

Domain Adaptive Nuclei Instance Segmentation and Classification via Category-aware Feature Alignment and Pseudo-labelling

Canran Li¹, Dongnan Liu¹, Haoran Li^{2,3}, Zheng Zhang⁴, Guangming Lu⁴,
Xiaojun Chang², and Weidong Cai¹

¹ School of Computer Science, University of Sydney, Australia

² ReLER, AAIL, University of Technology Sydney, Australia

³ Department of Data Science and Artificial Intelligence, Monash University,
Australia

⁴ School of Computer Science and Technology, Harbin Institute of Technology,
Shenzhen, China

cali5184@uni.sydney.edu.au

Abstract. Unsupervised domain adaptation (UDA) methods have been broadly utilized to improve the models' adaptation ability in general computer vision. However, different from the natural images, there exist huge semantic gaps for the nuclei from different categories in histopathology images. It is still under-explored how could we build generalized UDA models for precise segmentation or classification of nuclei instances across different datasets. In this work, we propose a novel deep neural network, namely Category-Aware feature alignment and Pseudo-Labeling Network (CAPL-Net) for UDA nuclei instance segmentation and classification. Specifically, we first propose a category-level feature alignment module with dynamic learnable trade-off weights. Second, we propose to facilitate the model performance on the target data via self-supervised training with pseudo labels based on nuclei-level prototype features. Comprehensive experiments on cross-domain nuclei instance segmentation and classification tasks demonstrate that our approach outperforms state-of-the-art UDA methods with a remarkable margin.

Keywords: Computational pathology · Nuclear segmentation · Nuclear classification · Unsupervised domain adaption · Deep learning.

1 Introduction

Automatic nuclei instance segmentation and classification are crucial for digital pathology with various application scenarios, such as tumour classification and cancer grading [1]. However, manual labelling is limited by high subjective, low reproducibility, and resource-intensive [2,3]. Although deep learning-based methods can achieve appealing nuclei recognition performance, they require sufficient labelled data for training [3,4,5,6,7]. By directly adopting the off-the-shelf deep learning models to a new histopathology dataset with a distinct distribution,

the models suffer from performance drop due to domain bias [8,9]. Recently, unsupervised domain adaption methods have been proposed to tackle this issue [10,11,12], and enable the learning models to transfer the knowledge from one labelled source domain to the other unlabelled target domain [13,14,15].

Several methods have recently been proposed for unsupervised domain adaptive nuclei instance segmentation in histopathology images [8,9]. Inspired by CyCADA [16], Liu et al. [8] first synthesize target-like images and use a nuclei inpainting mechanism to remove the incorrectly synthesized nuclei. The adversarial training strategies are then used separately at the image-, semantic- and instance-level with a task re-weighting mechanism. However, this work can only address the instance segmentation for nuclei within the same class. In Yang et al.’s work [9], firstly, local features are aligned by an adversarial domain discriminator, and then a pseudo-labelling self-training approach is used to further induce the adaptation. However, the performance improvement of this work relies on weak labels and fails to get good training results without target domain labels. In addition, although this work is validated on the cross-domain nuclei segmentation and classification, their adaptation strategies are class-agnostic. In the real clinical, the nuclei objects in the histopathology images belong to various classes, and the characteristics of the nuclei in different classes are also distinct [17]. In addition, the number of objects within each category is also imbalanced in the histopathology datasets [18]. To this end, previous UDA methods are limited for cross-domain nuclei segmentation and classification due to the lack of analysis on the nuclei classes when transferring the knowledge.

To address the aforementioned issues, in this work, we study cross-domain nuclei instance segmentation and classification via a novel class-level adaptation framework. First, we propose a category-aware feature alignment module to facilitate the knowledge transfer for the cross-domain intra-class features while avoiding negative transfer for the inter-class ones. Second, a self-supervised learning stage via nuclei-level feature prototypes is further designed to improve the model performance on the unlabelled target data. Extensive experiments indicated the effectiveness of our proposed method by outperforming state-of-the-art UDA methods on nuclei instance segmentation and classification tasks. Furthermore, the performance of our UDA method is comparable or even better than the fully-supervised upper bound under various metrics.

2 Methods

2.1 Overview

Our proposed model is based on Hover-Net [6], a state-of-the-art method for fully-supervised nuclei instance segmentation and classification. The framework is constructed by three branches sharing the same encoder for feature extraction and using three decoders for different tasks: 1) Nuclear pixel (NP) branch to perform binary classification of a pixel (nuclei or background); 2) Hover (HV) branch to predict the horizontal and vertical distances of nuclei pixels to their

centroid; 3) Nuclear classification (NC) branch to classify the nuclei types of pixels. The supervised Hover-Net loss function of our model is defined as \mathcal{L}_F :

$$\mathcal{L}_F = \mathcal{L}_{np} + \mathcal{L}_{hover} + \mathcal{L}_{nc} \quad (1)$$

The network architecture of our proposed model is shown in Fig. 1. Our proposed UDA framework is optimized in two stages. First, class-level feature alignment modules are proposed to alleviate the domain gap at the feature level. In the second stage, the pseudo-labelling process enhanced by the nuclei-level prototype is further proposed for self-supervised learning on the unlabelled target images.

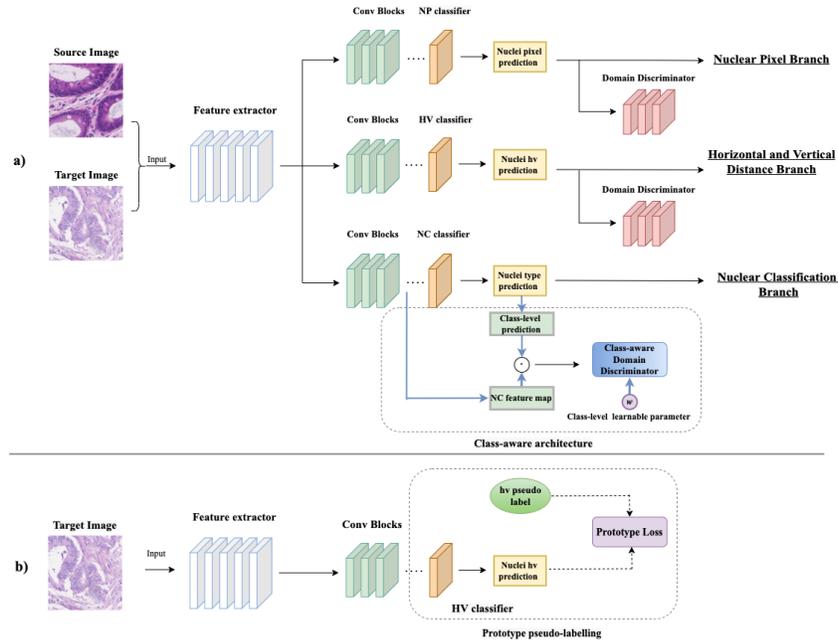


Fig. 1. Overview of the proposed category-aware prototype pseudo-labelling network.

2.2 Category-aware Feature Alignment

In the histopathology datasets with multi-class nuclei, there is a very large gap in the number of nuclear in each category. This may lead to many images in the dataset having only a few nuclear categories, and some nuclear types are absent from an image. In addition, the features of the nuclei from different classes also vary. Under this situation, the typical class-agnostic feature alignment strategies may lead to the negative transfer of the features in different categories. To tackle

this issue, for the NC branch on nuclei classification, we propose to conduct feature alignment for the cross-domain features within each class separately. Particularly, an adversarial domain discriminator D_c is introduced to adapt the features under the class c . Compared with directly employing a class-agnostic domain discriminator for the NC branch, the class-aware domain discriminators can avoid the misalignment across different classes and further encourage the knowledge transfer in classification learning.

The detailed paradigm is shown in Fig. 1a. First, we denote the features of the NC branch in the typical Hover-Net as F_{nc} , and the N-class prediction as P_{nc} . Note that P_{nc} contains N channels, and the predictions in each channel represent the classification results for each specific class. For each category c , we formulate the P_{nc} into a binary class prediction map P_{nc}^c , where the pixel value is set to 1 if it belongs to the nuclei in this category, otherwise set to 0. To incorporate the class-aware information into the feature alignment, we propose to generate the prototype features F_{pt}^c for the class c by dot-multiplying F_{nc} with the binary P_{nc}^c . In addition, if any P_{nc}^c is empty, we will not perform subsequent training on the nuclei in this class. In other words, we only deal with the nuclear types that exist in the input images. In each domain, the prototype features F_{pt}^c for the class c pass through the corresponding adversarial discriminators D_c for class-aware adaption at the feature level.

To avoid manually finetuning the trade-off weights for the adversarial loss in N categories of the NC branch, we let the overall framework automatically learn these weight parameters during training. The learnable weighted discriminator loss is formulated as follows:

$$\mathcal{L}_{NC}^{ca} = \omega_c^L \sum_{c=1} \mathcal{L}_c^{adv} \quad (2)$$

where the \mathcal{L}_c^{adv} denotes the adversarial training loss of D_c for class c , and ω_c^L is its corresponding learnable loss weight.

The overall domain discriminator loss function of our model is defined as:

$$\mathcal{L}_{dis} = \mathcal{L}_{NC}^{ca} + \mathcal{L}_{NP}^{adv} + \mathcal{L}_{HV}^{adv} \quad (3)$$

where \mathcal{L}_{NP}^{adv} and \mathcal{L}_{HV}^{adv} are the feature adaptation loss functions in the NP and HV branches, respectively. Particularly, we utilize adversarial domain discriminators on the cross-domain output features of the NP and HV branches for adaptation. Details of the supervised Hover-Net loss and the adversarial loss are shown in the Appendices. With the above loss terms, the overall loss function of the first stage approach can be written as:

$$\mathcal{L}_{s1} = \mathcal{L}_F + \mathcal{L}_{dis} \quad (4)$$

2.3 Nuclei-level Prototype Pseudo-labelling

Although the class-aware feature alignment modules can narrow the domain gaps, the lack of supervised optimization on the target images still limits the

model’s segmentation and classification performance. Therefore, in the second stage, we use the output of the first stage model as pseudo labels for self-supervised learning to further improve the model performance on the target images. The detailed training process can be referred to Fig. 1b.

Different from the traditional self-training process with the pseudo labels, we only train the target domain of the HV branch during the second stage pseudo-labelling process. In the extensive experiments, we noticed that the performance of the classification predictions on the target images from the first stage model is limited. In addition, the binary segmentation predictions lack object-wise information. Moreover, the model trained with all pseudo labels may not perform well due to the low quality of some pseudo-labelling classes. To avoid the disturbance from the less accurate and representative pseudo labels, we no longer consider classification and binary segmentation branches in the second stage but particularly focus on the predictions from the HV branch, where the feature maps describe the distance from each pixel to the nuclei’s centre point. Therefore, the features for each nuclear object can be regarded as a prototype at the object level, which contains morphological information such as the shape and size of the specific nuclear. By self-supervised learning with the pseudo labels on the predictions of the HV branch, the bias between nuclei objects can be further reduced. The overall loss function for the second stage is as follows:

$$\mathcal{L}_p = \frac{1}{N^p} \sum_{i=1}^{N^p} |x_i^p - \hat{y}_i^p|^2 \quad (5)$$

where x_i^p is the predicted features from the HV branch for each nuclear object p in the second stage and \hat{y}_i^p is the object features generated by the pseudo labels.

3 Experiments

3.1 Datasets and Evaluation Metrics

We conduct experiments on two datasets from Lizard [18], a large-scale colon tissue histopathology database at the 20x objective magnification for nuclei instance segmentation and classification under six types: epithelial, connective tissue, lymphocytes, plasma, neutrophils, and eosinophils.

In this work, DigestPath (Dpath) and CRAG are employed, where the images in Dpath are extracted from histological samples from four different hospitals in China, and the CRAG dataset contains images extracted from whole-slide images (WSIs) from University Hospitals Coventry and Warwickshire (UHCW). For both two datasets, we select 2/3 of the whole images for training, and the remaining 1/3 for testing and validation. Specifically, we use Dpath as the source domain, with 46 images for training and the rest 22 for validation. CRAG dataset is used as the target domain, with 42 images for training and the remaining 21 for testing. The training images are randomly cropped to 256×256 patches, and the data augmentation methods are applied, including flip, rotate, Gaussian blur and median blur.

For evaluation, we choose the same metrics as Hover-Net [6]. Dice, Aggregated Jaccard Index (AJI), Detection Quality (DQ), Segmentation Quality (SQ), and Panoptic Quality (PQ) are for nuclei instance segmentation. In addition, the F1-scores at the detection and classification levels are employed to evaluate the nuclei detection and classification performance.

3.2 Implementation Details

We utilize the Hover-Net framework with ResNet50 [19] pre-trained weights on ImageNet as our base architecture. In the first stage of our adaptation process, the model is trained in two steps following the Hover-Net [6]. In the first step, only the decoders are trained 50 epochs. In the second step, all layers are trained for another 50 epochs. We use Adam optimization in both steps, with an initial learning rate of 1e-4, which is then reduced to 1e-5 after 25 epochs. In the self-training pseudo-labelling part, the Adam optimizer with a learning rate of 1e-4 was used to train 50 epochs with a batch size of 20. Experiments were conducted on one NVIDIA GeForce 3090 GPU and implemented using PyTorch.

3.3 Comparison Experiments

We conduct a series of comparative experiments to compare the performance. The details are as follows: (1) Source Only [6]: original Hover-Net without adaptation. (2) PDAM [8]: a UDA nuclei instance segmentation method with pixel-level and feature-level adaptation. Since it can only be used for binary classification for objects, we only compare the results of nuclei instance segmentation with this method. (3) Yang et al. [9]: a UDA framework is proposed based on global-level feature alignment for nuclei classification and nuclear instance segmentation. In addition, a weakly-supervised DA method is also proposed using weak labels for the target images such as nuclei centroid. We only compare with the UDA method in this work for a fair comparison. (4) Fully-supervised: fully supervised training on the labelled target images, as the upper bound.

Table 1. Experimental results on UDA nuclei instance segmentation.

Methods	$D_{path} \rightarrow CRAG$				
	Dice	AJI	DQ	SQ	PQ
Source Only [6]	0.378	0.201	0.336	0.768	0.259
PDAM [8]	0.596	0.323	0.467	0.676	0.316
Yang <i>et al.</i> [9]	0.766	0.494	0.648	0.765	0.496
Baseline	0.750	0.455	0.604	0.759	0.458
Baseline+CA	0.772	0.502	0.661	0.773	0.510
Baseline+CA+PL	0.781	0.517	0.675	0.772	0.522
Proposed	0.785	0.519	0.681	0.769	0.524
Full-supervised	0.778	0.526	0.683	0.783	0.535

Evaluation on Nucleus Instance Segmentation. A comparison of the segmentation performance between our model and state-of-the-art methods is reported in Table 6. From the table, it can be observed that our instance segmentation effect is better than the two existing models. Compared with the source-only method, PDAM [8] has a performance improvement by aligning features at the panoptic level. The UDA method of Yang et al. [9] achieves good segmentation performance. Our method achieves the highest scores among all methods, with Dice and AJI being 2.3% and 1.6% higher than the previous methods, respectively. In comparison with the full-supervised model, the performance of our proposed UDA architecture is close to it. In particular, our Dice is higher than the results of the full-supervised model.

Table 2. Experimental results on UDA nuclei classification. F_c^1 , F_c^2 , F_c^3 , F_c^4 , F_c^5 and F_c^6 denote the F1 classification score for the Eosinophil, Epithelial, Lymphocyte, Plasma, Neutrophil and Connective tissue, respectively. F_{avg} denotes the average of all the F1-score for the classification under each category.

Methods	$Dpath \rightarrow CRAG$							
	Det	F_c^1	F_c^2	F_c^3	F_c^4	F_c^5	F_c^6	F_{avg}
Source Only [6]	0.490	0.022	0.389	0.324	0.195	0.038	0.161	0.188
Yang <i>et al.</i> [9]	0.736	0.037	0.670	0.330	0.371	0.017	0.428	0.309
Baseline	0.702	0.044	0.687	0.292	0.381	0.155	0.475	0.339
Baseline+CA	0.731	0.128	0.697	0.351	0.400	0.084	0.498	0.360
Baseline+CA+PL	0.772	0.110	0.725	0.327	0.383	0.352	0.558	0.409
Proposed	0.775	0.111	0.714	0.389	0.402	0.377	0.535	0.421
Full-supervised	0.748	0.167	0.724	0.388	0.419	0.428	0.545	0.445

Evaluation on Nucleus Classification. Table 5 reports the performance comparison of our method with other works on nuclei classification. Although Yang et al. [9] can achieve cross-domain nuclei classification and segmentation based on Hover-Net, their feature alignment modules are class-agnostic, which incurs the misalignment issues for the cross-domain features from different categories and further limits the performance. On the other hand, our proposed class-aware UDA framework has outperformed [9] by a large margin. Our model achieves a 4% improvement in the F1 detection score in the detection metric. In addition, our method improves the classification performance for all categories. Moreover, our method achieves significant improvements in classes with sparse samples like Neutrophil and Eosinophil, around 10% and 34%, respectively.

3.4 Ablation Studies

To test the validity of each component of our proposed model, we conducted ablation experiments. Firstly, based on our architecture, we kept only the feature-level domain discriminator on the three branches as our baseline method. Sec-

only, we kept only the class-aware structure, removing the pseudo labels based on the nuclei-level prototype (Baseline+CA). In addition, we also compared our nuclei-prototype pseudo-labelling process with the traditional one, which directly trains the model on the target images with all predictions in the first stage as the pseudo labels (Baseline+CA+PL).

Table 6 and Table 5 show the instance segmentation and classification performance of all the ablation methods. From the tables, we can observe that the class-aware structure substantially improves the classification performance under categories with sparse samples (e.g. Neutrophil and Eosinophil), with a higher than 10% improvement in the classification F1-score. This phenomenon illustrates the effectiveness of class-aware adaptation in transferring the knowledge between the multi-class datasets. We note that the class-aware structure also has an approximate 2% improvement in nuclei segmentation.

Self-supervised training also improved both instance segmentation and classification performance. In addition, we note that prototype loss has a higher than 4% improvement on the F1-score for nuclei detection and achieves a better nuclei segmentation performance. In addition, our proposed nuclei prototype pseudo-labelling process also outperforms the typical pseudo-labelling. Due to the inferior classification performance of the first stage model, training models with all the pseudo labels might bring the noise to the network optimization, and limit the overall performance. Visualization examples of the ablation studies are shown in Fig. 3.

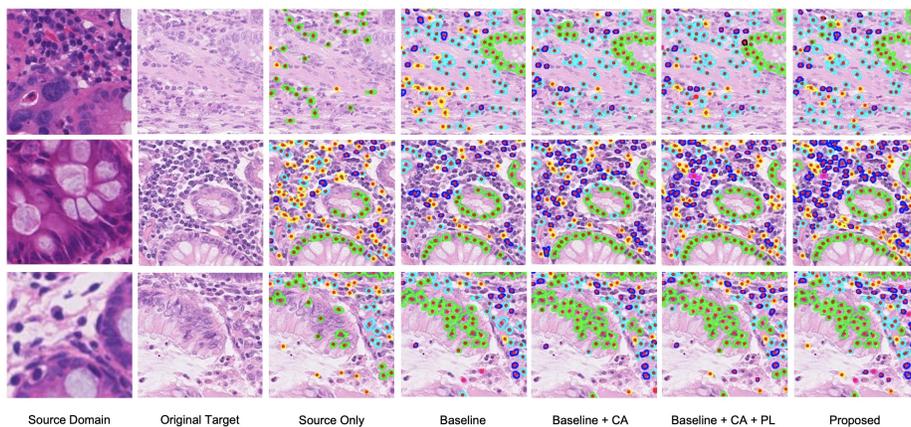


Fig. 2. Visualization predictions for the ablation experiments. Red: Eosinophil, Green: Epithelial, Yellow: Lymphocyte, Blue: Plasma, Magenta: Neutrophil, Cyan: Connective tissue. (Best to viewed in color and zoomed-in)

4 Conclusion

In this paper, we proposed a category-aware prototype pseudo-labelling architecture for unsupervised domain adaptive nuclear instance segmentation and classification. In our two-stage framework, category-aware feature alignment with learnable trade-off loss weights is proposed to tackle the class-imbalance issue and avoid misalignment during the cross-domain study. In addition, we proposed a nuclei-level prototype loss to correct the deviation in the second stage pseudo-labelling training, which further improves the segmentation and classification performance on the target images by introducing auxiliary self-supervision. Comprehensive results on various cross-domain nuclei instance segmentation and classification tasks demonstrate the prominent performance of our approach. Given the appealing performance of our method on the UDA nuclei instance segmentation tasks, we suggest that future directions can focus on the cross-domain multi-class object recognition tasks for other medical and general computer vision scenarios.

References

1. May M.: A better lens on disease: Computerized pathology slides may help doctors make faster and more accurate diagnoses. In *Scientific American*, vol. 302, pp. 74–77 (2010)
2. Lee H. and Kim J.: Segmentation of Overlapping Cervical Cells in Microscopic Images with Super-pixel Partitioning and Cell-Wise Contour Refinement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 63–69 (2016)
3. Naylor P., Laé M., Reyal F. and Walter T.: Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map. In *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 448–459 (2019)
4. Zhou Y., Onder O.F., Dou Q., Tsougenis E., Chen H. and Heng P.A.: CIA-Net: Robust Nuclei Instance Segmentation with Contour-Aware Information Aggregation. In *Information Processing in Medical Imaging. Lecture Notes in Computer Science*, vol. 11492, pp. 682–693 (2019)
5. Chen S., Ding C. and Tao D.: Boundary-Assisted Region Proposal Networks for Nucleus Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, pp. 279–288 (2020)
6. Graham S., Vu Q.D., Raza S.E.A., Azam A., Tsang Y.W., Kwak J.T. and Rajpoot N.: Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. In *Medical Image Analysis*, vol. 58 (2019)
7. Liu D., Zhang D., Song Y., Huang H. and Cai W.: Panoptic Feature Fusion Net: A Novel Instance Segmentation Paradigm for Biomedical and Biological Images. In *IEEE Transactions on Image Processing*, vol. 30, pp. 2045–2059 (2021)
8. Liu D., Zhang D., Song Y., Zhang F., O’Donnell L., Huang H., Chen M., and Cai W.: Pdam: A panoptic-level feature alignment framework for unsupervised domain adaptive instance segmentation in microscopy images. In *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 154–165 (2021)
9. Yang S., Zhang J., Huang J., Lovell B.C. and Han X.: Minimizing Labeling Cost for Nuclei Instance Segmentation and Classification with Cross-domain Images and

- Weak Labels. In Proceedings of the AAAI Conference on Artificial Intelligence, vol.35, no.1, pp. 697–705 (2021)
10. Chen C., Dou Q., Chen H., Qin J. and Heng P.A.: Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. In IEEE transactions on medical imaging, vol.39, no.7, pp. 2494–2505 (2020)
 11. Zhou Y., Huang L., Zhou T. and Shao L.: CCT-Net: Category-Invariant Cross-Domain Transfer for Medical Single-to-Multiple Disease Diagnosis. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8260–8270 (2021)
 12. Liu D., Zhang D., Song Y., Zhang F., O’Donnell L., Huang H., Chen M. and Cai W.: Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4243–4252 (2020)
 13. Kang G., Jiang L., Yang Y. and Hauptmann A.G.: Contrastive Adaptation Network for Unsupervised Domain Adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4893–4902 (2019)
 14. Long M., Cao Y., Wang J. and Jordan, M.: Learning transferable features with deep adaptation networks. In International Conference on Machine Learning, pp. 97–105 (2015)
 15. Zhang Q., Zhang J., Liu W., Tao D.: Category anchor-guided unsupervised domain adaptation for semantic segmentation. In Advances in Neural Information Processing Systems, 32 (2019)
 16. Hoffman J., Tzeng E., Park T., Zhu J.Y., Isola P., Saenko K., Efros A.A. and Darrell T.: Cycada: Cycle-consistent adversarial domain adaptation. In International Conference on Machine Learning, pp. 1989–1998 (2018)
 17. Irshad H., Veillard A., Roux L. and Racocceanu D.: Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. In IEEE reviews in biomedical engineering, vol. 7, pp. 97–114 (2013)
 18. Graham S., Jahanifar M., Azam A., Nimir M., Tsang Y.W., Dodd K., Hero E., Sahota H., Tank A., Benes K., Wahab N., Minhas F., Raza S.E.A., Daly H.E., Gopalakrishnan, K., Snead D. and Rajpoot, N.M.: Lizard: A Large-Scale Dataset for Colonic Nuclear Instance Segmentation and Classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 684–693 (2021)
 19. He K., Zhang X., Ren S. and Sun J.: Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)

Appendices

A Summary of the annotations in our experiment

Table 3. Details of the number of nuclei under each class in the datasets we used, DigestPath, CRAG and GlaS.

Class	DigestPath	CRAG	GlaS
Epithelial	70,789	99,124	31,986
Lymphocyte	49,932	27,634	9,763
Plasma	11,352	9,363	2,349
Neutrophil	2,262	1,673	90
Eosinophil	1,349	1,255	286
Connective	32,826	49,994	10,890
Total	168,510	189,043	55,364

B Loss function

B.1 The supervised Hover-net loss

The supervised Hover-Net loss function of our model is defined as \mathcal{L}_F :

$$\mathcal{L}_F = \mathcal{L}_{np} + \mathcal{L}_{hover} + \mathcal{L}_{nc} \quad (6)$$

where \mathcal{L}_{np} , \mathcal{L}_{hover} and \mathcal{L}_{nc} represent the loss with respect to the output at the NP, HV and NC branch, respectively.

For the NP and the NC branches, the loss is calculated by adding cross-entropy loss and dice loss:

$$\mathcal{L}_{np} = \mathcal{L}_{np}^{CE} + \mathcal{L}_{np}^{dice} \quad (7)$$

$$\mathcal{L}_{nc} = \mathcal{L}_{nc}^{CE} + \mathcal{L}_{nc}^{dice} \quad (8)$$

The cross entropy and dice losses are defined as:

$$\mathcal{L}^{CE} = -\frac{1}{N} \sum_{i=1} [p_i \log(q_i)] \quad (9)$$

$$\mathcal{L}^{dice} = 1 - \frac{2(q_i * p_i)}{q_i + p_i} \quad (10)$$

For the HV branch, we denoted \mathcal{L}^{mqe} as the mean squared error loss, and defined the \mathcal{L}^{hv} loss as the mean squared error between the horizontal and vertical gradients and the corresponding gradients of the ground truth:

$$\mathcal{L}_{hover} = \mathcal{L}^{mqe} + \mathcal{L}^{hv} \quad (11)$$

$$\mathcal{L}^{mqe} = \frac{1}{N} \sum_{i=1} (p_i - q_i)^2 \quad (12)$$

$$\mathcal{L}^{hv} = \frac{1}{M} \sum_{i=1} (p_{i,hor} - q_i)^2 - \frac{1}{M} \sum_{i=1} (p_{i,ver} - q_i)^2 \quad (13)$$

B.2 The domain discriminator loss

The overall domain discriminator loss function of our model is defined as \mathcal{L}_{dis} :

$$\mathcal{L}_{dis} = \mathcal{L}_{NC}^{ca} + \mathcal{L}_{NP}^{adv} + \mathcal{L}_{HV}^{adv} \quad (14)$$

where \mathcal{L}_{NC}^{ca} , \mathcal{L}_{NP}^{adv} and \mathcal{L}_{HV}^{adv} are the features adaptation loss functions in the NC, NP and HV branches, respectively.

The learnable weighted discriminator loss is formulated as follows:

$$\mathcal{L}_{NC}^{ca} = \omega_c^L \sum_{c=1} \mathcal{L}_c^{adv} \quad (15)$$

where the \mathcal{L}_c^{adv} denotes the adversarial training loss of D_c for class c , and ω_c^L is its corresponding learnable loss weight.

Concretely, we define the adversarial training loss as:

$$\mathcal{L}^{adv} = -\frac{1}{N} \sum_{i=1} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (16)$$

With the above loss terms, the overall loss function of the first stage approach can be written as:

$$\mathcal{L}_{s1} = \mathcal{L}_F + \mathcal{L}_{dis} \quad (17)$$

B.3 The prototype pseudo-labelling loss

The prototype pseudo-labelling loss function for the second stage is as follows:

$$\mathcal{L}_p = \frac{1}{N^p} \sum_{i=1}^{N^p} |x_i^p - \hat{y}_i^p|^2 \quad (18)$$

where x_i^p is the predicted features from the HV branch for each nuclear object p in the second stage and \hat{y}_i^p is the object features generated by the pseudo labels.

C Supplement on the experimental results

Table 4. Experimental results on UDA nuclei instance segmentation by transferring from *Dpath* to the *Glas* dataset.

	<i>Dpath</i> \rightarrow <i>GlaS</i>				
Methods	Dice	AJI	DQ	SQ	PQ
Source Only	0.494	0.264	0.359	0.732	0.262
PDAM	0.571	0.298	0.400	0.667	0.267
Yang <i>et al.</i>	0.639	0.294	0.377	0.735	0.275
Baseline	0.645	0.288	0.374	0.739	0.275
Proposed	0.651	0.296	0.385	0.734	0.281
Full-supervised	0.721	0.423	0.539	0.765	0.411

Table 5. Experimental results on UDA nuclei classification under the UDA scenario: *Dpath* \rightarrow *GlaS*

	<i>Dpath</i> \rightarrow <i>GlaS</i>							
Methods	Det	F_c^1	F_c^2	F_c^3	F_c^4	F_c^5	F_c^6	F_{avg}
Source Only	0.512	0.022	0.467	0.206	0.068	0.215	0.186	0.194
Yang <i>et al.</i>	0.565	0.000	0.621	0.141	0.152	0.018	0.240	0.195
Baseline	0.566	0.000	0.617	0.145	0.175	0.063	0.249	0.208
Proposed	0.572	0.000	0.647	0.131	0.192	0.168	0.323	0.244
Full-supervised	0.675	0.000	0.775	0.207	0.191	0.000	0.440	0.269

Table 6. Computational complexity analysis, which indicates our proposed modules can introduce performance gain but bring negligible auxiliary cost.

	<i>Dpath</i> \rightarrow <i>GlaS</i>	
Methods	Time cost	Number of Parameters
Source Only	35 s/iter	3.7M
Baseline	45 s/iter	3.7M
Proposed	45 s/iter	3.7M

D Visualization predictions for Dpath to GlaS experiment

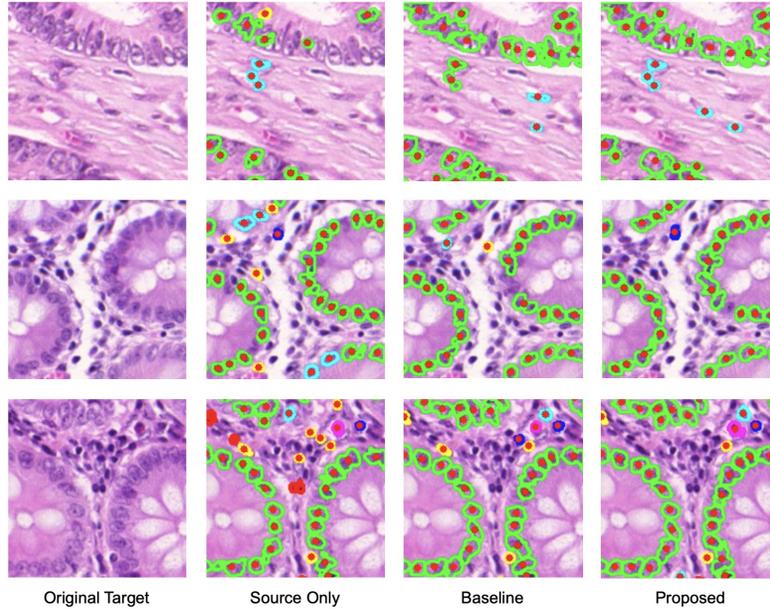


Fig. 3. Visualization predictions for the UDA experiment under the $D_{path} \rightarrow G_{laS}$ setting.

E Limitations

There are several limitations in our proposed method:

- 1) The pseudo labels for the nuclei prototype still contain noises, which limits the self-supervised learning performance.
- 2) The current method might incur performance drop under some specific cross-domain nuclei instance segmentation and classification tasks, such as from G_{laS} to D_{path} .