PDAM: A Panoptic-Level Feature Alignment Framework for Unsupervised Domain Adaptive Instance Segmentation in Microscopy Images

Dongnan Liu, *Member, IEEE*, Donghao Zhang, Yang Song, *Member, IEEE*, Fan Zhang, Lauren O'Donnell, Heng Huang, *Member, IEEE*, Mei Chen, *Senior Member, IEEE*, and Weidong Cai, *Member, IEEE*

***Abstract*— In this work, we present an unsupervised domain adaptation (UDA) method, named Panoptic Domain Adaptive Mask R-CNN (PDAM), for unsupervised instance segmentation in microscopy images. Since there currently lack methods particularly for UDA instance segmentation, we first design a Domain Adaptive Mask R-CNN (DAM) as the baseline, with cross-domain feature alignment at the image and instance levels. In addition to the image- and instance-level domain discrepancy, there also exists domain bias at the semantic level in the contextual information. Next, we, therefore, design a semantic segmentation branch with a domain discriminator to bridge the domain gap at the contextual level. By integrating the semantic- and instance-level feature adaptation, our method aligns the cross-domain features at the panoptic level. Third, we propose a task re-weighting mechanism to assign trade-off weights for the detection and segmentation loss functions. The task re-weighting mechanism solves the domain bias issue by alleviating the task learning for some iterations when the features contain source-specific factors. Furthermore, we design a feature similarity maximization mechanism to facilitate instance-level feature adaptation from the perspective of representational learning. Different from the typical feature alignment methods, our feature similarity maximization mechanism separates the domain-invariant and domain-specific features by enlarging their feature distribution dependency. Experimental results on three UDA instance segmentation scenarios with five datasets demonstrate the effectiveness of our proposed PDAM method, which outperforms state-of-the-art UDA methods by a large margin.**

***Index Terms*— Unsupervised domain adaptation, instance segmentation, microscopy images.**

I. Introduction

INSTANCE segmentation is necessary and important for microscopy image analysis, assigning a category label for each pixel and a unique instance label for each object of the same class. In the digital pathology study, nuclei instance segmentation enables pathologists to learn about the single nucleus structure, nuclei group spatial distribution, and mitosis counts, which serve as key factors for cancer diagnosis and prognosis [1]–[4]. For Electron Microscopy (EM) image analysis, mitochondria instance segmentation facilitates the pleomorphism learning of intracellular components and neural functions, which are important for cancer cell detection, Parkinson's disease gene recognition, and cell segmentation [5]–[7].

With the success of deep learning, recent deep neural network based methods are now prevalent in microscopy images instance segmentation [8]–[14]. Although these fully-supervised methods achieve state-of-the-art performance, their high accuracy heavily relies on massive annotated training images from the specific domains. When tested these off-the-shelf models on the images from new unseen domains, the performance suffers from a significant drop (as shown in Fig. 1) due to the domain bias towards the training domains [15], [16]. On the other hand, it is impractical to acquire sufficient annotations for each new domain, since the annotation process for microscopy images is time-consuming, labor-intensive, and error-prone [8], [17], [18].

Recently, unsupervised domain adaptation (UDA) methods have been proposed to tackle this dilemma, which transfer the knowledge learned from the labelled source domain to the unlabelled target domain [19]–[21]. UDA methods often work by reducing the cross-domain discrepancy at the content and appearance levels [18], [22]–[24]. With the content-level UDA methods, the feature extractors are forced to generate domain-invariant features by adversarial learning [20], [25], [26]. The appearance-level UDA methods, also known as image-to-image translation methods, align the

Manuscript received August 9, 2020; accepted September 3, 2020. Date of publication September 11, 2020; date of current version December 29, 2020. *(Corresponding author: Dongnan Liu.)*

Dongnan Liu, Donghao Zhang, and Weidong Cai are with the School of Computer Science, University of Sydney, Sydney, NSW 2008, Australia (e-mail: dliu5812@uni.sydney.edu.au; dzha9516@uni.sydney.edu.au; tom.cai@sydney.edu.au).

Yang Song is with the School of Computer Science and Engineering, University of New South Wales, Kensington, NSW 2052, Australia (e-mail: yang.song1@unsw.edu.au).

Fan Zhang and Lauren O'Donnell are with the Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115 USA (e-mail: fzhang@bwh.harvard.edu; odonnell@bwh.harvard.edu).

Heng Huang is with the Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15261 USA (e-mail: henghuanghh@gmail.com).

Mei Chen is with Microsoft Corporation, Redmond, WA 98052 USA (e-mail: may4mc@gmail.com).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2020.3023466

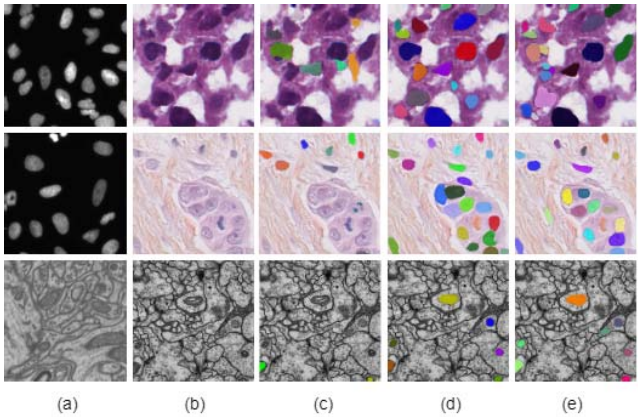

Fig. 1. Example images of our proposed framework. The images in the first two rows are the samples for adapting from the fluorescence microscopy dataset to the histopathology dataset. The third row contains the samples for adapting from two different EM datasets. In each row: (a) source images; (b) target images; (c) predictions from the models without domain adaptation; (d) predictions from our proposed PDAM method; (e) ground truth.

appearances of the cross-domain images by synthesizing the target-like images based on the source images [27]–[29]. However, these appearance-level UDA methods still suffer from domain bias between the synthesized and target images due to the imperfect translations [22], [24], [30], [31].

To reduce the cross-domain discrepancy in the content and appearance level, several methods have been proposed to align the features between the target and the synthesized target-like images [22], [30], [31]. These methods achieve state-of-the-art performance on various UDA tasks of classification, object detection, and semantic segmentation. However, none of these methods can be directly employed on the UDA instance segmentation tasks. Furthermore, although it is intuitive to consider directly replacing the Faster R-CNN in UDA object detection methods [26], [31], [32] with Mask R-CNN [33], we find such an approach only brings limited performance gain due to several challenges. First, existing UDA object detection methods can generate domain-invariant features at the image level (image contrast, brightness, etc.) and the object level (object scale, style, etc.) [26], [31], [32]. However, they cannot remove the domain-specific factors at the semantic level, such as the relationship between the foreground and background, and the spatial distribution of the objects. Second, the loss functions of the detection tasks are optimized according to the labeled source images. Therefore, the model biases towards the source domain if the feature extractors fail to generate domain-invariant features in some training iterations.

To solve the aforementioned challenges, we propose a Panoptic Domain Adaptive Mask R-CNN (PDAM) method, for UDA instance segmentation in microscopy images. Currently there is a lack of methods particularly designed for UDA instance segmentation. Therefore, we first design a Domain Adaptive Mask R-CNN (DAM) as the baseline, based on Mask R-CNN and domain adaptive Faster R-CNN [26]. Inspired by panoptic segmentation architectures [34], [35], the reconciliation of the semantic and instance segmentation, we then design a panoptic-level feature alignment module to learn domain-invariant features at both the instance and semantic levels. Specifically, we add an auxiliary semantic segmentation branch and a semantic-level feature adaptation module, to alleviate the domain shift in the semantic contextual features. In addition, we also design a task re-weighting mechanism to reset the importance of the loss functions for detection and segmentation according to whether the extracted features from the current training iteration being domain-invariant or not. During training, we down-weight the detection and segmentation losses if the extracted features are biased towards the source domain and up-weight the losses if the features are domain-invariant. Furthermore, we propose a feature similarity maximization mechanism to reduce the domain gap from the perspective of representation learning. Specifically, the similarity maximization is based on mutual information (MI), an entropy-based metric for dependency measurement between two distributions. Previous work enables unsupervised classification and facilitates the GAN training by maximizing MI between the latent layers and input (or output) layers in the CNN architectures [36], [37]. In this work, we propose to induce domain-invariant feature learning by enlarging the MI on the features between the source and target domains.

The preliminary version of this work was published as a conference paper in CVPR 2020 [24]. In this manuscript, we extend our previous work as follows:

- We propose a feature similarity maximization mechanism for UDA instance segmentation. This is a significant improvement since the feature distribution similarity is typically employed in the representation learning. In this work, we are making an early attempt to demonstrate its effectiveness in the UDA instance segmentation tasks, with experimental results and in-depth analysis.
- In addition to the UDA instance segmentation from the fluorescence microscopy to the histopathology images, we further apply our proposed method to UDA mitochondria instance segmentation in EM images. Compared with the state-of-the-art methods, our proposed PDAM method achieves better performance on various UDA instance segmentation tasks, which further indicates the strong generalization ability of the PDAM method.
- Compared to the previous conference version, we include more details about the motivations and insight into our method. Additionally, the presentation and writing of the overall paper are improved significantly for more precisely description.

In line with our previous work, our novel PDAM method is the first paradigm for UDA instance segmentation in medical images, to our best knowledge.

II. Related Work

Domain adaptation aims at transferring the knowledge learned from one source domain to another target domain via either supervised or unsupervised learning. For the supervised domain adaptation, the source domain is fully labeled while the target domain contains a small amount of annotations [38]–[40]. Typically, supervised domain adaptation methods are implemented via feature sharing [38], and

fine-tuning [39]. In [40], domain-specific batch normalization layers are employed to adapt an off-the-shelf model to a new domain with few annotations. However, the supervised domain adaptation methods still rely on the labeled target images. On the other hand, unsupervised domain adaptation requires no annotations from the target domain and is, therefore, more efficient. In this section, we review the literature of the unsupervised domain adaptation in detail since our proposed method is based on it.

A. Unsupervised Domain Adaptation for Natural Images

Under unsupervised domain adaptation setting, the source domain is fully labelled and the target domain is unlabelled [19]. 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 feature-level adaptation, adversarial learning for domain-invariant features [20], [21], Maximum Mean Discrepancy minimization (MMD) [41], and cross-domain covariance alignment [42] are widely employed for classification tasks. In addition, UDA is further employed for other tasks such as semantic segmentation [25], [43] and object detection [26], [31], [44]. In the semantic segmentation tasks, the segmentation results are forced to be domain-invariant, together with intermediate feature maps [25], [45]. Additionally, ADVENT [25] 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 [46], consisting of the image- and instance-level adaptions, is usually proposed for domain-invariant features of the whole image and each object [26], [31], [32]. 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 [27]–[29], [47], [48]. However, domain bias still exists because of imperfect translation [22], [30], [31]. To integrate the benefits of the feature-level and pixel-level adaptations, several methods have been proposed to learn domain-invariant features between the synthesized and real images [22], [30], [31].

B. Unsupervised Domain Adaptation for Medical Images

In addition to the general images, there are a wide range of UDA applications in medical image analysis [17], [18], [22], [24], [49]–[52]. For example, studies [49], [51] address the UDA histopathology images classification problems with GAN based architectures. PnP-AdaNet [23] is proposed for UDA semantic segmentation in CT images, generating domain-invariant features at the intermediate level and the output of the model using plug-and-play feature adaptation modules. TD-GAN [50] and SIFA [22] then incorporate the pixel- and feature-level alignment for more competitive UDA semantic segmentation performance. In addition to the UDA methods based on feature and appearance alignment, visual correspondence is exploited in [52] and [18] for UDA semantic segmentation in EM images via multiple instance learning.

Currently, there is a lack of methods for UDA instance segmentation. For unsupervised nuclei instance segmentation, Hou *et al.* [17] firstly proposed a histopathology image synthesis pipeline. Then, they train a GAN based model to refine the synthesized images and, meanwhile, generate nuclei segmentation predictions. However, this method's performance was limited by the domain shift between the synthesized and real histopathology images, due to the lack of feature-level adaptation. Additionally, this image synthesis pipeline was designed based on the characteristics of the histopathology images, which is domain-specific and cannot be directly adapted to other kinds of images. For a better performance on UDA instance nuclei segmentation for the histopathology images, our previous work [24] focused on aligning the cross-domain features at the panoptic level and alleviating the source-biased features learning. In this work, we further improve our previous method by facilitating instance-level feature adaptation from the view of representation learning. Additionally, our PDAM method is demonstrated to be effective on the UDA instance segmentation for other kind of microscopy images.

III. Method

Our PDAM method is trained on the real and synthesized images. Given images from the source and target domain, we firstly employ a CycleGAN [28] to synthesize target-like images from the source domain. Then we train the PDAM method using the synthesized images as the source domain, and the real target images as the target domain. In this section, we first introduce a pre-processing method to refine the synthesized images, named the auxiliary objects inpainting mechanism. Second, we present the detailed design of the PDAM method, including the baseline DAM, panoptic-level feature alignment, task re-weighting mechanism, and feature similarity maximization mechanism. Third, we show the definition of the overall objective function for the PDAM method.

A. Auxiliary Objects Inpainting Mechanism

In our previous work [24], we noticed that there occur unexpected nuclei objects on the synthesized histopathology images translated from the fluorescence microscopy images. As shown in Fig. 2, the synthesized histopathology images contain nuclei objects without corresponding annotations. In other words, these auxiliary generated nuclei are labeled as background in the ground truth. Subsequently, the model is forced to regard the redundant nuclei as background components when training the PDAM with these synthesized images. However, these auxiliary nuclei always have a similar appearance to other nuclei with annotations, as shown in Fig. 2 (c). To this end, the model is likely to predict some true nuclei as the background and results in false-negative predictions during inference. To avoid this problem, we designed an auxiliary objects inpainting mechanism to remove these undesired nuclei in the synthesized histopathology images. Denoting a raw synthesized histopathology image 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:

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

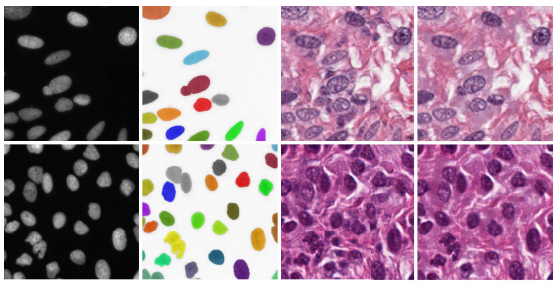

Fig. 2. Visual results for the effectiveness of the object inpainting mechanism on synthesizing the histopathology images. (a) original fluorescence microscopy patches; (b) corresponding nuclei annotations; (c) initial synthesized histopathology images from CycleGAN; (d) synthesized histopathology images after object inpainting mechanism.

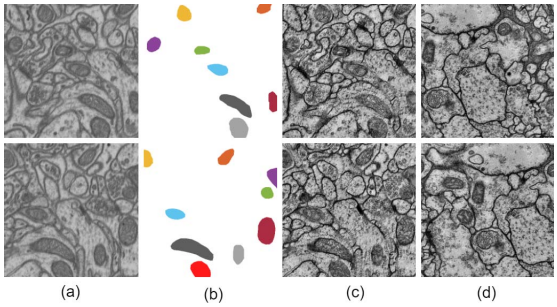

Fig. 3. Visual examples of the synthesized EM images in our experiments. (a) original source EM images; (b) corresponding mitochondria annotations; (c) synthesized target-like EM images; (d) real target EM images.

where $ostu(S_{raw})$ represents the binary segmentation output of S_{raw} based on Otsu threshold [53]. Given an image, Otsu threshold algorithm automatically generates a threshold value for the background and foreground segmentation. In M_{aux}, only auxiliary nuclei without annotation are retained. Then, we use a fast marching based method [54] to remove the unexpected nuclei, by replacing the pixel values of the labeled contents in M_{aux} with those of the unlabeled content. As a pre-processing step, the auxiliary objects inpainting mechanism in Equation 1 does not involve a learning process for optimization.

Due to the large domain gap between the fluorescence microscopy and histopathology images, the CycleGAN generates imperfect synthesized images. In this work, we further notice that the synthesized images from the CycleGAN are indeed of high quality when translating the images between two different EM datasets. As shown in Fig. 3, there are no auxiliary mitochondria object in the target-like synthesized images and the auxiliary objects inpainting mechanism is therefore not necessary under such a setting. This is because the domain gap between the two datasets in the same modality is smaller than that between datasets in different modalities.

B. Domain Adaptive Mask R-CNN

Currently, there is a lack of a UDA architecture specialized for instance segmentation. On the other hand, object detection methods can be extended to instance segmentation by including an auxiliary segmentation branch for each object [33]. Therefore, we firstly propose a Domain Adaptive Mask R-CNN (DAM) following the domain adaptive Faster R-CNN in [26], [32]. Specifically, the domain adaptive Faster R-CNN bridges the domain gap at the image and instance levels, by incorporating a Faster R-CNN with adversarial domain discriminators. During training, the domain discriminators predict the domain label for the features from the feature extractors, while the feature extractors aim at confusing the domain discriminators. In the domain adaptive Faster R-CNN, the input for the image-level domain discriminator are the features from the backbone. For the instance-level discriminator, the input are the instance features for object category and location prediction.

TABLE I
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

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	$k = (1, 1), s = 1, p = 0$	$2 \times 8 \times 8$

In this work, the DAM model regards the synthesized and target images as the source and target domains, respectively. For both the source and target input images, they share the same DAM architecture. The backbone of the DAM model is ResNet101 [55] and Feature Pyramid Network (FPN) [56], which obtains image-level features from multiple resolutions. In terms of the image-level feature alignment, the multi-resolution features from the backbone are first downsampled to the size of 8×8 with average pooling, and then summed together for the domain discriminator at the image level. Here we use patch-based predictions for the image-level domain discriminator to increase the training samples and avoid overfitting due to the small mini-batch size. For the instance-level feature alignment, the mask predictions are firstly flattened along the batch axis. Next, the flattened mask vectors pass through a fully connected layer, of which the output is the same size as the feature vectors from the bounding box branch. Third, the instance-level features are obtained by fusing the feature vectors from the box and mask branches via summation. The instance-level features of the source and target domains are then sent to another domain discriminator to generate domain-invariant features in each foreground instance. In addition, the adversarial training strategy for each domain discriminator is implemented by embedding a gradient reversal layer (GRL) before the first CNN (or FC) layer. The detailed network architecture and configuration of the image-level and instance-level domain discriminators are shown in Fig. 4 and Table I.

C. Panoptic Level Feature Alignment

Mask R-CNN is a proposal-based architecture for instance segmentation that processes each individual object, so it

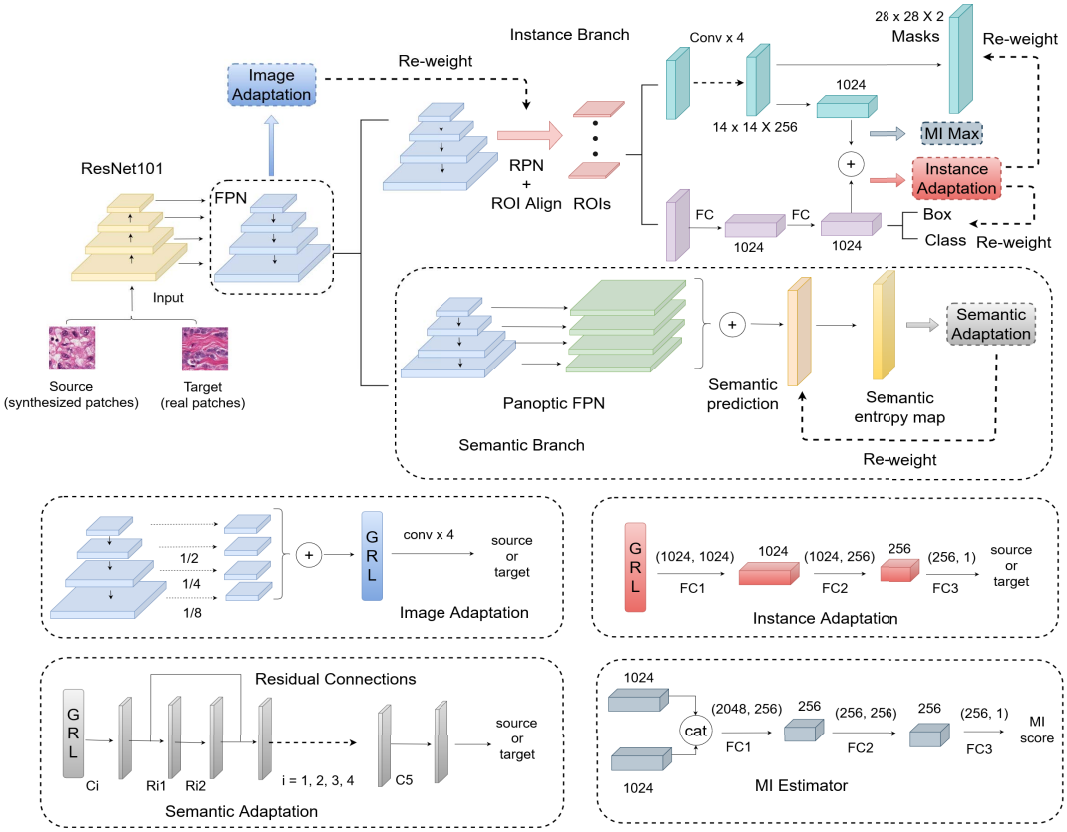

Fig. 4. Detailed illustration of Panoptic Domain Adaptive Mask R-CNN (PDAM). *Ci* and *FC* represent a convolution layer and a fully connected layer, respectively. *Ri*1 and *Ri*2 mean the first and second convolutional layers in the *ith* residual block, respectively. In the MI Estimator, *cat* represents the concatenation. *ReLU* and normalization layers after each convolutional block are omitted for brevity.

lacks a global understanding of the whole image [34]. Therefore, the DAM still suffers from domain shift due to the cross-domain discrepancy in the contextual information, such as the relationship between the foreground and background and the spatial arrangement of foreground objects. By incorporating the benefits from both semantic and instance segmentation, panoptic segmentation architecture [35] was previously proposed to process both semantic contextual and local instance features on the whole images. Inspired by this, we propose to learn domain-invariant features at the panoptic-level to alleviate the domain bias at the semantic level, by integrating the instance-level feature alignment module in DAM with a newly proposed semantic-level feature alignment module.

To align the cross-domain semantic-level features, we first design a semantic branch after the FPN for semantic segmentation predictions. The detailed architecture of the semantic branch in this work is similar to the Panoptic FPN [35]. Instead of the semantic segmentation predictions, the semantic entropy map passes through an adversarial domain discriminator to learn the domain-invariant features at the semantic level. Aligning the cross-domain entropy distributions helps minimize the entropy prediction in the target domain, which makes the model suitable for the target images [25]. Denoting the softmax semantic prediction as $P \in (0, 1)$, its entropy map is calculated using the Shannon entropy, defined as: $-plog(p)$.

TABLE II
THE PARAMETERS FOR EACH BLOCK IN THE SEMANTIC-LEVEL DISCRIMINATOR FOR PDAM. *k*, *s*, AND *p* FOLLOW THE SAME CONVENTION AS IN TABLE I

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		$2 \times 16 \times 16$

The detailed structure of the semantic-level feature discriminator is illustrated in Fig. 4 and Table II. Similar to the image-level discriminator, our semantic-level discriminator produces patch-based predictions for a stable training process.

D. Task Re-Weighting Mechanism

In the PDAM, the detection and segmentation learning is based on the synthesized images. During some training iterations, the unstable learning process of the adversarial domain discriminators could result in predicted features that are far from the decision boundaries and contain domain-specific

factors. Therefore, the model still suffers from some domain shift towards the source images.

To solve this problem, we propose a task re-weighting mechanism, by adding trade-off weights for detection and segmentation loss functions. To evaluate whether the extracted features are domain-invariant, we employ the predictions of the domain discriminator to calculate the trade-off weights. Denoting 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, the re-weighted task-specific loss L_{rw} is:

$$L_{rw} = min(\frac{p_t}{p_s}, \beta)L = min(\frac{1-p_s}{p_s}, \beta)L \tag{2}$$

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. 2, if the feature map before the task prediction is source-biased ($p_s \to 1$), this task loss function is then down-weighted, to prevent the model biasing toward the source domain. As illustrated in Fig. 4, the loss function for the region proposal network (RPN), semantic branch, and the instance branch are re-weighted by the prediction of the image-, semantic-, and instance-level domain discriminators, respectively.

E. Feature Similarity Maximization Mechanism

First, we define the features before the instance-level feature alignment module of the source and target domains as F_{is} and F_{it}, respectively. Instance-level feature adaptation is based on an assumption that the distributions of the label space for the source and target domains are the same. Therefore, the domain-invariant F_{is} and F_{it} should be similar to each other under the ideal domain adaptation circumstance. To measure the similarity, we employ the mutual information (MI), which is in proportion to the level of dependency between two distributions [36]. By maximizing the MI between F_{is} and F_{it}, our PDAM method further reduces the cross-domain discrepancy.

To incorporate the MI maximization into our CNN-based PDAM method, we employ the MI represented by the Jensen-Shannon formulation defined in [57], which is demonstrated to be effective for learning feature representations [37]. The MI between F_{is} and F_{it} is defined as:

$$\begin{aligned} I^{(JSD)}(F_{is}, F_{it}) = &-sp(-T(F_{is}, F_{it}; \omega) \\ &- sp(T(F'_{is}, F_{it}; \omega)) \end{aligned} \tag{3}$$

where F'_{is} is a marginal feature sampling of F_{is} by shuffling the F_{is} across the batch axis and $sp(f) = log(1 + e^f)$ is the softplus operation. $T(a, b; \omega)$ is a CNN-based MI estimator with weights of ω and inputs a and b. The detailed architecture of the $T(a, b; \omega)$ is shown in Fig. 4. We define the objective function of the feature similarity maximization as:

$$L_{FSM} = -I^{(JSD)}(F_{is}, F_{it}) \tag{4}$$

By minimizing the L_{FSM} with gradient descent optimization during training, our PDAM learns to maximize the instance-level mutual information $I^{(JSD)}(F_{is}, F_{it})$. This increases the amount of similar content between F_{is} and F_{it}, and further induces them to be domain-invariant. Note that we only employ the feature similarity mechanism for cross-domain feature alignment at the instance level, instead of the semantic and image level. For the semantic- and image-level features, the batch size is set to 1. Therefore, it is impossible to obtain the marginal feature sampling in Equation 3, of which the feature shuffling across the batch axis is an essential step. On the other hand, the batch size of the instance-level features is equal to the number of ROIs. To this end, the feature similarity maximization mechanism based on MI is only suitable for the instance-level features.

F. Objective Function

The PDAM is trained in an end-to-end fashion with the overall objective function defined as:

$$\begin{aligned} L_{pdam} = &\alpha_{img} L_{rpn} + \alpha_{ins} L_{det} + \alpha_{sem} L_{(sem-seg)} \\ &+ \alpha_{da}(L_{(img-da)} + L_{(sem-da)} + L_{(ins-da)}) \\ &+ \alpha_{fs} L_{FSM} \end{aligned} \tag{5}$$

L_{rpn} is the loss function for the RPN, including a smooth L1 regression loss for regression and a cross entropy loss for classification. L_{det} is the instance segmentation and detection loss for Mask R-CNN, which contains loss functions for instance classification, coordinates regression, and mask segmentation. $L_{(sem-seg)}$ is the cross entropy loss for semantic segmentation defined in Section III-C. $L_{(img-da)}$, $L_{(sem-da)}$, and $L_{(ins-da)}$ are cross entropy losses for domain classification at image, semantic, and instance levels. α_{img}, α_{ins}, and α_{sem} are calculated according to Eq. 2 for task re-weighting. α_{da} is updated as: $\alpha_{da} = \frac{2}{1+exp(-10t)} - 1$, where t is the training progress and $t \in [0, 1]$. L_{FSM} defined in Eq. 4 is for feature similarity maximization at the instance-level, and its trade-off weight α_{fs} is set as 0.1.

IV. Experiment

A. Dataset Description and Preparation

1) Adaptation Between Fluorescence Microscopy and Histopathology Images: Under this setting, we employ the fluorescence microscopy dataset BBBC039V1 [58] as the source domain and the histopathology datasets Kumar [8] and TNBC [9] as the target domain. BBBC039V1 contains 200 520 × 696 images for U2OS cells under a high-throughput chemical screen [58]. These images are grayscale, as they were acquired with the DNA channel staining of a single field of view. Kumar was obtained from The Cancer Genome Atlas (TCGA), containing 30 annotated 1000 × 1000 patches from 30 whole slide images of different patients at 40× magnification. These images are from 18 different hospitals and 7 different organs (breast, liver, kidney, prostate, bladder, colon, and stomach). In contrast to the disease diversity in Kumar, the TNBC dataset especially focuses on Triple-Negative Breast Cancer [9]. In the TNBC dataset, there are 50 annotated 512 × 512 patches from 11 different patients from the Curie Institute at 40× magnification. Example images of the three datasets are shown in Fig. 5.

With the BBBC039V1 dataset, we use the 100 training images and 50 validation images following the official data split. Before training, there are 4 steps for preprecessing. First, all images are normalized into range $[0, 255]$. Second, patches in size 256×256 are randomly cropped from the 100 training images, with data augmentation including rotation, scaling, and flipping. Third, the patches with fewer than 3 objects are removed. Fourth, all the patches are subtracted by 255 for the inverse. For validation, 50 images in the BBBC039V1 validation set are transferred to the synthesized histopathology images by CycleGAN and the auxiliary object inpainting mechanism. In the Kumar dataset, we have the same data split as the previous work in [8], [9], with 16 images for training and 14 for testing. As for the TNBC dataset, we use 8 cases with 40 images for training, and the remaining 3 cases with 10 images for testing. To preprocess the Kumar and TNBC datasets, 256×256 patches are randomly cropped from the training images, with basic data augmentation including flipping and rotation.

2) Adaptation Between Electron Microscopy Images From Different Sources: In addition to the UDA settings mentioned in our previous work [24], we further validate our PDAM method on the UDA instance mitochondria segmentation between different electron microscopy (EM) datasets. Specifically, we employ the EPFL dataset [59] as the source domain, which is obtained from the mouse brain hippocampus using Focused Ion Beam Scanning EM (FIB-SEM). EPFL is a 3D volume of $1024 \times 728 \times 165$ voxels at an isotropic resolution. To fit our 2D CNN architecture, we further split the volume into 165 2D patches of size 1024×728. The VNC [60] dataset for the target domain is obtained from the Drosophila melanogaster third instar larva Ventral Nerve Cord, utilizing serial section Transmission EM (ssTEM). Similar to the EPFL dataset, we split the VNC dataset of $1024 \times 1024 \times 20$ voxels into 20 2D 1024×1024 patches. Among the 20 2D images, we randomly split $2/3$ with 13 images for training and the remaining $1/3$ with 7 images for testing. Samples images of the EPFL and VNC datasets are shown in Fig. 6.

Among the 165 images in the EPFL dataset, we randomly select 132 images for training and use the remaining 33 for validation. Due to the large mitochondria size, we randomly crop 512×512 patches from the source EPFL and target VNC images, with basic data augmentation including horizontal and vertical flipping and rotation of 90°, 180°, and 270°. Then, we remove the patches containing fewer than 3 mitochondria objects. For validation, we transform the 33 EPFL validation images to the target-like synthesized images only using CycleGAN, as described in Section III-A.

B. Evaluation Metrics

To evaluate our method, we employ three commonly-used metrics at the pixel and object levels. For the object-level metrics, we use Aggregated Jaccard Index (AJI) [8] and Panoptic Quality (PQ) [34]. AJI extends the Jaccard Index for each object by considering the false-positive predictions (unclaimed detection). PQ was originally designed for panoptic segmentation, which multiplies the $F1$ score for object detection and the IoU score for instance segmentation. Therefore, PQ reflects the performance of the detection and segmentation and is widely employed for nuclei instance segmentation [61]. For the pixel-level evaluation, we employ the F1 score, which is the average harmonic mean between the precision and recall of the binary segmentation predictions.

C. Implementation Details

To initialize the PDAM method, the weights of the ResNet101 backbone are pretrained on the ImageNet classification task, while the weights for other layers are initialized with “Kaiming" initialization [62]. During training, the batch size is 1 and each batch contains 2 images, one from the source and the other from the target domain. Instead of the traditional batch normalization layers, we employ group normalization [63] layers due to the small batch size. The group number of the group normalization is set as 32 in our PDAM method, following the default setting in [63].

We employ Stochastic Gradient Descent (SGD) for PDAM optimization, with a weight decay of 0.001 and a momentum of 0.9. The initial learning rate of PDAM is 0.001, with linear warming up in the first 500 iterations. The learning rate is then decreased to 0.0001 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 [64] on two NVIDIA GeForce 1080Ti GPUs.

D. Adaptation Between Fluorescence Microscopy and Histopathology Images

1) In Comparison With Unsupervised Methods: To demonstrate the effectiveness of our proposed PDAM method on alleviating the large domain bias between the microscopy images from different modalities, we conduct UDA instance segmentation experiment by adapting from BBBC039V1 to Kumar and from BBBC039V1 to TNBC. We compare our PDAM method with several state-of-the-art UDA detection and segmentation methods, including CyCADA [30], Chen *et al.* [26], DDMRL [31], SIFA [22], and Liu *et al.* [24].

As CyCADA, Chen *et al.*, DDMR, and SIFA are originally designed for either object detection or semantic segmentation, we extend them to the instance segmentation scenario. For CyCADA, we extend it by integrating the CycleGAN with the DAM in Section III-B and it is employed as the UDA baseline of our PDAM method. Chen *et al.* was originally for UDA object detection based on Faster R-CNN, with feature alignment at the image and instance levels and a consistency regularization mechanism. To adapt it to UDA instance segmentation, we extend it by incorporating their consistency regularization into our DAM. DDMRL learns multi-domain-invariant features from various generated domains for UDA object detection. In this work, we use the source, synthesized, and the target images for the multi-domain setting. As the original DDMRL only aligned the cross-domain features at the image-level, we extend it using our DAM without the instance-level feature adaptations. SIFA is a UDA semantic

TABLE III
IN COMPARISON WITH OTHER UNSUPERVISED METHODS ON BOTH TWO HISTOPATHOLOGY DATASETS. RESULTS ARE PRESENTED AS MEAN VALUE WITH STANDARD DEVIATION IN THE PARENTHESES

	$BBBC039 \rightarrow Kumar$			$BBBC039 \rightarrow TNBC$		
Methods	AJI	F1	PQ	AJI	F1	PQ
w/o DA	0.3170(0.1388)	0.5076(0.1781)	0.2996(0.1386)	0.3379(0.0684)	0.5400(0.0874)	0.4013(0.0554)
CyCADA [30]	0.4447(0.1069)	0.7220(0.0802)	0.3515(0.1441)	0.4721(0.0906)	0.7048(0.0946)	0.4250(0.0759)
Chen *et al.* [26]	0.3756(0.0977)	0.6337(0.0897)	0.2663(0.1198)	0.4407(0.0623)	0.6405(0.0660)	0.3882(0.0821)
SIFA [22]	0.3924(0.1062)	0.6880(0.0882)	0.2799(0.1133)	0.4662(0.0902)	0.6994(0.0942)	0.4009(0.1306)
DDMRL [31]	0.4860(0.0846)	0.7109(0.0744)	0.4738(0.0872)	0.4642(0.0503)	0.7000(0.0431)	0.4771(0.0400)
Hou *et al.* [17]	0.4980(0.1236)	0.7500(0.0849)	0.3837(0.1576)	0.4775(0.1219)	0.7029(0.1262)	0.3738(0.1157)
Liu *et al.* [24]	0.5610(0.0718)	0.7882(0.0533)	0.5197(0.0865)	0.5672(0.0646)	0.7593(0.0566)	0.5207(0.0522)
PDAM	**0.5653 (0.0751)**	**0.7904 (0.0474)**	**0.5249 (0.0884)**	**0.5726 (0.0414)**	**0.7742 (0.0302)**	**0.5409 (0.0331)**

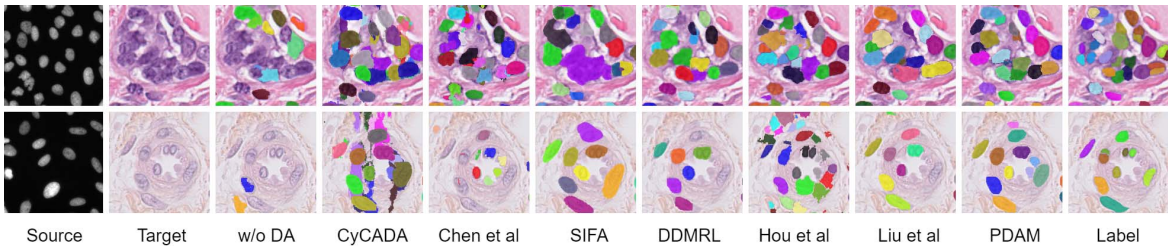

Fig. 5. Visual comparison on the experiments by adapting from the fluorescence microscopy to the histopathology datasets. The images of the first row are from Kumar dataset, and the second are from TNBC. The mask predictions are overlapped on the target testing images.

segmentation architecture for CT and MR images, with pixel- and feature-level adaptation. In our experiment, we firstly obtain semantic nuclei segmentation predictions from SIFA. For the instance segmentation predictions, we then use the watershed algorithm to separate the touching objects in these binary predictions. Additionally, we also compare with our preliminary work for UDA nuclei instance segmentation [24], referred to as Liu *et al.*.

In addition to the UDA methods, we also compare with Hou *et al.* [17], which was particularly proposed for unsupervised nuclei segmentation in the histopathology images. In Hou *et al.*, a data generator is firstly employed for the synthesized histopathology images from the randomly generated nuclei masks. Then, they employed the real and synthesized histopathology for nuclei segmentation by learning multiple tasks including image refinement, segmentation, and detection.

The quantitative and qualitative comparison results are shown in Table III and Fig. 5, respectively. As illustrated in Table III, our proposed PDAM method outperforms all the comparison methods in terms of AJI, F1, and PQ. Furthermore, we employ the one-tailed paired t-test to calculate the p-value between our PDAM method and each comparison method. Given all the p-values are under 0.01, our improvement is statistically significant. In Table III, w/o DA means directly training the fully supervised Mask R-CNN on the source BBBC039V1 dataset and testing it on the target histopathology dataset.

With the pixel-level adaptation and feature adaptation at the image and instance levels, the baseline CyCADA improves the Mask R-CNN without adaptation by $2\% \sim 22\%$ under all three metrics. Due to the effectiveness of the auxiliary objects inpainting mechanism, panoptic-level feature alignment, and task re-weighting mechanism, our previous work [24] then improves the baseline CyCADA significantly. In this work, we propose a feature similarity maximization mechanism to narrow the domain gap by enlarging the mutual contextual information between the source and target domain. Our current PDAM method outperforms the previous version under all three metrics and further improves the baseline CyCADA by $7\% \sim 17\%$.

Chen *et al.* 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 is not enough to reduce the domain gap and may even downgrade the object-level performance, such as PQ. With the appearance-level adaptation, all the other UDA methods avoid the influence of the large appearance dissimilarity and achieve better performance. Although DDMRL only adapts the features at the image level, its performance is still at the same level as CyCADA, by adapting knowledge across various domains. SIFA is a UDA semantic segmentation structure that alleviates the domain-bias at the image and semantic levels. In the UDA instance nuclei segmentation scenario, there exists a large number of nuclei objects in the histopathology images with complicated distributions and overlapping issues. The effectiveness of SIFA is therefore limited without any instance-level feature learning or adaptation. In the $BBBC039 \rightarrow Kumar$ experiment, Hou *et al.* outperforms other comparison methods due to their task-specific data generator for the synthesized histopathology images. However, we notice that their performance on the $BBBC039 \rightarrow TNBC$ experiment is not that competitive. As TNBC and Kumar are obtained from different sources, there still remains a domain gap between them, which makes the data generator in Hou *et al.* synthesize imperfect histopathology images. Although the pipeline in Hou *et al.* is effective for unsupervised nuclei segmentation in the histopathology images from the TCGA

TABLE IV
COMPARISON EXPERIMENTS BETWEEN OUR UDA METHOD AND FULLY SUPERVISED METHODS, FOR BBBC039V1 TO KUMAR EXPERIMENT. FOR CNN3 AND DIST, THE RESULTS OF PQ ARE UNKNOWN

	AJI			F1		
Methods	seen	unseen	all	seen	unseen	all
CNN3 [8]	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 [9]	0.5594(0.0598)	0.5604(0.0663)	0.5598(0.0781)	0.7756(0.0489)	0.8005(0.0538)	0.7863(0.0550)
Liu *et al.* [24]	0.5432(0.0477)	0.5848(0.0951)	0.5610(0.0982)	0.7743(0.0358)	0.8068(0.0698)	0.7882(0.0533)
PDAM	0.5463(0.0522)	0.5907(0.0974)	0.5653(0.0751)	0.7763(0.0304)	0.8092(0.0618)	0.7904(0.0474)
Upper bound [35]	0.5703(0.0480)	0.5778(0.0671)	0.5735(0.0855)	0.7796(0.0419)	0.8007(0.0511)	0.7886(0.0531)

database, their performance might be limited when validated on other histopathology images, such as TNBC. By contrast, our proposed PDAM method is effective for UDA instance segmentation for different histopathology datasets.

2) In Comparison With Fully Supervised Methods: In addition to the unsupervised methods, we compare our PDAM method with several fully supervised methods for nuclei instance segmentation on the Kumar dataset. With the same data split as CNN3 [8] and DIST [9], we directly compare our results with theirs. CNN3 is a contour-based architecture with three segmentation classes, including the foreground nuclei, background and the nuclei boundaries. DIST is a regression model based on the distance map. For the fully supervised upper bound, we select the Panoptic FPN [35], which incorporates a semantic segmentation branch with the Mask R-CNN. The Panoptic FPN is trained directly on the Kumar dataset with the same split and under the same setting as our PDAM method. We split the 16 testing images into two subsets: seen and unseen set. The seen set contains 8 images from 4 organs known to the training set and the unseen contains the remaining 6 images from 3 organs unknown to the training set.

As shown in Table IV, 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 at the local level, while CNN3 and DIST only process the image at a global semantic level. Even though our AJI is slightly lower than the fully supervised Panoptic FPN, we notice that our method works better when tested on the unseen testing set. Since our proposed PDAM method 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 remain 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 the fully supervised methods. Additionally, the PDAM method outperforms its previous version [24] on both seen and unseen testing set, which further indicates the promotion and robustness of our proposed feature similarity maximization mechanism.

E. Adaptation Between Electron Microscopy Images From Different Sources

To further demonstrate the generalization ability of our PDAM method, we validate it on UDA mitochondria instance segmentation for EM images. Specifically, we employ the EPFL and VNC datasets as the source and target domains, respectively. As introduced in Section IV-A, these two datasets are acquired from different species and by different institutions, and hence there exists a large domain discrepancy. As discussed in Section III-A, the synthesized images from CycleGAN do not contain unexpected objects without corresponding annotations. Therefore we exclude the nuclei inpainting mechanism from our PDAM method in the experiment under this scenario.

TABLE V
COMPARISON EXPERIMENTS FOR THE UDA INSTANCE MITOCHONDRIA SEGMENTATION FROM THE EPFL TO VNC DATASET. RESULTS ARE PRESENTED AS MEAN VALUE WITH STANDARD DEVIATION IN THE PARENTHESES

	AJI	F1	PQ
w/o DA	0.4062(0.0855)	0.6755(0.0366)	0.2880(0.0932)
UDA Baseline [30]	0.5209(0.0981)	0.7556(0.0510)	0.3864(0.1324)
Liu *et al.* [24]	0.5941(0.1106)	0.8086(0.0529)	0.4607(0.1663)
PDAM	**0.5974 (0.1001)**	**0.8336 (0.0294)**	**0.4808 (0.1649)**

As shown in Table V and Fig. 6, we compare our PDAM method with the Mask R-CNN without domain adaptation (w/o DA), CyCADA [30] as the UDA instance segmentation baseline, and our previous work [24](Liu *et al.*). All the comparison methods have the same settings as Section IV-D. Due to the domain shift, the Mask R-CNN without domain adaptation performs poorly on the target testing set. With the feature-level and appearance-level adaptation, the CyCADA lifts the performance significantly. By removing the domain shift from the semantic-level features and avoiding the influence from the source-biased features, our previous work outperforms the CyCADA (UDA baseline) by 5% $\sim$ 8%. In addition, our newly proposed feature similarity maximization mechanism further improves our previous work by enlarging the mutual contextual information in the domain-invariant feature between the source and target domains. The overall performance of our PDAM method in this scenario is consistent with Section IV-D, which indicates that our PDAM method is effective and robust for the UDA instance segmentation in various kinds of microscopy datasets under different settings.

V. DISCUSSION

A. Analysis of the Feature Similarity Maximization Mechanism

Based on our previous work [24], we further propose to alleviate the domain bias from the perspective of representation learning. The key insight of the feature similarity

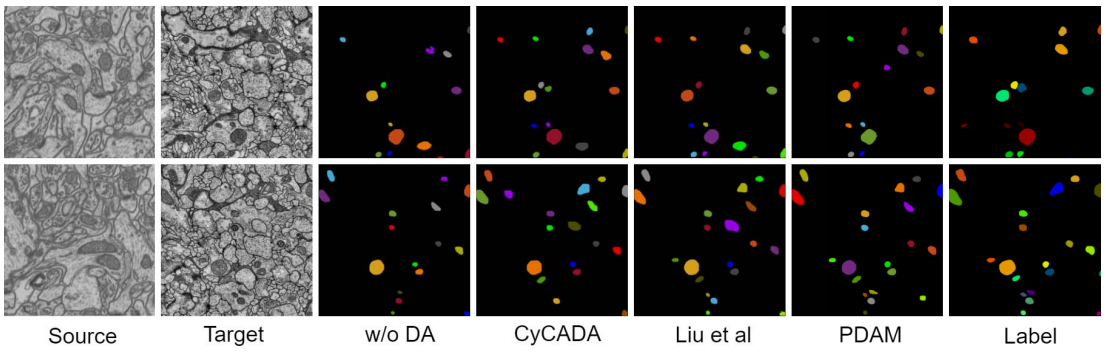

Fig. 6. Visual comparison on the experiments by adapting from the EPFL to the VNC dataset. The predictions are on the target testing images.

TABLE VI
EXPERIMENTAL RESULTS ON THE EFFECTIVENESS OF DIFFERENT DISTRIBUTION SIMILARITY MEASUREMENT METHODS FOR THE FEATURE SIMILARITY MAXIMIZATION MECHANISM

	$BBBC039V1 \rightarrow Kumar$			$BBBC039V1 \rightarrow TNBC$			$EPFL \rightarrow VNC$		
	AJI	$F1$	PQ	AJI	$F1$	PQ	AJI	$F1$	PQ
w/o FSM	0.5610(0.0718)	0.7882(0.0533)	0.5197(0.0865)	0.5672(0.0646)	0.7593(0.0566)	0.5207(0.0522)	0.5941(0.1106)	0.8086(0.0529)	0.4607(0.1663)
PDAM-Cosine	0.5573(0.0894)	0.7710(0.0684)	**0.5344 (0.0835)**	0.5706(0.0561)	0.7620(0.0486)	0.5374(0.0380)	0.5611(0.1015)	0.7917(0.0426)	**0.4997 (0.1646)**
PDAM-MI	**0.5653 (0.0751)**	**0.7904 (0.0474)**	0.5249(0.0884)	**0.5726 (0.0414)**	**0.7742 (0.0302)**	**0.5409 (0.0331)**	**0.5974 (0.1001)**	**0.8336 (0.0294)**	0.4808(0.1649)

TABLE VII
THE ABLATION STUDY FOR THE TASK RE-WEIGHTING MECHANISM

	$BBBC039V1 \rightarrow Kumar$			$BBBC039V1 \rightarrow TNBC$			$EPFL \rightarrow VNC$		
	AJI	$F1$	PQ	AJI	$F1$	PQ	AJI	$F1$	PQ
w/o TR	0.5239(0.0821)	0.7704(0.0595)	0.4777(0.0973)	0.5285(0.0483)	0.7472(0.0435)	0.5046(0.0487)	0.5265(0.0677)	0.7839(0.0496)	0.4446(0.1590)
PDAM	**0.5653 (0.0751)**	**0.7904 (0.0474)**	**0.5249 (0.0884)**	**0.5726 (0.0414)**	**0.7742 (0.0302)**	**0.5409 (0.0331)**	**0.5974 (0.1001)**	**0.8336 (0.0294)**	**0.4808 (0.1649)**

maximization mechanism is to induce the dependency of the domain-invariant features between the source and target domains. Previously, feature similarity maximization has been applied for feature disentanglement [65], [66], GAN training stabilization [36], and unsupervised learning [37]. In addition to the mutual information, cosine similarity is also effective for feature similarity measurement in recommendation system [67], and image instance segmentation [68]. In this section, we present our extensive experimental results on the effectiveness of different metrics for feature similarity in our PDAM method.

First, we incorporate the cosine similarity to the Feature Similarity Maximization Mechanism in the PDAM method. For the instance-level features maps F_{is} and F_{it} defined in Section III-E, we first flatten them along the batch axis, since the cosine similarity is employed between two vectors not tensors. The flattened instance-level feature vectors for the source and target domains are defined as F_{isf} and F_{itf}, respectively. Next, the cosine similarity between F_{isf} and F_{itf} is defined as:

$$Cosine(F_{isf}, F_{itf}) = \frac{F_{isf} \cdot F_{itf}}{\| F_{isf} \|_2 \cdot \| F_{itf} \|_2} \tag{6}$$

The $Cosine(F_{isf}, F_{itf})$ is in range $[-1, 1]$, while -1 indicates that F_{is} and F_{it} are at the opposite direction, and hence linearly dependent. To this end, we define the cosine similarity based feature similarity maximization mechanism as:

$$L_{FSM-cosine} = Cosine(F_{isf}, F_{itf}) \tag{7}$$

By minimizing the $L_{FSM-cosine}$ via SGD optimization, the $Cosine(F_{isf}, F_{itf})$ is pushed to -1, which induces F_{is} and F_{it} to be dependent on each other. Due to the necessity of having features flattened across the batch axis, the cosine similarity Feature Similarity Maximization Mechanism is not suitable for the semantic- and image-level features, given the batch size as 1.

The experimental results are presented in Table VI, where the w/o FSM method represents our previous work [24] without the feature similarity maximization mechanism, PDAM-Cosine and PDAM-MI are the PDAM method with the feature similarity maximization based on the cosine similarity and mutual information, respectively.

Compare with the PDAM-Cosine method, the PDAM-MI method achieves better performance except for the PQ in the $BBBC039V1 \rightarrow Kumar$ and $EPFL \rightarrow VNC$ experiments. However, the PDAM-MI method outperforms the w/o FSM method under all three metrics in different settings, while the PDAM-Cosine method sometimes even leads to a performance drop. As an essential step to calculate the cosine similarity, the tensor feature maps are flattened to vectors. Therefore, the spatial information in each ROI is deprecated and the PDAM-Cosine method is only capable to induce partial cross-domain instance-level features to be dependent. For other cross-domain instance-level features, there still exist domain-specific factors, which are harmful for the target learning and results in unstable improvement. On the other hand, the MI can be directly obtained from the tensor feature maps, which maintains sufficient spatial information for each ROI. Therefore, the PDAM-MI method is more stable and robust than the PDAM-Cosine method.

B. The Effectiveness of the Task Re-Weighting Mechanism

In order to present the Task Re-weighting Mechanism effects experimentally, we conduct an ablation study and the results are shown in Table VII. The w/o TR method is the

PDAM paradigm without the task re-weighting mechanism, which is implemented by setting the α_{img}, α_{ins}, and α_{sem} in Equation 5 to 1.0. The task re-weighting mechanism removes the source-specific factors for the detection and segmentation task learning, by down-weighting the task loss functions when the features are easy to be distinguished by the domain discriminators. To this end, removing the task re-weighting mechanism incurs source-biased predictions and results in the performance drop by when tested on the target images.

C. Computational Complexity Analysis

We conducted a computational complexity analysis between the PDAM method and our previous work [24]. To calculate the MI between a pair of cross-domain ROI in the feature similarity maximization mechanism, the FLOPs is about 2.36M, while the mask segmentation branch for each ROI has 1330M FLOPs. On the other hand, the MI estimator in the PDAM contains 0.59M parameters, while the overall amount of parameters of the PDAM model is about 82M. To this end, the computational cost of the MI-based feature similarity maximization mechanism is negligible. Furthermore, the PDAM method outperforms our previous work [24] under different metrics in various UDA settings, which demonstrates our proposed feature similarity maximization mechanism is effective and efficient.

VI. Conclusion

In this work, we propose a PDAM method for UDA instance segmentation in microscopy images by alleviating the domain bias at the appearance and panoptic feature level. To further improve our previous work [24], we propose a feature similarity maximization mechanism, which maximize the mutual information between the instance-level features of the source and target domains. In addition to adapting from fluorescence microscopy to the histopathology images in our previous work, we validate our method on the UDA instance segmentation between two EM datasets from different sources. Extensive experiments under the three UDA scenarios indicate that our method successfully transfers the domain-invariant information from the source to the target domain, by outperforming the state-of-the-art comparison methods significantly. As our PDAM method is the first architecture specifically designed for UDA instance segmentation and achieves promising performance on microscopy images, we will extend and validate our method for UDA instance segmentation for general images in the future.

References

[1] C. W. Elston and I. 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*, vol. 19, no. 5, pp. 403–410, Nov. 1991.

[2] 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*, vol. 64, no. 9, pp. 1914–1921, Nov. 1989.

[3] F. Clayton, "Pathologic correlates of survival in 378 lymph node-negative infiltrating ductal breast carcinomas. Mitotic count is the best single predictor," *Cancer*, vol. 68, no. 6, pp. 1309–1317, Sep. 1991.

[4] A. Madabhushi *et al.*, "Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: Comparison to oncotype DX," *J. Pathol. Informat.*, vol. 2, no. 2, p. 1, 2011.

[5] D.-H. Cho, T. Nakamura, and S. A. Lipton, "Mitochondrial dynamics in cell death and neurodegeneration," *Cellular Mol. Life Sci.*, vol. 67, no. 20, pp. 3435–3447, Oct. 2010.

[6] R. J. Giuly, M. E. Martone, and M. H. Ellisman, "Method: Automatic segmentation of mitochondria utilizing patch classification, contour pair classification, and automatically seeded level sets," *BMC Bioinf.*, vol. 13, no. 1, p. 29, Dec. 2012.

[7] A. Lucchi, K. Smith, R. Achanta, G. Knott, and P. Fua, "Supervoxel-based segmentation of mitochondria in EM image stacks with learned shape features," *IEEE Trans. Med. Imag.*, vol. 31, no. 2, pp. 474–486, Feb. 2012.

[8] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.

[9] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 448–459, Feb. 2019.

[10] D. Zhang *et al.*, "Panoptic segmentation with an end-to-end cell R-CNN for pathology image analysis," in *Proc. MICCAI*. Cham, Switzerland: Springer, 2018, pp. 237–244.

[11] D. Liu *et al.*, "Nuclei segmentation via a deep panoptic model with semantic feature fusion," in *Proc. IJCAI*, Aug. 2019, pp. 861–868.

[12] D. Liu, D. Zhang, Y. Song, H. Huang, and W. Cai, "Cell R-CNN v3: A novel panoptic paradigm for instance segmentation in biomedical images," 2020, *arXiv:2002.06345*. [Online]. Available: http://arxiv.org/abs/2002.06345

[13] A. Lucchi, C. Becker, P. M. Neila, and P. Fua, "Exploiting enclosing membranes and contextual cues for mitochondria segmentation," in *Proc. MICCAI*. Cham, Switzerland: Springer, 2014, pp. 65–72.

[14] A. Jorstad and P. Fua, "Refining mitochondria segmentation in electron microscopy imagery with active surfaces," in *Proc. ECCV*. Cham, Switzerland: Springer, 2014, pp. 367–379.

[15] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *Proc. CVPR*, Jun. 2011, pp. 1521–1528.

[16] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. NeurIPS*, 2014, pp. 3320–3328.

[17] L. Hou, A. Agarwal, D. Samaras, T. M. Kurc, R. R. Gupta, and J. H. Saltz, "Robust histopathology image analysis: To label or to synthesize?" in *Proc. CVPR*, Jun. 2019, pp. 8533–8542.

[18] R. Bermúdez-Chacón, O. Altingövde, C. Becker, M. Salzmann, and P. Fua, "Visual correspondences for unsupervised domain adaptation on electron microscopy images," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1256–1267, Apr. 2020.

[19] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[20] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. ICML*, 2015, pp. 1180–1189.

[21] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. CVPR*, Jul. 2017, pp. 7167–7176.

[22] C. Chen, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, "Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation," in *Proc. AAAI*, 2019, pp. 865–872.

[23] Q. Dou, C. Ouyang, C. Chen, H. Chen, and P.-A. Heng, "Unsupervised cross-modality domain adaptation of ConvNets for biomedical image segmentations with adversarial loss," in *Proc. IJCAI*, Jul. 2018, pp. 691–697.

[24] D. Liu *et al.*, "Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting," in *Proc. CVPR*, Jun. 2020, pp. 4243–4252.

[25] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, "ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proc. CVPR*, Jun. 2019, pp. 2517–2526.

[26] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster R-CNN for object detection in the wild," in *Proc. CVPR*, Jun. 2018, pp. 3339–3348.

[27] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image translation with conditional adversarial networks," in *Proc. CVPR*, Jul. 2017, pp. 5967–5976.

[28] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. ICCV*, Oct. 2017, pp. 2223–2232.

[29] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Proc. ECCV*. Cham, Switzerland: Springer, 2018, pp. 172–189.

[30] J. Hoffman *et al.*, "Cycada: Cycle-consistent adversarial domain adaptation," in *Proc. ICML*, 2018, pp. 1989–1998.

[31] T. Kim, M. Jeong, S. Kim, S. Choi, and C. Kim, "Diversify and match: A domain adaptive representation learning paradigm for object detection," in *Proc. CVPR*, Jun. 2019, pp. 12456–12465.

[32] Z. He and L. Zhang, "Multi-adversarial faster-RCNN for unrestricted object detection," in *Proc. ICCV*, Oct. 2019, pp. 6668–6677.

[33] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. ICCV*, 2017, pp. 2980–2988.

[34] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *Proc. CVPR*, Jun. 2019, pp. 9404–9413.

[35] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proc. CVPR*, Jun. 2019, pp. 6399–6408.

[36] M. I. Belghazi *et al.*, "Mutual information neural estimation," in *Proc. ICML*, 2018, pp. 531–540.

[37] R. D. Hjelm *et al.*, "Learning deep representations by mutual information estimation and maximization," in *Proc. ICLR*, 2019. [Online]. Available: https://openreview.net/forum?id=Bklr3j0cKX

[38] R. Bermúdez-Chacón, P. Márquez-Neila, M. Salzmann, and P. Fua, "A domain-adaptive two-stream U-Net for electron microscopy image segmentation," in *Proc. ISBI*, Apr. 2018, pp. 400–404.

[39] M. Ghafoorian *et al.*, "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation," in *Proc. MICCAI*. Cham, Switzerland: Springer, 2017, pp. 516–524.

[40] N. Karani, K. Chaitanya, C. Baumgartner, and E. Konukoglu, "A lifelong learning approach to brain MR segmentation across scanners and protocols," in *Proc. MICCAI*. Cham, Switzerland: Springer, 2018, pp. 476–484.

[41] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. ICML*, 2015, pp. 97–105.

[42] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proc. AAAI*, 2016, pp. 2058–2065.

[43] Y. Li, L. Yuan, and N. Vasconcelos, "Bidirectional learning for domain adaptation of semantic segmentation," in *Proc. CVPR*, Jun. 2019, pp. 6936–6945.

[44] N. Inoue, R. Furuta, T. Yamasaki, and K. Aizawa, "Cross-domain weakly-supervised object detection through progressive domain adaptation," in *Proc. CVPR*, Jun. 2018, pp. 5001–5009.

[45] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proc. CVPR*, Jun. 2018, pp. 7472–7481.

[46] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. NeurIPS*, 2015, pp. 91–99.

[47] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. NeurIPS*, 2017, pp. 700–708.

[48] F. Mahmood *et al.*, "Deep adversarial training for multi-organ nuclei segmentation in histopathology images," *IEEE Trans. Med. Imag.*, early access, Jul. 5, 2019, doi: 10.1109/TMI.2019.2927182.

[49] J. Ren, I. Hacihaliloglu, E. A. Singer, D. J. Foran, and X. Qi, "Adversarial domain adaptation for classification of prostate histopathology whole-slide images," in *Proc. MICCAI*. Cham, Switzerland: Springer, 2018, pp. 201–209.

[50] Y. Zhang, S. Miao, T. Mansi, and R. Liao, "Task driven generative modeling for unsupervised domain adaptation: Application to X-ray image segmentation," in *Proc. MICCAI*. Cham, Switzerland: Springer, 2018, pp. 599–607.

[51] Y. Huang, H. Zheng, C. Liu, X. Ding, and G. K. Rohde, "Epithelium-stroma classification via convolutional neural networks and unsupervised domain adaptation in histopathological images," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 6, pp. 1625–1632, Nov. 2017.

[52] R. Bermúdez-Chacón, C. Becker, M. Salzmann, and P. Fua, "Scalable unsupervised domain adaptation for electron microscopy," in *Proc. MICCAI*. Cham, Switzerland: Springer, 2016, pp. 326–334.

[53] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[54] A. Telea, "An image inpainting technique based on the fast marching method," *J. Graph. Tools*, vol. 9, no. 1, pp. 23–34, Jan. 2004.

[55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Jun. 2016, pp. 770–778.

[56] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. CVPR*, Jul. 2017, pp. 2117–2125.

[57] S. Nowozin, B. Cseke, and R. Tomioka, "F-GAN: Training generative neural samplers using variational divergence minimization," in *Proc. NeurIPS*, 2016, pp. 271–279.

[58] V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, "Annotated high-throughput microscopy image sets for validation," *Nature Methods*, vol. 9, no. 7, p. 637, 2012.

[59] A. Lucchi, Y. Li, and P. Fua, "Learning for structured prediction using approximate subgradient descent with working sets," in *Proc. CVPR*, Jun. 2013, pp. 1987–1994.

[60] S. Gerhard, J. Funke, J. Martel, A. Cardona, and R. Fetter. (Nov. 2013). *Segmented Anisotropic ssTEM Dataset of Neural Tissue*. [Online]. Available: https://figshare.com/articles/dataset/Segmented_anisotropic_ssTEM_dataset_of_neural_tissue/856713, doi: 10.6084/m9.figshare.856713.v1.

[61] S. Graham *et al.*, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101563.

[62] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. ICCV*, Dec. 2015, pp. 1026–1034.

[63] Y. Wu and K. He, "Group normalization," in *Proc. ECCV*. Cham, Switzerland: Springer, 2018, pp. 3–19.

[64] A. Paszke *et al.*, "Automatic differentiation in Pytorch," in *Proc. NeurIPS Autodiff Workshop*, 2017.

[65] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets," in *Proc. NeurIPS*, 2016, pp. 2172–2180.

[66] R. T. Chen, X. Li, R. B. Grosse, and D. K. Duvenaud, "Isolating sources of disentanglement in variational autoencoders," in *Proc. NeurIPS*, 2018, pp. 2610–2620.

[67] J. Ma, C. Zhou, P. Cui, H. Yang, and W. Zhu, "Learning disentangled representations for recommendation," in *Proc. NeurIPS*, 2019, pp. 5711–5722.

[68] C. Payer, D. Štern, T. Neff, H. Bischof, and M. Urschler, "Instance segmentation and tracking with cosine embeddings and recurrent hourglass networks," in *Proc. MICCAI*. Cham, Switzerland: Springer, 2018, pp. 3–11.