$, pixel-wise conditional probabilities $Pr_n(\hat{y}_n^1|y_n, X)$ and $Pr_n(\hat{y}_n^2|y_n, X)$, as illustrated by the schematic graphical model for recovering clean labels from two noisy datasets in Figure 2.5(a). We note the symmetry of the proposed deep learning framework, which enables the straightforward generalization when $S > 2$. For each of the three components in Figure 2.5(a), we follow the construction in [8, 11, 22] and the NTN to have the corresponding u-net architectures with the deep network framework shown in Figure 2.5(b). The main difference among these three deep network models are the constraints applied to their outputs of the last layers:

$$\sum_{y_n=1}^{C} Pr_n(y_n|X) = 1, \quad \sum_{\hat{y}_n^1=1}^{C} Pr_n(\hat{y}_n^1|y_n, X) = 1, \quad \sum_{\hat{y}_n^2}^{C} Pr_n(\hat{y}_n^2|y_n, X) = 1, \tag{2.6}$$

which guarantee the legitimacy of the modeled probability distribution functions.

We note that the clean label model $Pr(Y|X)$ has to be combined with the noise transition network models $Pr_n(\hat{y}_n^1|y_n, X)$ and $Pr_n(\hat{y}_n^2|y_n, X)$ for training as we do not observe the ground-truth segmentations. The integration of the three components in Figure 2.5(a) is motivated by the label-flip noise model in the noise-tolerant image classification framework in [14] and the introduced asymmetric Bernoulli noise (ABN) model in [11].

Model Inference

Due to the unobserved clean segmentation labels, training three different components given X and noisy segmentations $\hat{Y}^1$ and $\hat{Y}^2$ is an iterative procedure to maximize the

following three log-likelihood functions based on the model (2.4):

$$\mathcal{L}_s = \frac{1}{N}\sum_{n\in\mathcal{I}} log \sum_{y_n=1}^{C} Pr((\hat{y}_n^s)_{obs}|y_n, X; \theta_s) Pr(y_n|X;\theta_3), \quad s = 1, 2, \tag{2.7}$$

$$\mathcal{L}_3 = \frac{1}{N}\sum_{n\in\mathcal{I}} log \sum_{y_n=1}^{C} Pr((\hat{y}_n^1)_{obs}|y_n, X; \theta_1) Pr((\hat{y}_n^2)_{obs}|y_n, X; \theta_2) Pr(y_n|X;\theta_3) \tag{2.8}$$

where θ_1, θ_2 and θ_3 are the corresponding network parameters of two transition probability networks and the clean label prediction network; $(\hat{y}_n^1)_{obs}$ and $(\hat{y}_n^2)_{obs}$ denote observed noisy labels; and $N = |\mathcal{I}|$. We alternate the order of optimization with respect to θ_1 and θ_2 for minimizing (2.6) and θ_3 for minimizing (2.7). Similar to [11, 14], we consider Y as latent variables and maximize the likelihood functions by the EM algorithm:

E-step: Given deep network paramters $\theta_1^{(t)}$, $\theta_2^{(t)}$ and $\theta_3^{(t)}$ for three component networks at each iteration, the posterior probabilities of the latent segmentation label $Pr(y_n|(\hat{y}_n^s)_{obs}, X)$ and $Pr(y_n|(\hat{y}_n^1)_{obs}, (\hat{y}_n^2)_{obs}, X)$ for the corresponding likelihood functions (2.6) and (2.7) can be updated as follows:

$$Pr^{(t)}(y_n|(\hat{y}_n^s)_{obs}, X) = \frac{Pr((\hat{y}_n^s)_{obs}|y_n, X; \theta_s^{(t)}) Pr(y_n|X;\theta_3^{(t)})}{\sum_{y_n=1}^{C} Pr(\hat{y}_{obs}^s|y_n, X; \theta_s^{(t)}) Pr(y_n|X;\theta_3^{(t)})}, \quad s = 1, 2, \tag{2.9}$$

$$Pr^{(t)}(y_n|(\hat{y}_n^1)_{obs}, (\hat{y}_n^2)_{obs}, X) = \frac{Pr((\hat{y}_n^1)_{obs}|y_n, X; \theta_1^{(t)}) Pr((\hat{y}_n^2)_{obs}|y_n, X; \theta_2^{(t)}) Pr(y_n|X;\theta_3^{(t)})}{\sum_{y_n=1}^{C} Pr((\hat{y}_n^1)_{obs}|y_n, X; \theta_1^{(t)}) Pr((\hat{y}_n^2)_{obs}|y_n, X; \theta_2^{(t)}) Pr(y_n|X;\theta_3^{(t)})}. \tag{2.10}$$

M-step: With the estimated posterior probabilities, we update the corresponding network parameters through optimizing the expected complete likelihood functions. In practice, we cannot guarantee the optimality of M-step updates due to our deep network modeling. We implement gradient descent and backprogation in the corresponding component networks

to update parameters as follows:

$$\nabla\theta_s^{(t+1)} \leftarrow \frac{1}{N}\sum_{n\in\mathcal{I}}\sum_{y_n=1}^{C} Pr^{(t)}(y_n|(\hat{y}_n^s)_{obs}, X)\frac{\partial logPr((\hat{y}_n^s)_{obs}|y_n, X;\theta_s)}{\partial\theta_s}, \quad s=1,2, \tag{2.11}$$

$$\nabla\theta_3^{(t+1)} \leftarrow \frac{1}{N}\sum_{n\in\mathcal{I}}\sum_{y_n=1}^{C} Pr^{(t)}(y_n|(\hat{y}_n^1)_{obs}, (\hat{y}_n^2)_{obs}, X)\frac{\partial logPr(y_n|X;\theta_3)}{\partial\theta_3}. \tag{2.12}$$

For transition probability networks, we only observe one noisy label for each pixel and we can only unambiguously derive $Pr((\hat{y}_n^s)_{obs}|y_n, X;\theta_s)$. For the other transition probabilities, we simply set them to be $[1 - Pr((\hat{y}_n^s)_{obs}|y_n, X;\theta_s)]/(C-1)$.

For the complete procedure of ANTN model inference, we first initialize the clean label prediction network by training with the mixture of noisy datasets, then train each transition probability network with the corresponding noisy labels as described in the EM algorithm. After these two steps, we iteratively train the component networks by alternating the optimization with a fixed number of interval epochs for each of them until convergence.

3. Experiments

We evaluate the effectiveness of ANTN and NTN by comparing them with off-the-shelf and deep-learning image segmentation algorithms on both synthetic and histo-images.

3.1 Datasets

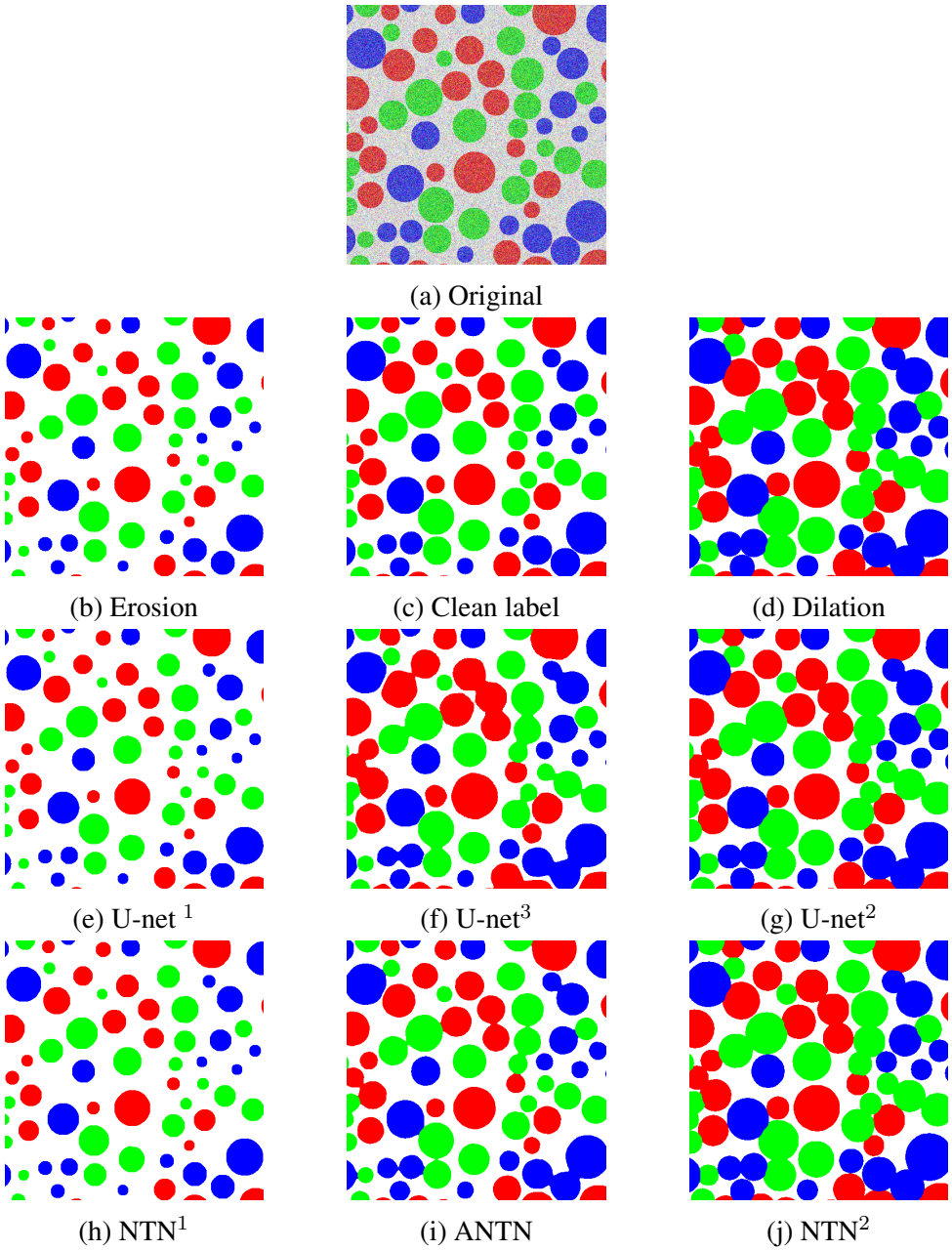
To quantitatively evaluate performance of both ANTN and NTN and compare them with other segmentation algorithms, we first create a synthetic image set with the corresponding simulated noisy segmentations. After the experiment on synthetic data, we then apply the ANTN and NTN to a set of histo-images, obtained from a study of Duchenne Muscular Dystrophy (DMD) disease [23], for performance evaluation.

Synthetic Data:

We generate 135 472×472 synthetic images for quantitative performance evaluation. First, we randomly simulate red, green, and blue circular objects with different radii uniformly distributed from 15 to 40 pixels in each image. Hence, there are four classes required to be segmented: red, green, and blur circular objects as well as white background regions. For each of RGB channels, the corresponding intensities for pixels in each class follow a Gaussian distribution with the mean 200 and standard variation 50. An example of the generated synthetic images and the corresponding ground truth for its object segmentation are shown in Figures 3.1(a) and (c). To further create different types of noisy segmentation labels, we erode and dilate the ground-truth segmentation by a rectangle structural element with the width and length set to 5 pixels, with the generated noisy labels given in Figures 3.1(b) and (d) for the corresponding image example.

Histopathological Images:

We also have obtained 11 samples of ultra-high resolution histo-images for studying DMD [23]. They are split into 472×472 sub-images and preprocessed by a stain normalization method [24]. Some of the preprocessed sub-images are shown in Figure 3.4(a). For these images, we are interested in quantifying the percentage of fibrosis (stained blue) and muscle (stained pink) to estimate the seriousness of the disease [23, 25]. Hence, the segmentation task is to segment fibrosis (blue), muscle (pink), and other tissue types (white). We have applied two simple off-the-shelf segmentation algorithms: K-Means [26] and Otsu thresholding [27] on all the sub-images and we consider the obtained segmentation labels as the noisy segmentation labels in deep-learning methods including ANTN and NTN. The K-Means clustering are implemented based on the Euclidean distances of pixels represented in L*a*b* color space. Three centroids for the corresponding three desired clusters (shown as red, blue, and white regions in Figure 3.4(a)) are randomly initialized for three times. The K-Means clustering solution with the lowest within-cluster distance is considered as the final segmentation for a given split image. Figure 3.4(b) shows K-Means clustering segmentation results. It can be seen that K-Means may provide bad segmentations as illustrated in the last group of image, due to color distribution inhomogeneity. For Otsu's method, we search for the optimal threshold for segmentation based on two histograms: one of the pixel intensity and the other of the intensity ratio between the blue and red channels. The threshold by intensity helps separate both muscle (pink) and fibrosis (blue) from the rest (white) of a given image. The threshold by the blue/red channel ratio separates fibrosis from muscle. As shown in the third column in Figure 3.4(c), the corresponding segmentation results are quite noisy due to the large color intensity variation within the image.



(a) Original

(b) Erosion (c) Clean label (d) Dilation

(e) U-net [1] (f) U-net[3] (g) U-net[2]

(h) NTN[1] (i) ANTN (j) NTN[2]

Figure 3.1: Synthetic image and corresponding segmentations.

3.2 Performance evaluation on synthetic data

For synthetic data, we compare the performance of ANTN and NTN with the popular deep-learning segmentation architecture u-net [8] taking noisy segmentation labels as the

ground truth for training. 35 synthetic images and their corresponding segmentations are used for training. For ANTN, we first initialize the clean label prediction network (a u-net with the architecture illustrated in chapter two) by training with a mixture of two noisy datasets for the first 100 epochs, then train both transition probability networks (two similar u-nets) by the proposed EM-algorithm with the corresponding erosion and dilation noisy segmentations in next 200 epochs. Finally, we iteratively train the whole network setting the alternating interval to be 10 epochs for next 200 epochs. We keep the learning rate at 10^{-4} for the first 450 epochs and 10^{-5} for the last 50 epochs. For competing methods, we directly train the u-net considering either erosion, dilation, or their mixture as the ground-truth segmentation. With erosion and dilation noisy labels, the training procedure converges for 200 epochs. With the mixture of noisy labels, it converges for 100 epochs. For NTN, in addition to training the original u-net layers, we also train the label-flip-noise transition layer with the corresponding noisy labels by weight decay of 10^{-4} to diffuse the label-flip-noise transition probability from identity to approximate the average noise transition probability matrix for 150 more epochs [14]. We do not train NTN with the mixture of noisy labels as it can only take one single type of noisy labels [14]. Training of the u-net with different noisy labels can be considered as the intermediate steps of ANTN and NTN model inference.

We provide the examples of the corresponding segmentation results in Figures 3.1(e)-(j), in which u-net^1, u-net^2, and u-net^3 represent the u-nets trained with the corresponding erosion, dilation, and mixture of noisy segmentations; NTN^1 and NTN^2 represent the NTNs trained with the corresponding erosion and dilation noisy segmentions. It is clearly that the u-net or NTN often can not correctly segment the corresponding objects without appropriate modeling of segmentation noise with erosion and dilation bias. In Figures 3.1(i), it is clear that our ANTN performs the best due to the adaptive integration of label-flip-noise transitions. In addition, the performance improvement may also come from

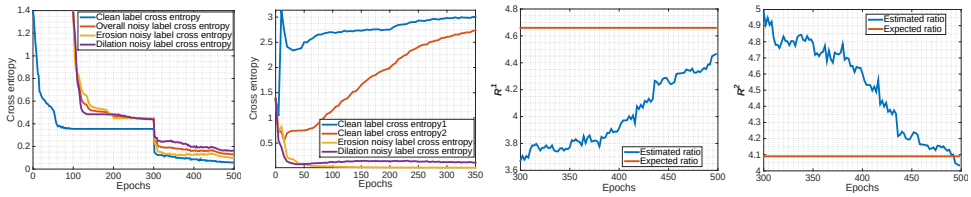


Figure 3.2: (a) Cross entropy evaluation for the u-net^3 and ANTN. (b) Cross entropy evaluation for the u-net^1,u-net^2 and NTNs. (c) Estimated clean-label ratio for erosion dataset. (d) Estimated clean-label ratio for dilation dataset

the integration of multiple types of noisy labels with the capability of borrowing signal strengths. We further quantitatively evaluate segmentation accuracy by the synthetic test dataset of 100 images and the result is shown in Table 3.1, clearly showing that ANTN achieves the best performance.

Method	U-Net^1	U-Net^2	U-Net^3	NTN^1	NTN^2	ANTN
Accuracy	81.63%	83.71%	93.38%	82%	82.49%	**97.71%**

Table 3.1: Accuracy comparison of three networks.

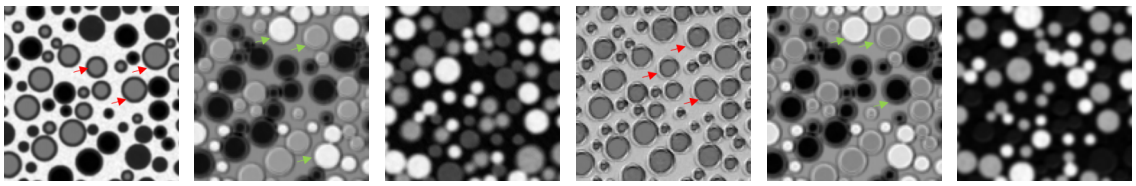
In order to show the convergence of our training procedure for ANTN and NTN, we analyze the trends of the cross entropy between the intermediate segmentation labels during training and the clean ground-truth labels, as well as the noisy labels taken for training. From Figure 3.2(a), we observe that the training of the clean label network in ANTN converges around 100 epochs with the clean-label cross entropy reaching the plateau. Note that the intermediate results at this point is also the final results of u-net^3 training with the mixture of noisy labels. After that, we implement EM algorithm to train two noise transition probability networks. Clearly, the change of the noisy-label cross entropy indicates that the training of two transition probability networks converges in the next 200 epochs.

During the next iterative training procedure, we observe the corresponding cross entropy values drop drastically and then continuously decrease till convergence. Figure 3.2(b) shows the corresponding cross entropy changes during u-net as well as NTN training with either erosion or dilation noisy datasets. The training for u-net stops at 200 epochs which also serves the initialization of NTN training before the noise transition layer training. We can see that the clean-label cross entropy diverges gradually though the noisy-label cross entropy decreases till convergence. This is because no component in u-net models potential segmentation noise.

To further validate the convergence and effectiveness of ANTN, we compare the ratio R of the estimated clean labels to the corresponding sth type of noisy labels during training with the actual ratio of clean labels to noisy labels for the corresponding erosion or dilation training outputs, as shown in Figures 3.2(c) and (d):

$$R^s = \frac{\sum_{n\in\mathcal{I}} I(\arg\max_u Pr(y_n = u|(\hat{y}_n^1)_{obs}, (\hat{y}_n^2)_{obs}, X) = (\hat{y}_n^s)_{obs})}{\sum_{n\in\mathcal{I}} I(\arg\max_u Pr(y_n = u|(\hat{y}_n^1)_{obs}, (\hat{y}_n^2)_{obs}, X) \neq (\hat{y}_n^s)_{obs})}, \quad s = 1, 2. \quad (3.1)$$

From Figures 3.2(c) and (d), the estimated ratios indeed approach the actual ratios in the training data with the corresponding trend indicating the learned ANTN models the noise transitions better and better during the iterative training stage.



(a) u-net^1 (b) u-net^2 (c) u-net^3 (d) NTN^1 (e) NTN^2 (f) ANTN

Figure 3.3: Representative learned feature maps by different networks.

We also show the representative feature maps generated by different networks in Figure 3.3. The feature maps generated by u-net[1], u-net[2], NTN[1] and NTN[2] are inhomogeneous and less confident in the surrounding regions of circular objects and these regions are pointed out by red arrow and green arrow respectively. It reveals that the erosion noise and dilation noise cause the u-net[1], u-net[2], NTN[1] and NTN[2] being ambiguous in the eroded region (pointed by red arrow) and dilated region (pointed by green arrow) during learning. On the contrary, both the ANTN and u-net[3] have homogeneous and confident boundary surrounding regions of circular objects showing that ANTN and u-net[3] are less affected by dilation noise or erosion noise. This is consistent with the fact that ANTN and u-net[3] perform better than other methods in synthetic data experiment.

Finally, we check the noisy transition matrices learned by NTN and the average transition matrices for ANTN, compared to the expected noisy transition matrices obtained by clean and noisy training data. We emphasize that the noisy transition matrix in ANTN is pixel-wise and dependent on image features, we compute the average transition matrices by simply averaging pixel-wise transition probabilities across training images. Clearly, ANTN can better approximate the noise transition by visual comparison in Figure 3.4.

3.3 Performance evaluation on histopathological images

Method	1	2	3	4	5	6	7	8	9	10	11
K-Means	0.3501	0.6183	0.2594	0.3432	0.2748	0.2177	0.6241	0.4196	0.4211	0.5025	0.2335
Otsu	0.3158	0.4928	0.3050	0.3529	0.3129	0.2558	0.5480	0.3653	0.4219	0.4995	0.2502
U-net[1]	0.2854	0.5123	0.2429	0.3254	0.2444	0.2271	0.4847	0.3506	0.3827	0.4742	0.1603
U-net[2]	0.2940	0.5058	0.2580	0.3233	0.2504	0.2308	0.5018	0.3505	0.3932	0.5152	0.1959
U-net[3]	0.3150	0.2917	0.2831	0.2810	0.2578	0.2467	0.2898	0.3424	0.2709	0.3036	0.7437
NTN[1]	0.2848	0.4978	0.2484	0.3239	0.2470	0.2280	0.4955	0.3520	0.3857	0.4823	0.1594
NTN[2]	0.3128	0.5066	0.2861	0.3330	0.2737	0.2473	0.5225	0.3584	0.4213	0.5500	0.2281
ANTN	0.2751	0.2790	0.2676	0.2663	0.2472	0.2332	0.2788	0.3113	0.2670	0.2966	0.2311

Table 3.2: Performance comparison of different methods on 11 original histo-images

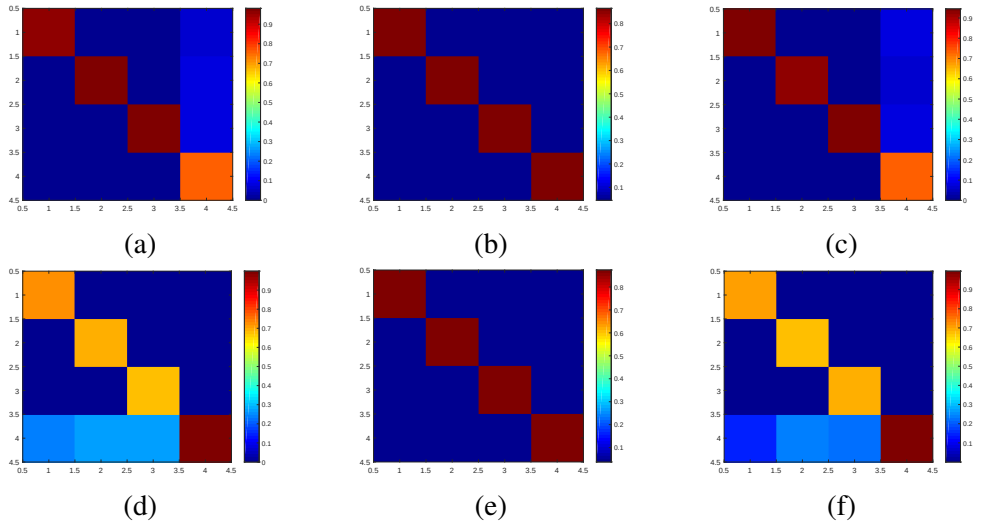
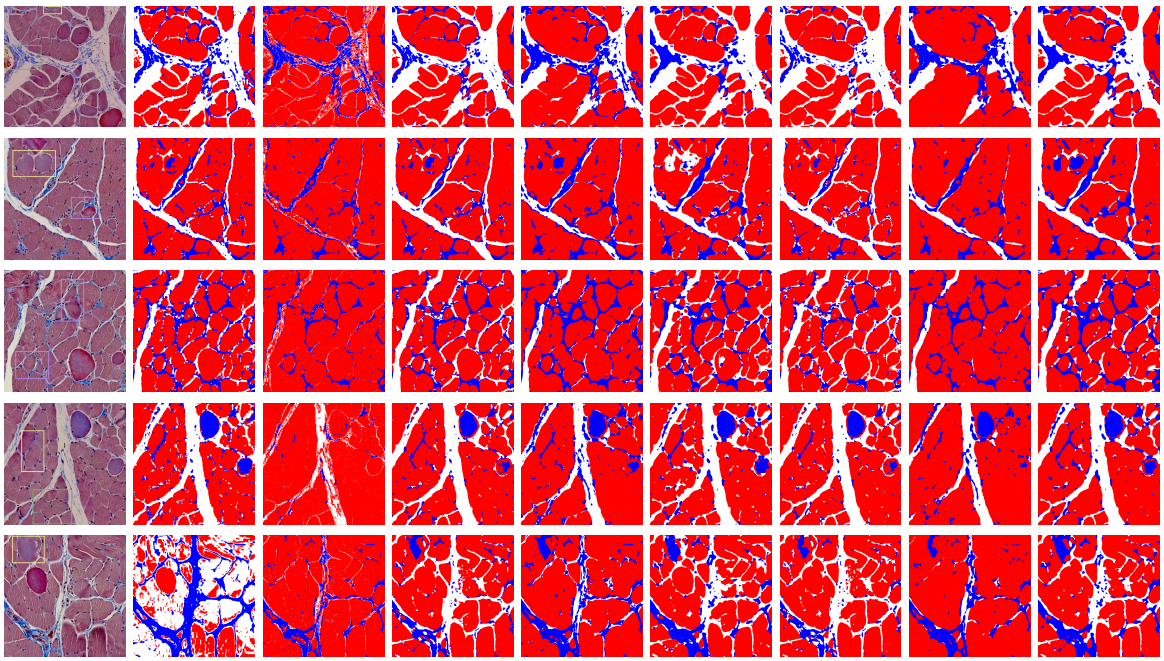


Figure 3.4: Comparison between the expected transition matrices (a), learned transition matrices by NTN (b), and learned average transition matrices by ANTN (c). The first and second rows represent the learned matrices for the erosion and dilation labels respectively.

With the experiments in synthetic data, we further implement ANTN and NTN to DMD histo-images and compare segmentation results with both original K-Means and Otsu thresholding results and the results from previously evaluated deep-learning methods.

It is difficult to obtain ground-truth pixel-by-pixel segmentation labels when studying histo-images in practice which essentially motivates the presented work as the existing deep-learning methods often rely on clean segmentation labels for model inference. The goal of ANTN and NTN is to enable a new deep-learning model framework to incorporate noisy labels for training. For this set of experiments, we select 26 sub-images from one of 11 DMD histo-images with their corresponding K-Means and Otsu segmentation results as noisy segmentation labels. The example sub-images together with the corresponding segmentation results are shown in Figures 3.4(a), (b), and (c). As we observe empirically, K-Means often performs better than Otsu segmentation for our images in the successful cases. Model inference of u-net, NTN, and ANTN has been done similarly as for synthetic data. Note that u-net^1, u-net^2 and u-net^3 now represent the u-net trained with the

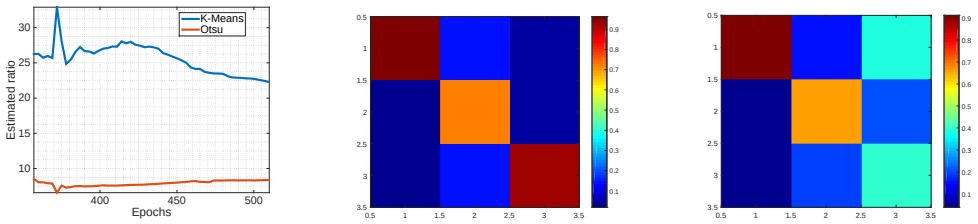


(a) Original (b) K-Means (c) Otsu (d) U-net^1 (e) U-net^2 (f) U-net^3 (g) NTN^1 (h) NTN^2 (i) ANTN

Figure 3.5: Original histo-images and corresponding segmentation results

corresponding K-Means, Otsu thresholding, and mixture of noisy segmentations. NTN^1 and NTN^2 represent the NTN trained with the corresponding K-Means and Otsu noisy segmentions. With the learning rate 10^{-4}, training the u-net with the single type of noisy segmentations converges in 400 epochs and training with the mixture converges around 157 epochs. For NTN, we initialize the training with the corresponding u-net and then diffuse the noise transition layer by weight decay of 10^{-4} for 150 epochs. For ANTN, we initialize the clean label prediction network with the trained u-net^3 then further train two transition probability networks for 200 epochs. The consequent iterative adaptive training converges around 155 epochs with the same 10 epochs for the alternating interval as described earlier.

We provide the corresponding segmentation results from u-net, NTN, and ANTN in Figures 3.4 and the zoomed details of the marked box in Figure 3.4(a) are shown in Figure



(a) (b) (c)

Figure 3.6: (a) Estimated ratio of intermediate output labels to noisy labels. (b) Estimated transition matrix for K-Means noisy dataset. (c) Estimated transition matrix for Otsu noisy dataset.

3.6. In Figure 3.6, the segmentations highlighted by green box show that ANTN achieves the most homogeneous and coherent segmentations of fibrosis and muscle, the segmentations highlighted by green box show the better fibrosis segmentation from ANTN even within necrotic muscle bundles (the fifth and the sixth row), and the segmentations highlighted by purple box indicate ANTN works great as well when the fine segmentation is expected. Without ground-truth segmentation, we follow the way in [28, 29] for quantitative evaluation based on the *entropy* U within segmented regions in RGB color space and *disparity* D across regions in L*a*b* color space. In the segmented region j of histo-image, we denote the number of pixel with intensity m in channel c to be $L_j(m)$ and the total number of pixel in region j to be S_j, thus $\frac{L_j^c(m)}{S_j}$ can be taken as the probability that a pixel has the intensity m in region j for channel c. Averaging the three channel and all the region, we can derive the expected entropy H for segmentation:

$$H = G * \sum_{j=1} (\frac{S_j}{S_I}) H_j \tag{3.2}$$

where

$$G = \sqrt{\sum_{b=1}^{MaxArea} [N(b)]^{1+1/b}}, \tag{3.3}$$

$$H_j = -\frac{1}{3}\sum_{c=1}^{3}\sum_{m=1}^{255}\frac{L_j^c(m)}{S_j} log \frac{L_j^c(m)}{S_j}. \tag{3.4}$$

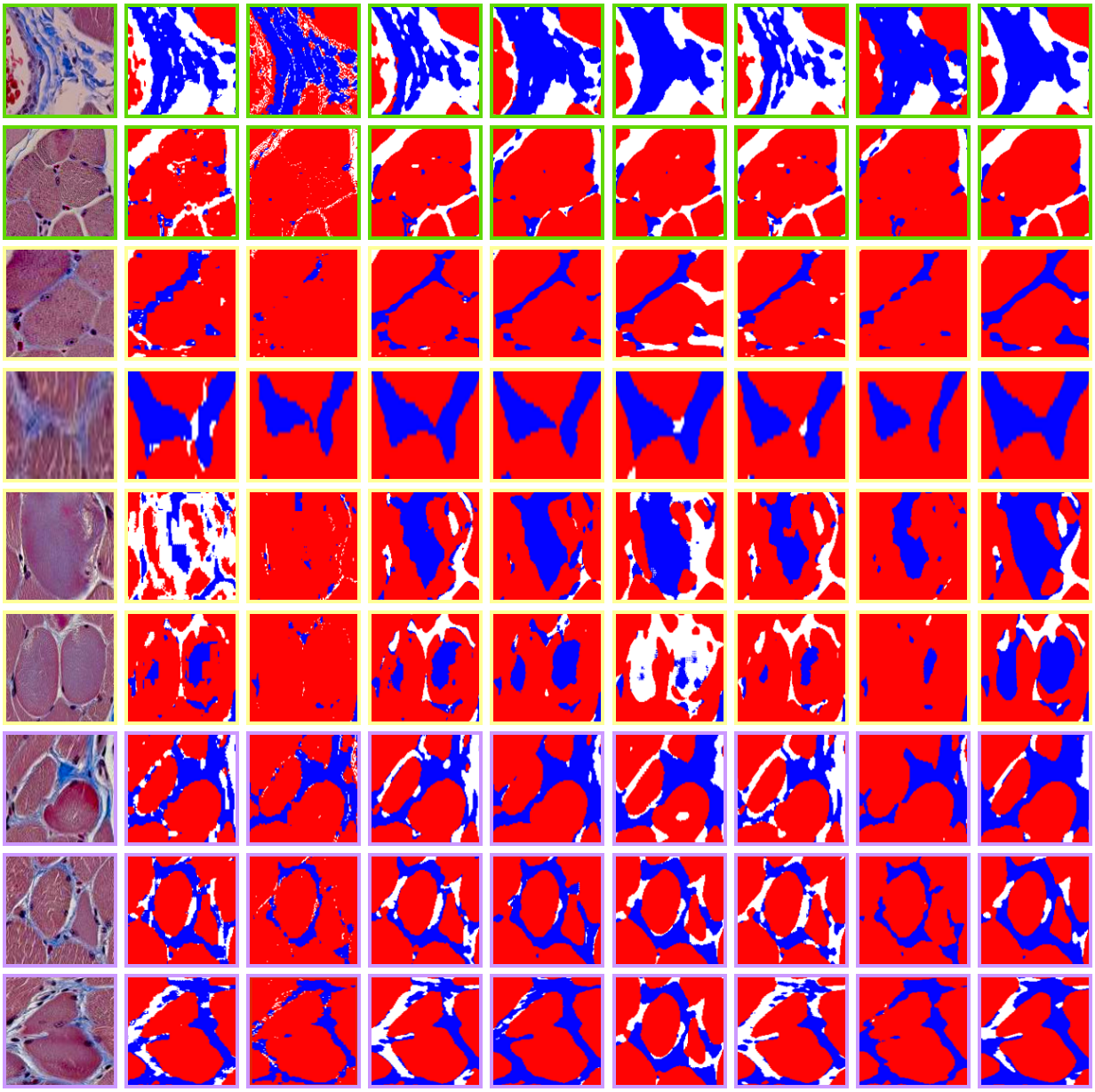
G is the penalized term for over segmentation with $N(b)$ representing the amount of regions having b pixel, and H_j is the region-wise entropy. For the *disparity* D across regions, we compute the average intensity $A_i^c = \frac{\sum_{k\in\mathcal{R}_i} X_k^c}{N_i}$, in which X_k^c is the corresponding channel intensity for pixel k; $\mathcal{R}_i$ denotes the set of pixels belonging to the ith cluster; and N_i is the total number of pixels in the ith cluster. Let $P_i = \frac{N_i}{\sum_{j=1}^{3} N_j}$. We have *disparity*:

$$D = A_1^2 P_1 - A_2^3 P_2. \tag{3.5}$$

Note that D is computed by the weighted average intensity differences only between red and blue regions with the corresponding channels as we are mostly interested in muscle and fibrosis in DMD histo-images [30]. Clearly, the smaller the H and the larger the D are, the better the segmentation is. Hence, we evaluate the segmentation results quantitatively by $E = \frac{H}{D}$. The comparison of E values for 11 original histo-image groups (each includes 100 split images and the training images are from the third group) is given in Table 2, in which we have highlighted the entities of the best (red) and the second best (green) for each group. Clearly, ANTN with noisy training samples is outperforming all the other methods in 7 of 11 samples.

We also investigate the estimated ratio similarly as for synthetic data based on the intermediate outputs during ANTN training by noisy segmentations from either K-Means or Otsu algorithm. It is observed that the ratio with respect to K-Means is much larger

than that with Otsu (Figure 3.5(a)). Besides, the corresponding average transition matrices after convergence are shown in Figure 3.5(b) and (c). Clearly, the average label-flip noise transition matrix trained for K-Means segmentation has diagonal entry values closer to 1 compared to that for Otsu segementation. This tells that K-Means segmentation results match better with the segmentation results derived by ANTN, being consistent with the fact that K-Means achieves better segmentation results compared to Otsu thresholding.



(a) Original (b) K-Means (c) Otsu (d) U-net^1 (e) U-net^2 (f) U-net^3 (g) NTN^1 (h) NTN^2 (i) ANTN

Figure 3.7: Zoomed details of the regions (scaled to same size) marked in the Figure 3.4(a) ■ inhomogeneous regions required homogeneous segmentation, ■ faint blue regions required blue segmentation, ■ complicated/fine regions required accurate segmentation.

4. SUMMARY AND CONCLUSIONS

In this thesis, we aim to tackle the difficulty of applying deep learning in histo-image segmentation when the clean ground-truth is unavailable so that clinicians will be free from the time-consuming work in manual pixel annotation. The core idea is taking the noisy segmentations from the off-the-shelf algorithms as training set to train the deep learning network to generate clean segmentation. To adapt the deep learning network for noisy label training, we propose the noise-tolerant network (NTN) and adaptive noise-tolerant network (ANTN) based on the U-Net architecture. While the NTN considers label noise only depends on the class of clean label, the ANTN not only considers more appropriately that the label noise depends on both the class of clean label and the appearance of image, but also can integrate multiple noisy datasets into training. The experiments on synthetic images and histo-images show that the ANTN has the best performance among other deep learning algorithms and the off-the-shelf algorithms.

Some problems are still being remained for the ANTN model. For instance, ANTN assigns the transition probability to each pixel of image but such pixel-wise allocation should be unnecessary due to the similar features in the region of pixels. Besides, ANTN has not consider the correlation of label noise process between pixels either. We will focus on both above problems and experiment on more benchmark datasets in the future.

REFERENCES

[1] A. J. Mendez, P. G. Tahoces, M. J. Lado, M. Souto, and J. J. Vidal, "Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms," *Medical Physics*, 1998.

[2] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren, "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer," *BMC medical imaging*, 2006.

[3] O. Sertel, J. Kong, U. V. Catalyurek, G. Lozanski, J. H. Saltz, and M. N. Gurcan, "Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading," *Journal of Signal Processing Systems*, 2009.

[4] K. Nguyen, A. Sarkar, and A. K. Jain, "Structure and context in prostatic gland segmentation and classification," in *Proc. Int'l Conf. Medical Image Computing and Computer-Assisted Intervention*, 2012.

[5] C. Demir, S. H. Gultekin, and B. Yener, "Augmented cell-graphs for automated cancer diagnosis," *Bioinformatics*, 2005.

[6] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Proc. Int'l Conf. Medical Image Computing and Computer-Assisted Intervention*, 2013.

[7] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, "Dcan: Deep contour-aware networks for object instance segmentation from histology images," *Medical image analysis*, 2017.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int'l Conf. Medical Image Computing and Computer-Assisted Intervention*, 2015.

[9] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *Proc. Int'l Conf. Medical Image Computing and Computer-Assisted Intervention*, 2013.

[10] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images," *IEEE Trans. on Medical Imaging*, 2016.

[11] V. Mnih and G. E. Hinton, "Learning to label aerial images from noisy data," in *Proc. Int'l Conf. Machine Learning*, 2012.

[12] T. Xiao, T. Xia, Y. Yang, C. Huang, and X. Wang, "Learning from massive noisy labeled data for image classification," in *Proc. Conf. Computer Vision and Pattern Recognition*, 2015.

[13] A. Veit, N. Alldrin, G. Chechik, I. Krasin, A. Gupta, and S. Belongie, "Learning from noisy large-scale datasets with minimal supervision," *arXiv preprint arXiv:1701.01619*, 2017.

[14] S. Sukhbaatar, J. Bruna, M. Paluri, L. Bourdev, and R. Fergus, "Training convolutional networks with noisy labels," *arXiv preprint arXiv:1406.2080*, 2014.

[15] S. Reed, H. Lee, D. Anguelov, C. Szegedy, D. Erhan, and A. Rabinovich, "Training deep neural networks on noisy labels with bootstrapping," *arXiv preprint arXiv:1412.6596*, 2014.

[16] P. Kakar and A. Y.-S. Chia, "If you can't beat them, join them: Learning with noisy data," in *Proc. ACM Conf. Multimedia*, 2015.

[17] B. Frénay and M. Verleysen, "Classification in the presence of label noise: A survey," *IEEE Trans. on Neural Networks and Learning Systems*, 2014.

[18] J. R. Quinlan, "Induction of decision trees," *Machine learning*, 1986.

[19] X. Zhu and X. Wu, "Class noise vs. attribute noise: A quantitative study," *Artificial Intelligence Review*, 2004.

[20] A. Kalai and R. A. Servedio, "Boosting in the presence of noise," in *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, 2003.

[21] A. Sarma and D. D. Palmer, "Context-based speech recognition error detection and correction," in *Proceedings of HLT-NAACL 2004: Short Papers*, 2004.

[22] B. De Brabandere, X. Jia, T. Tuytelaars, and L. Van Gool, "Dynamic filter networks," in *Proc. Neural Information Processing Systems*, 2016.

[23] W. Klingler, K. Jurkat-Rott, F. Lehmann-Horn, and R. Schleip, "The role of fibrosis in duchenne muscular dystrophy," *Acta Myologica*, 2012.

[24] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, 2001.

[25] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, 2009.

[26] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1979.

[27] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, 1975.

[28] H.-C. Chen and S.-J. Wang, "The use of visible color difference in the quantitative evaluation of color image segmentation," in *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing*, 2004.

[29] H. Zhang, J. E. Fritts, and S. A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," *CVIU*, 2008.

[30] R. Hunter, "Photoelectric color difference meter," *Josa*, 1958.