Unsupervised Instance Segmentation in Microscopy Images via Panoptic Domain Adaptation and Task Re-weighting

Dongnan Liu[1] Donghao Zhang[1] Yang Song[2] Fan Zhang[3] Lauren O'Donnell[3]
Heng Huang[4] Mei Chen[5] Weidong Cai[1]

[1]School of Computer Science, University of Sydney, Australia
[2]School of Computer Science and Engineering, University of New South Wales, Australia
[3]Brigham and Women's Hospital, Harvard Medical School, USA
[4]Department of Electrical and Computer Engineering, University of Pittsburgh, USA
[5]Microsoft Corporation, USA

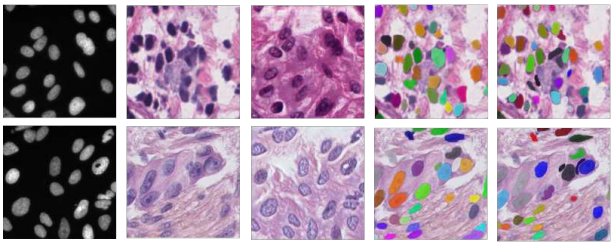
{dliu5812, dzha9516}@uni.sydney.edu.au, yang.song1@unsw.edu.au
{fzhang, odonnell}@bwh.harvard.edu, henghuanghh@gmail.com
may4mc@gmail.com, tom.cai@sydney.edu.au

Abstract

Unsupervised domain adaptation (UDA) for nuclei instance segmentation is important for digital pathology, as it alleviates the burden of labor-intensive annotation and domain shift across datasets. In this work, we propose a Cycle Consistency Panoptic Domain Adaptive Mask R-CNN (CyC-PDAM) architecture for unsupervised nuclei segmentation in histopathology images, by learning from fluorescence microscopy images. More specifically, we first propose a nuclei inpainting mechanism to remove the auxiliary generated objects in the synthesized images. Secondly, a semantic branch with a domain discriminator is designed to achieve panoptic-level domain adaptation. Thirdly, in order to avoid the influence of the source-biased features, we propose a task re-weighting mechanism to dynamically add trade-off weights for the task-specific loss functions. Experimental results on three datasets indicate that our proposed method outperforms state-of-the-art UDA methods significantly, and demonstrates a similar performance as fully supervised methods.

1. Introduction

Nuclei instance segmentation in histopathology images is an important step in the digital pathology workflow. Pathologists are able to diagnose and prognose cancers according to mitosis counts, the morphological structure of each nucleus, and spatial distribution of a group of nuclei [7, 25, 5, 1, 34]. Currently, supervised learning-based methods for nuclei instance segmentation are prevalent as they are efficient while preserving high accuracy [24, 35, 3, 9, 33, 50, 29, 28]. However, their performance heavily relies on large-scale training data, which requires expertise for annotation. This process is time-consuming and labor-intensive due to the complicated cellular structures, as shown in Fig. 1(b), and large image sizes. For example, annotating a histopathology dataset with 50 images and 12M pixels costs a pathologist 120 to 230 hours [16]. Moreover, in real clinical studies, even one whole slide image in $40\times$ objective magnification contains 1B pixels [10]. Therefore, investigating methods without depending on histopathology annotations is necessary. It can help pathologists to reduce the workload, and tackle the issue of lacking histopathology annotations.



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

Figure 1. Example images of our proposed framework. (a) fluorescence microscopy images; (b) real histopathology images; (c) our synthesized histopathology images; (d) nuclei segmentation generated by our proposed UDA method; (e) ground truth.

The recently proposed unsupervised domain adaptation

(UDA) methods tackle this issue by conducting supervised learning on the source domain and obtain a good performance model for the target domain without annotations [36, 8, 45]. Currently, UDA reduces distances between the distribution of feature maps of the source and target domains. In addition, some other methods focus on the pixel-to-pixel translation from the source domain images to the target ones, for aligning cross-domain image appearances [20, 52]. For these methods, there still remain some differences in the distributions between the synthesized and real images, due to the imperfect translations [15, 2, 21].

To incorporate the benefits of the image translation and the UDA methods, several works have been proposed to learn the domain-invariant features between the target and the synthesized target-like images [15, 21, 2]. Such methods achieve state-of-the-art performance on UDA classification, object detection, and semantic segmentation tasks. However, currently there is a lack of UDA methods specifically designed for instance segmentation, and directly extending the existing UDA methods on object detection [4, 21, 14] to the UDA nuclei instance segmentation task still suffers from challenges. First, existing UDA object detection methods focus on alleviating the domain bias at the image level (image contrast, brightness, etc.) and the instance level (object scale, style, etc.) [21, 4, 14]. They ignore the domain shift at the semantic level, such as the relationship between the foreground and background, and the spatial distribution of the objects. Second, these UDA object detection methods are multi-task learning paradigms, which optimize different loss functions simultaneously. If the feature extractors fail to generate domain-invariant features in some training iterations, then back-propagating the weights according to the task loss functions in these iterations causes the model bias towards the source domain.

To solve the aforementioned problems in UDA nuclei instance segmentation tasks in histopathology images, we propose a Cycle-Consistent Panoptic Domain Adaptive Mask R-CNN (CyC-PDAM) model. As none of the previous UDA methods are specially designed for instance segmentation, we extend the CyCADA [15] to an instance segmentation version based on Mask R-CNN [11], as our baseline. In our CyC-PDAM, we firstly propose a simple nuclei inpainting mechanism to remove the auxiliary nuclei in the synthesized histopathology images. Second, inspired by the panoptic segmentation architectures [23, 22], we propose a semantic-level adaptation module for domain-invariant features based on the relationship between the foreground and the background. By reconciling the domain-invariant features at the semantic and instance levels, our proposed CyC-PDAM achieves panoptic-level domain adaptation. Furthermore, a task re-weighting mechanism is proposed to reset the importance for each task loss. During training, the specific task losses are down-weighted if the features for task predictions are not domain-invariant and source-biased, and up-weighted if the features are hard to differentiate.

To prove the effectiveness of our proposed CyC-PDAM architecture, experiments have been conducted on three public datasets for unsupervised nuclei instance segmentation of histopathology images on two different datasets by unsupervised domain adaptation from a fluorescence microscopy image dataset. Unlike histopathology images, no structures are similar to the nuclei in the background of fluorescence microscopy images, due to the differences between image acquisition techniques, as shown in Fig. 1(a). It is much easier to obtain manual annotation for the fluorescence microscopy images compared with histopathology images, therefore it is chosen as our source domain.

Our contribution is summarized as follows: (1) We propose a CyC-PDAM model for UDA nuclei instance segmentation in histopathology images. To our best knowledge, this is the first UDA instance segmentation method. (2) A simple nuclei inpainting mechanism is proposed to remove false-positive objects in the synthesized images. (3) Our CyC-PDAM produces domain-invariant features at the panoptic level, by integrating the instance-level adaptation with a newly proposed semantic-level adaptation module. (4) A task re-weighting mechanism is proposed to alleviate the domain bias towards the source domain. (5) Compared with state-of-the-art UDA methods, our proposed CyC-PDAM paradigm outperforms them by a large margin. Moreover, it achieves competitive performance compared with state-of-the-art fully supervised methods for nuclei segmentation.

2. Related Work

2.1. Domain Adaptation for Natural Images

Domain adaptation aims at transferring the knowledge learned from one labeled domain to another without annotation [36]. Recently, UDA methods have reduced the cross-domain discrepancies based on the content in the feature level and the appearance in the pixel level. For the feature-level adaptation, adversarial learning for domain-invariant features [8, 45], Maximum Mean Discrepancy minimization (MMD) [32], local pattern alignment [48], and cross-domain covariance alignment [42] are widely employed for classification tasks. In addition, domain adaptation is further employed for other tasks such as semantic segmentation [46, 26] and object detection [4, 21, 19, 47]. In semantic segmentation tasks, the segmentation results are forced to be domain-invariant, together with intermediate feature maps [26, 46, 44]. Additionally, ADVENT [46] further minimized the Shannon entropy for the semantic segmentation predictions in source and target domains to alleviating the cross-domain discrepancy. For object detection, a domain adaptive Faster R-CNN [40], consisting of

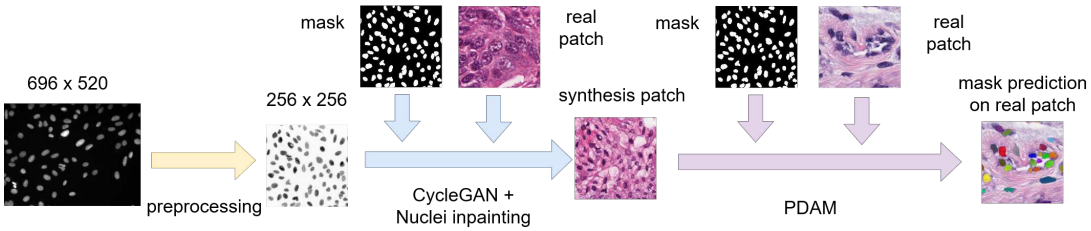


Figure 2. Overall architecture for our proposed CyC-PDAM architecture. The annotations of the real histopathology patches are not used during training.

the image- and instance-level adaptions, was usually proposed for domain-invariant features of the whole image and each object [4, 21, 14]. On the other hand, image-to-image translation addresses the domain adaptation problems in the pixel level by generating target-like images and training task-specific fully supervised models on them [30, 17, 20, 52, 33, 37]. However, domain bias still exists because of imperfect translation. Moreover, several methods have been proposed to align the feature-level adaptation with the pixel-level one, by learning domain-invariant features between the target images and the synthesized images [15, 21, 2].

2.2. Domain Adaptation for Medical Images

Unsupervised domain adaptation for medical image analysis has rarely been explored [39, 51, 2, 18, 16]. [39] and [18] solve the UDA histopathology images classification problems with GAN based architectures. In addition, DAM [6] is proposed to generate domain-invariant intermediate features and model predictions, for UDA semantic segmentation in CT images. With the help of cycle-consistency reconstruction, TD-GAN [51] and SIFA [2] are proposed for semantic segmentation on different medical images, with both pixel- and feature-level adaptations. However, none of them is designed for UDA nuclei instance segmentation. Even though Hou *et al*. [16] proposed to train a GAN based refiner and a nuclei segmentation model with the synthesized histopathology images for unsupervised nuclei instance segmentation, their paradigm only contains pixel-level adaptation and is still not capable for minimizing the domain gap in the feature level. In this work, we therefore propose a CyC-PDAM paradigm for UDA nuclei instance segmentation, which alleviates the domain bias issue in the pixel and feature levels.

3. Methods

Our proposed architecture is based on CyCADA and we fuse CyCADA with the instance segmentation framework Mask R-CNN. Furthermore, we improve it with nuclei inpainting mechanism, panoptic-level domain adaptation, and task re-weighting mechanism. Fig. 2 illustrates the overall architecture of our approach.

3.1. CyCADA with Mask R-CNN

Name	Hyperparamaters	Output size
Input		$256 \times 8 \times 8$
Conv1	$k = (3, 3), s = 1, p = 1$	$256 \times 8 \times 8$
Conv2	$k = (3, 3), s = 1, p = 1$	$512 \times 8 \times 8$
Conv3	$k = (3, 3), s = 1, p = 1$	$512 \times 8 \times 8$
Conv4	$k = (1, 1), s = 1, p = 0$	$2 \times 8 \times 8$

Table 1. The parameters for each block in the image-level discriminator for PDAM. k, s, and p denote the kernel size, stride, and padding of the convolution operation, respectively.

As there is no UDA architectures targeting instance-level segmentation, we firstly design a domain adaptive Mask R-CNN. The backbone of the Mask R-CNN in this work is constructed with ResNet101 [12] and Feature Pyramid Network (FPN) [27]. Inspired by the previous UDA methods for object detection [4, 21], we add one discriminator after FPN for the image-level adaptation, and the other after the instance branch for instance-level adaptation, as shown in Fig. 3. For the image-level adaptation, the multi-resolution feature maps of the FPN output are firstly downsampled to the size 8×8 with average pooling, and then summed together for the image-level discriminator. The image-level discriminator consists of 4 convolutional layers (details in Table 1) and a gradient reversal layer (GRL) for adversarial learning. In the instance-level adaptation, the $14 \times 14 \times 256$ feature map in the mask branch is downscaled to the size $2 \times 2 \times 256$ with average pooling and then resized to 1024×1, to sum with the 1024×1 feature from the bounding box branch. The instance-level discriminator consists of 3 fully connected layers and a GRL, whose input is the summation of features mentioned above.

3.2. Nuclei Inpainting Mechanism

Even though CycleGAN is effective for synthesizing histopathology-like images, due to the large domain gap and nuclei number incompatibility between the source and target domains, the label space for the generated images sometimes changes after transferring from the source domain. For example, there are redundant and undesired nuclei in the synthesized images shown in Fig. 4. If these images are directly used to train the task-specific CNN with the origi-

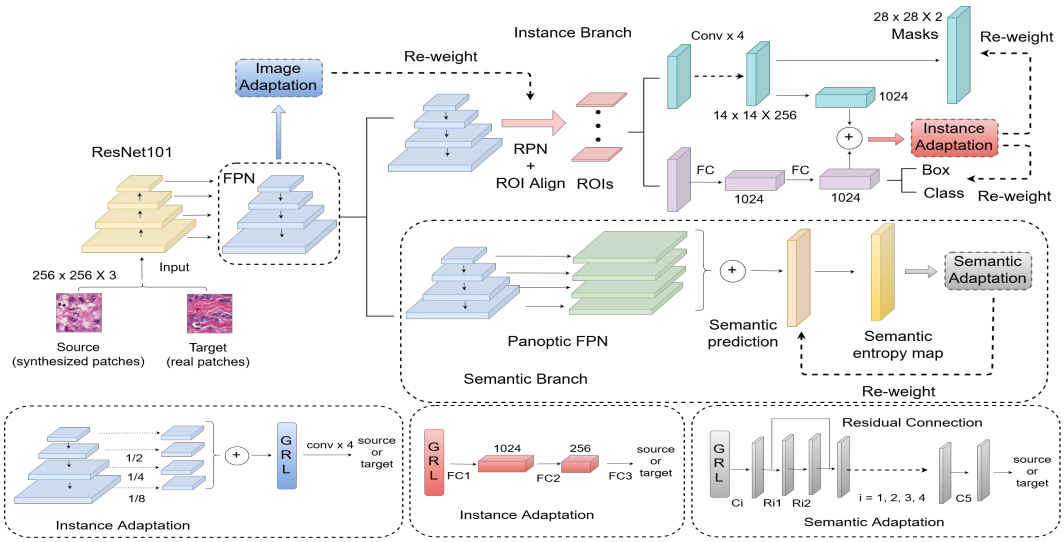


Figure 3. Detailed illustration of Panoptic Domain Adaptive Mask R-CNN (PDAM). Ci and FC represent a convolution layer, and a fully connected layer, respectively. $Ri1$ and $Ri2$ mean the first and second convolutional layers in the ith residual block, respectively. $ReLU$ and normalization layers after each convolutional block are omitted for brevity.

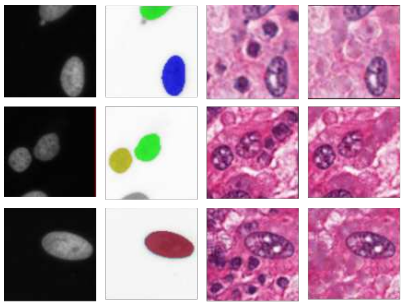


Figure 4. Visual results for the effectiveness of nuclei inpainting mechanism. (a) original fluorescence microscopy patches; (b) corresponding nuclei annotations; (c) initial synthesized images from CycleGAN; (d) final synthesized images after nuclei inpainting mechanism.

nal labels, the model is forced to regard redundant nuclei as background, even though they appear as real nuclei.

Therefore, we propose an auxiliary nuclei inpainting mechanism to remove the nuclei which only appear in the synthesized images without corresponding annotations. Denoting a raw synthesized histopathology image by CycleGAN as S_{raw} and its corresponding mask as M, we first obtain the mask predictions M_{aux} of all the auxiliary generated nuclei, formulated as:

$$M_{aux} = (otsu(S_{raw}) \cup M) - M \tag{1}$$

where $ostu(S_{raw})$ represents a binary segmentation method for S_{raw} based on Otsu threshold. In M_{aux}, only auxiliary nuclei without annotation is labeled. Then, we get the newly synthesized image S_{inp} after removing these nuclei, which can be represented as:

$$S_{inp} = inp(S_{raw}, M_{aux}) \tag{2}$$

where inp is a fast marching based method for inpainting objects [43], by replacing the pixel values for the auxiliary nuclei labeled in M_{aux} with them for the unlabeled background. Fig. 4 illustrates the visual effectiveness of our proposed nuclei inpainting mechanism. However, some background materials are labeled as false positive predictions in M_{aux}. Directly inpainting them makes the texture and appearance of synthesized images unrealistic, and enlarges the domain gap between the synthesized and real images. However, the image-level adaptation is able to address this issue by alleviating the domain bias on global visual information, such as curve, texture, and illumination. Our nuclei inpainting mechanism is time-efficient, which takes 0.09 second to process one single 256×256 synthesized histopathology patch, on average.

3.3. Panoptic Level Domain Adaptation

We define the semantic-level features of an image as the relationship between its foreground and background. In addition to the image- and feature-level domain bias, the domain shift at the semantic level also exists. Due to the differences in the nuclei objects and background between the synthesized and real histopathology images, domain adaptive Mask R-CNN mentioned in Sec. 3.1 suffers from domain bias in the semantic-level features, as the Mask R-CNN only focuses on the local features for each object and lacks a semantic view of the whole image. Inspired by the previous panoptic segmentation architecture, which unified the semantic and instance segmentation to

process the global and local features of the images, we propose a semantic-level adaptation to induce the model to learn domain-invariant features based on the relationship between the foreground and background. By incorporating the semantic- and instance-level adaptation, our panoptic domain adaptive method reduces the cross-domain discrepancies in a global and local view.

As shown in Fig. 3, a semantic branch for semantic segmentation prediction is added to the output of the FPN. Our semantic branch has the same implementation as [22]. As the fluorescence microscopy images and histopathology images can both be acquired from tissue samples and they can show complementary and correlated information, the semantic segmentation label spaces of the synthesized and real histopathology images have a strong similarity. In addition, aligning the cross-domain entropy distributions helps to minimize the entropy prediction in the target domain, which makes the model suitable for the target images [46]. Therefore, we use the Shannon entropy [41] of the softmax semantic predictions to induce the domain-invariant features to learn at the semantic level. Denoting the softmax semantic prediction as P and $P \in (0, 1)$, its Shannon entropy is defined as: $-plog(p)$.

Fig. 3 and Table 2 indicate the detailed structure of the discriminator for semantic level adaptation. We employ residual connected CNN blocks to avoid gradient vanishing [12, 13]. To make the adversarial learning more stable, instead of bilinear interpolation, we use stride convolutional layers for upsampling. Finally, the domain label is predicted as a 16×16 patch. Due to the small mini-batch size, the patch-based domain label prediction increases the number of training samples, to avoid overfitting.

Name	Hyperparamaters	Output size
Input		$2 \times 256 \times 256$
C1	$k = (7, 7), s = 2, p = 3$	$64 \times 128 \times 128$
R11 and R12	$k = (3, 3), s = 1, p = 1$	$64 \times 128 \times 128$
C2	$k = (5, 5), s = 2, p = 2$	$128 \times 64 \times 64$
R21 and R22	$k = (3, 3), s = 1, p = 1$	$128 \times 64 \times 64$
C3	$k = (5, 5), s = 2, p = 2$	$256 \times 32 \times 32$
R31 and R32	$k = (3, 3), s = 1, p = 1$	$256 \times 32 \times 32$
C4	$k = (5, 5), s = 2, p = 2$	$512 \times 16 \times 16$
R41 and R42	$k = (3, 3), s = 1, p = 1$	$512 \times 16 \times 16$
C5	$k = (1, 1), s = 1, p = 0$	$2 \times 16 \times 16$
Output		$2 \times 16 \times 16$

Table 2. The parameters for each block in the semantic-level discriminator for PDAM. k, s, and p follow the same convention as in Table 1.

3.4. Task Re-weighting Mechanism

In the previous UDA methods, the task-specific loss functions (segmentation, classification, and detection) are based on the source domain predictions. Even though several adversarial domain discriminators are employed to ensure the predicted feature maps are domain-invariant, the cross-domain discrepancies of these feature maps are still large in some training iterations, where the features are far from the decision boundaries of the domain discriminators. If the task-specific losses are updated to optimize the models with these easily-distinguished features, the models will bias towards the source images when testing it with the target data. To this end, we propose a task re-weighting mechanism to add a trade-off weight for each task-specific loss function according to the prediction of the domain discriminator. Denote the probability of the feature map before the final task prediction belonging to the source and target domains as p_s and p_t, respectively, and the task-specific loss function as L, then the re-weighted task-specific loss L_{rw} is:

$$L_{rw} = min(\frac{p_t}{p_s}, \beta)L = min(\frac{1 - p_s}{p_s}, \beta)L \tag{3}$$

where β is a threshold value to avoid the $\frac{1-p_s}{p_s}$ becoming large and making the model collapse, when $p_s \to 0$. According to Eq. 3, if the feature map deciding the task prediction belongs to the source domain ($p_s \to 1$), the loss function is then down-weighted, to alleviate the source-bias feature learning of the model. As illustrated in Fig. 3, the loss function for the region proposal network (RPN), semantic branch, and the instance branch are re-weighted by the prediction at the image-, semantic-, and instance-level domain discriminators, respectively.

3.5. Network Overview and Training Details

In our proposed CyC-PDAM, the CycleGAN has the same implementation as its original work [52]. When training the CycleGAN, the initial learning rate was set to 0.0001 for the first $1/2$ of the total training iterations, and linearly decayed to 0 for the other $1/2$.

The PDAM is trained with a batch size of 1 and each batch contains 2 images, one from the source and the other from the target domain. Due to the small batch size, we replace traditional batch normalization layers with group normalization [49] layers, with the default group number 32 as in [49].

The overall loss function of PDAM is defined as:

$$\begin{aligned} L_{pdam} = \alpha_{img}L_{rpn} + \alpha_{ins}L_{det} + \alpha_{sem}L_{(sem-seg)} \\ + \alpha_{da}(L_{(img-da)} + L_{(sem-da)} + L_{(ins-da)}) \end{aligned} \tag{4}$$

where L_{rpn} is the loss function for the RPN, L_{det} is the loss of class, bounding box, and instance mask prediction of Mask R-CNN, $L_{(sem-seg)}$ is the cross entropy loss for semantic segmentation, $L_{(img-da)}$, $L_{(sem-da)}$ and $L_{(ins-da)}$ are cross entropy losses for domain classification at image, semantic and instance levels. α_{img}, α_{ins}, and α_{isem} are calculated according to Eq. 3 for task re-weighting. In our experiment, we set β as 2. α_{da} is updated as:

$$\alpha_2 = \frac{2}{1 + exp(-10t)} - 1 \quad (5)$$

where t is the training progress and $t \in [0, 1]$. Thus α_{da} is gradually changed from 0 to 1, to avoid the noise from the unstable domain discriminators in the early training stage.

During training, the PDAM is optimized by SGD, with a weight decay of 0.001 and a momentum of 0.9. The initial learning rate is 0.002, with linear warming up in the first 500 iterations. The learning rate is then decreased to 0.0002 when it reaches $3/4$ of the total training iteration. During inference, only the original Mask R-CNN architecture is used with the adapted weight and all of the hyperparameters for testing are fine-tuned on the validation set. All of our experiments were implemented with Pytorch [38], on two NVIDIA GeForce 1080Ti GPUs.

4. Experiments

4.1. Datasets Description and Evaluation Metrics

Our proposed architecture was validated on three public datasets, referred to as Kumar [24], TNBC [35], and BBBC039V1 [31], respectively. Among them, Kumar and TNBC are histopathology datasets, while BBBC039V1 is a fluorescence microscopy dataset. Kumar was acquired from The Cancer Genome Atlas (TCGA) at $40\times$ magnification, containing 30 annotated 1000×1000 patches from 30 whole slide images of different patients. All these images are from 18 different hospitals and 7 different organs (breast, liver, kidney, prostate, bladder, colon, and stomach). In contrast to the disease variability in Kumar, the TNBC dataset especially focuses on Triple-Negative Breast Cancer (TNBC) [35]. In TNBC, there are 50 annotated 512×512 patches from 11 different patients from the Curie Institute at $40\times$ magnification. BBBC039V1 is about U2OS cells under a high-throughput chemical screen [31]. It contains 200 520×696 images about bioactive compounds, with the DNA channel staining of a single field of view.

For evaluation, we employ three commonly used pixel- and object-level metrics. Aggregated Jaccard Index (AJI) is an extended Jaccard Index for object-level evaluation [24], and object-level F1 score is the average harmonic mean between the precision and recall for each object. For pixel-level evaluation, we employ pixel-level F1 score for binarization predictions.

4.2. Experiment Setting

We conducted our experiments on two nuclei segmentation tasks: adapting from BBBC039V1 to Kumar, and from BBBC039V1 to TNBC. As the source domain in two experiments, 100 training images and 50 validation images from BBBC039V1 are used, following the official data split[1]. The annotations for Kumar and TNBC are not used during training the UDA architecture, only for evaluation.

The preprocessing for source fluorescence microscopy images has 3 steps. First, all images are normalized into range $[0, 255]$. Second, $10K$ patches in size 256×256 are randomly cropped from the 100 training images, with data augmentation including rotation, scaling, and flipping to avoid overfitting. Third, the patches with fewer than 3 objects are removed. For better synthesizing target-like histopathology images, we finally inverse the pixel value of foreground nuclei and background for all source fluorescence microscopy patches. For validation, 50 images in the BBBC039V1 validation set are transferred to synthesized histopathology images by CycleGAN and nuclei inpainting mechanism.

For the Kumar dataset as the target domain, we have the same data split as previous work in [24, 35], with 16 images for training, and 14 for testing. When training the model, totally $10K$ patches in size 256×256 are randomly cropped from the 16 training histopathology images, with basic data augmentation including flipping and rotation, to avoid overfitting. As for TNBC, we use 8 cases with 40 images for training, and the remaining 3 cases with 10 images for testing. To train the model with TNBC, $10K$ 256×256 patches are randomly extracted from the training images with basic data augmentation including flipping and rotation.

4.3. Comparison Experiments

4.3.1 Comparison with Unsupervised Methods

In this section, our proposed CyC-PDAM is compared with several state-of-the-art UDA methods, including CyCADA [15], Chen *et al*. [4], SIFA [2], and DDMRL [21]. As the original CyCADA focuses on classification and semantic segmentation, we extend it with Mask R-CNN for UDA instance segmentation, as described in Sec. 3.1. Chen *et al*. [4] are originally for UDA object detection based on Faster R-CNN, by adapting the features at the image and instance levels. For UDA instance segmentation, we replace the original VGG16 based Faster R-CNN with the same Mask R-CNN in our architecture, and the original image- and instance-level adaptation in [4] with ours in Sec. 3.1. SIFA [2] is a UDA semantic segmentation architecture for CT and MR images, with a pixel- and feature-level adaptation. In our experiment, we add the watershed algorithm to separate the touching objects in the semantic segmentation prediction of SIFA, for a fair comparison. DDMRL [21] learns multi-domain-invariant features from various generated domains for UDA object detection and it is extended for instance segmentation, in a similar way as CyCADA [15] and Chen *et al*. [4]. In addition, we also compared with Hou *et al*. [16], which is particularly designed for unsupervised nuclei segmentation in histopathology images. They trained a multi-task (segmentation, detection, and refinement) CNN

[1] https://data.broadinstitute.org/bbbc/BBBC039/

	$BBBC039 \rightarrow Kumar$			$BBBC039 \rightarrow TNBC$		
Methods	AJI	Pixel-F1	Object-F1	AJI	Pixel-F1	Object-F1
CyCADA [15]	0.4447 ± 0.1069	0.7220 ± 0.0802	0.6567 ± 0.0837	0.4721 ± 0.0906	0.7048 ± 0.0946	0.6866 ± 0.0637
Chen *et al*. [4]	0.3756 ± 0.0977	0.6337 ± 0.0897	0.5737 ± 0.0983	0.4407 ± 0.0623	0.6405 ± 0.0660	0.6289 ± 0.0609
SIFA [2]	0.3924 ± 0.1062	0.6880 ± 0.0882	0.6008 ± 0.1006	0.4662 ± 0.0902	0.6994 ± 0.0942	0.6698 ± 0.0771
DDMRL [21]	0.4860 ± 0.0846	0.7109 ± 0.0744	0.6833 ± 0.0724	0.4642 ± 0.0503	0.7000 ± 0.0431	0.6872 ± 0.0347
Hou *et al*. [16]	0.4980 ± 0.1236	0.7500 ± 0.0849	0.6890 ± 0.0990	0.4775 ± 0.1219	0.7029 ± 0.1262	0.6779 ± 0.0821
Proposed	$\mathbf{0.5610 \pm 0.0718}$	$\mathbf{0.7882 \pm 0.0533}$	$\mathbf{0.7483 \pm 0.0525}$	$\mathbf{0.5672 \pm 0.0646}$	$\mathbf{0.7593 \pm 0.0566}$	$\mathbf{0.7478 \pm 0.0417}$

Table 3. In comparison with other unsupervised methods on both two histopathology datasets.

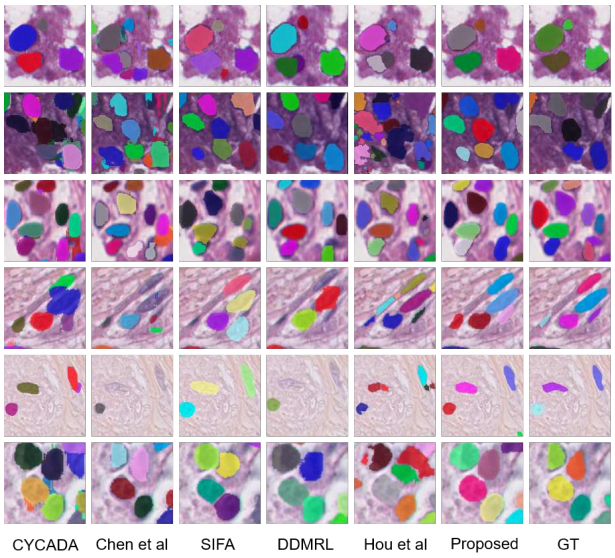


Figure 5. Visualization results for the comparison experiments. The first 3 rows are from Kumar dataset, and the last 3 rows are from TNBC.

architecture with their synthesized histopathology images from randomly generated binary nuclei masks.

Table 3 shows that our proposed method outperforms all the comparison methods by a large margin, on different histopathology datasets. In addition, the one-tailed paired t-test is employed to prove that all of our improvements are statistically significant, with all the p-values under 0.05. Chen *et al*. [4] learns the domain-invariant features at the image and instance levels. However, due to the large differences between the fluorescence microscopy and real histopathology images, feature-level adaptation only is not enough to reduce the domain gap. With pixel-level adaptation on appearance, all the other methods achieve better performance. Compared with the baseline method CyCADA [15], our CyC-PDAM has a large improvement of $6 - 12\%$, due to the effectiveness of our proposed nuclei inpainting mechanism, panoptic-level adaptation, and task re-weighting mechanism. SIFA [2] focuses on domain-invariant features in the image and semantic levels, with a UDA semantic segmentation structure. As there exists a large number of nuclei objects in the histopathology images, the effectiveness of SIFA is still limited without any instance-level learning or adaptation. Although DDMRL [21] only adapts the features at the image level, its performance is still at the same level as CyCADA, by adapting knowledge across various domains. Among all the comparison methods, Hou *et al*. [16] achieves the second-best performance. Due to the effectiveness of panoptic-level feature adaptation and task re-weighting mechanism, our method still outperforms it under all three metrics, in both two experiments. Fig. 5 are visualization examples of all the comparison methods.

	AJI	Pixel-F1	Object-F1
w/o NI	0.5042 ± 0.1034	0.7336 ± 0.0839	0.6958 ± 0.0832
w/o TR	0.4969 ± 0.0972	0.7654 ± 0.0678	0.6923 ± 0.0778
w/o SEM	0.5046 ± 0.1065	0.7470 ± 0.0754	0.6965 ± 0.0805
proposed	0.5610 ± 0.0718	0.7882 ± 0.0533	0.7483 ± 0.0525

Table 4. Ablation study on BBBC039V1 to Kumar experiment. NI, TR, and SEM represent the nuclei inpainting mechanism, task re-weighting mechanism, and semantic branch, respectively.

4.3.2 Ablation Study

In order to test the effectiveness of each component in our proposed CyC-PDAM, ablation experiments are conducted on the Kumar dataset. Based on our CyC-PDAM, we remove the nuclei inpainting mechanism, task re-weighting mechanism, and semantic branch for panoptic-level adaptation and train the ablated models with the same setting and dataset as Sec. 4.3.1. Table 4 and Fig. 6 show the detailed results of the ablation experiment. As shown in Fig. 6, the method without nuclei inpainting mechanism (w/o NI) tends to ignore some nuclei, which increases the false-negative predictions. Moreover, we notice that there are also false split and merged predictions for w/o NI model. It is because the increasing false negative predictions are harmful to the spatial distribution of all the objects, which further affects the effectiveness of the semantic-level adaptation. Among the predictions of the method without task re-weighting mechanism (w/o TR), there exist some objects with irregular sizes. The task re-weighting mechanism prevents the model from being influenced by the domain-specific features in the source domain, and removing it, therefore, incurs source-biased predictions. Compared with our method, the model without semantic-branch (w/o SEM) is not able to learn domain-invariant features at the semantic level, including the spatial distribution of the nuclei objects

	AJI			Pixel-F1		
Methods	seen	unseen	all	seen	unseen	all
CNN3 [24]	0.5154 ± 0.0835	0.4989 ± 0.0806	0.5083 ± 0.0695	0.7301 ± 0.0590	0.8051 ± 0.1006	0.7623 ± 0.0946
DIST [35]	0.5594 ± 0.0598	0.5604 ± 0.0663	0.5598 ± 0.0781	0.7756 ± 0.0489	0.8005 ± 0.0538	0.7863 ± 0.0550
Proposed	0.5432 ± 0.0477	0.5848 ± 0.0951	0.5610 ± 0.0982	0.7743 ± 0.0358	0.8068 ± 0.0698	0.7882 ± 0.0533
Upper bound [22]	0.5703 ± 0.0480	0.5778 ± 0.0671	0.5735 ± 0.0855	0.7796 ± 0.0419	0.8007 ± 0.0511	0.7886 ± 0.0531

Table 5. Comparison experiments between our UDA method and fully supervised methods, for BBBC039V1 to Kumar experiment. For CNN3 and DIST, the results of object-level F1 are unknown.

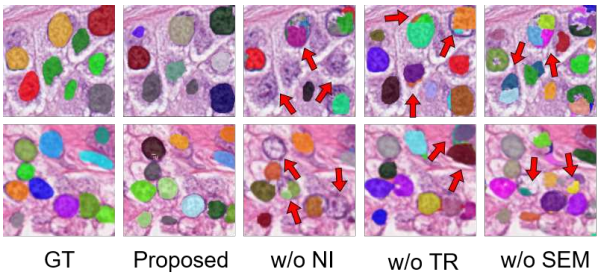


Figure 6. Visualization results for the ablation experiment. NI: nuclei inpainting mechanism; TR: task re-weighting mechanism; SEM: semantic branch.

and the detailed information in the background. Therefore, there not only remain falsely split and merged predictions, but also false-positive and imperfect segmentation results. As shown in Table 4, the segmentation accuracy under three metrics decreases by $4 - 6\%$ after removing each module. In addition, the one-tailed paired t-test is employed to calculate the p-value between our proposed method and the other ablated methods. After adding each of the three modules, the improvements are statistically significant ($P < 0.05$), which further demonstrates the effectiveness of our proposed method.

4.3.3 Comparison with Fully Supervised Methods

As our data split in Kumar dataset is the same as several state-of-the-art methods for fully supervised nuclei segmentation, we compare their original reported results with ours. Table 5 illustrates the comparison results between our proposed UDA architecture and other fully supervised methods. CNN3 [24] is a contour-based nuclei segmentation architecture, which considers nuclei boundaries as the third class, in addition to the foreground and background classes. DIST [35] is a regression model based on the distance map. For Panoptic FPN [22], we directly train it using the same set of 16 real histopathology patches as CNN3 and DIST and it is employed as the upper bound of our unsupervised method. The testing images for Kumar are divided into two subsets: one contains 8 images from 4 organs known to training set, referred to as seen, and the other contains 6 images from 3 organs unknown to the training set, referred to as unseen.

As shown in Table 5, the performance of our proposed UDA architecture is superior to the fully supervised CNN3 and DIST. It is because our proposed method is able to process each ROI on the local level, while CNN3 and DIST only process the image at a global semantic level. By adapting the semantic-level features of the foreground and the background, the performance of our method is at the same level as the fully supervised Panoptic FPN for the pixel-level F1-score. Even though our AJI is slight lower than the fully supervised Panoptic FPN, we notice that our method works better when tested on the unseen testing set. This is because our proposed CyC-PDAM focuses on learning the domain-invariant features and avoids being influenced by the domain bias of testing images from unseen organs. These results show that, although there remains large differences between the fluorescence microscopy images and histopathology images, our proposed UDA architecture still successfully narrows the domain gap between them, and achieves even better performance compared with fully supervised methods requiring histopathology nuclei annotations.

5. Conclusion

In this work, we propose a CyC-PDAM architecture for UDA nuclei segmentation in histopathology images. We firstly design a baseline architecture for UDA instance segmentation, including appearance-, image-, and instance-level adaptation. Next, a nuclei inpainting mechanism is designed to remove the auxiliary objects in the synthesized images, to further avoid false-negative predictions. In the feature-level adaptation, a semantic branch is proposed to adapt the features with respect to the foreground and background, and incorporating semantic- and instance-level adaptation enables the model to learn domain-invariant features at the panoptic level. In addition, a task re-weighting mechanism is proposed to reduce the bias. Extensive experiments on three public datasets indicate our proposed method outperforms the state-of-the-art UDA methods by a large margin and reaches the same level as the fully supervised methods. From a larger perspective, the UDA instance segmentation problems are not limited to histopathology image analysis. With the promising performance close to fully supervised methods in this work, we suggest that our proposed method can also contribute to other general image analysis applications.

References

[1] Ajay Basavanhally, Michael Feldman, Natalie Shih, Carolyn Mies, John Tomaszewski, Shridar Ganesan, and Anant Madabhushi. Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: Comparison to oncotype dx. *Journal of pathology informatics*, 2, 2011.

[2] Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng-Ann Heng. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In *Association for the Advancement of Artificial Intelligence (AAAI)*, pages 865–872, 2019.

[3] Hao Chen, Xiaojuan Qi, Lequan Yu, Qi Dou, Jing Qin, and Pheng-Ann Heng. Dcan: Deep contour-aware networks for object instance segmentation from histology images. *Medical Image Analysis*, 36:135–146, 2017.

[4] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster R-CNN for object detection in the wild. In *Computer Visionand Pattern Recognition (CVPR)*, pages 3339–3348, 2018.

[5] Frederic Clayton. Pathologic correlates of survival in 378 lymph node-negative infiltrating ductal breast carcinomas. mitotic count is the best single predictor. *Cancer*, 68(6):1309–1317, 1991.

[6] Qi Dou, Cheng Ouyang, Cheng Chen, Hao Chen, and Pheng-Ann Heng. Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 691–697, 2018.

[7] Christopher W Elston and Ian O Ellis. Pathological prognostic factors in breast cancer. I. the value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 19(5):403–410, 1991.

[8] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning (ICML)*, 2015.

[9] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58:101563, 2019.

[10] David A Gutman, Jake Cobb, Dhananjaya Somanna, Yuna Park, Fusheng Wang, Tahsin Kurc, Joel H Saltz, Daniel J Brat, Lee AD Cooper, and Jun Kong. Cancer digital slide archive: an informatics resource to support integrated in silico analysis of TCGA pathology data. *Journal of the American Medical Informatics Association*, 20(6):1091–1098, 2013.

[11] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Computer Visionand Pattern Recognition (CVPR)*, pages 770–778, 2016.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Eu-ropean Conference on Computer Vision (ECCV)*, pages 630–645. Springer, 2016.

[14] Zhenwei He and Lei Zhang. Multi-adversarial faster-rcnn for unrestricted object detection. In *International Conference on Computer Vision (ICCV)*, pages 6668–6677, 2019.

[15] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. *International Conference on Machine Learning (ICML)*, 2018.

[16] Le Hou, Ayush Agarwal, Dimitris Samaras, Tahsin M Kurc, Rajarsi R Gupta, and Joel H Saltz. Robust histopathology image analysis: To label or to synthesize? In *Computer Vision and Pattern Recognition (CVPR)*, pages 8533–8542, 2019.

[17] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *European Conference on Computer Vision (ECCV)*, pages 172–189, 2018.

[18] Yue Huang, Han Zheng, Chi Liu, Xinghao Ding, and Gustavo K Rohde. Epithelium-stroma classification via convolutional neural networks and unsupervised domain adaptation in histopathological images. *IEEE Journal of Biomedical and Health Informatics*, 21(6):1625–1632, 2017.

[19] Naoto Inoue, Ryosuke Furuta, Toshihiko Yamasaki, and Kiyoharu Aizawa. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 5001–5009, 2018.

[20] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976. IEEE, 2017.

[21] Taekyung Kim, Minki Jeong, Seunghyeon Kim, Seokeon Choi, and Changick Kim. Diversify and match: A domain adaptive representation learning paradigm for object detection. In *Computer Vision and Pattern Recognition (CVPR)*, pages 12456–12465, 2019.

[22] Alexander Kirillov, Ross Girshick, Kaiming He, and Piotr Dollár. Panoptic feature pyramid networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 6399–6408, 2019.

[23] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 9404–9413, 2019.

[24] 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 Transactions on Medical Imaging*, 36(7):1550–1560, 2017.

[25] V Le Doussal, M Tubiana-Hulin, S Friedman, K Hacene, F Spyratos, and M Brunet. Prognostic value of histologic grade nuclear components of Scarff-Bloom-Richardson (SBR). an improved score modification based on a multivariate analysis of 1262 invasive ductal breast carcinomas. *Cancer*, 64(9):1914–1921, 1989.

[26] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos. Bidirectional learning for domain adaptation of semantic segmentation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 6936–6945, 2019.

[27] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2117–2125, 2017.

[28] Dongnan Liu, Donghao Zhang, Yang Song, Heng Huang, and Weidong Cai. Cell r-cnn v3: A novel panoptic paradigm for instance segmentation in biomedical images. *arXiv preprint arXiv:2002.06345*, 2020.

[29] Dongnan Liu, Donghao Zhang, Yang Song, Chaoyi Zhang, Fan Zhang, Lauren ODonnell, and Weidong Cai. Nuclei segmentation via a deep panoptic model with semantic feature fusion. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 861–868. AAAI Press, 2019.

[30] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 700–708, 2017.

[31] Vebjorn Ljosa, Katherine L Sokolnicki, and Anne E Carpenter. Annotated high-throughput microscopy image sets for validation. *Nature Methods*, 9(7):637–637, 2012.

[32] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I Jordan. Learning transferable features with deep adaptation networks. *International Conference on Machine Learning (ICML)*, 2015.

[33] Faisal Mahmood, Daniel Borders, Richard Chen, Gregory N McKay, Kevan J Salimian, Alexander Baras, and Nicholas J Durr. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Transactions on Medical Imaging*, 2018.

[34] Sidra Nawaz and Yinyin Yuan. Computational pathology: Exploring the spatial dimension of tumor ecology. *Cancer letters*, 380(1):296–303, 2016.

[35] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, 2018.

[36] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2009.

[37] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2337–2346, 2019.

[38] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. *NeurIPS 2017 Autodiff Workshop*, 2017.

[39] Jian Ren, Ilker Hacihaliloglu, Eric A Singer, David J Foran, and Xin Qi. Adversarial domain adaptation for classification of prostate histopathology whole-slide images. In *International Conference On Medical Image Computing Computer Assisted Intervention (MICCAI)*, pages 201–209. Springer, 2018.

[40] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 91–99, 2015.

[41] Claude Elwood Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[42] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *Association for the Advancement of Artificial Intelligence (AAAI)*, 2016.

[43] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of Graphics Tools*, 9(1):23–34, 2004.

[44] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 7472–7481, 2018.

[45] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 7167–7176, 2017.

[46] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2517–2526, 2019.

[47] Tao Wang, Xiaopeng Zhang, Li Yuan, and Jiashi Feng. Few-shot adaptive faster r-cnn. In *Computer Vision and Pattern Recognition (CVPR)*, pages 7173–7182, 2019.

[48] Jun Wen, Risheng Liu, Nenggan Zheng, Qian Zheng, Zhefeng Gong, and Junsong Yuan. Exploiting local feature patterns for unsupervised domain adaptation. In *Association for the Advancement of Artificial Intelligence (AAAI)*, volume 33, pages 5401–5408, 2019.

[49] Yuxin Wu and Kaiming He. Group normalization. In *European Conference on Computer Vision (ECCV)*, pages 3–19, 2018.

[50] Donghao Zhang, Yang Song, Dongnan Liu, Haozhe Jia, Siqi Liu, Yong Xia, Heng Huang, and Weidong Cai. Panoptic segmentation with an end-to-end cell r-cnn for pathology image analysis. In *International Conference On Medical Image Computing Computer Assisted Intervention (MICCAI)*, pages 237–244. Springer, 2018.

[51] Yue Zhang, Shun Miao, Tommaso Mansi, and Rui Liao. Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation. In *International Conference On Medical Image Computing Computer Assisted Intervention (MICCAI)*, pages 599–607. Springer, 2018.

[52] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017.