Unsupervised Instance Segmentation in Microscopy Images via Panoptic Domain Adaptation and Task Re-weighting (Supplementary material)

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

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

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

This document is the supplementary material for our submitted CVPR 2020 paper. First, we prove more visual examples of the synthesized images of our propsoed CyC-PDAM. Next, we add some statistical analysis for the experiment results. We then compare our CyC-PDAM with the method without domain adaptation, which further indicates the effectiveness of our proposed method. Finally, more implementation details are updated.

1. Visualization Examples of the Synthesized Images

Although histopathology images contain more complex structures than fluorescence microscopy images, our proposed CyC-PDAM is able to narrow the domain gap between the two modalities with pixel-level domain adaptation by synthesizing target-like histopathology images, from the microscopy images. Fig. 1 contains several synthesized samples for the Kumar and TNBC datasets.

2. Statistical Analysis

As mentioned in our paper, we conduct one-tailed-paired t-test for statistically significance analysis in Sec. 4.3.1 and Sec. 4.3.2. Table 1 contains the p-values between each unsupervised comparison method and our proposed method on the Kumar and TNBC datasets. All the p-values are under 0.05 and most of them are under 0.01, which indicates that our improvement compared to other methods is statistically significant. Table 3 contains the p-values for the ablation experiment. Under three metrics, all the p-values are under 0.01, which demonstrate that after adding each proposed module, the performance of the model is improved by a large margin.

3. Improvement Compared with the Method without Domain Adaptation

To further demonstrate the effectiveness of our proposed UDA architecture, we test the improvement of the UDA compared with the method without domain adaptation, on both Kumar and TNBC datasets. As shown in Table 2, w/o DA means directly training the fully supervised Mask R-CNN model with the fluorescence microscopy images in the BBBC039V1 dataset, and testing it on the testing set of the histopathology image dataset. With the pixel-level adaptation, image-level and instance-level feature adaptation, the performances under three metrics of the baseline CyCADA are improved 11% to 22%. Furthermore, with our proposed nuclei inpainting mechanism, panoptic-level feature adaptation, and task re-weighting mechanism on CyC-PDAM, the performances are further lifted 6% to 12%. Fig. 2 and Fig. 3 show the qualitative comparison results.

4. More Implementation Details

In this section, we introduce more implementation details of our proposed model. Our overall paradigm contains two parts. For the CycleGAN in the data generator, the weights of the model are initialized with normal distribution initialization. For the initialization of the end-to-end PDAM, the weights of the ResNet101 backbone are

	$BBBC039 \rightarrow Kumar$ (p-value)			$BBBC039 \rightarrow TNBC$ (p-value)		
Methods	AJI	Pixel-F1	Object-F1	AJI	Pixel-F1	Object-F1
CyCADA [4]	3.51×10^{-6}	5.55×10^{-4}	1.88×10^{-6}	5.45×10^{-5}	2.50×10^{-3}	2.27×10^{-4}
Chen et al [2]	1.03×10^{-7}	1.44×10^{-7}	6.64×10^{-5}	4.95×10^{-5}	3.72×10^{-5}	1.03×10^{-5}
SIFA [1]	3.94×10^{-7}	6.50×10^{-5}	8.74×10^{-3}	1.03×10^{-4}	1.91×10^{-3}	4.86×10^{-4}
DDMRL [6]	5.67×10^{-5}	9.36×10^{-6}	2.23×10^{-5}	7.73×10^{-4}	9.33×10^{-3}	4.80×10^{-4}
Hou et al [5]	4.92×10^{-3}	5.98×10^{-3}	1.83×10^{-3}	4.08×10^{-3}	2.92×10^{-2}	2.53×10^{-3}

Table 1. The p-value for the comparison methods on the Kumar and TNBC datasets.

	$BBBC039 \rightarrow Kumar$			$BBBC039 \rightarrow TNBC$		
Methods	AJI	Pixel-F1	Object-F1	AJI	Pixel-F1	Object-F1
w/o DA	0.3170 ± 0.1388	0.5076 ± 0.1781	0.5107 ± 0.1459	0.3379 ± 0.0684	0.5400 ± 0.0874	0.5796 ± 0.0731
UDA baseline [4]	0.4447 ± 0.1069	0.7220 ± 0.0802	0.6567 ± 0.0837	0.4721 ± 0.0906	0.7048 ± 0.0946	0.6866 ± 0.0637
Proposed	$\mathbf{0.5610 \pm 0.0718}$	$\mathbf{0.7882 \pm 0.0533}$	$\mathbf{0.7483 \pm 0.0525}$	$\mathbf{0.5672 \pm 0.0646}$	$\mathbf{0.7593 \pm 0.0566}$	$\mathbf{0.7478 \pm 0.0417}$

Table 2. In comparison with the method without domain adaptation, and the UDA baseline (CyCADA).

	AJI	Pixel-F1	Object-F1
w/o NI	1.69×10^{-3}	1.08×10^{-3}	7.58×10^{-4}
w/o TR	1.05×10^{-4}	2.42×10^{-2}	4.91×10^{-5}
w/o SEM	1.00×10^{-3}	1.38×10^{-3}	1.78×10^{-4}

Table 3. p-values for the ablation study on BBBC039V1 to Kumar experiment. NI, TR, and SEM represent the nuclei inpainting mechanism, task re-weighting mechanism, and semantic branch, respectively.

pretrained on the ImageNet classification task, while the weights for other layers are initialized with "Kaiming" initialization [3].

The CycleGAN of our model is from the official CycleGAN repository[1], with the generators of 9 residual connected CNN blocks, and the discriminators of 3 CNN layers. All the Mask R-CNN models mentioned in the experiments have the same implementation, with the official repository [7]. More specifically, for the BBBc039V1 to Kumar experiment, the number of ROIs after RPN is set to 500 and 8000, during training and testing, respectively, because the testing image size (1000×1000) is about 16 times of the training size (256×256). For the BBBc039V1 to TNBC experiment, the numbers of ROIs after RPN during training and testing are set to 300 and 1200, respectively, as the size of the 512×512 testing images is 4 times of the 256×256 training ones.

5. Discussions on the Nuclei Scale Variation

Nuclei scale variation is very common and challenging in digital pathology and our CyC-PDAM successfully solves it in the unsupervised nuclei segmentation for histopathology images. We have discussed the scales of the nuclei predictions in the ablation study section. In addition, we notice in Fig. 5 of the paper, there also remain nuclei predictions in irregular sizes for all the comparison methods. With the panoptic-level feature adaptation, the model learns the domain-invariant features for each object including its texture, scale, and location. The task re-weighting mechanism induces the model to pay more attention to the domain-invariant features, which prevents the model from learning the source-biased scale features and makes the model suitable for the various object scales in the target domain. By outperforming fully supervised methods when tested on the unseen organ images with unknown nuclei scales, our method is further validated to be effective for nuclei segmentation at various scales in the histopathology images.

[1] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

6. More Analysis on the Ablation Study Results

We provide some intuitive explanation for the ablation study in our paper in Sec. 4.3.2. NI successfully removes the auxiliary nuclei in the synthesized images and the model without NI tends to treat some foreground objects as the background during training. By ignoring some foreground objects, the model learns inaccurate domain-invariant semantic-level information, which hurts the feature-level adaptation. In addition, the false-negative predictions are harmful to the segmentation and detection task learning. Therefore NI has a similar importance to TR and SEM, without involving the learning process.

7. The Importance and the Novelty of Our Work

Our CyC-PDAM has significant differences from the existing methods. Firstly, this is the first work on the UDA instance segmentation, which unifies the UDA detection and segmentation and has not been investigated by previous work. Second, our nuclei inpainting mechanism is originally designed to remove the auxiliary generated nuclei and preserve the true nuclei in the synthesized images based on the labels. However, none of the previous work has discussed or solved the auxiliary nuclei issue in the synthesized

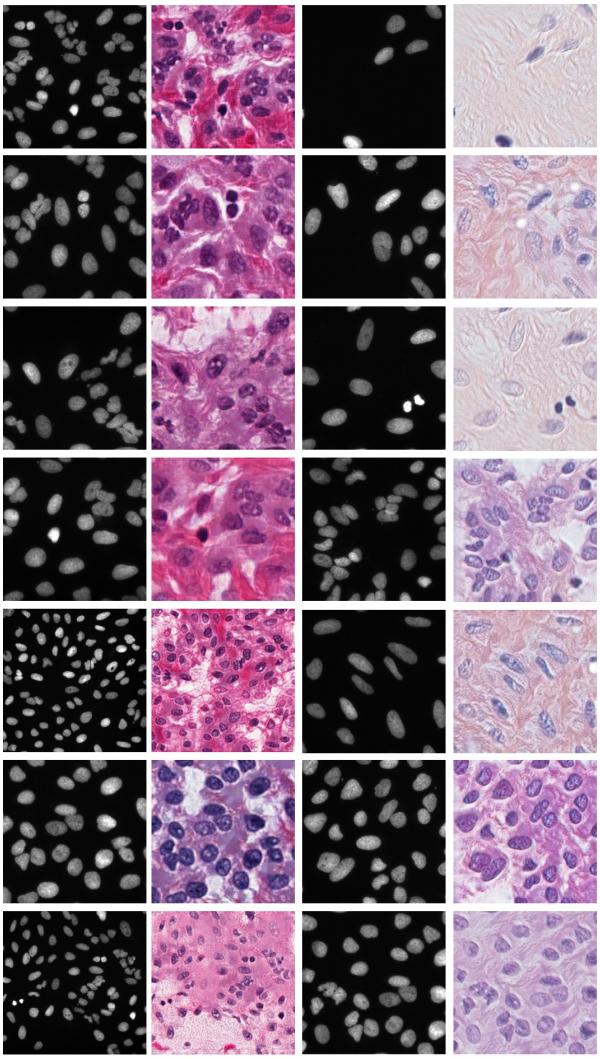

Figure 1. Visualization examples for the synthesized histopathology images. (a) and (c) fluorescence microscopy images; (b) synthesized histopathology images for the Kumar dataset; (d) synthesized histopathology images for the TNBC dataset.

histopathology images. Third, we are the first to combine the panoptic segmentation idea with UDA for instance segmentation. Panoptic segmentation is currently only used in fully supervised tasks and we take the first step to use this idea for UDA tasks. Fourth, our task re-weighting mechanism is originally proposed to prevent the model from learning the source-biased domain-specific features. This is an important problem in UDA study, and, to the best of our knowledge, we are the first to solve it by resetting the importance of each task loss. Furthermore, our extensive experiments demonstrate that our methods outperform the SOTA UDA methods significantly, which indicates the novelty and importance of our proposed paradigm in the UDA study.

References

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

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

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1026–1034, 2015.

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

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

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

[7] Francisco Massa and Ross Girshick. maskrcnn-benchmark: Fast, modular reference implementation of Instance Segmentation and Object Detection algorithms in PyTorch. `https://github.com/facebookresearch/maskrcnn-benchmark`, 2018.

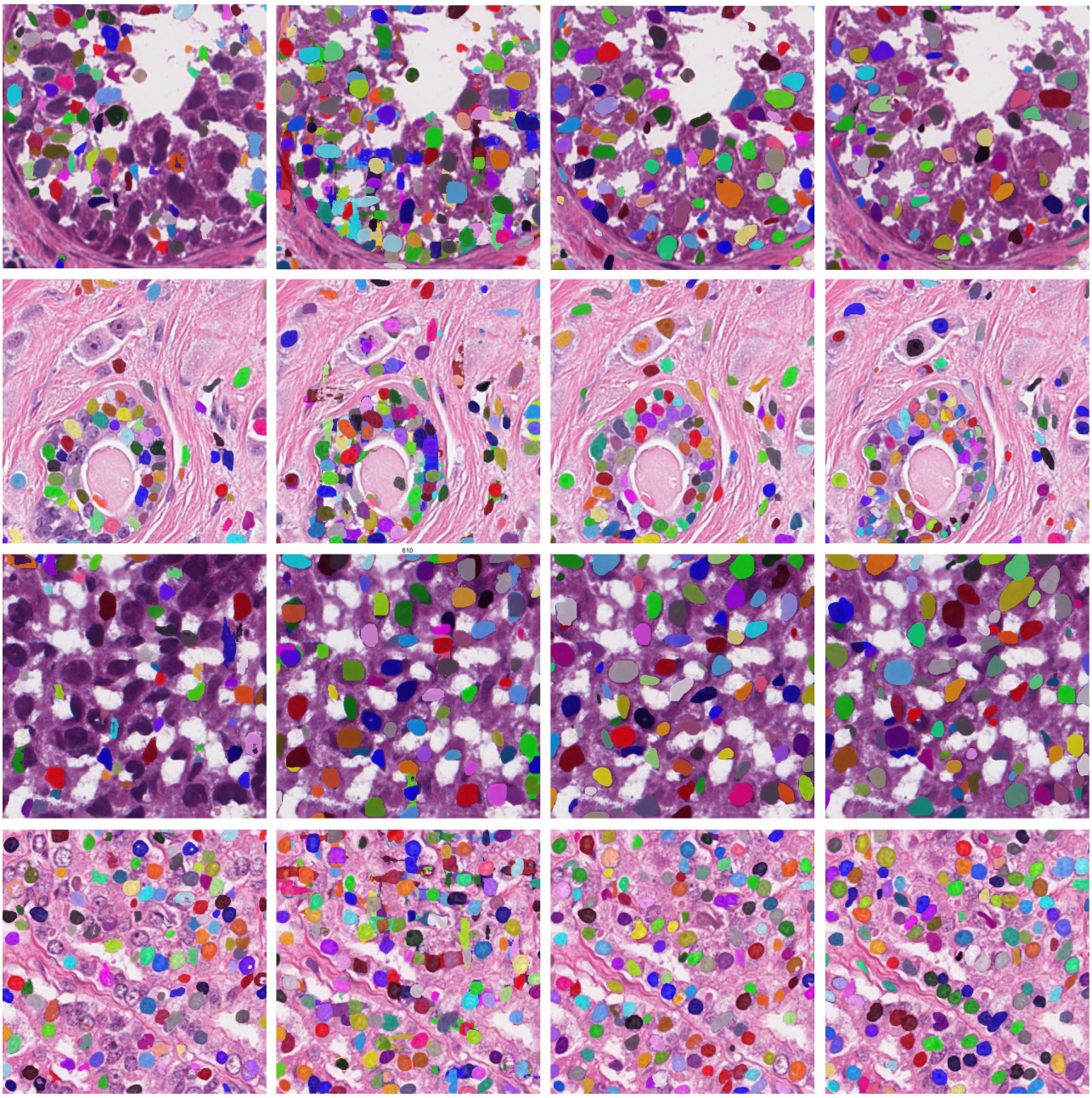

Figure 2. Visualization examples for the comparison experiment with the two baseline on the Kumar dataset.

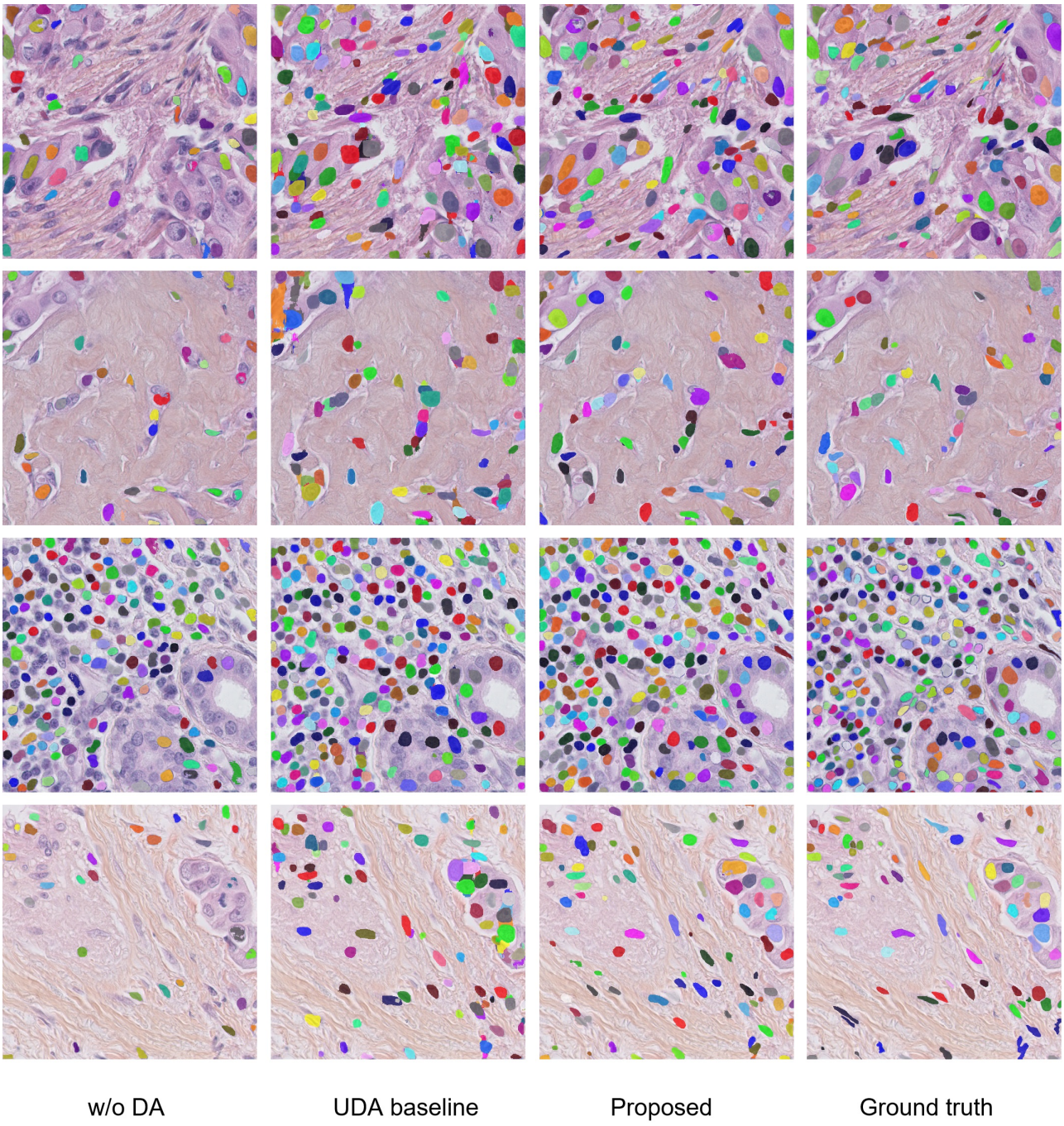

Figure 3. Visualization examples for the comparison experiment with the two baseline on the TNBC dataset.