NOISE-TOLERANT DEEP LEARNING FOR HISTOPATHOLOGICAL IMAGE SEGMENTATION

Weizhi Li[⋆] *Xiaoning Qian*[⋆] *Jim Ji*[⋆]

[⋆] Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA

ABSTRACT

Inhomogeneous color distribution and intensity impose major difficulty in fully automated histopathological image (histo-image) segmentation. In this paper, we propose a novel deep learning framework for histo-image segmentation. We innovate a noise-tolerant layer to the output layer of a deep learning image segmentation framework U-Net, which alleviates the requirement of accurately segmented training images and enables "unsupervised" histo-image segmentation by taking noisy segmentation results of traditional image segmentation algorithms as the training outputs. We implement noise-tolerant U-Net for histo-image segmentation to study Duchenne Muscular Dystrophy (DMD), a muscle degenerative disease. Performance comparison with traditional algorithms and the original U-Net demonstrates the great potential of the proposed noise-tolerant U-Net for histo-image segmentation.

Index Terms— Image segmentation, deep learning, U-Net, noisy labels

1. INTRODUCTION

Histopathological images (histo-images) have been the standard modality in clinical research for quantification of potential disease biomarkers providing the critical disease progression information [1]. For example, to assess the progression of recessive X-linked disease Duchenne Muscular Dystrophy (DMD), fibrosis has been considered an important biomarker since the change of the proportion of muscle and fibrosis through histo-image analysis provides prognostic clue of the disease [2, 3, 4]. When studying fibrosis in DMD, histo-images are usually of Hematoxylin-Eosin (H&E) staining, with muscle stained red, fibrosis stained blue, and the remaining parts, such as connective tissue, kept white [5]. It is hence critical to accurately segment muscle and fibrosis for reliable quantitative disease prognosis.

However, inhomogeneous color intensity and distribution has been a major obstacle hampering accurate histo-image segmentation. Although such color inhomogeneity caused by staining and scanning can be reduced by image normalization methods [1], fully automated image segmentation by traditional algorithms still does not provide satisfactory accuracy [6]. For instance, traditional adaptive thresholding based segmentation method cannot easily generate satisfactory results due to the difficulty in obtaining the optimal threshold in varying color intensity across histo-images [7]. K-means clustering [8] for color groups of interest can totally fail in the cases where there are not enough similar pixels corresponding to a specific cluster of interest, for example, fibrosis in DMD.

Deep learning has had huge success in image classification, as witnessed from AlexNet, VGGNet to ResNet [9, 10, 11]. It also can achieve the highest accuracies in several biomedical image segmentation challenges [12, 13]. For example, an end-to-end convolutional neural network (CNN) architecture – U-Net – has obtained accurate neuronal structure segmentation in electron microscopic images [12]. The main reason of overwhelming performance by deep learning is its capability of deriving better and richer image features in a data-driven fashion [9, 10, 11]. This motivates the deep learning histo-image segmentation based on invariant features accounting for image appearance variation. In [13], U-Net has been extended to a multi-level framework that achieves top performance in histo-image segmentation challenges.

These existing CNN-based image segmentation methods take supervised solutions and require manually labeled segmentations for training. However, this is contradictory to the essential aim of computer-aided histo-image analysis as manual histo-image annotation, especially considering pixel-based labeling for segmenting images in such an ultrahigh resolution, is often time-consuming and error-prone [1, 6, 14]. Although there is a recent trend of unsupervised histo-image analysis [15], they are mostly for classification of the whole histo-image, for example for cancer prognosis. To best of our knowledge, there does not exist unsupervised deep learning histo-image segmentation, as required in our DMD histo-image analysis.

As manually segmented histo-images are difficult to obtain, we modified the original U-Net framework to a noise-tolerant U-Net (Figure 1) for "unsupervised" histo-image segmentation. With proposed noise-tolerant U-Net, we can take

The authors would like to thank Dr. Joe Kornegay and Dr. Jay Griffin for helpful discussions, thank Dr. Sharla Birch and Mr. Stephen McConnell for providing the histo-images used for training and testing in this work, and thank Dr. Yoonsuck Choe for helpful comments.

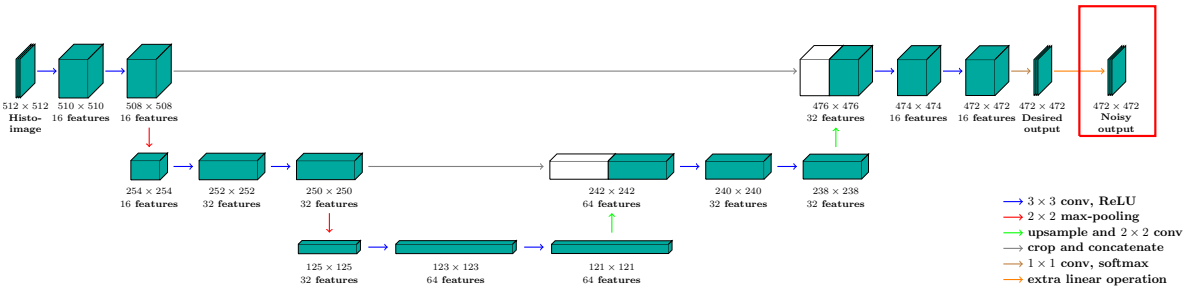

Fig. 1. Schematic illustrations of U-Net (without the extra linear layer in the red box) and our noise-tolerant U-Net (with the extra layer). The sizes of input images or feature maps with the corresponding numbers of features are denoted under each box.

noisy segmentation results, for example from Otsu's method or K-Means clustering, to train the network and still achieve high segmentation accuracy in testing images demonstrated in DMD histo-image segmentation.

2. METHOD

In this section, before presenting our noise-tolerant U-Net, we first briefly describe the necessary image pre-processing steps for DMD histo-images.

2.1. Image pre-processing

Due to the typical ultrahigh resolution (227nm/pixel) of histo-images, we first split the whole histo-image slide into sub-images (split images) for further analysis [1, 6, 14]. In this paper, we use the function `ndpisplit` provided in [16] to split the whole slide into 512×512 split images. In order to reduce potentially imbalanced color distributions across split images, we further adopt the stain normalization method [17] to normalize the split images. Three examples of pre-processed split images are given in the first column of Figure 3.

2.2. U-Net

U-Net is an image-to-image deep learning framework shown to be effective in biomedical image segmentation [12, 13]. 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 1). 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 upsampled 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 1.

U-Net is a supervised deep learning framework, requiring accurate segmentation labels for training. However, for our DMD histo-images, 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.

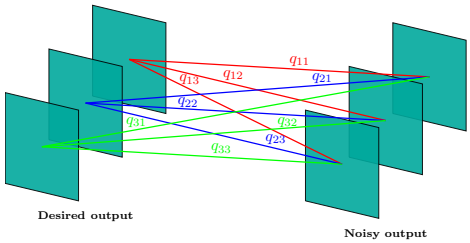

Fig. 2. Illustration of the "noise-tolerant" layer

2.3. Noise-tolerant U-Net

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 so that the performance will be robust to potentially noisy training segmentations. The main difference of our noise-tolerant U-Net from the original U-Net is an additional "transition" layer after the softmax output layer to allow noisy labels for training (marked in the red box in Figure 1). This is motivated by the "label flip noise model" in a recent noise-tolerant AlexNet-based image classification framework [18] that addresses a similar noisy label problem. The difference is that our noise-tolerant U-Net is for pixel-wise labeling in histo-image segmentation but the method in [18] is for the whole image classification.

Figure 2 illustrates the added layer to make the U-Net noise-tolerant, specifically designed for DMD histo-image segmentation. As we are interested in segmenting muscle, fibrosis, and other tissue based on red, blue, and white stains, histo-image segmentation can be considered as three-class classification problem. The softmax output layer assigns a corresponding label, denoted by Y_s^k, to a pixel k for the desired true segmentation. The added noise-tolerant layer allows noisy output labels Y_n based on Y_s for a given image. Let $q_{ij} = Pr(Y_n^k = j | Y_s^k = i)$ denotes the corresponding transition probability from label i to label j for pixel k. The parameters in this added transition layer can be represented by a 3×3 transition matrix $Q = \left(q_{ij}\right)_{3\times 3}$ with the con-

straints: $0 \leq q_{ij} \leq 1$ and $\sum_j q_{ij} = 1, \forall i$. Here, we consider the label flip noise is systematic and inherent to specific segmentation algorithms by which the training segmentations are derived [19].

Given w training images $X = \{X_1, \ldots, X_w\}$ and the corresponding noisy segmentation labels Y_n, the training of the added transition layer is motivated by the minimization of the cross-entropy loss function as in [18]:

$$L = -\frac{1}{K}\sum_{k=1}^{K} log\big[\sum_{i=1}^{3} Pr(Y_n^k = j|Y_s^k = i)Pr(Y_s^k = i|X)\big]$$
$$= -\frac{1}{K}\sum_{k=1}^{K} log\big[\sum_{i=1}^{3} q_{ij}Pr(Y_s^k = i|X)\big], \quad (1)$$

where K is the total number of pixels in X. By the total probability theorem, it is clear that this is equivalent to the maximum likelihood estimates of involved parameters in the modified U-Net with noisy segmentation Y_n as $L = -\frac{1}{K}\sum_{k=1}^{K} log[Pr(Y_n^k|X)]$. The training of the other layers simply follows the back-propagation procedure for the original U-Net. More importantly, we can rewrite

$$L = -\frac{1}{K}\sum_{i=1}^{3}\sum_{k\in\mathcal{S}_i} log\big[Pr^i(Y_n^k = j|X;Q)\big],$$

where $\mathcal{S}_i$ is the set of pixels that have the true label i, and $Pr^i(Y_n^k = j|X;Q)$ denotes the full model prediction probability for pixel k in $\mathcal{S}_i$. Asymptotically when $K \to \infty$, $L \to -\sum_{i=1}^{3}\sum_{j=1}^{3} q_{ij}^* log[Pr^i(Y_n = j|X;Q)] \geq -\sum_{i=1}^{3}\sum_{j=1}^{3} q_{ij}^* log(q_{ij}^*)$, achieving the minimum when $Pr^i(Y_n = j|X;Q) \to q_{ij}^*$ which is actual flip transition probability. Denote the confusion matrices for desired and noisy segmentations by $C_s = \big(c_{ij}^s\big)$ and $C_n = \big(c_{ij}^n\big)$ respectively, where $c_{ij}^s = \frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{S}_i} Pr^i(Y_s^k = j|X)$ and $c_{ij}^n = \frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{S}_i} Pr^i(Y_n^k = j|X;Q)$. It is clear $C_n = C_sQ$. If we know the actual label flip transition matrix $Q = Q^*$, minimizing L will asymptotically force $c_{ij}^n = Pr^i(Y_n = j|X;Q) \to q_{ij}^*$ hence $C_n = C_sQ^* \to Q^*$ forcing C_s converging to identity. Therefore, training the noise-tolerant U-net using noisy segmentations with actual transition matrix Q^* directly forces the softmax layer to predict the true labels.

Last but not least, minimizing L does not guarantee Q converging to Q^* [18]. 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 [18], we use the ridge regularization and fix the corresponding weight decay parameter to 10^{-4} in our experiments.

3. EXPERIMENTS

From 510 512×512 spit images of five original DMD histoimages (one group includes 110 split images and the other four includes 100 split images), we have held out all 110 split images from one group as training candidates for U-Net and all unseen split images would be for performance evaluation and comparison in our experiments.

We first apply K-Means clustering 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) 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. The second column in Figure 3 shows K-Means clustering segmentation results. It can be seen that K-Means may provide bad segmentations as illustrated in the bottom split image example, 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 (red) 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, the corresponding segmentation results are quite noisy due to the large color intensity variation within the image.

In order to train the original U-Net, we contacted a clinical expert to select ten "clean" segmentations by K-Means clustering in 110 training candidates as we do not have completely labeled images. With the learning rate 10^{-4}, we trained the U-Net first using all of the ten segmented images and then using only one of these ten images. It took 7,000 iterations (700 epochs for ten iterations/epoch) to converge when using ten training images, and 4,000 iterations (4,000 epochs for one iteration/epoch) for one training image. The corresponding segmentation results are illustrated in the fourth and fifth columns in Figure 3, respectively. By consulting with the expert, these results are better than both K-Means and Otsu segmentations. We also note that the expert considered the results from the U-Net trained with ten "clean" segmentations typically have underestimated fibrosis regions while those from the U-Net trained with one image provides overestimated fibrosis, which may be more desirable for subsequent image quantification.

We further evaluated the segmentations by our noise-tolerant U-Net. We randomly selected the segmentation results of 110 training candidates from Otsu's method as the "noisy" segmentation for training. As overestimated fibrosis is clinically preferred, we evaluated both the original U-Net and noise-tolerant U-Net with a single random training image to make the comparison fair. We have the same setup for training the original U-Net as described earlier. For noise-tolerant U-Net training, we fixed the transition matrix Q to be the identity matrix for the first 3,200 iterations to train the first 15 layers so that $C_n = C_sQ \to Q^*$ and then diffused Q approaching Q^* with weight decay. We find that for one training image, after another 800 iterations for the noise-tolerant layer

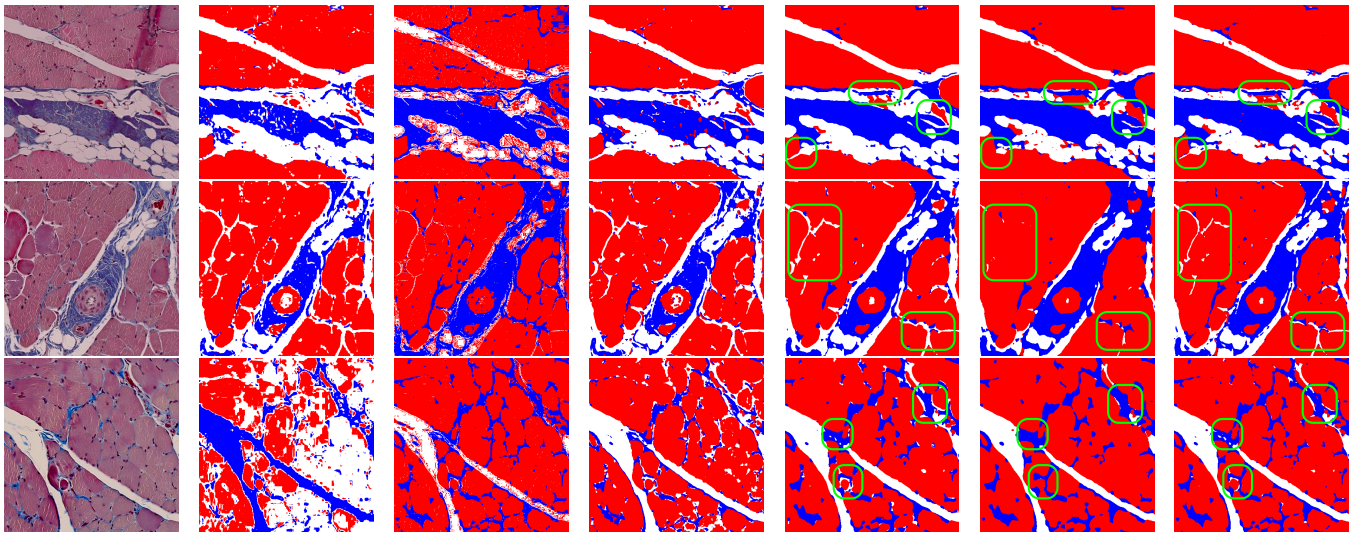

Fig. 3. Segmentation results. Column 1: original split images; Column 2: K-Means, Column 3: Otsu; Column 4: U-Net trained with ten "clean" segmented images: Column 5: U-Net trained with one "clean" segmented image; Column 6: U-Net trained with one "noisy" segmented image; Column 7: Noise-tolerant U-Net trained with one "noisy" segmented image.

training, the resulting Q can help recover "clean" segmentation. The comparison of segmentation results by the original and the noise-tolerant U-Net are shown in the sixth and seventh columns in Figure 3. It is clear visually that both of them perform better than K-Means and Otsu's method, moreover noise-tolerant U-Net outperforming the original U-Net without the noise-tolerant layer, especially at the places marked in green boxes. The segmentation results by noise-tolerant U-Net, when trained with "noisy" segmentation, are in fact consistent with the results by the original U-Net trained with "clean" segmentation. This clearly shows the potential of our noise-tolerant U-Net in histo-image segmentation, especially when manual segmentation is difficult to obtain.

To provide quantitative comparison of segmentation results, we follow the instruction of [20, 21] to evaluate based on the *uniformity* U within clustered regions and *disparity* D across regions in L*a*b* color space since we do not have the ground-truth. For channel c of histo-images, 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

$$\begin{aligned} U &= \sum_{i=1}^{3}\left\{\frac{1}{3}\sum_{c=1}^{3}\left[\frac{\sum_{k\in\mathcal{R}_i}(X_k^c - A_i^c)^2}{N_i}\right]\right\}^{\frac{1}{2}} P_i; \\ D &= A_1^2 P_1 - A_2^3 P_2. \end{aligned}$$

Notice 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 [22]. Clearly, the smaller the U and the larger the D are, the better the segmentation is. Hence, we evaluate the segmentation results quantitatively by $E = \frac{U}{D}$. The comparison of E values for five original histo-image groups (each includes 100 split images and the training images are from the third group) is given in Table 1, in which we have highlighted the entities within one standard deviation over the least E for each group. Clearly, noise-tolerant U-Net with noisy training samples and U-Net with clean training samples are outperforming all the other methods for comparison. Again, even without manual segmentations for training, our proposed noise-tolerant U-Net can achieve good histo-image segmentation for further analysis.

	1	2	3	4	5
KM	0.1240	**0.1455**	0.2141	0.1876	0.1317
OS	**0.1081**	0.1735	0.2450	0.1932	0.1660
UN	**0.0983**	**0.1451**	**0.1908**	**0.1724**	**0.1241**
UN*	0.1059	0.1611	**0.1993**	0.1857	0.1649
NTUN	**0.0976**	**0.1486**	**0.1870**	**0.1731**	**0.1315**

Table 1. Performance comparison by E for five groups: KM stands for K-Means; OS for Otsu; UN for U-Net trained with "clean" segmentation, UN* for U-Net trained with "noisy" segmentation; and NTUN for noise-tolerant U-Net trained with "noisy" segmentation.

4. CONCLUSION

We have proposed a noise-tolerant version of the original U-Net, which enables "unsupervised" deep learning for reliable segmentation of histo-images. Our preliminary experimental results show clear advantages of noise-tolerant U-Net over the original U-Net and other traditional histo-image segmentation algorithms. Our future research will focus on more thorough performance evaluation and exploring more flexible formulations of additional noise-tolerant layers into CNN-based image segmentation algorithms, which will greatly facilitate the analysis of histo-images in clinical research.

5. REFERENCES

[1] 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*, vol. 2, pp. 147–171, October 2009.

[2] A. Fischmann, P. Hafner, M. Gloor, M. Schmid, A. Klein, U. Pohlman, T. Waltz, R. Gonzalez, T. Haas, O. Bieri, and Fischer D., "Quantitative MRI and loss of free ambulation in Duchenne muscular dystrophy," *Journal of Neurology*, vol. 260, pp. 969–974, April 2013.

[3] H. Akima, D. Lott, C. Senesac, J. Deol, S. Germain, I. Arpan, R. Bendixen, H.L. Sweeney, G. Walter, and K. Vandenborne, "Relationships of thigh muscle contractile and non-contractile tissue with function, strength, and age in boys with Duchenne muscular dystrophy," *Neuromuscular Disorders*, vol. 22, pp. 16–25, January 2012.

[4] W. Klingler, K. Jurkat-Rott, F. Lehmann-Horn, and R. Schleip, "The role of fibrosis in Duchenne muscular dystrophy," *Acta Myologica*, vol. 31, pp. 184, December 2012.

[5] M.T. McCann, J. Majumdar, C. Peng, C.A. Castro, and J. Kovačević, "Algorithm and benchmark dataset for stain separation in histology images," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 3953–3957.

[6] S. Kothari, J.H. Phan, R.A. Moffitt, T.H. Stokes, S.E. Hassberger, Q. Chaudry, A.N. Young, and M.D. Wang, "Automatic batch-invariant color segmentation of histological cancer images," in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2011, pp. 657–660.

[7] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, pp. 23–27, June 1975.

[8] 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)*, vol. 28, pp. 100–108, January 1979.

[9] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, September 2014.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, December 2015.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[13] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P. Heng, "DCAN: Deep contour-aware networks for object instance segmentation from histology images," *Medical Image Analysis*, vol. 36, pp. 135–146, February 2017.

[14] 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*, vol. 55, pp. 169–183, April 2009.

[15] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, I. Eric, and C. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 1626–1630.

[16] C. Deroulers, D. Ameisen, M. Badoual, C. Gerin, A. Granier, and M. Lartaud, "Analyzing huge pathology images with open source software," *Diagnostic Pathology*, vol. 8, pp. 92, June 2013.

[17] P. Shirley, "Color transfer between images," *IEEE Corn*, vol. 21, pp. 34–41, September/October 2001.

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

[19] B. Frénay and M. Verleysen, "Classification in the presence of label noise: a survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, pp. 845–869, May 2014.

[20] H. Chen and S. Wang, "The use of visible color difference in the quantitative evaluation of color image segmentation," in *2004 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2004, pp. iii–593.

[21] H. Zhang, J.E. Fritts, and S.A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," *Computer Vision and Image Understanding*, vol. 110, pp. 260–280, May 2008.

[22] R.S. Hunter, "Photoelectric color difference meter," *Josa*, vol. 48, pp. 985–995, December 1958.