Minimizing Labeling Cost for Nuclei Instance Segmentation and Classification with Cross-domain Images and Weak Labels

Siqi Yang[1*], **Jun Zhang**[2*], **Junzhou Huang** [2†], **Brian C. Lovell**[1] , **Xiao Han**[2†]

[1]The University of Queensland, Australia
[2]Tencent AI Lab, Shenzhen, China

siqi.yang@uq.net.au, junejzhang@tencent.com, joehhuang@tencent.com, lovell@itee.uq.edu.au, haroldhan@tencent.com

Abstract

Nucleus instance segmentation and classification in histopathological images is an essential prerequisite in pathology diagnosis/prognosis. However, nucleus annotations (e.g., segmentation and labeling) require domain experts, and annotating nuclei at pixel-level is time-consuming and labor-intensive. Moreover, nuclei from different cancer types vary in shapes and appearances. These inter-cancer variations require careful annotations for specific cancer types. Therefore, to minimize the labeling cost, we propose a novel application that considers each cancer type as an individual domain and apply domain adaptation techniques to improve the segmentation/classification performance among different cancer types. Unlike the previous studies that focus on unsupervised or weakly-supervised domain adaptation independently, we would like to discover what kinds of labeling can achieve the most cost-effective domain adaptation performance in nucleus instance segmentation and classification. Specifically, we propose a unified framework that is applicable to different level annotations: no annotations, image-level, and point-level annotations. Cyclic adaptation with pseudo labels and adversarial discriminator are utilized for unsupervised domain alignment. Image-level or point-level annotations are additionally adopted to supervise the nucleus classification and refine the pseudo labels. Experiments demonstrate the effectiveness and efficacy of the proposed framework (jointly using unsupervised and weakly supervised learning) on adapting the segmentation and classification model from one cancer type to 18 other cancer types.

1 Introduction

Segmenting nuclei and classifying each nucleus into a specific category (*e.g.,* neoplastic, epithelial, or inflammatory) are of importance and could assist the diagnose in digital pathology, such as survival prediction (Lu et al. 2018) and cancer recurrence prediction (Corredor et al. 2019). However, it often tasks hours to annotate an pathology 512×512 image patch (cropped from 40x magnification whole-slide

image) with full nucleus masks in a pixel-level. Currently, only a few datasets (*e.g.,* CoNSep (Graham et al. 2019) and PanNuke (Gamper et al. 2019)) for nucleus instance segmentation and classification are publicly available. To minimize the labeling cost, previous works in medical image analysis resort to the techniques of unsupervised domain adaptation or weakly supervised learning, but these techniques are utilized independently.

Unsupervised domain adaptation (UDA) methods minimize labeling costs by utilizing the cross-domain data and aligning the distribution shift between the labeled source domain data and unlabeled target domain data. In the area of medical image analysis, several studies have been proposed to tackle the UDA for image classification (Ren et al. 2018; Huang et al. 2017), semantic segmentation (Zhang et al. 2018a; Chen et al. 2019) and instance segmentation with binary classification (Liu et al. 2020).

Recent UDA methods for object detection and semantic segmentation have used two main mechanisms: adversarial training to align domain distribution and pseudo-label self-training. Adversarial alignment does not guarantee the alignment of class conditional distributions and thus some methods propose to select source-like images (Saito et al. 2019) or instances (Zhu et al. 2019a). Pseudo labels are the confident predictions by the source-trained model and thus are source-like samples. These two mechanisms mainly focus on aligning source-like samples on the target domain, and we denote them as easy-transfer samples. On the contrary, the hard-transfer samples, predictions with low probability, are ignored, and the target-specific characteristics on hard-transfer samples are not investigated.

On the other hand, as annotating the centers of nuclei only or annotating image-level labels require much fewer efforts, weakly supervised learning has been studied for nucleus instance segmentation (Tian et al. 2020; Qu et al. 2019; Yoo, Yoo, and Paeng 2019). However, these weakly supervised methods do not consider domain distribution alignment and thus could not make full use of the labeled data from different domains.

Both annotating weak labels and utilizing labeled cross-domain images are the effective efforts to minimize the labeling cost, but no work integrates them for nucleus instance segmentation and classification. As such, we propose

---

to utilize weak labels to help unsupervised domain adaptation, *i.e.,* weakly supervised domain adaptation (WDA), to achieve the trade-off between accuracy and labeling expenses. Moreover, we propose a unified framework that is applicable to both unsupervised and weakly-supervised domain adaptation. In this way, different levels of annotations, including no annotations, image-level, and point-level annotations, can be flexibly applied according to user requirements. With the help of weak labels, hard-transfer samples in UDA could be properly processed.

In our framework, for the unsupervised domain adaptation, an adversarial domain discriminator is used to align local features, *e.g.,* texture and color. To achieve the semantic-aware adaptation, we adopt the cyclic adaptation with pseudo labels (Yang et al. 2020). Besides the unsupervised learning, we introduce weak labels to assist the domain adaptation in two aspects. First, we utilize weak labels to improve the accuracy of nucleus classification. We add point-level supervision or image-level supervision during the model training. Additionally, we use the weak labels to refine the pseudo labels, as the pseudo labels might be noisy and class-imbalanced. Second, we propose to utilize the weak labels to align the easy-transfer and hard-transfer samples. To harvest the hard-transfer samples, we propose to classify the easy/hard-transfer samples by weak labels and minimize the distribution gap via adversarial alignment. For the easy-transfer samples, we use the pseudo labels to train the segmentation and classification. In contrast, for the hard-transfer samples, we only use the point-level/image-level labels to supervise the classification.

Our major contributions are summarized as follows:

- To the best of our knowledge, we are the first to jointly investigate the UDA and WDA for nucleus instance segmentation and classification.

- We propose a unified framework that is applicable to both UDA and WDA with different forms of annotations, *i.e.,* pixel-level and image-level annotations. Our framework can be easily tailored to a semi-supervised task.

- To harvest the hard-transfer samples ignored by the pseudo-label self-training, we propose using weak labels to determine and align the hard-transfer examples.

2 Related Work

Unsupervised Domain Adaptation. Towards the UDA for image classification, a vast number of methods have been proposed to minimize the domain discrepancy by aligning the feature distributions from source and target domains. For example, maximum mean discrepancy (Long et al. 2015, 2016), correlation alignment (Sun and Saenko 2016), joint distribution discrepancy loss (Long et al. 2017) and adversarial training (Ganin and Lempitsky 2015; Tzeng et al. 2014, 2017; Zhu et al. 2017).

UDA has also been explored for semantic segmentation and two adaptation techniques are widely-used: adversarial domain alignment (Hoffman et al. 2018; Chang et al. 2019; Wu et al. 2018; Zhang et al. 2018b) and pseudo-label self-training (Zou et al. 2018; Saleh et al. 2018; Lian et al.

2019; Vu et al. 2019; Du et al. 2019). The pseudo-label self-training based methods generate pixel-level pseudo labels and then use them to fine-tune the segmentation model. In addition to minimizing the inter-domain gap, Pan *et al.* (Pan et al. 2020) proposed to minimize the domain gap within the samples of the target domain. They determined the easy/hard-transfer samples by the entropy of predictions and used an adversarial domain discriminator to align them. However, their hard-transfer samples do not obtain enough supervision, and the hard-to-transfer information in those samples is still ignored. To overcome this issue, we propose to utilize weak labels to determine and learn the hard-transfer samples. In medical image analysis, UDA is also studied for the segmentation in CT images (Dou et al. 2018; Chen et al. 2019; Ouyang et al. 2019), X-ray images (Zhang et al. 2018a).

UDA for object detection has received much attention in the past two years (Chen et al. 2018; Zhu et al. 2019b; Saito et al. 2019; He and Zhang 2019; Kim et al. 2019; Xie et al. 2019; Hsu et al. 2020; Zhuang et al. 2020; Yang et al. 2020). Most methods conduct domain adaptation at both image-level and instance-level. Some methods (Saito et al. 2019; Zhu et al. 2019b) select source-like target images/instances for training via adversarial loss and some methods (Hsu et al. 2020; Yang et al. 2020) generate pseudo labels.

There is only one recent work (Liu et al. 2020) has tackled the UDA for instance segmentation. They first utilized CycleGAN (Zhu et al. 2017) to synthesize target-like images and then used an inpainting mechanism to remove the artifacts produced by CycleGAN. After that, adversarial domain discriminators are employed for image-level and instance-level adaptation. However, they only performed binary classification for instance classification, *i.e.,* nuclei/non-nuclei, while we aim to categorize the nuclei into four types. Also, our proposed framework does not rely on pixel-level translation like CycleGAN and could be complementary to it. None of the above methods tackle the instance segmentation and classification like ours.

Weakly Supervised Segmentation. Different types of weak labels has been used for semantic segmentation, including image-level (Pinheiro and Collobert 2015; Bearman et al. 2016; Chang et al. 2020), point-level (Bearman et al. 2016; Tian et al. 2020; Yoo, Yoo, and Paeng 2019), video-level (Tsai, Zhong, and Yang 2016; Chen et al. 2020), scribble (Lin et al. 2016; Vernaza and Chandraker 2017) and bounding box (Papandreou et al. 2015; Dai, He, and Sun 2015; Khoreva et al. 2017). In these works' settings, the training and test data are from the same domain, and only the weak labels are provided for the training data. However, the training does not consider domain alignment and therefore, could not make use of the fully-annotated data from the other domains.

Weakly-supervised Domain Adaptation. Several works proposed to use weak labels to help the domain adaptation for detection and segmentation (Paul et al. 2020; Dong et al. 2019; Paul et al. 2020). Inoue *et al.* (Inoue et al. 2018) proposed to use image-level annotations to select the instance-level pseudo labels that belong to the existing categories for
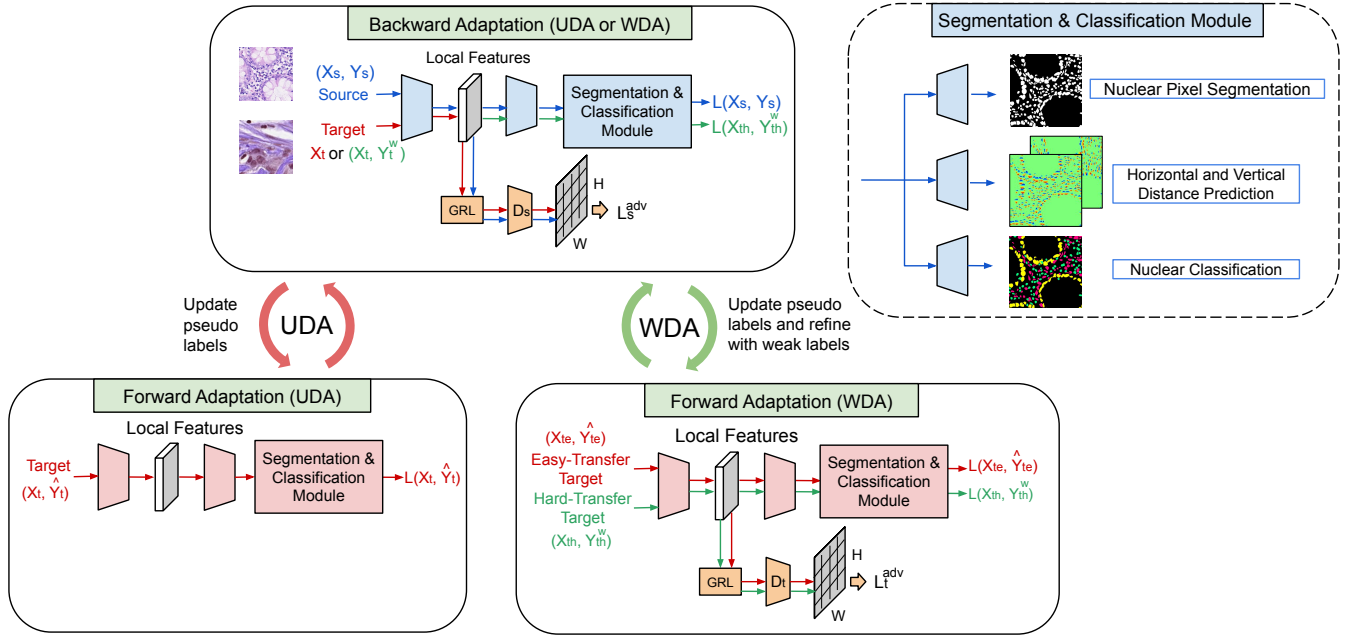
Figure 1: The proposed framework is applicable to unsupervised domain adaptation (UDA) and weakly-supervised domain adaptation (WDA). In UDA, two steps of forward adaptation and backward adaptation are computed iteratively with pseudo labels updated. For WDA, weak labels Y^W are utilized for minimizing weakly supervision loss L^W and pseudo label refinement. In forward adaptation for WDA, weak labels are used to determine and align easy/hard-transfer samples. For nucleus instance segmentation and classification, the decoder consists of three branches for nuclear pixel segmentation, prediction of horizontal and vertical distances to nuclear centroids, and nucleus classification.

self-training. Dong *et al.* (Dong et al. 2019) proposed that pseudo labeling tends to ignore hard-transfer classes, and therefore they normalized the class-wise probability according to the proportion of each class. However, the proportions of classes are still determined by pseudo labels. The most relevant work to us is (Paul et al. 2020). They designed a domain discriminator for each category and used the pseudo or weak labels to select the category-specific discriminator for semantic-aware feature alignment. Both their method and ours try to propose a unified framework for UDA and WDA. Different from their semantic segmentation task, the instance segmentation and classification task we tackled is more challenging. It is because in addition to classifying each pixel, instances need to be separated correctly. Our proposed pseudo-label self-training and hard-transfer samples alignment demonstrate better performance on this task.

3 A Unified Framework for UDA and WDA

As shown in Fig. 1, we integrate the UDA and WDA into a unified framework based on the pseudo-label self-learning. In this section, we first describe the problem we aim to solve and provide an overview of the proposed framework. We then introduce the details of UDA and WDA.

3.1 Problem Definition

In this work, we tackle both the unsupervised domain adaptation (UDA) and weakly-supervised domain adaptation

(WDA) for nucleus instance segmentation and classification. In domain adaptation, the source domain data and target domain data are the related data but from different distributions. In the settings of UDA, we have N_s labeled images $\{\mathcal{X}_s, \mathcal{Y}_s\} = \{X_s^i, Y_s^i\}_{i=1}^{N_s}$ in source domain and N_t non-labeled images $\mathcal{X}_t = \{X_t^i\}_{j=1}^{N_t}$. Note that in nucleus instance segmentation and classification, each pixel k is annotated with its instance id $z_k \in \{1, ..., M\}$ and nuclear type $c_k \in \{1, ..., C\}$, where M is the number of nucleus instances and C is the number of nuclear types. In WDA, we have weak labels for the target images, including point-level annotations $\mathcal{X}_t^P$ and image-level annotations $\mathcal{X}_t^I$. For each image, the point-level annotations are a set of centroids to the nuclei with their spatial locations and corresponding nuclear types. Image-level annotation for each image is a multi-hot vector that represents the available categories in the image.

3.2 Overview

In our framework, we employ the Hover-Net (Graham et al. 2019) as the base model for nucleus instance segmentation and classification . In Hover-Net, nucleus instance segmentation and classification are learnt simultaneously. The network consists of a shared encoder and three decoders for different tasks: 1) nuclear pixel segmentation, 2) prediction of horizontal and vertical distances to the nuclear centroids, and 3) nuclear type classification. The overview of network

architecture is shown in Fig 1.

Unsupervised Domain Adaptation. For the unsupervised domain adaptation, we extend the state-of-the-art UDA method (Yang et al. 2020) for object detection to our tasks of instance segmentation and classification. The training procedure is composed of multiple cycles. In each cycle, the training comprises two steps, and the network parameters are shared between these two steps.

Step 1: Backward Adaptation. Motivated by the theorems proposed by Ben-David *et al.* (Ben-David et al. 2010, 2007), this step performs domain adaptation by jointly minimizing the error on the source domain and the domain discrepancy across domains. In Hover-Net, the full supervision loss on the source domain includes the losses of nuclear pixel segmentation, distance prediction, and nuclear type classification. We denote it as $\mathcal{L}_s^F$ for simplicity. Details of the loss are shown in the supplementary materials.

To minimize the domain discrepancy, the gradient reversal layer (GRL) proposed by Ganin and Lempitsky (Ganin and Lempitsky 2015) is adopted, where the gradients of the domain classifier are reversed for domain confusion. The adversarial training loss is as follows:

$$\mathcal{L}_s^{adv}(\mathcal{X}_s, \mathcal{X}_t) = \frac{1}{2} \frac{1}{N_s HW} \sum_{i,h,w} (D_b(g_s^{(i)})^{(h,w)})^2$$
$$+ \frac{1}{2} \frac{1}{N_t HW} \sum_{j,h,w} (1 - D_b(g_t^{(j)}))^{(h,w)})^2 \,, \quad (1)$$

where g represents the features output from a specific layer of the encoder, H and W are height and width of the output feature map of the domain discriminator D. Then the training objective of this step is to minimize the following loss:

$$\mathcal{L}_b^U = \mathcal{L}_s^F(\mathcal{X}_s, \mathcal{Y}_s) + \lambda_s^{adv} \mathcal{L}_s^{adv}(\mathcal{X}_s, \mathcal{X}_t) \,. \quad (2)$$

Step 2: Forward Adaptation. In this step, in order to learn the target-specific information, pseudo labels $\hat{\mathcal{Y}}_t$ are used to refine the model obtained from the previous step. The predictions with high confidence generated from the model in the previous step are used as pseudo labels. The network is then optimized by the full supervision loss $\mathcal{L}_t^F$. Then the loss of this step for UDA is defined as

$$\mathcal{L}_f^U = \mathcal{L}_t^F(\mathcal{X}_t, \hat{\mathcal{y}}_t). \quad (3)$$

Weakly Supervised Domain Adaptation. Since no annotations are available on the target domain, the problem of unsupervised domain adaptation is challenging and there is still performance gap compared with the fully supervised method. With a small increase of labeling cost, we propose to use the weak labels to aid the UDA. First, we utilize the weak labels to supervise the classification of target samples. We introduce the supervision loss of the weak labels $\mathcal{L}_t^W$ into the training of Backward Adaptation. With the aid of weak labels, the training objective of Backward Adaptation in Eq. 2 can be re-written as

$$\mathcal{L}_b^W = \mathcal{L}_s^F(\mathcal{X}_s, \mathcal{Y}_s) + \lambda_s^{adv} \mathcal{L}_s^{adv}(\mathcal{X}_s, \mathcal{X}_t) + \lambda^W \mathcal{L}_t^W(\mathcal{X}_t, \mathcal{Y}_t^W).$$
$$(4)$$

Second, we use the weak labels to refine the class predictions in the pseudo labels as the pseudo labels might be noisy and class-imbalanced. Third, we propose to utilize the weak labels to determine and align the easy-transfer samples $\mathcal{X}_{te}$ and hard-transfer samples $\mathcal{X}_{th}$. More details of the WDA are present in the following sections.

3.3 Weak Labels for Supervising Nucleus Classification

Although the UDA method could achieve performance gain on both segmentation and classification tasks, the noisy and class-imbalanced pseudo labels may lead to the ignorance of infrequent classes without supervision. In order to guide the nucleus classification, we introduce weak supervision loss to supervise the training of classification networks. Different forms of annotations require different loss functions, and we illustrate them respectively in this section.

Point-level Supervision Loss. For an image X, the point-level annotation Y^P is a set of centroids of nuclei with their corresponding locations and types: $Y^P = \{[z^{(m)}, y_P^{(m)}]\}_{m=1}^M$. Here, $z^{(m)}$ is the location of the z_{th} centroid point, $y_P^{(m)}$ is an one-hot vector of its type, and M is the number of centroid points. Let us denote the probability vector at pixel $z^{(m)}$ as $p^{(m)}$. We can compute the cross-entropy loss over the centroid points for an image as follows:

$$\mathcal{L}^P(X, Y^P) = -\frac{1}{M} \sum_m y_P^{(m)} \log(p^{(m)}) \,. \quad (5)$$

Image-level Supervision Loss. The image-level annotation for each image Y^I is a multi-hot vector that represents the available categories in the image: $Y^I = \{y_I^{(c)} | c = 1, ..., C\}$. To obtain the image-level prediction, we compute the global average pooling across the spatial locations on the probability map of each class c. It represents the probability that category c exists in the image and we denote it as $a^{(c)}$. To obtain accurate predictions of the existing categories, we compute binary cross-entropy loss for each category and the image-level supervision loss is:

$$\mathcal{L}_t^I(X, Y^I) = \sum_{c=1}^C -y_I^{(c)} \log(a^{(c)}) - (1 - y_I^{(c)}) \log(1 - a^{(c)}) \,.$$
$$(6)$$

Pseudo Label Refinement. Since the pseudo labels might be noisy and class-imbalanced, we propose to utilize the weak labels to refine them. In this work, we specifically use the location and type information provided by the point-level annotations. Given a predicted instance, we measure the distance between its centroid and the ground truth centroids. If the distance is smaller than 6 pixels at 20x or 12 pixels at 40x, then we consider this instance as a true positive and otherwise false positive in detection. We discard the false positives and correct the predicted type of the true positives if the types are wrong. In this way, the performance of nucleus classification can be improved.

3.4 Weak Labels for Learning Hard-transfer Samples

Within the target domain, there also exist variations among samples. However, the pseudo labeling focuses on aligning source-like samples on target domain, and we denote them as easy-transfer samples. It is because the iterative self training with pseudo labels is equivalent to the EM algorithm (Sener et al. 2016). The hard-transfer samples, predictions with low probability, are ignored and the target-specific characteristics on hard-transfer samples are not investigated. To obtain good adaptation result, we propose that it is necessary to make use of the hard-transfer samples. To this end, several questions have been raised: 1) how to determine the hard samples; 2) how to train the hard samples.

A recent work (Pan et al. 2020) have proposed to use the entropy of predictions to determine the easy/hard samples and then align the entropy maps of them via a domain discriminator. However, aligning the entropy maps is agnostic to categories. Also, relying on the pseudo labels on easy-transfer samples still has the class-imbalanced problem. To overcome these problems, we propose to utilize the weak labels to further investigate the hard-transfer samples.

We use the point-level annotations to determine the easy/hard-transfer samples. More specifically, we calculate the recall of the predictions. If the recall of an image is larger than a certain threshold, we consider it as an easy-transfer sample, x_{te}; otherwise, a hard-transfer sample, x_{th}. For the easy-transfer samples, we use the pixel-wise pseudo labels to guide the training of segmentation with the full supervision loss, $\mathcal{L}_{te}^F$. For the hard-transfer samples, there might be no predictions or the pseudo labels are not reliable. We use the point-level annotations with the point-level supervision loss $\mathcal{L}_{th}^W$ to supervise the training instead.

In addition, to minimize the feature distributions between easy-transfer and hard-transfer samples, we propose to use an adversarial discriminator. The loss of adversarial training is as follows:

$$\mathcal{L}_t^{adv}(\mathcal{X}_{te}, \mathcal{X}_{th}) = \frac{1}{2}\frac{1}{N_{te}HW}\sum_{i,h,w}(D_f(g_{te}^{(i)})^{(h,w)})^2$$
$$+ \frac{1}{2}\frac{1}{N_{th}HW}\sum_{j,h,w}(1 - D_f(g_{th}^{(j)}))^{(h,w)})^2 , \quad (7)$$

where H and W are height and width of the output feature map of the domain discriminator, and g represents the features output from a specific layer in the encoder.

Then the objective of Forward Adaptation in WDA is :
$$\mathcal{L}_f^W = \mathcal{L}_{te}^F(\mathcal{X}_{te}, \hat{\mathcal{Y}}_{te}) + \mathcal{L}_{th}^W(\mathcal{X}_{he}, \mathcal{Y}_{he}^W) + \lambda_t^{adv}\mathcal{L}_t^{adv}(\mathcal{X}_{te}, \mathcal{X}_{th}) .$$
$$(8)$$

4 Experiments

In this section, we evaluate our proposed domain adaptation framework for the nucleus instance segmentation and classification. We demonstrate the results of UDA and WDA respectively, and provide ablation studies to show the effects of using weak labels.

Datasets. To the best of our knowledge, there are only two public datasets that provide annotations for both instance segmentation and classification: CoNSep (Graham et al.

2019) and PanNuke (Gamper et al. 2019). The domain adaptation scenario in this work is to adapt the segmentation and classification model from one cancer type (source domain) to the other 18 cancer types (target domain). We use CoNSep as source domain and PanNuke as target domain. Details of these datasets are shown below:

CoNSep (Graham et al. 2019). The colorectal nuclear segmentation and phenotype (CoNSep) dataset contains H&E stained image tiles of a single cancer type, *i.e.,* colorectal adenocarcinoma (CRA). It is made of 41 image tiles at 40x objective magnification with $24,319$ nuclei annotated by pathologists. The size of images is 1000×1000 pixels. The images were from 16 patients and were scanned at University Hospitals Coventry and Warwickshire, UK. This dataset categorizes the nuclei into 4 types: epithelial, inflammatory, spindle-shaped and miscellaneous. We use its training set with labels as our source domain.

PanNuke (Gamper et al. 2019). PanNuke contains annotated images for 19 cancer types, which were sampled from TCGA and a hospital in UK. All images were resized at 40x objective magnification and cropped into 256×256 pixels. It has over 200k nuclei annotated for both segmentation and classification. The annotations were first generated semi-automatically by a model trained with limited label data and then corrected by clinical experts. The nuclei are grouped into 5 categories: neoplastic, non-neoplastic epithelial, inflammatory, connective and dead. PanNuke divides the data into three splits. We use the first, second and third split as training, evaluation and test set of target domain.

We note that there is a mismatch of categories between these two datasets. For the normal epithelial and neoplastic epithelial, CoNSep groups them into a single class, while PanNuke categorizes them into two classes. Moreover, besides the neoplastic epithelial, the neoplastic class in PanNuke also includes tumors from other cell types, including sarcomas from soft tissue; melanomas from melanocytes; lymphomas from lymphoid cells. In order to map the categories from CoNSep to PanNuke, we group the neoplastic class and non-neoplastic epithelial class into one class: neoplastic and epithelial.

Metrics. For instance segmentation, we evaluate the segmentation performance in terms of Dice, Aggregated Jaccard Index (AJI), Detection Quality (DQ), Segmentation Quality (SQ), and Panoptic Quality (PQ), respectively.

The nucleus detection is evaluate with F1-score. The detected nuclei with more than 50% overlap with ground-truth nuclei are regarded as the true positive detections; and false positive detections, otherwise.

For instance classification, we use the same F1-score as HoverNet (Graham et al. 2019) to evaluate the classification performance for each nuclear type.

Methods in Comparison. We list the details of the methods in comparison as follows:

- Source Only: The model is the Hover-Net trained on the source data without adaptation.

- GRL (Ganin and Lempitsky 2015): The local features output the encoder of Hover-Net are fed into a domain discriminator, and a gradient reversal layer is used to flip the

Method	Detection	Classification			
		Neo and Epithelial	Inflammatory	Connective	Dead Cells
Source Only	0.630	0.259	0.232	0.273	0.018
GRL (Ganin and Lempitsky 2015)	0.692	0.330	0.247	0.224	0.017
Paul *et al.* (UDA) (Paul et al. 2020)	0.700	0.372	0.267	0.245	0.016
Ours (UDA)	0.700	0.381	0.277	0.233	0.019
Paul *et al.* (WDA: image) (Paul et al. 2020)	0.688	0.400	0.262	0.217	0.015
Paul *et al.* (WDA: point) (Paul et al. 2020)	0.705	0.416	0.285	0.257	**0.024**
Ours (WDA: image)	0.702	0.392	0.279	0.255	0.016
Ours (WDA: point)	**0.735**	**0.462**	**0.290**	**0.300**	0.023
Full Supervision	0.792	0.623	0.431	0.445	0.094

Table 1: F1-score for detection and classification.

Method	Dice	AJI	DQ	SQ	PQ
Source Only	0.576	0.387	0.461	0.657	0.342
GRL	0.723	0.509	0.587	**0.756**	0.450
Paul *et al.* (UDA)	0.731	0.501	0.600	0.751	0.446
Ours (UDA)	0.740	0.516	0.602	0.753	0.460
Paul *et al.* (WDA: image)	0.723	0.504	0.578	0.748	0.440
Paul *et al.* (WDA: point)	0.742	0.520	0.610	0.748	0.464
Ours (WDA: image)	0.738	0.511	0.592	0.748	0.449
Ours (WDA: point)	**0.743**	**0.534**	**0.628**	0.741	**0.471**
Full Supervision	0.824	0.652	0.756	0.813	0.622

Table 2: Results of instance segmentation on the adaptation from CoNSep to PanNuke.

Method	D	P	W	H	Dice	AJI	DQ	SQ	PQ
Source Only					0.630	0.259	0.232	0.273	0.018
UDA (ours)	✓				0.729	0.509	0.587	**0.756**	0.450
	✓	✓			0.740	0.516	0.602	0.753	0.460
WDA:image (ours)	✓		✓		0.735	0.514	0.591	0.740	0.449
	✓	✓	✓		0.738	0.511	0.592	0.748	0.449
WDA:point (ours)	✓		✓		0.731	0.508	0.588	0.749	0.447
	✓	✓	✓		0.741	0.523	0.612	0.748	0.464
	✓	✓	✓	✓	**0.743**	**0.534**	**0.628**	0.741	**0.471**

Table 3: Ablation studies for instance segmentation. D: the domain discriminator for low-level feature alignment, P: pseudo labeling, W: weak supervision loss and H: hard-transfer samples alignment.

gradients of the discriminator. It is considered as a module of our framework, and we denote it as Domain Discriminator (DD) in the following sections.

- Paul *et al.* (Paul et al. 2020): This method is the stat-of-the-art UDA and WDA method for semantic segmentation, where the image/point-level labels are used to select the corresponding category-specific domain discriminator. In their UDA settings, image-level predictions are served as pseudo weak labels. To make a fair comparison, we adapt their category-specific domain discriminators to the Hover-Net.

- Ours: The UDA of our method is a cyclic adaptation with pseudo labels and a domain discriminator for local features (DD). The WDA model with image-level labels is the UDA with weak supervision loss. The WDA with point-level labels is the UDA with weak supervision loss and hard-transfer sample alignment.

Training Details. The nucleus instance segmentation and classification model is designed based on the Hover-Net (Graham et al. 2019). We initialize the model with the weights of ResNet50 (He et al. 2016) pre-trained on ImageNet. The framework is implemented with TensorFlow. We use Adam optimizer with a learning-rate of 10^{-4} to train the decoder for the first 50 epochs and all layers for another 50 epochs with a batch size of 16. More implementation details are shown in supplementary materials.

4.1 Evaluation on Nucleus Instance Detection

As shown in Table 1, our method with point supervision (WDA:point) achieves the best detection performance com-

pared with all other methods based on UDA and WDA. Particularly, our method consistently outperforms the state-of-the-art method of Paul *et al.* (Paul et al. 2020) in both UDA and WDA. Compared with the best performance of Paul *et al.* (Paul et al. 2020) using WDA: point, Our WDA: point achieves 3% improvement in terms of F1-score of detection.

Different from the point-level supervision, methods based on WDA with image-level supervision do not achieve any significant improvements (even worse) compared with the UDA based methods. The reason should be that the proposed image-level label (*i.e.,* multi-hot vector that indicates available categories) is not related to nucleus detection task.

4.2 Evaluation on Nucleus Instance Classification

The classification results of all methods are summarized in Table 1. Similar to the detection performance, our method achieves superior classification performance compared with the state-of-the-art methods (*i.e.,* Paul *et al.*). Our F1 for the neoplastic-and-epithelial class and the connective class are 4.6% and 4.3% higher than those of Paul *et al.*.

Notably, with the help of image-level supervision, both our method and Paul *et al.* can improve the classification performance in most classification categories, in comparison with using UDA strategy only.

4.3 Evaluation on Nucleus Instance Segmentation

Table 2 illustrates the segmentation results of all methods. Except for the evaluation metric of SQ, our method based on WDA with point-level supervision achieves the best segmentation performance compared with all the other meth-

Method	D	P	W	H	Detection	Classification			
						Neo and Epithelial	Inflammatory	Connective	Dead Cells
Source Only					0.630	0.259	0.232	0.273	0.018
UDA (ours)	✓				0.692	0.330	0.247	0.224	0.017
	✓	✓			0.700	0.380	0.277	0.233	0.019
WDA: image (ours)	✓		✓		0.708	0.367	0.247	0.245	0.017
	✓	✓	✓		0.702	0.392	0.279	0.255	0.016
WDA: point (ours)	✓		✓		0.689	0.412	**0.301**	0.265	0.021
	✓	✓	✓		0.710	0.432	0.299	0.269	0.024
	✓	✓	✓	✓	**0.735**	**0.462**	0.290	**0.300**	**0.023**

Table 4: Ablation studies for nucleus detection and classification. F1-scores are shown for evaluation. D: the domain discriminator for low-level feature alignment, P:pseudo labeling, W: weak supervision loss and H: hard-transfer samples alignment.
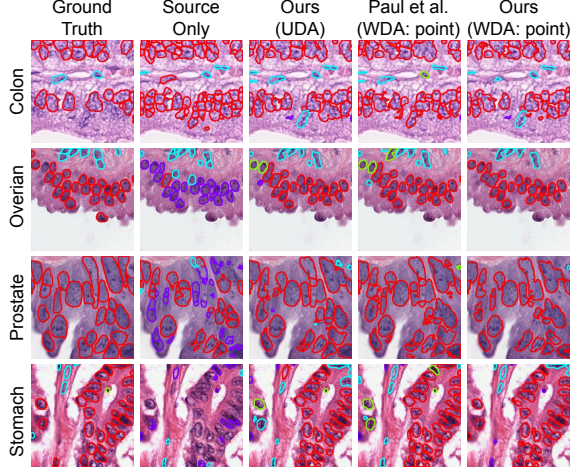

Figure 2: Examples of the nucleus segmentation and classification. Red: neoplastic and epithelial, green: inflammatory, blue: connective and purple: dead.

ods. Compared with the model training with source only, domain adaptation based methods improved the segmentation by around 15% in terms of Dice. Substantial improvement can also be observed in other evaluation metrics (*e.g.,* AJI, DQ, SQ, and PQ). Compared with the UDA based methods, the WDA based methods achieved comparable segmentation performance in terms of Dice. However, around 2% improvements could be observed on the other segmentation evaluation metrics, such as AJI, DQ, and PQ.

4.4 Ablation Studies

Effects of Weak Supervision Loss To evaluate the effects of the weak supervision loss, we only use a single adversarial domain discriminator as a baseline domain adaption method. Table 3 and 4 shows the segmentation, detection, and classification performance of all ablation methods. From the results, we have the following observations: 1) By adding the weak supervision of image-level loss, the classification is improved compared with the methods using UDA only. 2) By adding the weak supervision of point-level loss, almost all tasks have substantial performance improvement compared with the methods using UDA only. It demonstrates that our proposed framework of jointly using UDA and WDA can help increase the performance with relatively low cost of annotation.

In our experiments, we have also trained the model with both image-level and point-level supervision. However, the results are close to the result of using point-level supervision only. It probably because the point-level annotation is more specific and has contained the image-level information.

Effects of Pseudo Labels As discussed above, simply combing the weak supervision loss with a domain discriminator could increase the performance of nucleus classification. However, the improvement in segmentation is limited.

In UDA, with the pseudo labels, the Dice for evaluating instance segmentation is increased from 0.729 to 0.740. The other evaluation metrics in segmentation also indicate the improvement of using pseudo labels. Similar results are also obtained in the task of instance detection and classification.

With the combination of the pseudo labels and weak labels, both the segmentation and classification can be improved. It has better segmentation performance when compared with the model with weak supervision loss only, and higher classification accuracy when compared with pseudo labels only.

Effects of Aligning Hard-Transfer Samples In Table 3 and 4, we show the segmentation and classification performance with/without aligning hard-transfer samples. With the alignment of hard-transfer samples, the AJI is increased by 1%. For nucleus classification, the effects of aligning hard-transfer samples are more significant. The F1 of detection is increased by 2.5%. And the F1 for the neoplastic and epithelial class is increased by 3%.

5 Conclusions

In this paper, we perform a competitive study to investigate the importance of using UDA and WDA for nucleus instance segmentation and classification. We propose a unified framework that is applicable to both UDA and WDA with different forms of annotations, *i.e.,* no annotation, image-level annotation, and pixel-level annotation in the target domain data. Moreover, we propose to use weak labels to help align the hard-transfer examples. The experimental results suggest our proposed framework can achieve state-of-the-art performance in nucleus instance detection, classification, and segmentation performance, respectively.

Acknowledgements

References

Bearman, A.; Russakovsky, O.; Ferrari, V.; and Fei-Fei, L. 2016. What's the point: Semantic segmentation with point supervision. In *Proceedings of the European Conference on Computer Vision*.

Ben-David, S.; Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; and Vaughan, J. W. 2010. A theory of learning from different domains. *Machine Learning* 79(1): 151–175.

Ben-David, S.; Blitzer, J.; Crammer, K.; and Pereira, F. 2007. Analysis of representations for domain adaptation. In *Advances in neural information processing systems*.

Chang, W.-L.; Wang, H.-P.; Peng, W.-H.; and Chiu, W.-C. 2019. All about structure: Adapting structural information across domains for boosting semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Chang, Y.-T.; Wang, Q.; Hung, W.-C.; Piramuthu, R.; Tsai, Y.-H.; and Yang, M.-H. 2020. Weakly-Supervised Semantic Segmentation via Sub-Category Exploration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Chen, C.; Dou, Q.; Chen, H.; Qin, J.; and Heng, P.-A. 2019. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Chen, Y.; Li, W.; Sakaridis, C.; Dai, D.; and Van Gool, L. 2018. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Chen, Y.-W.; Tsai, Y.-H.; Lin, Y.-Y.; and Yang, M.-H. 2020. VOSTR: Video Object Segmentation via Transferable Representations. *International Journal of Computer Vision* .

Corredor, G.; Wang, X.; Zhou, Y.; Lu, C.; Fu, P.; Syrigos, K.; Rimm, D. L.; Yang, M.; Romero, E.; Schalper, K. A.; et al. 2019. Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non–small cell lung cancer. *Clinical cancer research* 25(5): 1526–1534.

Dai, J.; He, K.; and Sun, J. 2015. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*.

Dong, J.; Cong, Y.; Sun, G.; and Hou, D. 2019. Semantic-transferable weakly-supervised endoscopic lesions segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*.

Dou, Q.; Ouyang, C.; Chen, C.; Chen, H.; and Heng, P.-A. 2018. Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. In *Proceedings of the International Joint Conference on Artificial Intelligence*.

Du, L.; Tan, J.; Yang, H.; Feng, J.; Xue, X.; Zheng, Q.; Ye, X.; and Zhang, X. 2019. Ssf-dan: Separated semantic feature based domain adaptation network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*.

Gamper, J.; Koohbanani, N. A.; Benet, K.; Khuram, A.; and Rajpoot, N. 2019. PanNuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In *European Congress on Digital Pathology*.

Ganin, Y.; and Lempitsky, V. 2015. Unsupervised Domain Adaptation by Backpropagation. In *International Conference on Machine Learning*.

Graham, S.; Vu, Q. D.; Raza, S. E. A.; Azam, A.; Tsang, Y. W.; Kwak, J. T.; and Rajpoot, N. 2019. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis* 58: 101563.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

He, Z.; and Zhang, L. 2019. Multi-adversarial faster-rcnn for unrestricted object detection. In *Proceedings of the IEEE International Conference on Computer Vision*.

Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.-Y.; Isola, P.; Saenko, K.; Efros, A.; and Darrell, T. 2018. Cycada: Cycle-consistent adversarial domain adaptation. In *International Conference on Machine Learning*.

Hsu, H.-K.; Yao, C.-H.; Tsai, Y.-H.; Hung, W.-C.; Tseng, H.-Y.; Singh, M.; and Yang, M.-H. 2020. Progressive domain adaptation for object detection. In *The IEEE Winter Conference on Applications of Computer Vision*.

Huang, Y.; Zheng, H.; Liu, C.; Ding, X.; and Rohde, G. K. 2017. 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.

Inoue, N.; Furuta, R.; Yamasaki, T.; and Aizawa, K. 2018. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Khoreva, A.; Benenson, R.; Hosang, J.; Hein, M.; and Schiele, B. 2017. Simple does it: Weakly supervised instance and semantic segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*.

Kim, T.; Jeong, M.; Kim, S.; Choi, S.; and Kim, C. 2019. Diversify and match: A domain adaptive representation learning paradigm for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Lian, Q.; Lv, F.; Duan, L.; and Gong, B. 2019. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach. In *Proceedings of the IEEE International Conference on Computer Vision*.

Lin, D.; Dai, J.; Jia, J.; He, K.; and Sun, J. 2016. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Liu, D.; Zhang, D.; Song, Y.; Zhang, F.; O'Donnell, L.; Huang, H.; Chen, M.; and Cai, W. 2020. Unsupervised Instance Segmentation in Microscopy Images via Panoptic Domain Adaptation and Task Re-weighting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Long, M.; Cao, Y.; Wang, J.; and Jordan, M. I. 2015. Learning transferable features with deep adaptation networks. *International Conference on Machine Learning* .

Long, M.; Zhu, H.; Wang, J.; and Jordan, M. I. 2016. Unsupervised domain adaptation with residual transfer networks. In *Advances in neural information processing systems*.

Long, M.; Zhu, H.; Wang, J.; and Jordan, M. I. 2017. Deep transfer learning with joint adaptation networks. In *International Conference on Machine Learning*.

Lu, C.; Romo-Bucheli, D.; Wang, X.; Janowczyk, A.; Ganesan, S.; Gilmore, H.; Rimm, D.; and Madabhushi, A. 2018. Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. *Laboratory Investigation* 98(11): 1438–1448.

Ouyang, C.; Kamnitsas, K.; Biffi, C.; Duan, J.; and Rueckert, D. 2019. Data efficient unsupervised domain adaptation for cross-modality image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*.

Pan, F.; Shin, I.; Rameau, F.; Lee, S.; and Kweon, I. S. 2020. Unsupervised Intra-domain Adaptation for Semantic Segmentation through Self-Supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Papandreou, G.; Chen, L.-C.; Murphy, K. P.; and Yuille, A. L. 2015. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*.

Paul, S.; Tsai, Y.-H.; Schulter, S.; Roy-Chowdhury, A. K.; and Chandraker, M. 2020. Domain Adaptive Semantic Segmentation Using Weak Labels. In *Proceedings of the European Conference on Computer Vision*.

Pinheiro, P. O.; and Collobert, R. 2015. From image-level to pixel-level labeling with convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Qu, H.; Wu, P.; Huang, Q.; Yi, J.; Riedlinger, G. M.; De, S.; and Metaxas, D. N. 2019. Weakly supervised deep nuclei segmentation using points annotation in histopathology images. In *International Conference on Medical Imaging with Deep Learning*.

Ren, J.; Hacihaliloglu, I.; Singer, E. A.; Foran, D. J.; and Qi, X. 2018. Adversarial domain adaptation for classification of prostate histopathology whole-slide images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*.

Saito, K.; Ushiku, Y.; Harada, T.; and Saenko, K. 2019. Strong-Weak Distribution Alignment for Adaptive Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Saleh, F. S.; Aliakbarian, M. S.; Salzmann, M.; Petersson, L.; and Alvarez, J. M. 2018. Effective use of synthetic data for urban scene semantic segmentation. In *Proceedings of the European Conference on Computer Vision*.

Sener, O.; Song, H. O.; Saxena, A.; and Savarese, S. 2016. Learning transferrable representations for unsupervised domain adaptation. In *Advances in neural information processing systems*.

Sun, B.; and Saenko, K. 2016. Deep coral: Correlation alignment for deep domain adaptation. In *Proceedings of the European Conference on Computer Vision*.

Tian, K.; Zhang, J.; Shen, H.; Yan, K.; Dong, P.; Yao, J.; Che, S.; Luo, P.; and Han, X. 2020. Weakly-Supervised Nucleus Segmentation Based on Point Annotations: A Coarse-to-Fine Self-Stimulated Learning Strategy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*.

Tsai, Y.-H.; Zhong, G.; and Yang, M.-H. 2016. Semantic co-segmentation in videos. In *Proceedings of the European Conference on Computer Vision*.

Tzeng, E.; Hoffman, J.; Saenko, K.; and Darrell, T. 2017. Adversarial Discriminative Domain Adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Tzeng, E.; Hoffman, J.; Zhang, N.; Saenko, K.; and Darrell, T. 2014. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474* .

Vernaza, P.; and Chandraker, M. 2017. Learning random-walk label propagation for weakly-supervised semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Vu, T.-H.; Jain, H.; Bucher, M.; Cord, M.; and Pérez, P. 2019. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Wu, Z.; Han, X.; Lin, Y.-L.; Gokhan Uzunbas, M.; Goldstein, T.; Nam Lim, S.; and Davis, L. S. 2018. Dcan: Dual channel-wise alignment networks for unsupervised scene adaptation. In *Proceedings of the European Conference on Computer Vision*.

Xie, R.; Yu, F.; Wang, J.; Wang, Y.; and Zhang, L. 2019. Multi-level Domain Adaptive learning for Cross-Domain Detection. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*.

Yang, S.; Wu, L.; Wiliem, A.; and Lovell, B. C. 2020. Unsupervised Domain Adaptive Object Detection using Forward-Backward Cyclic Adaptation. *arXiv preprint arXiv:2002.00575* .

Yoo, I.; Yoo, D.; and Paeng, K. 2019. Pseudoedgenet: Nuclei segmentation only with point annotations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*.

Zhang, Y.; Miao, S.; Mansi, T.; and Liao, R. 2018a. Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*.

Zhang, Y.; Qiu, Z.; Yao, T.; Liu, D.; and Mei, T. 2018b. Fully convolutional adaptation networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*.

Zhu, X.; Pang, J.; Yang, C.; Shi, J.; and Lin, D. 2019a. Adapting object detectors via selective cross-domain alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Zhu, X.; Pang, J.; Yang, C.; Shi, J.; and Lin, D. 2019b. Adapting Object Detectors via Selective Cross-Domain Alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Zhuang, C.; Han, X.; Huang, W.; and Scott, M. R. 2020. iFAN: Image-Instance Full Alignment Networks for Adaptive Object Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Zou, Y.; Yu, Z.; Vijaya Kumar, B.; and Wang, J. 2018. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European Conference on Computer Vision*.