Adaptive Noise-Tolerant Network for Image Segmentation

Weizhi Li*

Abstract

Unlike image classification and annotation, for which deep network models have achieved dominating superior performances compared to traditional computer vision algorithms, deep learning for automatic image segmentation still faces critical challenges. One of such hurdles is to obtain ground-truth segmentations as the training labels for deep network training. Especially when we study biomedical images, such as histopathological images (histo-images), it is unrealistic to ask for manual segmentation labels as the ground truth for training due to the fine image resolution as well as the large image size and complexity. In this paper, instead of relying on clean segmentation labels, we study whether and how integrating imperfect or noisy segmentation results from off-the-shelf segmentation algorithms may help achieve better segmentation results through a new Adaptive Noise-Tolerant Network (ANTN) model. We extend the noisy label deep learning to image segmentation with two novel aspects: (1) multiple noisy labels can be integrated into one deep learning model; (2) noisy segmentation modeling, including probabilistic parameters, is adaptive, depending on the given testing image appearance. Implementation of the new ANTN model on both the synthetic data and real-world histo-images demonstrates its effectiveness and superiority over off-the-shelf and other existing deep-learning-based image segmentation algorithms.

1 Introduction

Many deep learning models operate in a supervised nature and have had enormous successes in many applications, including image classification and annotation [8, 11, 13, 16, 17, 18]. However, when we have only unlabeled data, for example, for biomedical image analysis, the existing Convolutional Neural Network (CNN) based methods may not apply. In order to overcome such limitations, a crowd-sourcing strategy [1] has been proposed to collect manual annotations from different people when getting labels from medical experts is difficult and then integrate the cues from these noisy crowd-sourcing annotations to infer the ground truth [10]. In image classification, these noisy crowd-sourcing labels are considered to be Noisy At Random (NAR) [5] and several NAR models [11, 16] have been proposed by modeling the label-flip noise independent of input images.

In this paper, we focus on deep learning for image segmentation when it is difficult to obtain clean pixel-wise segmentation labels. There are indeed many existing traditional image segmentation algorithms that can provide segmentation results with reasonable quality

*Work was done by the author while affiliated with Texas A&M University in 2017.

and can be taken as noisy training labels. However, for these segmentation labels, the NAR assumption may not hold any more, since the label-flip noise is not only dependent on the true object class but also the image features of the corresponding classes [5]. Hence, it is more proper to consider the Noisy Not At Random (NNAR) model [5] and develop adaptive image-dependent label-flip noise transition models. Motivated by the recent dynamic filter networks [9] that adaptively adjust deep network parameters accordingly to the input features, we propose an Adaptive Noise-Tolerant Network (ANTN) for image segmentation. The graphical probabilistic model and architecture of the network are shown in Figure 1. In ANTN, we explicitly model the probabilistic dependency between the input image, the ground-truth segmentation, and noisy segmentation results from off-the-shelf image segmentation algorithms. 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.

We develop an EM (Expectation-Maximization) based model inference algorithm and apply ANTN for image segmentation with both synthetic and histo-images. Performance comparison with off-the-shelf and deep learning algorithms shows the effectiveness and superiority of ANTN over other competing algorithms.

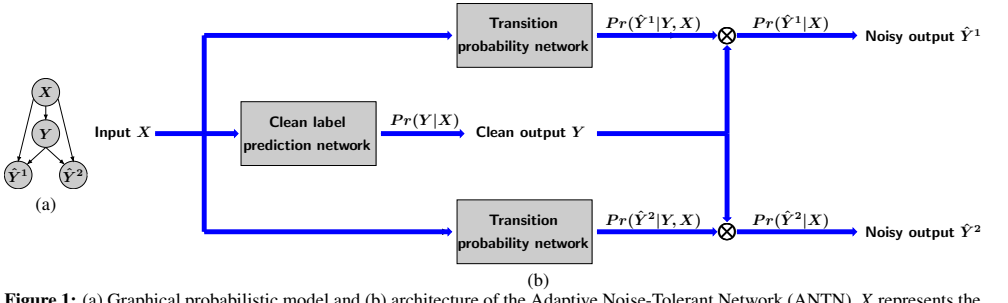


Figure 1: (a) Graphical probabilistic model and (b) architecture of the Adaptive Noise-Tolerant Network (ANTN). X represents the input image, Y represents the ground-truth segmentation, $\hat{Y}^1$ and $\hat{Y}^2$ represent noisy segmentations.

2 Related Work

Hinton et al. [11] have pioneered to use the deep network to incorporate the label-flip noise in aerial image labeling. They assumed the label-flip noise is only dependent on the true label and adopted an EM algorithm to train network model parameters iteratively, considering the true labels as latent variables. Benoît et al. [5] pointed out it is more realistic that mislabeling is dependent on input features and Xiao et al. [18] took such an assumption into consideration and integrated 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, 18], Veit et al. [17] recently 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, 17, 18] require a small clean dataset to assist model inference. Sukhbaatar et al. [16] proposed 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 instable. Reed et al. [13] introduced "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. [8] added a regularization term on the coefficients of hidden units during training to obtain stable results.

Most of the existing noisy label deep network models [8, 11, 13, 16, 17, 18] are on image classification or patch-labeling. The authors in [2] extended the label-flip noise model [16] to image segmentation and devised a Noise-Tolerant Network (NTN) based on the architecture of u-net [15] assuming that the label noise is only dependent on the true label. For image segmentation, it may be more appropriate to model label transition probabilities adaptively as discussed earlier. In addition, when different types of noisy labels are available, there is still no existing method to flexibly integrate them in the literature to the best of our knowledge. Our ANTN specifically addresses these two problems for deep-learning image segmentation.

3 Method

We study whether and how integrating multiple noisy datasets from off-the-shelf segmentation algorithms may help achieve better segmentation results. We develop the ANTN model for image segmentation with two novel aspects: (1) multiple noisy labels can be naturally integrated; (2) noisy segmentation modeling is adaptive, with image-dependent label transition probabilities. First, we introduce the probabilistic model of image segmentation when multiple noisy segmentations are available. Based on the model, we construct the adaptive deep learning framework, motivated by the recent dynamic filter network [9]. We note that we focus on the model with two noisy datasets but the model can be generalized to the settings with more than two noisy datasets in a straightforward manner due to the symmetry of our proposed framework. We also provide the model inference procedure to train the proposed adaptive deep learning model.

3.1 Model

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 this paper. We can model the relationships between input images and noisy segmentations based on the following general probabilistic model:

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

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 [2, 11, 16], (1) the clean pixel-wise labels indexed by n: y_n's, are conditionally independent given X: $Pr(Y|X) = \Pi_{n \in \mathcal{I}} Pr(y_n|X)$; 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: $Pr(\hat{Y}|Y,X) = \Pi_{n \in \mathcal{I}} Pr(\hat{y}_n|y_n)$. Hence, the log-likelihood with one set of noisy labels for X can be written as:

$$L = \log Pr(\hat{Y}|X) = \sum_{n \in \mathcal{I}} \log \Big[\sum_{y_n=1}^{C} Pr(\hat{y}_n|y_n) Pr(y_n|X) \Big], \quad (2)$$

with which a Noise-Tolerant Network (NTN) with the u-net architecture [2] has been developed to recover clean segmentations.

In the proposed ANTN model (1), 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 [4], we can rewrite the probabilistic model (1):

$$Pr(\hat{Y}^1,\hat{Y}^2|X)=\prod_{n\in\mathcal{I}}\sum_{y_n=1}^{C}Pr_n(\hat{y}_n^1|y_n,X)Pr_n(\hat{y}_n^2|y_n,X)Pr(y_n|X). \tag{3}$$

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 1(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 1(a), we follow the construction in [2, 4, 11, 15] to have the corresponding u-net architectures with the deep network framework shown in Figure 1(b). The main difference among these three deep network models are the constraints applied to their outputs of the last layers:

$$\sum_{y_n=1}^{C}Pr_n(y_n|X)=1,\quad \sum_{\hat{y}_n^1=1}^{C}Pr_n(\hat{y}_n^1|y_n,X)=1,\quad \sum_{\hat{y}_n^2}^{C}Pr_n(\hat{y}_n^2|y_n,X)=1, \tag{4}$$

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 1(a) is motivated by the label-flip noise model in the noise-tolerant image classification framework in [16] and the introduced asymmetric Bernoulli noise (ABN) model in [11].

3.2 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 (3):

$$\mathcal{L}_s=\frac{1}{N}\sum_{n\in\mathcal{I}}log\sum_{y_n=1}^{C}Pr((\hat{y}_n^s)_{obs}|y_n,X;\theta_s)Pr(y_n|X;\theta_3),\quad s=1,2, \tag{5}$$

$$\mathcal{L}_3=\frac{1}{N}\sum_{n\in\mathcal{I}}log\sum_{y_n=1}^{C}Pr((\hat{y}_n^1)_{obs}|y_n,X;\theta_1)Pr((\hat{y}_n^2)_{obs}|y_n,X;\theta_2)Pr(y_n|X;\theta_3) \tag{6}$$

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 (5) and θ_3 for minimizing (6). Similar to [11, 16], 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 (5) and (6) can be updated as follows:

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

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

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:

$$\nabla\theta_s^{(t+1)}\leftarrow\frac{1}{N}\sum_{n\in\mathcal{I}}\sum_{y_n=1}^{C}Pr^{(t)}(y_n|(\hat{y}_n^s)_{obs},X)\frac{\partial logPr((\hat{y}_n^s)_{obs}|y_n,X;\theta_s)}{\partial\theta_s},\quad s=1,2, \tag{7}$$

$$\nabla\theta_3^{(t+1)}\leftarrow\frac{1}{N}\sum_{n\in\mathcal{I}}\sum_{y_n=1}^{C}Pr^{(t)}(y_n|(\hat{y}_n^1)_{obs},(\hat{y}_n^2)_{obs},X)\frac{\partial logPr(y_n|X;\theta_3)}{\partial\theta_3}. \tag{8}$$

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.

4 Experimental Results

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

4.1 Datasets

To quantitatively evaluate ANTN and compare it with other segmentation algorithms, we first create a synthetic image set with the corresponding simulated noisy segmentations. With the validated performance improvement over existing algorithms, we then apply ANTN to a set of histo-images, obtained from a study of Duchenne Muscular Dystrophy (DMD) disease [9], 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 2(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 2(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 [9]. They are split into 472×472 sub-images and preprocessed by a stain normalization method [14]. Some of the preprocessed sub-images are shown in Figure 5(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 [6, 9]. 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 [7] and Otsu thresholding [12] on all the sub-images and we consider the obtained segmentation labels as the noisy segmentation labels in deep-learning methods including ANTN. Segmentation examples are shown in Figures 5(b) and (c).

4.2 Performance evaluation on synthetic data

For synthetic data, we compare the performance of ANTN with the existing deep-learning segmentation methods: (1) U-net [15] that is a CNN taking noisy segmentation labels as the ground truth for training; (2) Noise-tolerant u-net (NTN) [2, 16] that models the segmentation noise independent of image features. 35 synthetic images and their corresponding segmentations are used for training. For our ANTN, we first initialize the clean label prediction network (a u-net with the same architecture as in [2, 15]) 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 to diffuse the label-flip-noise transition probability from identity to approximate the average noise transition probability matrix for 150 more epochs [2, 16]. We do not train NTN with the mixture of noisy labels as it can only take one single type of noisy labels [2, 16]. Training of the u-net with different noisy labels can be considered as the intermediate steps of ANTN and NTN model inference.

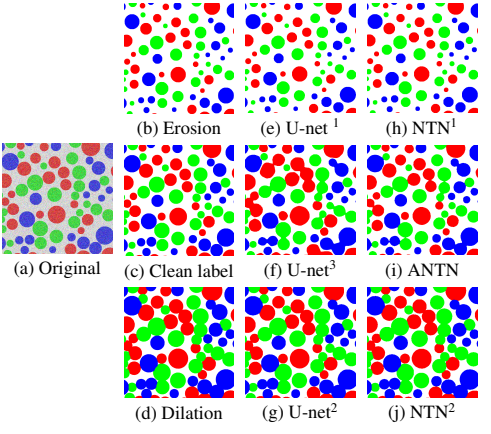
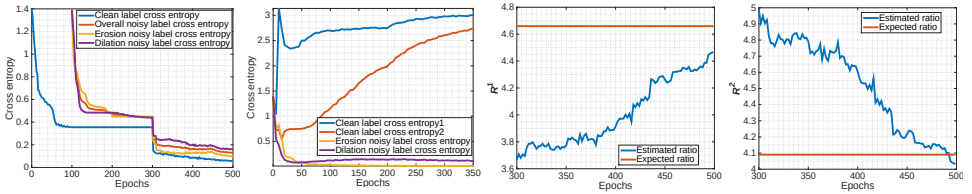


Figure 2: Synthetic image and corresponding segmentations.



(a) (b) (c) (d)

Figure 3: (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

We provide the examples of the corresponding segmentation results in Figures 2(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 [2, 16] often can not correctly segment the corresponding objects without appropriate modeling of segmentation noise with erosion and dilation bias. In Figures 2(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 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 get the highest accuracy of **97.71%** for ANTN followed by **93.38%** by u-net[3] with mixture training being superior over other methods, which have obained the accuracy all below **85%**.

In order to show the convergence of our training procedure, 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(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(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(c) and (d):

$$R^s = \frac{\sum_{n\in\mathcal{I}} I(\arg\max_u Pr(y_n = u|(\hat{y}_n^1)_{obs},(\hat{y}_n^2)_{obs},X) = (\hat{y}_n^s)_{obs})}{\sum_{n\in\mathcal{I}} I(\arg\max_u Pr(y_n = u|(\hat{y}_n^1)_{obs},(\hat{y}_n^2)_{obs},X) \neq (\hat{y}_n^s)_{obs})}, \quad s = 1,2. \tag{9}$$

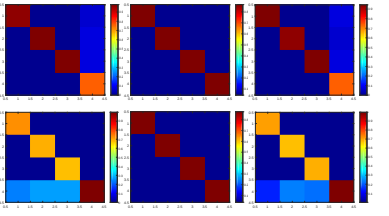
From Figures 3(c) and (d), the estimated ratios indeed approach the actual ratios in the training data with the corresponding trend indicating the learned ANTN models the noise transitions better and better during the iterative training stage.

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 4.

4.3 Experiments on histopathological images

With the promising results with synthetic data, we further implement ANTN 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.



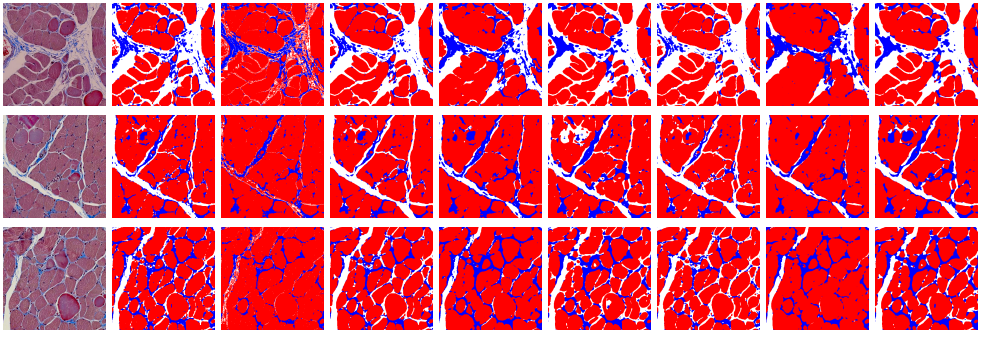
(a) (b) (c)

Figure 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.

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. ANTN enables 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 5(a), (b), and (c). As we observe empirically, K-Means often performs better than Otsu segmentation for our images. 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 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 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 5(d)-(i), which visually demonstrates that ANTN achieves the most homogeneous and coherent segmentations of fibrosis, muscle, and other tissue type regions based on the stains. Without ground-truth segmentation, we further evaluate the ratio of uniformity within segmented region to disparity aross segmented region of the original image intensities in $L*a*b*$ space as suggested in [8, 19], and the smaller the ratio is, the better the segmentation is. The performance comparison results for all 11 original histo-images are given in Table 1 and the entries within one standard deviation from the best segmentation results are highlighted in the table. Using the mixtures of K-Means and Otsu segmentation as training labels, u-net^3 and ANTN obtain better segmentation results compared to the other methods overall. This shows that integrating different types of noisy labels for deep network model training may help improve the performance. More importantly, ANTN achieves the best

segmentation performance for 9 of 11 images.



(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 5: Original histo-images and corresponding segmentation results

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 1: Performance comparison of different methods on 11 original histo-images.

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 6(a)). Besides, the corresponding average transition matrices after convergence are shown in Figure 6(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, indicating that K-Means achieves better segmentation results compared to Otsu thresholding. This again agrees with our empirical observation from the beginning.

5 Conclusions

We have proposed a novel adaptive noise-tolerant network (ANTN) to integrate multiple noisy datasets for image segmentation. ANTN models the feature-dependent transition probabilities adaptively from multiple off-the-shelf segmentation algorithms that help generate noisy labels for training. Based on the extensive performance evaluation on both synthetic data and real-world histo-images, it is clear that ANTN is a promising automated deep-learning image segmentation method that can take noisy

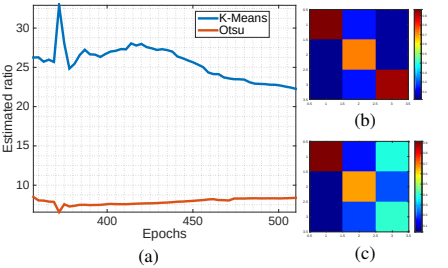


Figure 6: (a) Estimated ratio of intermediate output labels to noisy labels. (b) Estimated transition matrix for K-Means noisy dataset. (c) Estimated transition matrix for Otsu noisy dataset.

or “weak” segmentation results and further improve segmentation performance by borrowing signal strengths from multiple weak labels.

References

[1] Shadi Albarqouni, Christoph Baur, Felix Achilles, Vasileios Belagiannis, Stefanie Demirci, and Nassir Navab. AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE Trans. on Medical Imaging*, 2016.

[2] Authors. "noise-tolerant deep learning for histopathological image segmentation", icip 2017 submittion id 1664, supplied as additional material icip2017.pdf.

[3] Hsin-Chia Chen and Sheng-Jyh 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.

[4] Bert De Brabandere, Xu Jia, Tinne Tuytelaars, and Luc Van Gool. Dynamic filter networks. In *Proc. Neural Information Processing Systems*, 2016.

[5] Benoît Frénay and Michel Verleysen. Classification in the presence of label noise: A survey. *IEEE Trans. on Neural Networks and Learning Systems*, 2014.

[6] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2009.

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

[8] Pravin Kakar and Alex Yong-Sang Chia. If you can't beat them, join them: Learning with noisy data. In *Proc. ACM Conf. Multimedia*, 2015.

[9] Werner Klingler, Karin Jurkat-Rott, Frank Lehmann-Horn, and Robert Schleip. The role of fibrosis in duchenne muscular dystrophy. *Acta Myologica*, 2012.

[10] Balaji Lakshminarayanan and Yee Whye Teh. Inferring ground truth from multi-annotator ordinal data: A probabilistic approach. *arXiv preprint arXiv:1305.0015*, 2013.

[11] Volodymyr Mnih and Geoffrey E Hinton. Learning to label aerial images from noisy data. In *Proc. Int'l Conf. Machine Learning*, 2012.

[12] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 1975.

[13] Scott Reed, Honglak Lee, Dragomir Anguelov, Christian Szegedy, Dumitru Erhan, and Andrew Rabinovich. Training deep neural networks on noisy labels with bootstrapping. *arXiv preprint arXiv:1412.6596*, 2014.

[14] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 2001.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proc. Int'l Conf. Medical Image Computing and Computer-Assisted Intervention*, 2015.

[16] Sainbayar Sukhbaatar, Joan Bruna, Manohar Paluri, Lubomir Bourdev, and Rob Fergus. Training convolutional networks with noisy labels. *arXiv preprint arXiv:1406.2080*, 2014.

[17] Andreas Veit, Neil Alldrin, Gal Chechik, Ivan Krasin, Abhinav Gupta, and Serge Belongie. Learning from noisy large-scale datasets with minimal supervision. *arXiv preprint arXiv:1701.01619*, 2017.

[18] Tong Xiao, Tian Xia, Yi Yang, Chang Huang, and Xiaogang Wang. Learning from massive noisy labeled data for image classification. In *Proc. Conf. Computer Vision and Pattern Recognition*, 2015.

[19] Hui Zhang, Jason E Fritts, and Sally A Goldman. Image segmentation evaluation: A survey of unsupervised methods. *CVIU*, 2008.