Journal of Machine Learning Research 1 (2000) 1-48 Submitted 4/00; Published 10/00

Unsupervised Domain Adaptation for the Histopathological Cell Segmentation through Self-Ensembling

***

Editor: **

Abstract

Histopathological images are generally considered as the golden standard for clinical diagnosis and cancer grading. Accurate segmentation of cells/nuclei from histopathological images is a critical step to obtain reliable morphological information for quantitative analysis. However, cell/nuclei segmentation relies heavily on well-annotated datasets, which are extremely labor-intensive, time-consuming, and expensive in practical applications. Meanwhile, one might want to fine-tune pretrained models on certain target datasets. But it is always difficult to collect enough target training images for proper fine-tuning. Therefore, there is a need for methods that can transfer learned information from one domain to another without additional target annotations. In this paper, we propose a novel framework for cell segmentation on the unlabeled images through the unsupervised domain adaptation with self-ensembling. It is achieved by applying generative adversarial networks (GANs) for the unsupervised domain adaptation of cell segmentation crossing different tissues and staining methods. Images in the source and target domain can be differentiated through the learned discriminator. Meanwhile, we present a self-ensembling model to consider the source and the target domain together as a semi-supervised segmentation task to reduce the differences of outputs which can achieve better performance on different pathological domains. Additionally, we introduce conditional random field (CRF) as post-processing to preserve the local consistency on the outputs. We validate our proposed segmentation method with unsupervised domain adaptation on three public cell segmentation datasets captured from different types of tissues, which achieved superior performance in comparison with state-of-the-art.

Keywords: Unsupervised Domain Adaptation, Cell Segmentation, Semi-supervised Segmentation, Self-Ensembling

1. Introduction

The field of computer-aided digital pathology plays an extremely important role in the understanding, diagnosis, and treatment on varieties of cancers (Chang et al., 2012). Plenty of tissue biopsies are performed every year. Within all these pathological images, it is self-evident that the importance of cell is almost second to none as it is the basic component of human beings. After decades of modern cytological and histopathologic study, cellular stains such as hematoxylin that stains nuclei (Titford, 2005) have been developed and turned into digital images. Although the interpretation and determination of cells and aberrant phenotypes in these pathological images could be done by professional doctors and pathologists today, it still takes them a large amount of time, labor, and fund to do so. To solve the problems above, there is a critical need for more accurate, robust, low-cost, and computationally efficient nuclei segmentation methods.

©2000 .

Nowadays, cell segmentation in histopathology images has been extensively studied with a variety of deep learning methods. Inspired by the advance of Fully Convolutional Networks (FCN) (Long et al., 2015), there is a variety of deep learning methods proposed in the segmentation field like U-Net (Ronneberger et al., 2015), DeepLab (Chen et al., 2017), and UNet++ (Zhou et al., 2018). For example, Chen et al. (2016) proposed a deep contour-aware network (DCAN) to establish better segmentation by a multi-task learning framework that learns not only probability maps but also clear contours. To train these fully-supervised methods, sufficient annotations are required with both images and corresponding pixel-level ground-truth labels (Kumar et al., 2017). However, well-annotated datasets for cell segmentation are extremely limited in clinical diagnosis. Meanwhile, collecting unlabeled data and then manual labeling the data is also an expensive, time-consuming, and tedious process. Thus, the need for training a deep model for cell segmentation without manual annotations appears, where the domain adaptation techniques could be the solution. More recently, a few unsupervised domain adaptation approaches for cell segmentation have been proposed. Liu et al. (2020) proposed a method based on Cycada (Hoffman et al., 2018) which adds a task re-weighting mechanism along with a nuclei inpainting mechanism to make the framework perform better on data from different organs. Haq and Huang (2020) proposed a framework based on GAN (Goodfellow et al., 2014) along with a reconstruction network to do segmentation on unlabeled data from different organs.

Despite the above methods have already taken cell segmentation across different datasets into consideration, there are still multiple challenges in the implementation of cell segmentation with domain adaptation. Firstly, current methods may lack robustness and stable performance when segmenting cross-domain cells. This limitation is especially critical for cell segmentation tasks, where the annotations are hard to obtain even in the source domain. Secondly, existing works only consider the domain adaptation of cell segmentation across one or two datasets within the same tissue or organ, which cannot be widely applicable in clinical diagnosis.

Taking the challenges above into account, this paper proposes a novel framework for cross-domain cell segmentation. Particularly, we first apply a semi-supervised method with self-ensembling. The framework can extract features from the source domain with data augmentation to improve the segmentation performance. Then the unsupervised domain adaptation can effectively transfer cell segmentation results across different tissues or segmentation using different staining methods. The contribution of this paper can be summarized in three aspects: 1) we develop a new framework that contains semi-supervised segmentation with domain adaptation on the target domain; 2) we introduce a semi-supervised framework to do data augmentation in the feature extraction part and a CRF (Boykov and Kolmogorov, 2004) module in the post-processing part to improve the robustness of the cell segmentation. 3) extensive and comprehensive experiments are carried out on three datasets from different tissues to demonstrate the effectiveness of our proposed methods.

2. Methodology

Overview: Figure. 1 presents the overview of our framework, which consists of three parts: Semi-supervised segmentation network (S), Discriminator (D), and Reconstructor (R). Firstly, we introduce a semi-supervised segmentation network by self-ensembling to

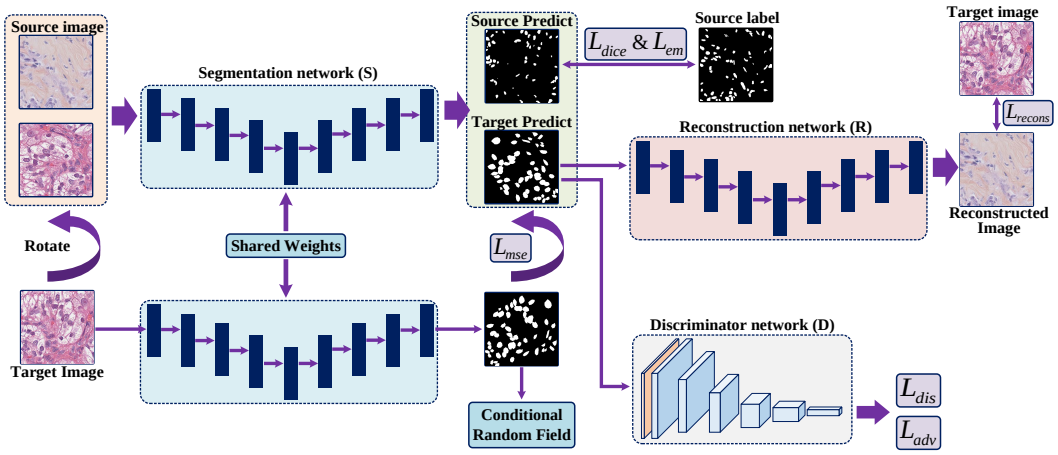

Figure 1: The complete architecture of our unsupervised domain adaptation framework. The segmentation network generates segmentation predictions. The reconstruction network reconstructs input images. The discrimination network distinguishes source domain predictions from target domain predictions.

ensure the model could learn from both source and target domain. Then, we transfer the segmentation from the source domain to the target domain by employing a discriminator which discriminate the outputs of the source and target. The reconstruction network is used to reconstruct the original images for ensuring that the predictions match the related images. Finally, we use a conditional random field (CRF) (Krähenbühl and Koltun, 2011) to constrain the regional consistency of the results. After the model training, the testing patches can be sequentially processed to obtain the segmentation results.

Semi-supervised Segmentation: Formally, in the cell segmentation problem, the histopathological image patches from the datasets are input X of size $H \times W \times 3$. Then, we want to predict the segmentation output $\widehat{Y}$ of size $H \times W \times 1$. In the source domain, we also have binary masks with pixel-wise ground-truth label Y of size $H \times W \times 1$ in our framework.

The segmentation network takes images X as input and uses the segmentation predictions $\widehat{Y}$ of the same width and height as output, i.e., $\widehat{Y} = S(X)$. We train S to generate the predictions $\widehat{Y}_s$ through a semi-supervised segmentation method while using the Y_s as the ground-truth input of the source domain. As for the target domain, we could not compute pixel-level loss for segmentation since there is no label for segmentation of the target domain in the unsupervised domain adaptation problem. For small datasets, the GAN method sometimes may not be able to obtain enough information from the features in the segmentation network. Therefore, using the entropy minimization loss can control the weight of the labeled examples, increase the confidence of the segmentation output, and make the model relatively more stable. In practice, the dice-coefficient loss and the entropy minimization loss are more effective than the binary cross-entropy loss, which is normally

used in cell segmentation tasks. So we use both the dice-coefficient loss and the entropy minimization loss as our segmentation loss:

$$L_{\text{dice}} = 1 - \frac{2.Y_s'.\widehat{Y}_{\text{s}}'}{Y_s' + \widehat{Y}_{\text{s}}'} \tag{1}$$

$$L_{em} = -\frac{1}{H * W} \sum_{h=0}^{H} \sum_{w=0}^{W} \widehat{Y}_s \log(\widehat{Y}_s) \tag{2}$$

where Y_s' and $\widehat{Y}_{\text{s}}'$ are flatten Y_s and $\widehat{Y}_s$ respectively.

However, the segmentation above just uses a fully-supervised segmentation method in the source domain. In practice, the images from the source and the target domain may differ a lot in staining result, clarity, and direction. Therefore, the data augmentation and the robustness of model transferring should be taken into consideration. To improve the data augmentation and the segmentation performance, we apply a self-ensembling method by using rotate transformation in a generalized form. To achieve this, we optimize the consistency loss with a teacher model, which shares its weights with the student model. The key point in the teacher-student learning-based semi-supervised segmentation network is based on the smoothness assumption. For example, data points close to each other in the image space are more likely to be close in the label spaceLaine and Aila (2016). To be specific, the semi-supervised segmentation tasks can be learned by optimizing a mean square error loss:

$$L_{mse} = ||\widehat{Y}_{s,\pi} - \widehat{Y}_s||^2 \tag{3}$$

where $\widehat{Y}_{s,\pi}$ is the prediction of the source image which goes through the transformation-consistent regularization named π.

In above, the target data goes through the model twice to get two predictions under different perturbations. In this case, the teacher model and the student model share their weights. In other words, the model assumes a dual role as a teacher and as a student. As a student, it learns as the fully-supervised segmentation method before. Meanwhile, as a teacher, it generates targets to be used by itself as a student for learning through optimizing the mean square error loss. By doing so, the self-ensembling method is applied to the segmentation model to make the performance more stable, which may do more favor in the diagnosis and other medical applications.

The overall segmentation loss function is then defined as:

$$L_{seg} = L_{em} + \lambda L_{mse} + L_{dice} \tag{4}$$

where L_{em}, L_{dice} and L_{mse} are supervised term (the former two) and regularization term. The time-dependent warming up function λ is a weighting factor for supervised loss and regularization loss. This weighting function is a Gaussian ramp-up curve that could slowly drop down the weight of the mean square error loss as:

$$\lambda = k * e^{||e-E||} \tag{5}$$

where E denotes the training epoch, k scales the maximum value of the weighting function, and e defines when the weight comes to the peak. In our experiments, we empirically set k to 1.0 and e to 30.

Training S with the annotated source data teaches S to make accurate predictions for source images. However, at this stage the segmentation network still generates incorrect predictions for target images as there are discrepancies between the source and the target. Therefore, the S needs to generate target domain predictions as much as close to the source domain predictions by making the distribution of target predictions $\widehat{Y}_t$ closer to source predictions $\widehat{Y}_s$. Thus, we define the adversarial loss as:

$$L_{adv}(X_t) = -\frac{1}{H' \times W'} \sum_{h',w'} \log(D(\widehat{Y}_t)) \tag{6}$$

where $\widehat{Y}_t = S(X_t)$, and H' and W' are the height and width of discriminator output $D(\widehat{Y}_t)$. This adversarial loss helps S to fool the discriminator so that it could consider $\widehat{Y}_t$ as a source domain segmentation prediction.

Overall, this semi-supervised segmentation network can be treated as the generator module of a GAN framework (Goodfellow et al., 2014). Accordingly, to make the predictions closer to each other in both the source and target domain, we also need a discriminator.

Discriminator: We introduce a discriminator D into the framework, where it can take source domain prediction or target domain prediction as its input and then distinguish whether the input comes from the source domain or the target domain. To train D, we use a cross-entropy loss as:

$$L_{dis}(\widehat{Y}) = -\frac{1}{H' \times W'} \sum_{h',w'} z \cdot \log(D(\widehat{Y})) + (1-z) \cdot \log(D(\widehat{Y})) \tag{7}$$

where z=0 when D takes target domain prediction as its input, and z=1 when input comes from source domain prediction.

Moreover, it is possible that these target predictions are not well-correlated with the target input images. A network for reconstructing images from the predictions to a similar appearance as input can ensure the correlation between the input image and the corresponding segmentation prediction.

Reconstructor: To ensure that the target domain predictions spatially correspond to the target domain images, we use a Reconstruction network R in our framework. We consider the segmentation network S as an encoder and the Reconstruction network R as a decoder that could reconstruct target images from the corresponding predictions. It can take the predictions $\widehat{Y}_t$ as its inputs and produce the reconstructed image as the output $R(\widehat{Y}_t)$. We calculate the reconstruction loss as:

$$L_{recons}(X_t) = \frac{1}{H \times W \times C} \sum_{h,w,c} (X_t - R(\widehat{Y}_t)^2) \tag{8}$$

where $R(\widehat{Y}_t)$ is the output of the Reconstructor for input $\widehat{Y}_t$, and H, W, C are the height, width, and number of channels of the input image X_t.

Overall, we optimize the following total loss when training our framework:

$$L(X_S, X_t) = L_{seg}(X_S) + \lambda_{adv} L_{adv}(X_t) + \lambda_{recons} L_{recons}(X_t) + L_{dis}(\widehat{Y}) \tag{9}$$

where the λ_{adv} and λ_{recons} are the weights to balance above losses.

At this point, our results have learned information from both the labeled source domain and the unlabeled target domain. Because neighbouring voxels share substantial spatial context, the segmentation results produced by the CNN are smooth. However, local minimization training and noise in the images may still result in some spurious outputs, like small isolated regions or holes in the predictions. And the lack of regional spatial information could also have great impact on the performance. To solve this problem, we employ a fully connected conditional random field (Krähenbühl and Koltun, 2011) as a post-processing step to achieve more structured predictions and constrain the spatial consistency of the results. For an input image X and its segmentation prediction $\widehat{Y}$, the Gibbs energy in the CRF model is given by:

$$E\left(\widehat{Y}\right) = \sum_{i} \psi_u\left(\widehat{Y}_i\right) + \sum_{ij, i \neq j} \psi_p\left(\widehat{Y}_i, \widehat{Y}_j\right) \tag{10}$$

The unary potential is the negative log-likelihood $\psi_u\left(\widehat{Y}_i\right) = -logP(\widehat{Y}_i|X)$, where in this case $P(\widehat{Y}_i|X)$ is the model's output for pixel i. The pairwise potentials in our model have the form:

$$\psi_p\left(\widehat{Y}_i, \widehat{Y}_j\right) = \mu\left(\widehat{Y}_i, \widehat{Y}_j\right) \sum_{m=1}^{K} w^{(m)} k^{(m)}\left(f_i, f_j\right) \tag{11}$$

where each $k^{(m)}$ is a Gaussain kernel $k^{(m)}(f_i, f_j) = exp(-\frac{1}{2}(f_i - f_j)^T \Lambda^{(m)}(f_i - f_j))$. The vectors f_i and f_j are feature vectors for pixels i and j in an arbitrary feature space, $w^{(m)}$ are linear combination weights, μ is a label compatibility function. Each kernel $k^{(m)}$ is characterized by a symmetric, positive-definite precision matrix $\Lambda^{(m)}$, which defines its shape. The contrast-sensitive two-kernel potentials are used for our image segmentation and labeling, consisting of the appearance kernel and the smoothness kernel. They are defined in terms of the color vectors I_i and I_j and positions p_i and p_j :

$$k(f_i, f_j) = w^{(1)} exp(-\frac{|p_i - p_j|^2}{2\Theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\Theta_\beta^2}) + w^{(2)} exp(-\frac{|p_i - p_j|^2}{2\Theta_\gamma^2}) \tag{12}$$

The appearance kernel means that nearby pixels which have similar color are likely to be in the same class. The degrees of nearness and similarity are controlled by parameters Θ_α and Θ_β. The smoothness kernel removes small regions that isolated with big ones. Finally, the weights $w^{(1)}$ and $w^{(2)}$ define the relative strength of the two factors.

3. Experiment

Datasets: There are three datasets used in our experiments. Irshad et al. (2014) released the KIRC dataset which consists of 463 images of 400×400 pixel size and comes from whole slide images (WSI) of Kidney Renal Clear cell carcinoma (KIRC). Naylor et al. (2018) released the TNBC dataset that consists of 50 images of 512×512 pixel size and comes from slides of Triple Negative Breast Cancer Cell (TNBC). Hou et al. (2020) released the TCIA dataset that consists of 1356 images of 256×256 pixel size which comes from slides of 14 different cancer types. We only use 97 images from Bladder Urothelial Carcinoma (BLCA)

Table 1: Segmentation results of unsupervised domain adaptation under four compared methods with three histopathological image datasets.

Source Domain	TNBC				TCIA			
Target Domain	KIRC		TCIA		KIRC		TNBC	
	Iou	Dice	Iou	Dice	Iou	Dice	Iou	Dice
DA_ADV (Dong et al., 2018)	0.3276	0.4911	0.4480	0.6082	0.2914	0.4484	0.4078	0.5712
CBST (Zou et al., 2018)	0.3031	0.4627	0.4015	0.5638	0.2853	0.4413	0.3901	0.5547
CellSegUDA (Haq and Huang, 2020)	0.5487	0.7039	0.5802	0.7188	0.5539	0.7067	0.4109	0.6078
Our framework$_{withoutCRF}$	0.551	0.7023	0.5689	0.7233	0.5566	0.718	0.4023	0.6051
Our framework$_{withoutEMloss}$	0.5523	0.7025	0.5661	0.7211	0.5665	0.7251	0.4271	0.6295
Our framework$_{withoutself-ensembling}$	0.5607	0.7189	0.5791	0.7301	0.5653	0.7233	0.4308	0.6246
Our framework	**0.5683**	**0.7234**	**0.6049**	**0.7413**	**0.5737**	**0.7261**	**0.5263**	**0.6796**

among the TCIA dataset. All the datasets are extracted at 40x magnification.

Settings: In our framework, we use U-Net (Ronneberger et al., 2015) as our segmentation and reconstruction baseline. We use 80% of the images for training, 10% for validation and 10% for evaluation. During the training, we employ Adam optimizer (Kingma and Ba, 2014) to optimize the losses with learning rates of 0.0001, 0.001, and 0.001 used in the segmentation network, discriminator, and reconstructor respectively. We use 0.001 and 0.01 as λ_{adv} and λ_{recons} respectively and train in total 500 epochs. All experiments are carried out by using Pytorch on a Linux system with 2 RTX 2080Ti and take about 14 GB memory of the graphic cards for 12 hours.

Results: In our experiment, we use TNBC (Naylor et al., 2018) and TCIA (Hou et al., 2020) datasets as the source domain respectively and the other two datasets as the target domain respectively. Besides, our proposed method is compared with 3 recently proposed related methods. The first one is DA-ADV, a UDA method based on the GAN method which also uses a discriminator like ours and proposed by Dong et al. (2018). The second one is CBST, another UDA method proposed by Zou et al. (2018). This is a popular classbalanced self-training framework by generating pseudo labels, which is a different method to transfer the domain from ours. The third one is CellSegUDA, an unsupervised adversarial domain adaptation method proposed by Haq and Huang (2020). This method is based on the GAN method and achieved excellent performance on cell segmentation tasks. We also perform ablation experiments to validate the effectiveness of each component in our framework.

To evaluate the segmentation accuracy of the nuclei instances, we use both the Dice and IOU metrics as our validation standards. As shown in Table 1, compared with the former three baseline methods, our framework's performance has a significant improvement under both the IOU and the Dice metrics. Since the self-ensembling method can enhance robustness and the GAN network provides better performance on the target domain, our framework works better in cell segmentation of UDA tasks in comparison with the other three state-of-the-arts. Moreover, in the ablation experiments, with individual components removed, the performance of the framework falls, which could prove that every component could contribute to the performance of the framework.

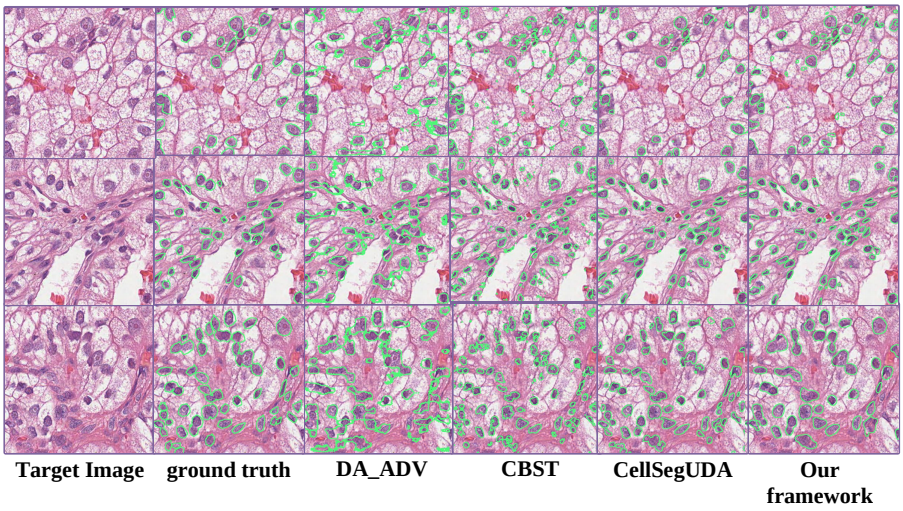

Figure 2: The visualization results for segmentation from TNBC to KIRC. From up to down are different images we randomly choose from the target domain. From left to right are the results of different method.

Figure 2 shows the visualization of segmentation results of our proposed method and 3 compared methods, which indicates that there are significant improvements in our proposed method. It is shown that our framework performs better in different types of histopathological images. Moreover, in the results of the target domain, our framework works well in a blurry image and get a larger and more accurate segmentation result, which benefits from the robustness provided by the self-ensembling method.

4. Conclusions

In this paper, we propose a novel unsupervised domain adaptation method for the cell segmentation across different tissues and datasets. The method is based on the GAN framework and particularly suitable for the segmentation of whole-slide images from different tissues and in different stain forms. Moreover, we improve the segmentation performance by introducing a teacher-student learning-based semi-supervised segmentation network, which could help with the data augmentation and to overcome the problem of low robustness that usually appears in domain adaptation. Meanwhile, we use the conditional random field method as our post-processing step to achieve more structured predictions and constrain the spatial consistency of the results. Based on this work, we may apply self-supervised methods in this field and study how to make the domain adaptation more accurate in the future.

References

Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on pattern analysis and machine intelligence*, 26(9):1124–1137, 2004.

Hang Chang, Ju Han, Alexander Borowsky, Leandro Loss, Joe W Gray, Paul T Spellman, and Bahram Parvin. Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association. *IEEE TMI*, 32(4):670–682, 2012.

Hao Chen, Xiaojuan Qi, Lequan Yu, and Pheng-Ann Heng. Dcan: deep contour-aware networks for accurate gland segmentation. In *CVPR*, pages 2487–2496, 2016.

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE TPAMI*, 40(4):834–848, 2017.

Nanqing Dong, Michael Kampffmeyer, Xiaodan Liang, Zeya Wang, Wei Dai, and Eric Xing. Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio. In *International conference on medical image computing and computer-assisted intervention*, pages 544–552. Springer, 2018.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, pages 2672–2680, 2014.

Mohammad Minhazul Haq and Junzhou Huang. Adversarial domain adaptation for cell segmentation. In *Medical Imaging with Deep Learning*, 2020.

Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *ICML*, pages 1989–1998. PMLR, 2018.

Le Hou, Rajarsi Gupta, John S Van Arnam, Yuwei Zhang, Kaustubh Sivalenka, Dimitris Samaras, Tahsin M Kurc, and Joel H Saltz. Dataset of segmented nuclei in hematoxylin and eosin stained histopathology images of 10 cancer types. *arXiv preprint arXiv:2002.07913*, 2020.

Humayun Irshad, Laleh Montaser-Kouhsari, Gail Waltz, Octavian Bucur, JA Nowak, Fei Dong, Nicholas W Knoblauch, and Andrew H Beck. Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd. In *Pacific Symposium on Biocomputing*, pages 294–305. World Scientific, 2014.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems*, 24:109–117, 2011.

Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE TMI*, 36(7):1550–1560, 2017.

Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016.

Dongnan Liu, Donghao Zhang, Yang Song, Fan Zhang, Lauren O'Donnell, Heng Huang, Mei Chen, and Weidong Cai. Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting. In *CVPR*, pages 4243–4252, 2020.

Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015.

Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE TMI*, 38(2):448–459, 2018.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.

M Titford. The long history of hematoxylin. *Biotechnic & Histochemistry*, 80(2):73–78, 2005.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11. Springer, 2018.

Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision (ECCV)*, pages 289–305, 2018.