From Whole Slide Imaging to Microscopy: Deep Microscopy Adaptation Network for Histopathology Cancer Image Classification

Yifan Zhang[1,2], Hanbo Chen[2], Ying Wei[2], Peilin Zhao[2], Jiezhang Cao[1], Xinjuan Fan[3], Xiaoying Lou[3], Hailing Liu[3], Jinlong Hou[2], Xiao Han[2], Jianhua Yao[2], Qingyao Wu[1(✉)], Mingkui Tan[1(✉)], and Junzhou Huang[2(✉)]

[1] South China University of Technology, Guangzhou, China
{qyw,mingkuitan}@scut.edu.cn
[2] Tencent AI Lab, Shenzhen, China
jzhuang@uta.edu
[3] Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

Abstract. Deep learning (DL) has achieved remarkable performance on digital pathology image classification with whole slide images (WSIs). Unfortunately, high acquisition costs of WSIs hinder the applications in practical scenarios, and most pathologists still use microscopy images (MSIs) in their workflows. However, it is especially challenging to train DL models on MSIs, given limited image qualities and high annotation costs. Alternatively, directly applying a WSI-trained DL model on MSIs usually performs poorly due to huge gaps between WSIs and MSIs. To address these issues, we propose to exploit deep unsupervised domain adaptation to adapt DL models trained on the labeled WSI domain to the unlabeled MSI domain. Specifically, we propose a novel Deep Microscopy Adaptation Network (DMAN). By reducing domain discrepancies via adversarial learning and entropy minimization, and alleviating class imbalance with sample reweighting, DMAN can classify MSIs effectively even without MSI annotations. Extensive experiments on colon cancer diagnosis demonstrate the effectiveness of DMAN and its potential in customizing models for each pathologist's microscope.

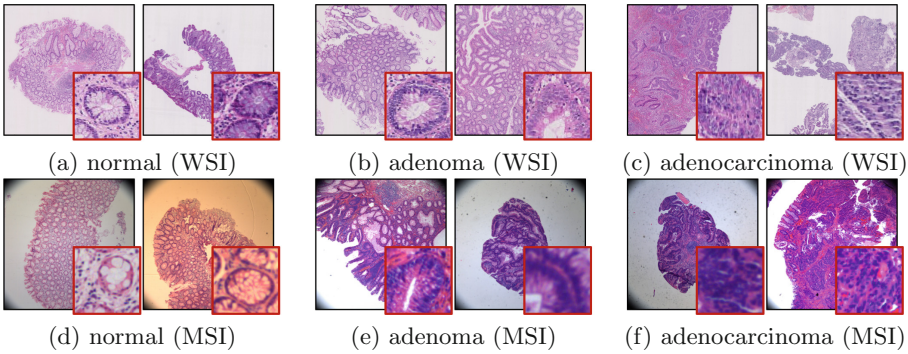
Keywords: Histopathology image classification · Unsupervised domain adaptation · Deep learning · Microscopy image · While slide image

1 Introduction

Histopathology image is a gold standard for clinical diagnosis of cancer [1,2]. By examining processed tissue slides, pathologists are able to identify abnormal

Y. Zhang, H. Chen and Y. Wei are co-first authors.

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11764, pp. 360–368, 2019.
https://doi.org/10.1007/978-3-030-32239-7_40



(a) normal (WSI) (b) adenoma (WSI) (c) adenocarcinoma (WSI)

(d) normal (MSI) (e) adenoma (MSI) (f) adenocarcinoma (MSI)

Fig. 1. Examples of the inter-domain discrepancy between WSIs (a–c) and MSIs (d–f), and the intra-domain discrepancy in each category. Bottom right of each subfigure shows the zoomed view.

tissues, pinpoint cancer types, and differentiate cancer stages at the cell level. Notably, this task is highly labor intensive and requires extensive expertise. Hence, computer-aided diagnosis (CAD) has been in great demand [3–5].

Despite success of deep learning (DL) [6–8] in medical image analysis, it hinges on massive annotated images for training [2]. Thanks to digital slide scan devices and hence a growing number of labeled whole slide images (WSIs), remarkable performance has been achieved on digital pathology image classification (DPIC) tasks, such as colon cancer diagnosis [9] and survival analysis [10].

However, due to high acquisition costs of WSIs, pathologists still largely rely on microscopes in their workflows. This brings the need for CAD systems for microscopy images (MSIs), where MSIs captured by digital cameras are fed into computers as a stream for analysis. Unfortunately, it is even more challenging to train DL models on MSIs due to: (1) variances of microscope devices in, *e.g.,* light sources, scope shades and view fields; (2) varied preferences of pathologists, *e.g.,* some pathologists prefer darker light; (3) wild imaging environments with white noises, motion blurs and losses of focus. To combat these challenges, one attractive option is to customize models for each pathologist with a small number of MSIs gathered on his or her own device. However, annotation costs for such customized MSIs are inevitably expensive. To this end, we are motivated to leverage labeled data in the WSI domain to improve the performance in the unlabeled MSI domain. This problem, known as deep unsupervised domain adaptation (DUDA) [11–13], remains largely unexplored for MSIs [1].

DUDA from WSIs to MSIs poses three challenges. As shown in Fig. 1, the first challenge is the inter-domain discrepancy, mainly derived from different imaging devices and techniques [1]. In this regard, directly applying a WSI-trained model to MSIs tends to perform poorly and be impractical. The second challenge lies in the intra-domain discrepancy, which originates from inconsistent data preparations, such as tissue collections, sectioning and staining [14]. Such inconsistency would result in intra-class inhomogeneity and raise the difficulty

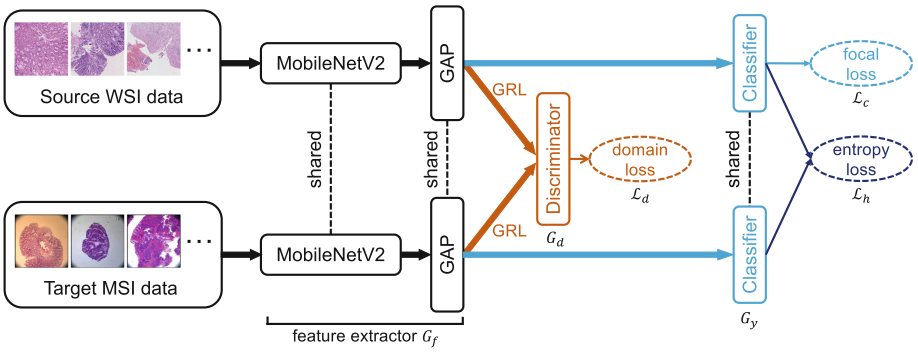


Fig. 2. The scheme of Deep Microscopy Adaptation Network, where GAP means Global Average Pooling [6] and GRL denotes Gradient Reverse Layer [11]. Considering the practical requirement of low computational cost and high inference speed, we implement MobileNetV2 [6], a highly efficient deep convolutional network, to extract features and predict classes.

of DPIC [15]. Lastly, class imbalance makes it particularly difficult to classify minor but important categories [2], *e.g.,* the adenocarcinoma in Fig. 1. However, most DUDA methods for general images [11–13] and for WSIs [14,16,17] neglect the intra-domain discrepancy and class imbalance, hence performing poorly in this task.

To solve these challenges, we propose a novel Deep Microscopy Adaptation Network (DMAN) algorithm. Specifically, we minimize both domain discrepancies via adversarial learning and entropy minimization, while alleviating class imbalance with sample reweighting. In this way, DMAN is able to learn domain-invariant and discriminative features which contribute to DPIC of MSIs with only labeled WSIs. Notably, this approach eliminates the need for annotating MSIs and helps to customize models for each pathologist's microscope, thus showing great potential for real-world applications.

2 Method

2.1 Problem Definition

This paper studies deep unsupervised domain adaptation (DUDA) from WSIs to MSIs. In this case, WSIs are well-labeled as the source domain, while the target domain (*i.e.,* MSIs) is fully unlabeled. Formally, we formulate source data as $\mathcal{D}_s = \{\mathbf{x}_i^s, y_i^s\}_{i=1}^{n_s}$, where $\mathbf{x}_i^s$ denotes the i-th WSI with y_i^s as its label, and n_s is the total number of WSIs. Meanwhile, target data is denoted as $\mathcal{D}_t = \{\mathbf{x}_j^t\}_{j=1}^{n_t}$, where $\mathbf{x}_j^t$ is the j-th MSI and n_t is the number of unlabeled MSIs. Both domains have the same label space with C classes. The goal of this paper is to learn a deep neural network, $G_y(G_f(\mathbf{x})) \to y$, which learns domain-invariant features $G_f(\mathbf{x})$ for both domains, so that the classifier $G_y(\cdot)$ trained with labeled WSIs can

also apply to unlabeled MSIs. To this end, we propose a novel Deep Microscopy Adaptation Network (DMAN).

2.2 Deep Microscopy Adaptation Network

As shown in Fig. 2, DMAN consists of three key components: a deep convolutional network G_f for feature representation, a classifier G_y for prediction, and a domain discriminator G_d to discriminate features of WSIs from those of MSIs.

To overcome the three challenges described in Sect. 1, *i.e.,* inter-domain discrepancy, intra-domain discrepancy, and class imbalance for classification, we train DMAN with the following three strategies, respectively. First, we adopt domain adversarial learning [13] to enforce G_f to learn domain-invariant features, so that the inter-domain discrepancy is minimized. On one hand, a domain discriminator G_d is trained by minimizing a domain classification loss $\mathcal{L}_d$ to adequately distinguish feature representations between WSIs and MSIs; on the other hand, the feature extractor G_f is trained to confuse the discriminator by maximizing $\mathcal{L}_d$. Following [11,14], we implement the adversarial optimization with a gradient reversal layer (GRL) which reverses the gradient $\nabla\mathcal{L}_d$ when backpropagating $\nabla\mathcal{L}_d$ to G_f. Second, to alleviate the intra-domain discrepancy and encourage intra-class homogeneity, we train the base network (G_f, G_y) by minimizing the prediction entropy $\mathcal{L}_h$. Third, we also exploit a class classification loss $\mathcal{L}_c$, namely focal loss, to train the imbalance-aware and discriminative base network (G_f, G_y). We summarize the overall optimization procedure as follows:

$$\begin{aligned}(\hat{\theta}_f, \hat{\theta}_y) &= \underset{\theta_f,\theta_y}{\operatorname{argmin}} \underbrace{-\alpha\mathcal{L}_d}_{\text{domain loss}} + \underbrace{\beta\mathcal{L}_h}_{\text{entropy loss}} + \underbrace{\mu\mathcal{L}_c}_{\text{focal loss}}, \\ (\hat{\theta}_d) &= \underset{\theta_d}{\operatorname{argmin}} \underbrace{\alpha\mathcal{L}_d}_{\text{domain loss}}, \end{aligned} \tag{1}$$

where θ_f, θ_y and θ_d indicate parameters of G_f, G_y and G_d, respectively. Moreover, α, β and μ are trade-off parameters. Note that the whole training process can be implemented with standard backpropagation in an end-to-end manner.

We next detail domain classification loss $\mathcal{L}_d$ in Sect. 2.3, entropy loss $\mathcal{L}_h$ in Sect. 2.4 and class classification loss $\mathcal{L}_c$ in Sect. 2.5.

2.3 Adversarial Learning for Inter-domain Adaptation

Diverse imaging devices and techniques intrinsically result in huge domain discrepancies between WSIs and MSIs. To resolve the discrepancies, as we mentioned above, we resort to domain adversarial learning. Specifically, as shown in Eq. (1), the key is to train the domain discriminator G_d and the feature extractor G_f via a minimax optimization problem regarding the domain loss. To this end, most existing methods [11,13] adopt the generative adversarial loss [7] as the domain classification loss $\mathcal{L}_d$:

$$-\frac{1}{n_t}\sum_{\mathbf{x}_j^t\in\mathcal{D}_t}\log\left(G_d(G_f(\mathbf{x}_j^t))\right) - \frac{1}{n_s}\sum_{\mathbf{x}_i^s\in\mathcal{D}_s}\log\left(1-G_d(G_f(\mathbf{x}_i^s))\right), \tag{2}$$

where the domain label of the source domain is 0, and that of the target domain is 1. Such a loss with sigmoid cross-entropy only evaluates domain classification correctness but fails to measure the domain distance [18]. As a result, maximizing this loss may not guarantee an effective G_f capable of learning highly domain-invariant features. Instead, inspired by LSGAN [18], we propose to use the least square loss for inter-domain adaptation:

$$\mathcal{L}_d = \frac{1}{n_t} \sum_{\mathbf{x}_j^t \in \mathcal{D}_t} \big(G_d(G_f(\mathbf{x}_j^t))-1\big)^2 + \frac{1}{n_s} \sum_{\mathbf{x}_i^s \in \mathcal{D}_s} \big(G_d(G_f(\mathbf{x}_i^s))\big)^2. \tag{3}$$

This loss, directly matching a domain label with the prediction without sigmoid, preserves domain distance. In this sense, it not only improves domain confusion but also stabilizes training, thus boosting performance of DUDA.

2.4 Entropy Minimization for Intra-class Homogeneity

Color, scale, resolution and intensity variations are common in WSIs and MSIs. These variations intrinsically lead to intra-class inhomogeneity in DPIC [14], and thus degrade performance of existing deep algorithms [2]. To solve this issue, we propose to impose the following entropy loss:

$$\mathcal{L}_h = -\frac{1}{n_t + n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s, \mathcal{D}_t} \sum_{k=1}^{C} G_y^k(G_f(\mathbf{x}_i)) \log \big(G_y^k(G_f(\mathbf{x}_i))\big), \tag{4}$$

where C denotes the class number and $G_y^k(G_f(\cdot))$ indicates the predicted probability of class k. Note that, the conditional entropy is a measure of class overlaps, so minimizing entropy encourages small overlaps among classes but high compactness within a class [19], and thus alleviate intra-class inhomogeneity.

2.5 Focal Loss for Class-Imbalance Classification

We next detail class classification loss $\mathcal{L}_c$ based on only labeled WSIs $(\mathbf{x}_i^s, y_i^s)$. Most existing DL models for medical image classification adopt cross-entropy:

$$-\frac{1}{n_s} \sum_{(\mathbf{x}_i^s, y_i^s) \in \mathcal{D}_s} \sum_{k=1}^{C} y_{i,k}^s \log \big(G_y^k(G_f(\mathbf{x}_i^s))\big), \tag{5}$$

where $y_{i,k}^s$ means the label of the i-th source sample regarding class k. Although cross-entropy performs well in many class-balanced classification tasks, it ignores the class-imbalance issue, which is quite common in DPIC [1]. One possible solution to this issue is cost-sensitive learning, which assigns uneven misclassification costs for different classes. One critical problem here is how to define costs for multiple classes, since the ground-truth distribution of classes is unknown and complicated in real-world DPIC tasks.

Our solution is motivated by an observation that class imbalance intrinsically makes classification of minority classes more difficult. As a result, the predicted probabilities of minority classes would be lower than those of majority classes [8]. Hence, if we assign higher costs for the predictions with low probabilities but assign low costs for those with high probabilities, we can redress the imbalance better. Inspired by this observation, we propose to use the focal loss [8] as follows:

$$\mathcal{L}_c = -\frac{1}{n_s} \sum_{(\mathbf{x}_i^s, y_i^s) \in \mathcal{D}_s} \sum_{k=1}^{C} y_{i,k}^s (1 - G_y^k(G_f(\mathbf{x}_i^s)))^\gamma \log \left(G_y^k(G_f(\mathbf{x}_i^s)) \right), \tag{6}$$

where γ is a hyperparameter to determine the degree to which classification focuses on minority classes. To avoid excessive cost mitigation on majority classes, a balanced value of γ is expected. By minimizing this loss, DMAN focuses more on minority classes and thus deals with class imbalance better.

3 Experimental Results

We evaluate our method on the following settings.

Dataset: 303 H&E stained histopathology slides, diagnosed as 3 types of colon polyps (normal, adenoma, and adenocarcinoma), are provided by the Sixth Affiliated Hospital of Sun Yat-sen University. WSI of each slide is acquired in 40× magnification scale (229 nm/pixel) by the Hamamatsu NanoZoomer 2.0-RS scanner. ROIs corresponding to the 3 types are then annotated by experts on WSI scans with our in-house tool. MSIs of 30 slides are acquired with microscope in 10× magnification scale (FOV: 2.73×2.73 mm^2, matrix size: 2048×2048). Specifically, we focus on 10× magnification scale, since it is the preferred scale for pathologist's diagnosis. For consistency, WSIs are down-sampled to the same resolution as MSIs. Then, a sliding window crops MSIs and annotated WSI regions into patches of size 512. The label of each patch is defined by the annotation in its center. WSI and MSI patches acquired from 15 slides are taken as the testing set, and the rest serves as the training set. Data statistics are shown in Table 1.

Table 1. Statistics of dataset

Domain	Training set				Test set			
	normal	adenoma	adenocarcinoma	Total	normal	adenoma	adenocarcinoma	Total
WSI	36,047	3,627	3,080	42,754	1,944	201	223	2,368
MSI	2,696	1,042	1,084	4,822	1,110	487	713	2,310

Baselines: We first compare DMAN with two baselines directly trained on WSIs, including **Source-only-C** (with cross-entropy loss) and **Source-only** (with focal loss). We also evaluate two other baselines that harness the collective

power of WSIs and annotated MSIs, including **Fine-tuning** (finetuning the model on WSIs with 500 labeled MSIs), and **Mix** (training with a mixture of WSIs and 500 labeled MSIs). We also compare DMAN with several state-of-the-art DUDA methods, including **DDC** [12], **DANN** [11], and **ADDA** [13]. For fair comparison, all methods employ the same network structure to DMAN but with different losses and optimization rules. For completeness, we also evaluate two variants of DMAN, *i.e.,* **DMAN-F** (using focal loss as the domain loss) and **DMAN-H** (without the entropy loss).

Implementation Details and Metrics: We implement DMAN with Tensorflow. The discriminator consists of three fully connected layers with 1024, 1024 and 1 hidden units, respectively. We use Adam optimizer with the batch size 16 and fixed learning rate 10^{-5} on a single GPU. Moreover, we set $\mu = 1$, $\gamma = 2$, $\alpha = 0.1$ and $\beta = 0.1$. Ablation studies are not included due to space limit. We use Accuracy, mean Precision, mean Recall and F1-measure as the **metrics**.

3.1 Evaluation on Digital Pathology Image Classification

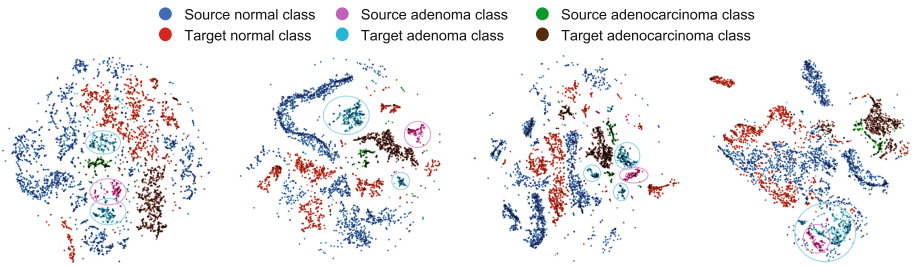
From the results in Table 2, we are able to draw several conclusions. **First,** we observe apparent discrepancies between WSIs and MSIs, evidenced by the performance gap between testing Source-only on WSIs and on MSIs. Moreover, the superiority of Source-only over Source-only-C on MSIs confirms the effectiveness of the focal loss in alleviating class imbalance. **Second,** fine-tuning with limited labeled MSIs improves slightly over Source-only due to over-fitting. Although Mix performs better than Fine-tuning, it also suffers from over-fitting and performs unsatisfactorily. **Third,** all DUDA methods outperform Source-only. Notably, DMAN and DANN (without labeled MSIs) even outperform Mix (with labeled MSIs). These observations demonstrate positive contributions of DUDA to DPIC tasks. **Lastly,** DMAN-based methods outperform all other baselines. This validates not only the superiority of our method but also its potential for customizing to each pathologist's microscope. Moreover, the improvement of DMAN over DMAN-F and DMAN-H validates the necessity of least square domain loss and entropy loss.

3.2 Visualization of Feature Representations

We visualize t-SNE embeddings of learned features after the GAP layer. Take the adenoma class as an example. Figure 3(a) displays the domain discrepancy between WSIs and MSIs. Figure 3(b–c) show that DDC and DANN fail to reduce the inter-domain discrepancy and suffer from intra-class inhomogeneity. Figure 3(d) displays that DMAN can well resolve both inter-domain and intra-domain discrepancies.

Table 2. Comparisons on DPIC in terms of four metrics (%), where Labeled Set and Test Set indicate the labeled set for training and evaluation, respectively.

Methods	Labeled set	Test set	Accuracy	Precision	Recall	F1-measure
Source-only	WSI	WSI	93.67	83.77	93.03	87.16
Source-only-C	WSI	MSI	82.21	87.78	76.15	78.55
Source-only	WSI	MSI	83.72	88.48	78.59	81.10
Fine-tuning	Both	MSI	85.02	83.47	85.40	84.15
Mix	Both	MSI	87.53	86.68	87.16	86.91
DDC	WSI	MSI	87.14	85.96	88.60	86.96
ADDA	WSI	MSI	85.11	86.55	83.51	84.72
DANN	WSI	MSI	88.01	88.60	87.17	87.82
DMAN-F	WSI	MSI	88.57	89.27	87.83	88.50
DMAN-H	WSI	MSI	89.09	90.17	88.21	89.10
DMAN	WSI	MSI	**90.48**	**90.67**	**90.35**	**90.50**



(a) Source-only (b) DDC (c) DANN (d) DMAN

Fig. 3. t-SNE plots in terms of classes and domains, where clusters of the adenoma are circled. The closer samples across domains with the same class are, the more effective domain adaptation is.

4 Conclusion

We have proposed the novel DMAN[1] for adapting WSI-trained networks to predict MSIs. In detail, we propose to reduce inter-domain discrepancy with adversarial learning and diminish intra-domain discrepancy using entropy minimization. Moreover, we exploit the focal loss to effectively alleviate the class-imbalance issue. In this way, DMAN conduct DPIC of MSIs effectively based on only labeled WSIs. Promising experiments demonstrate the effectiveness of DMAN and its potential in customizing models to each pathologist's microscope.

[1] This work was partially supported by National Natural Science Foundation of China (NSFC) (61876208, 61502177 and 61602185), Guangdong Provincial Scientific and Technological Fund (2017B090901008, 2017A010101011, 2017B090910005, 2018B010107001), Pearl River S&T Nova Program of Guangzhou 201806010081, CCF-Tencent Open Research Fund RAGR20170105, Program for Guangdong Introducing Innovative and Entrepreneurial Teams 2017ZT07X183.

References

1. Xing, F., Xie, Y., Su, H., Liu, F., Yang, L.: Deep learning in microscopy image analysis: a survey. TNNLS **29**, 1–19 (2018)
2. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A., Ciompi, F., Snchez, C.I.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
3. Becker, C., et al.: Domain adaptation for microscopy imaging. TMI **34**, 1125–1139 (2015)
4. Bermúdez-Chacón, R., Becker, C., Salzmann, M., Fua, P.: Scalable unsupervised domain adaptation for electron microscopy. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 326–334. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_38
5. Heimann, T., Mountney, P., John, M., Ionasec, R.: Learning without labeling: domain adaptation for ultrasound transducer localization. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8151, pp. 49–56. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40760-4_7
6. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: inverted residuals and linear bottlenecks. In: CVPR, pp. 4510–4520 (2018)
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Bengio, Y.: Generative adversarial nets. In: NeurIPS, pp. 2672–2680 (2014)
8. Lin, T.Y., et al.: Focal loss for dense object detection. In: ICCV (2017)
9. Armin, M.A., et al.: Visibility map: a new method in evaluation quality of optical colonoscopy. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 396–404. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_49
10. Li, R., Yao, J., Zhu, X., Li, Y., Huang, J.: Graph CNN for survival analysis on whole slide pathological images. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 174–182. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_20
11. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: ICML, pp. 1180–1189 (2015)
12. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: maximizing for domain invariance. arXiv:1412.3474 (2014)
13. Tzeng, E., et al.: Adversarial discriminative domain adaptation. In: CVPR (2017)
14. Lafarge, M.W., Pluim, J.P.W., Eppenhof, K.A.J., Moeskops, P., Veta, M.: Domain-adversarial neural networks to address the appearance variability of histopathology images. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS 2017. LNCS, vol. 10553, pp. 83–91. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_10
15. Mangin, J.F.: Entropy minimization for automatic correction of intensity nonuniformity. In: Workshop on MMBIA (2000)
16. Wollmann, T., Eijkman, C.S., Rohr, K.: Adversarial domain adaptation to improve automatic breast cancer grading in lymph nodes. In: ISBI, pp. 582–585 (2018)
17. Ren, J., Hacihaliloglu, I., Singer, E.A., Foran, D.J., Qi, X.: Adversarial domain adaptation for classification of prostate histopathology whole-slide images. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 201–209. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_23
18. Mao, X., et al.: Least squares generative adversarial networks. In: ICCV (2017)
19. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. In: NeurIPS (2005)