Published in final edited form as:

Med Image Comput Comput Assist Interv. 2019 October ; 11764: 740–749. doi:

10.1007/978-3-030-32239-7_82.

Adversarial Domain Adaptation and Pseudo-Labeling for Cross-Modality Microscopy Image Quantification

Fuyong Xing^{1,2}, Tell Bennett^{2,3}, Debashis Ghosh^{1,2}

¹Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus

²Data Science to Patient Value, University of Colorado Anschutz Medical Campus

³Department of Pediatrics, University of Colorado Anschutz Medical Campus

Abstract

Cell or nucleus quantification has recently achieved state-of-the-art performance by using convolutional neural networks (CNNs). In general, training CNNs requires a large amount of annotated microscopy image data, which is prohibitively expensive or even impossible to obtain in some applications. Additionally, when applying a deep supervised model to new datasets, it is common to annotate individual cells in those target datasets for model re-training or fine-tuning, leading to low-throughput image analysis. In this paper, we propose a novel adversarial domain adaptation method for cell/nucleus quantification across multimodality microscopy image data. Specifically, we learn a fully convolutional network detector with task-specific cycle-consistent adversarial learning, which conducts pixel-level adaptation between source and target domains and completes a cell/nucleus detection task. Then we generate pseudo-labels on target training data using the detector trained with adapted source images and further fine-tune the detector towards the target domain to boost the performance. We evaluate the proposed method on multiple cross-modality microscopy image datasets and obtain a significant improvement in cell/nucleus detection compared to the reference baselines and a recent state-of-the-art deep domain adaptation approach. In addition, our method is very competitive with the fully supervised models trained with all real target training labels.

1 Introduction

Convolutional neural networks (CNNs) have recently produced excellent cell detection performance [14]. Due to domain shifts, CNNs trained on one microscopy image dataset (e.g., bright-field microscopy) might not be applicable to another (e.g., phase-contrast microscopy). It is common to annotate images in the target domain for model re-training. However, individual cell annotations are very expensive. Although one can fine-tune a pretrained CNN, it may still be difficult to collect enough annotated target data for proper fine-tuning in real applications [21]. Thus, it is necessary to design methods that can transfer learned knowledge from one to another domain without additional target data annotations.

Unsupervised domain adaptation is one type of method to address domain shifts without target data labeling [4,2]. Many previous unsupervised deep domain adaptation methods align feature distributions from source and target domains [16,4,21]. In particular, adversarial domain adaptation has been applied to medical image segmentation and

classification [12,17]. However, these approaches that conduct deep feature matching might not effectively capture low-level image appearance variance. Additionally, it might be difficult to fully align feature distributions for structured output tasks [8] such as cell detection. On the other hand, some recent work [19,7] focuses on pixel-space adaptation for general image analysis using generative adversarial networks (GANs) [5]. These methods work well only for limited domain shifts, which might not hold in cross-modality microscopy image datasets, or formulate a pixel-wise classification model, which might not be optimal for cell detection [23,11].

In this paper, we propose a novel adversarial domain adaptation method for cross-modality cell/nucleus detection (see Figure 1). Instead of conducting feature-space adaptation, we directly translate raw source images to target ones and learn a cell detector with task-specific cycle-consistent adversarial learning. In order to encode topological information in the label space, cell detector is modeled with structured regression in a fully convolutional network (FCN). In addition, we further fine-tune the detector using real target training images with generated pseudo-labels (i.e., estimated cell locations), which do not need any manual target image annotations. In this way, our domain adaptation is able to take advantage of both accurate cell annotations in the source domain and high-quality training images in the target domain for cell localization. To the best of our knowledge, this is the first work using GAN-based domain adaptation for cell detection in cross-modality microscopy images.

2 Methodology

An overview of the proposed domain adaptation method is shown in Figure 1. During the training phase, the neural networks are learned with an adversarial loss, a cycle-consistent loss and a task-specific structured regression loss. The generators and discriminators interact with each other to adapt source images to the target domain, while the detector learns to locate individual cells with adapted source images, which appear as if drawn from the target domain. Next, we apply the detector to target training data for pseudo-label generation and then further fine-tune the detector with these *artificial annotations on real target images*. During the testing phase, for each new target image, the detector predicts one heat map, where local maxima indicate the locations of cell centers.

2.1 Task-Specific Cycle-Consistent Adversarial Learning

Problem Definition—Let $(X^s, Y^s) = \{(x_i^s, y_i^s)\}_{i=1}^{N^s}$ denote source training data with N^s images $x_i^s \in \mathbb{R}^{w \times h \times c}$ and corresponding structured labels $y_i^s \in \mathbb{R}^{w \times h}$ (w, h and c are the width, height and channel of images or labels), and $X^t = \{x_i^t\}_{i=1}^{N^t}$ represent target training data with N^t unannotated images $x_i^t \in \mathbb{R}^{w' \times h' \times c'}$. With (X^s, Y^s) and X^t , our goal is to learn mappings between source S and target T domains such that source images X^s can be translated to target style ones $\hat{X}^s$, and build a cell detector R based on these translated source images, which is applicable to real target data.

Pixel-Level Domain Adaptation Modeling—To close the gap between source and target domains, we need to map data from one domain to the other such that a discriminator fails in differentiating the two domains. Instead of relying on feature-matching adaptation, we conduct pixel-space adaptation by directly translating raw images between source and target domains. The reason is that feature-space adaptation assumes there exists a shared, cross-domain feature space, but this might not hold for structured prediction tasks [8], such as cell detection. Additionally, in order to preserve the content of source images during image translation, we apply a cycle-consistent constraint [24] to our generative adversarial learning for domain adaptation. This cycle-consistent adversarial learning has two generator-discriminator pairs, which enable the model to convert adapted images back into the domain where they come from. Meanwhile, it does not require paired source-target training images, which are generally not available in cross-modality cell detection.

Formally, given unpaired training data from source S and target T domains, our generative adversarial domain adaptation model learns two generators, $G_{ST}: S \rightarrow T$ and $G_{TS}: T \rightarrow S$. Generator G_{ST} adapts source images to the target domain so that its associated discriminator D_T is unable to distinguish between adapted source images and real target images. Similarly, G_{TS} maps data from the target to source domain, aiming to fool its corresponding discriminator D_S . This is achieved by using the adversarial loss [5] in both directions

$$\mathcal{L}_{GAN}(G_{ST}, D_T) = \mathbb{E}_{\mathbf{x}^t \sim \mathbf{X}^t} [\log D_T(\mathbf{x}^t)] + \mathbb{E}_{\mathbf{x}^s \sim \mathbf{X}^s} [\log(1 - D_T(G_{ST}\mathbf{x}^s))], \quad (1)$$

$$\mathcal{L}_{GAN}(G_{TS}, D_S) = \mathbb{E}_{\mathbf{x}^s \sim \mathbf{X}^s} [\log D_S(\mathbf{x}^s)] + \mathbb{E}_{\mathbf{x}^t \sim \mathbf{X}^t} [\log(1 - D_S(G_{TS}\mathbf{x}^t))], \quad (2)$$

where $\mathbf{x}^s/\mathbf{x}^t$ is a sample from the source/target domain. Differing from the vanilla GAN taking random noise as input [5], our generator G_{ST}/G_{TS} synthesizes realistic target/source images from source/target images, i.e., adapting source/target data to the target/source domain. In order to enforce the cycle consistency [24] such that the reconstruction of adapted images is identical to their original versions, i.e., $G_{TS}(G_{ST}(\mathbf{x}^s)) \approx \mathbf{x}^s$ and $G_{ST}(G_{TS}(\mathbf{x}^t)) \approx \mathbf{x}^t$, we adopt an ℓ_1 penalty on the reconstruction error

$$\begin{aligned} \mathcal{L}_{cycle}(G_{ST}, G_{TS}) = & \mathbb{E}_{\mathbf{x}^s \sim \mathbf{X}^s} [\|G_{TS}(G_{ST}(\mathbf{x}^s)) - \mathbf{x}^s\|_1] \\ & + \mathbb{E}_{\mathbf{x}^t \sim \mathbf{X}^t} [\|G_{ST}(G_{TS}(\mathbf{x}^t)) - \mathbf{x}^t\|_1]. \end{aligned} \quad (3)$$

In addition to the generators and discriminators, our model also contains a structured regressor R , which maps input images into the label space for individual cell detection. Compared to pixel-wise classification, structured regression is able to exploit additional context information during training for more robust prediction [11,23]. Due to lack of data annotations in the target domain, we train this regressor with the adapted source images $G_{ST}(\mathbf{x}^s)$. Because the cycle-consistent adversarial learning preserves the content of source images during image translation, while making the adapted source images look similar to those from the target domain, we expect that a regressor trained with $G_{ST}(\mathbf{x}^s)$ and

corresponding labels $\mathbf{y}^s$ can be applicable to the target domain. Specifically, this task-specific regression loss is chosen as a weighted mean squared error

$$\mathcal{L}_{reg}(G_{ST}, R) = \mathbb{E}_{(\mathbf{x}^s, \mathbf{y}^s) \sim (\mathbf{X}^s, \mathbf{Y}^s)} \left[\left\| (\mathbf{y}^s + \alpha \bar{\mathbf{y}}^s \mathbf{1})^{1/2} \odot (R(G_{ST}(\mathbf{x}^s)) - \mathbf{y}^s) \right\|_F^2 + \left\| (\mathbf{y}^s + \alpha \bar{\mathbf{y}}^s \mathbf{1})^{1/2} \odot (R(\mathbf{x}^s) - \mathbf{y}^s) \right\|_F^2 \right], \quad (4)$$

where the label $\mathbf{y}^s$ is a continuous-valued, non-negative proximity map and each pixel value measures the proximity of this pixel to its closest real cell center [23], higher values for closer positions. $\bar{\mathbf{y}}^s$ is the mean value of $\mathbf{y}^s$ and $\mathbf{1}$ is a matrix with all entries equal to one. α controls the contributions from different image regions and $\odot$ denotes the element-wise multiplication. Note that we train this cell detector with both adapted and original source images [3].

With the loss functions defined above, our full objective is formulated as

$$\mathcal{L}(G_{ST}, G_{TS}, D_S, D_T, R) = \mathcal{L}_{GAN}(G_{ST}, D_T) + \mathcal{L}_{GAN}(G_{TS}, D_S) + \lambda \mathcal{L}_{cycle}(G_{ST}, G_{TS}) + \gamma \mathcal{L}_{reg}(G_{ST}, R), \quad (5)$$

where λ and γ weight the importance of cycle consistency and regression, respectively. We aim to learn the detector R by solving the following problem

$$\arg \min_{\left\{ G_{ST}, G_{TS}, R \right\}} \max_{\left\{ D_S, D_T \right\}} \mathcal{L}(G_{ST}, G_{TS}, D_S, D_T, R). \quad (6)$$

In our implementation, G_{ST} and G_{TS} are FCNs [10] with 9 residual blocks [6]. D_S and D_T are chosen as the 70×70 PatchGAN network [9], which is computationally efficient. For regressor R , we use the fully residual convolutional network [23] that has 9 residual blocks [6] and 4 long-range skip connections [18].

2.2 Learning with Pseudo-Labels

With pixel-space domain adaptation, we are able to translate source images into target-style ones in terms of color and texture, and the detector trained with these translated images would be applicable to real target data. However, it might be difficult for cycle-consistent adversarial learning to effectively handle very large domain shifts between source and target domains. We believe that the detector can be further improved by using real target training images. Because individual cell annotations are not available in the target domain, we propose to fine-tune the detector with pseudo-labels, which are generated by applying the detector to cell location estimation/prediction on target training images.

For each target training image $\mathbf{x}^t$, the detector R predicts one identical-sized heat map $\hat{\mathbf{y}}^t$, where larger values mean these pixels are more close to cell centers. In order to mitigate the effects of image noise, we suppress the pixels with values less than $\eta \cdot \max(\hat{\mathbf{y}}^t)$, where $\eta \in$

[0,1], by assigning them zeros and then locate cell centers by seeking local maxima on the suppressed prediction map. With these detected centers, we follow [23] to generate (pseudo) labels $\tilde{y}^t$ for images x^t by using a function of the distance transform, and fine-tune the detector R using the regression loss with $(x^t, \tilde{y}^t)$ pairs. During the testing phase, we apply the final detector to cell detection on target images and locate cells by seeking local maxima in the corresponding predicted maps, where the low-valued pixels are also suppressed.

3 Experiments

Datasets

We extensively evaluate the proposed method on three microscopy image datasets, which are acquired using different staining techniques or imaging protocols/modalities. The first dataset [20] consists of 100 hematoxylin and eosin (H&E) stained colon cancer histology images, each of which is acquired with bright-field microscopy at 20 $\times$ magnification and has a dimension of 500 $\times$ 500 $\times$ 3. The second dataset contains 114 pancreatic neuroendocrine tumor bright-field microscopy images (at 20 $\times$ magnification), and they use immunohistochemical (IHC) Ki67 staining and have a dimension of 500 $\times$ 500 $\times$ 3. The third dataset [1] is composed of 22 cervical cancer HeLa cell line images, which are generated with phase-contrast (PC) microscopy. This dataset does not use either H&E or IHC staining and each image has size 400 $\times$ 400 $\times$ 3. All the datasets provide gold-standard annotations of cell/nucleus centers. Following [1,23], we randomly split each dataset into two halves, one for training and the other for testing. We use the metrics [23] for detection evaluation, i.e., precision, recall and F_1 score.

Implementation Details

We empirically set the parameter values as: $\alpha = 5$ in Equation (4), $\lambda = 10$, $\gamma = 1$ in Equation (5) and $\eta = 0.4$ for pixel suppression during model inference. We closely follow [24] to set the other hyperparameters and use the Adam solver [13] with a learning rate of 2×10^{-4} for adversarial model training. We apply a stage-wise training strategy to model learning [7]. With pseudo-labels, we fine-tune the detector with 10^5 iterations. We use data augmentation such as randomly cropping, rotation and mirroring during model training.

Domain Adaptation Evaluation

Table 1 lists the F_1 score of domain adaptation on cross-modality microscopy image datasets. The reference baselines are the regression detector models directly trained with source data only. For a comparison, we evaluate two variants of our method: (1) pseudo-labeling (PL), which fine-tunes the baseline models with estimated cell locations (i.e., pseudo-labels) obtained by running the baselines on target training data, and (2) pixel-space adaptation (PSA), which is trained with adapted source images but no further fine-tuning with pseudo-labels. We also report the results of a recent state-of-the-art deep adversarial domain adaptation method, ADDA [21], and it is outperformed by our method on all the datasets, particularly in the H&E $\rightarrow$ PC experiment, where there is a large dataset shift between the source and target domains. We note that both PL and PSA outperform the baselines, and PSA provides a significant improvement by increasing the average F_1 score

from 52.2% to 74.6%. Our method (i.e., PSA + PL) achieves the best performance for all the domain adaptation settings, further improving the average score to 79.5% and closing the gap to the supervised models trained with all real target data labels only. In particular, our method approaches the ideal performance for H&E→IHC and H&E→PC. This indicates that our method is effective at adapting H&E stained microscopy images to other modality domains. Qualitative results of cell detection are provided in Figures S1 ~ S3 at the Supplementary Material.

We also extend and evaluate the proposed method by using an additional subset of real target training annotations for model training. Figure 2 shows that using only a subset of target annotations, our method further improves cell detection and provides equivalent performance to the supervised models trained with the full target annotations only. In this scenario, it can significantly reduce human effort for target data labeling. Additionally, we further notice that the improvement grows slowly when more real target training labels are used, probably because the detector approaches its model capacity with more real data.

Comparison with State-of-the-art Fully Supervised Methods

We compare the proposed method to several recent fully supervised deep learning models, FCN-8s [15], U-Net [18] and FCRNA/FCRNB [22], which are trained with all real target training annotations, as shown in Table 2. We see that our method outperforms the classification models, i.e. FCN-8s and U-Net, on all the three datasets. In addition, our method is very competitive with the fully supervised regression models, i.e. FCRNA/FCRNB, especially for the IHC and PC datasets. This confirms the effectiveness of our method that requires no real target labels.

4 Conclusion

We propose a novel deep generative adversarial domain adaptation method for cell/nucleus detection in cross-modality microscopy image data. It adapts raw source images to a target domain with cycle-consistent adversarial learning and trains a structured regression-based object detector with the adapted source images. Additionally, it further fine-tunes the detector with pseudo-labels of target training data. Extensive experiments on multiple microscopy image datasets demonstrate the effectiveness of the proposed method.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement.

This research was supported by the National Cancer Institute of the National Institutes of Health under Award Number R21CA237493.

References

1. Arteta C, Lempitsky V, Noble JA, Zisserman A: Learning to detect cells using non-overlapping extremal regions In: Ayache N, Delingette H, Golland P, Mori K (eds.) MICCAI 2012. vol. 7510, pp. 348–356. Springer, Berlin, Heidelberg (2012)

2. Bermúdez-Chacón R, Becker C, Salzmann M, Fua P: Scalable unsupervised domain adaptation for electron microscopy In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W (eds.) MICCAI 2016. vol. 9901, pp. 326–334. Springer, Cham (2016)
3. Bousmalis K, et al.: Unsupervised pixel-level domain adaptation with generative adversarial networks. In: CVPR pp. 3722–3731 (2017)
4. Ganin Y, et al.: Domain-adversarial training of neural networks. JMLR 17(1), 2096–2030 (2016)
5. Goodfellow I, et al.: Generative adversarial nets. In: NIPS. pp. 2672–2680 (2014)
6. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. In: CVPR pp. 770–778 (2016)
7. Hoffman J, Tzeng E, Park T, Zhu JY, Isola P, Saenko K, Efros A, Darrell T: CyCADA: Cycle-consistent adversarial domain adaptation. In: ICML pp. 1989–1998 (2018)
8. Hong W, Wang Z, Yang M, Yuan J: Conditional generative adversarial network for structured domain adaptation. In: CVPR pp. 1335–1344 (2018)
9. Isola P, Zhu J, Zhou T, Efros AA: Image-to-image translation with conditional adversarial networks. In: CVPR pp. 5967–5976 (2017)
10. Johnson J, Alahi A, Li LF: Perceptual losses for real-time style transfer and super-resolution In: Leibe B, Matas J, Sebe N, Welling M (eds.) ECCV 2016. vol. 9906, pp. 694–711. Springer, Cham (2016)
11. Kainz P, et al.: You should use regression to detect cells In: Navab N, Hornegger J, Wells WM, Frangi AF (eds.) MICCAI 2015. vol. 9351, pp. 276–283. Springer International Publishing, Cham (2015)
12. Kamnitsas K, et al.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks In: Niethammer M, Styner M, Aylward S, Zhu H, Oguz I, Yap PT, Shen D (eds.) IPMI 2017. vol. 10265, pp. 597–609. Springer, Cham (2017)
13. Kingma DP, Ba J: Adam: A method for stochastic optimization. In: ICLR pp.1–15 (2015)
14. Litjens G, et al.: A survey on deep learning in medical image analysis. MIA 42, 60–88 (2017)
15. Long J, Shelhamer E, Darrell T: Fully convolutional networks for semantic segmentation. In: CVPR pp. 3431–3440 (2015)
16. Long M, Cao Y, Wang J, Jordan M: Learning transferable features with deep adaptation networks. In: ICML pp. 97–105 (2015)
17. Ren J, Hacihaliloglu I, Singer EA, Foran DJ, Qi X: Adversarial domain adaptation for classification of prostate histopathology whole-slide images In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G (eds.) MICCAI 2018. vol. 11071, pp. 201–209. Springer, Cham (2018)
18. Ronneberger O, Fischer P, Brox T: U-net: Convolutional networks for biomedical image segmentation In: Navab N, Hornegger J, Wells WM, Frangi AF (eds.) MICCAI 2015. vol. 9351, pp. 234–241. Springer, Cham (2015)
19. Shrivastava A, et al.: Learning from simulated and unsupervised images through adversarial training. In: CVPR pp. 2242–2251 (2017)
20. Sirinukunwattana K, Raza SEA, Tsang YW, Snead DRJ, Cree IA, Rajpoot NM: Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE TMI 35(5), 1196–1206 (2016)
21. Tzeng E, Hoffman J, Saenko K, Darrell T: Adversarial discriminative domain adaptation. In: CVPR pp. 2962–2971 (2017)
22. Xie W, Noble JA, Zisserman A: Microscopy cell counting with fully convolutional regression networks. In: DLMIA Workshop pp. 1–8 (2015)
23. Xie Y, Xing F, Shi X, Kong X, Su H, Yang L: Efficient and robust cell detection: A structured regression approach. MIA 44, 245–254 (2018)
24. Zhu JY, Park T, Isola P, Efros AA: Unpaired image-to-image translation using cycle-consistent adversarial networkss. In: ICCV pp. 2223–2232 (2017)

Fig. 1.

The overview of the proposed adversarial domain adaptation method for cell detection. G_{ST} and D_T are the source-to-target generator and its associated discriminator, respectively. G_{TS} is the target-to-source generator, and R is the regression-based cell detector. The GAN loss is used to learn the adversarial networks, and the cycle loss ensures that adapted source images can translate into the original images. The detector R is learned with adapted source images and then further fine-tuned with pseudo-labels on target images. The dashed gray arrow represents the transition from the training to testing stage. Here source and target images are generated using bright-field and phase-contrast microscopy, respectively. For clarity, we do not draw the target cycle.

Fig. 2.

F₁ score of cell detection with respect to different numbers (percentage) of real target training annotations. Red dashed lines represent the supervised models trained with the full (i.e. 100%) target training annotations only.

Table 1.

F_1 score (%) of domain adaptation (source $\rightarrow$ target) on different datasets. The last row denotes the fully supervised models trained with all target training data only, and the last column represents the average across all six settings in each row.

Method	H&E $\rightarrow$ IHC	H&E $\rightarrow$ PC	IHC $\rightarrow$ H&E	IHC $\rightarrow$ PC	PC $\rightarrow$ H&E	PC $\rightarrow$ IHC	Avg.
ADDA[21]	79.4	56.9	66.9	84.9	51.3	69.9	68.2
Baseline	77.8	16.7	54.0	33.7	60.0	70.7	52.2
PL	78.7	26.9	57.0	40.2	62.5	71.9	56.2
PSA	78.9	83.4	68.9	88.6	55.4	72.5	74.6
Ours	83.1	92.5	70.5	89.3	63.0	78.5	79.5
Target only	87.4	96.1	81.4	96.1	81.4	87.4	88.3

Table 2.

Comparison with state-of-the-art fully supervised methods in terms of precision (P), recall (R), and F_1 score (%). For our method that does not use any real target annotations, we report the performance of the experimental settings, $IHC \rightarrow H\&E$, $H\&E \rightarrow IHC$ and $H\&E \rightarrow PC$ for the H&E, IHC and PC datasets, respectively.

Method	H&E			IHC			PC		
	P	R	F_1	P	R	F_1	P	R	F_1
FCN-8s [15]	57.3	79.5	66.6	97.0	55.0	70.2	70.2	93.1	80.0
U-Net [18]	39.5	85.2	54.0	88.5	64.0	74.3	82.7	80.4	81.5
FCRNA [22]	72.4	87.7	79.3	83.1	89.5	86.2	97.3	94.5	95.9
FCRNB [22]	68.8	97.0	80.5	82.4	92.6	87.2	98.0	91.5	94.6
Ours	81.2	62.3	70.5	73.9	94.8	83.1	92.7	92.3	92.5