Dual Adaptive Pyramid Network for Cross-Stain Histopathology Image Segmentation

Xianxu Hou[1,2⋆], Jingxin Liu[1,2,3⋆(✉)], Bolei Xu[1,2], Bozhi Liu[1,2], Xin Chen[4], Mohammad Ilyas[5], Ian Ellis[5], Jon Garibaldi[4], and Guoping Qiu[1,2,4]

[1] College of Information Engineering, Shenzhen University, Shenzhen, China
[2] Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen, China
[3] Histo Pathology Diagnostic Center, Shanghai, China
[4] School of Computer Science, University of Nottingham, Nottingham, United Kingdom
[5] School of Medicine, University of Nottingham, Nottingham, United Kingdom
jingxin.liu@outlook.com

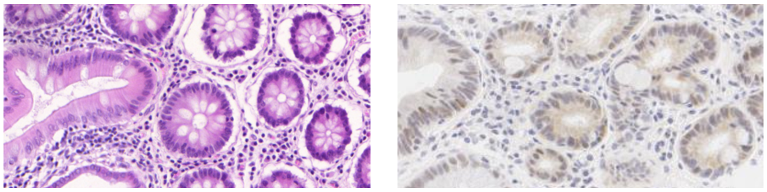
Abstract. Supervised semantic segmentation normally assumes the test data being in a similar data domain as the training data. However, in practice, the domain mismatch between the training and unseen data could lead to a significant performance drop. Obtaining accurate pixel-wise label for images in different domains is tedious and labor intensive, especially for histopathology images. In this paper, we propose a dual adaptive pyramid network (DAPNet) for histopathological gland segmentation adapting from one stain domain to another. We tackle the domain adaptation problem on two levels: 1) the image-level considers the differences of image color and style; 2) the feature-level addresses the spatial inconsistency between two domains. The two components are implemented as domain classifiers with adversarial training. We evaluate our new approach using two gland segmentation datasets with H&E and DAB-H stains respectively. The extensive experiments and ablation study demonstrate the effectiveness of our approach on the domain adaptive segmentation task. We show that the proposed approach performs favorably against other state-of-the-art methods.

Keywords: Gland Segmentation · Histopathology · Domain Adaptation

1 Introduction

Deep convolutional neural networks (DCNNs) have achieved remarkable success in the field of medical image segmentation [5], which aims to identify and segment specific regions, such as organs or lesions in MR images, and cellular structures or tumor regions in pathological images. Although excellent performance has been achieved on benchmark dataset, deep segmentation models have

[⋆] Equal contribution



(a) H&E (b) DAB-H

Fig. 1. Image examples of different histopathological stains. (a) Hematoxylin and Eosin; (b) Diaminobenzidene and Hematoxylin.

poor generalization capability to unseen datasets [10] due to the domain shift between the training and test data.

Such domain shift is commonly observed especially in histopathology image analysis. For instance, the Hematoxylin and Eosin (H&E) stained colon image has significantly different visual appearances from that stained by Diaminobenzidene and Hematoxylin (DAB-H) (Fig. 1). Thus, the model trained on one (source) dataset would not generalize well when applied to the other (target) dataset. Although fine-tuning the model with labelled target data could possibly alleviate the impact of domain shift, manually annotating is a time-consuming, expensive and subjective process in medical area. Therefore, it is of great interest to develop algorithms to adapt segmentation models from a source domain to a visually different target domain without requiring additional labels in the target domain.

Domain adaptation algorithms have been developed to address the domain-shift problem. The main insight behind these methods is trying to align visual appearance or feature distribution between the source and target domains. Zhang *et al.* [11] render the source image with the target domain "style", and then learn domain-invariant representations in an adversarial manner. AdapSeg [9] is developed to align the two domain images in the structured output space. CyCADA [3] unifies adversarial adaptation methods together with cycle-consistent image translation techniques.

In this paper, we propose a DCNN-based domain adaptation algorithm for histopathology image segmentation, referred to as Dual Adaptive Pyramid Network (DAPNet). The proposed DAPNet is designed to reduce the discrepancy between two domains by incorporating two domain adaptation components on image level and feature level. The image-level adaptation considers the overall difference between source and target domain like image color and style, while feature-level adaptation addresses the spatial inconsistency of the two domains. In particular, each component is implemented as a domain classifier with an adversarial training strategy to learn domain-invariant features.

The contribution of this work can be summarized as follows. First, we develop a deep unsupervised domain adaptation algorithm for histopathology image segmentation. Second, we propose two domain adaptation components to alleviate the domain discrepancy at the image and feature levels based on pyramid features. Third, we conduct extensive experiments and our proposed DAPNet outperforms other state-of-the-art methods.

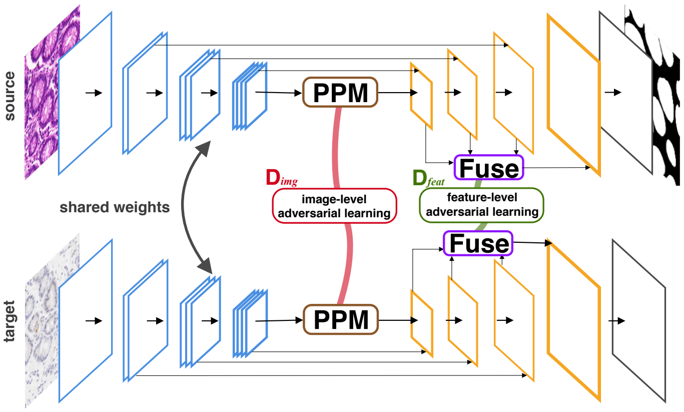

source
target
PPM
PPM
Fuse
Fuse
shared weights
D_{img}
image-level adversarial learning
D_{feat}
feature-level adversarial learning

Fig. 2. Overview of our DAPNet. Both source and target domain images are fed to the segmentation network. The training procedure optimizes the segmentation loss based on the source ground truth, and two domain classification losses of image-level and feature-level adversarial learning modules to make the segmentation output close to the image labels of the source domain.

2 Method

In this work, we aim to learn gland segmentation model from images with a certain stain type and apply the learned model to a different stain scenario. The training data is used as the source domain $\mathcal{S}$ while the test data with a different stain type is regarded as the target domain $\mathcal{T}$. In the $\mathcal{S}$ domain, we have access to the stained images X_S as well as the corresponding ground-truth labels Y_S. In the target domain $\mathcal{T}$, we only have the unlabelled stained images X_T.

2.1 Model Overview

The overview of the proposed DAPNet is illustrated in Fig. 2. It contains a semantic segmentation network G and two adversarial learning modules D_{img} and D_{feat}. During training, both the source images x_s and target images x_t are fed into the network G as inputs. The source images and the corresponding labels are used to optimize G for the segmentation task, while both source and target images are used for optimizing domain adaptation losses by adversarial learning with D_{img} and D_{feat}.

2.2 Segmentation Network

As shown in Fig. 2, our segmentation network consists of 3 components. First a dilated ResNet-18 [2] is used as backbone to encode the input images. In order to achieve larger receptive field of our model, we apply a Pyramid Pooling Module (PPM) from PSPNet [12] on the last layer of the backbone network.

The PPM separates the feature map into different pooled representations with varied pyramid levels. The different levels of features are then upsampled and concatenated as the pyramid pooling global feature. Furthermore, we adopt skip connections from U-Net [7] and a pyramid feature fusion architecture to achieve final segmentation. The decoded feature maps are upsampled to the same spatial resolution and merged by concatenation in a pyramidal way. The output feature maps undergo a 1×1 convolutional layer to reduce the dimension of channel to 512. Our method involves downsampling pyramid feature extraction and upsampling pyramid feature fusion. However, the CyCADA needs to first map source training data into the target domain in pixel level.

The segmentation task is learned by minimizing both standard cross-entropy loss and Dice coefficient for images from the source domain:

$$\mathcal{L}_{seg} = \mathbb{E}_{x_s \sim X_S}[-y_s log(\widetilde{y}_s)] + \alpha \mathbb{E}_{x_s \sim X_S}[-\frac{2y_s\widetilde{y}_s}{y_s + \widetilde{y}_s}] \tag{1}$$

where y_s stands for ground-truth labels, $\widetilde{y}_s$ stands for predicted labels and α is the trade-off parameter.

2.3 Domain Adaptation

Image-level Adaptation. In this work, image-level representation refers to the PPM outputs of the segmentation network G. Image-level adaptation helps to reduce the shift by the global image difference such as image color and image style between the source and target domains. To eliminate the domain distribution mismatch, we employ a discriminator D_{img} to distinguish PPM features between source images and target images. At the same time, D_{img} also guides the training of segmentation network in an adversarial manner. In particular, we employ PatchGAN [4], a fully convolutional neural operating on image patches, from which we can get a two-dimensional feature map as the discriminator outputs. The loss for training D_{img} is formulated as follows:

$$\mathcal{L}_{img} = \mathbb{E}_{x_t \sim X_T}[logD_{img}(p_t)] + \mathbb{E}_{x_s \sim X_S}[log(1 - D_{img}(p_s))] \tag{2}$$

where p_s and p_t denote the PPM outputs of the segmentation network G for source domain and target domain.

Feature-level Adaptation. The feature-level representation refers to the fused feature maps before feeding into the final segmentation classifier. Aligning the feature-level representations helps to reduce the segmentation differences in both global layout and local context. Similar to image-level adaptation, we also train a domain classifier D_{feat} formulated as a PatchGAN to align the feature-level distribution. Let us denote the final fused feature representation as f_s and f_t for source domain and target domain respectively. The loss for D_{feat} is written as follows:

$$\mathcal{L}_{feat} = \mathbb{E}_{x_t \sim X_T}[logD_{feat}(f_t)] + \mathbb{E}_{x_s \sim X_S}[log(1 - D_{feat}(f_s))] \tag{3}$$

2.4 Overall Training Objective

We integrate the segmentation module for source images and the two domain adaptation modules to train all the networks G, D_{img} and D_{feat} jointly. The overall objective function can be formulated as follows:

$$\min_{G} \max_{D_{img},D_{feat}} \mathcal{L}_{seg}(x_s, y_s) + \lambda_1 \mathcal{L}_{img}(x_s, x_t) + \lambda_2 \mathcal{L}_{feat}(x_s, x_t) \tag{4}$$

where λ_1 and λ_2 are two trade-off parameters. The min-max game is optimized by adversarial training and G is used to achieve segmentation for images in target domain during test.

3 Experiments and Results

3.1 Datasets

Two colorectal cancer gland segmentation datasets with different stains are used to evaluate our model. **Warwick-QU** dataset [8], introduced in gland segmentation challenge in MICCAI 2015, consists of 165 H&E stained images cropped from whole slide images (WSIs). The WSIs are acquired in 20× optical magnification. In our experiments, the dataset is separated into training and test sets with 85 and 80 images respectively. **GlandVision** dataset [1] contains 20 DAB-H stained colon images with size of 1280×1024, which were captured with 10× optical magnification. We randomly select 14 images for training and the rest for test. It is noted that those two datasets are labelled with different strategies. The masks in Warwick-QU cover the whole glandular structures, while GlandVision only considers the lumen regions.

3.2 Implementation details

Our DAPNet employs 3×3 kernel for convolutional operations followed by a batch normalization layer. We train all the models using Adam optimization with a batch size of 4 for 300 epochs. We randomly crop image patches of size 256×256 for training. The initial learning rate is 10^{-3}, which is kept the same for the first 150 epochs and linearly decayed to zero over the next 150 epochs. The hyper-parameters α, λ_1 and λ_2 are set to 1, 0.002 and 0.005 respectively. Our method is based on LSGAN [6], which replaces the negative log likelihood objective by a least square loss. This loss achieves a more stable model training and generates higher quality results.

3.3 Results

We evaluate the performance of our DAPNet for gland segmentation in both adaptive directions. In particular, we denote Warwick-QU (source) to GlandVision (target) as Warwick-QU → GlandVision and vice versa, and the test images

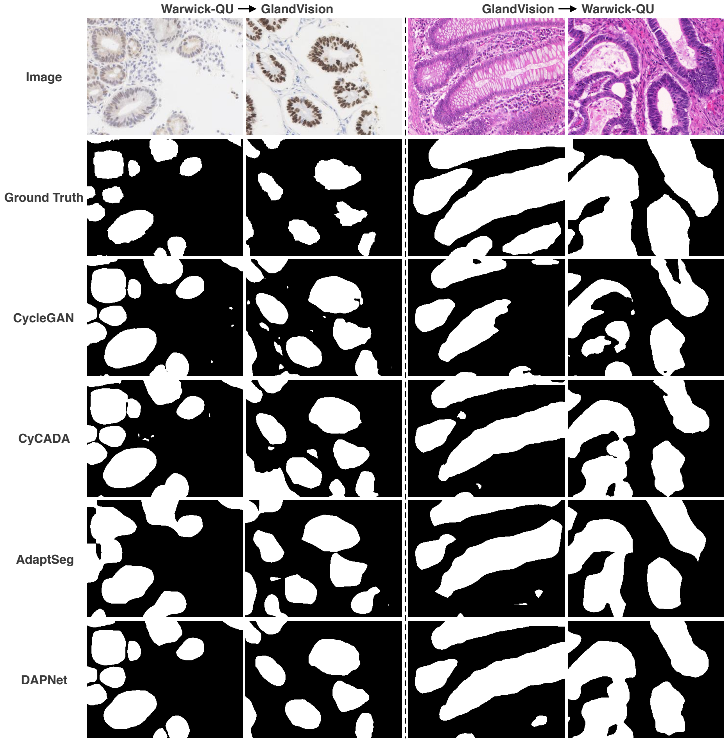


Fig. 3. Qualitative results of gland segmentation adapting from Warwick-QU to GlandVision dataset (left two columns) and vice versa (right two columns).

in the target domain are used for evaluation. Extensive experiments including comparisons to the state-of-the-art methods and ablation study are provided.

We compare our DAPNet with three state-of-the-art unsupervised domain adaptation methods: CycleGAN [13], CyCADA [3] and AdaptSeg [9]. The comparison with CycleGAN is achieved by two stages. We first use CycleGAN transforms the source domain images to target domain, and then use the transformed images along with the corresponding label in the source domain to train the segmentation network G. We report the segmentation results using Pixel Accuracy (Acc.) and the Intersection over Union (IoU) in Table 1. We can observe that our model DAPNet outperforms all the other methods for domain adaptation between WarwickQU and GlandVision in both directions. We have repeated the model training and testing for 3 times with random parameter initializations and the same hyper-parameters. All tests have shown that our proposed method consistently outperforms other methods with statistical significance (paired t-test with $p<0.01$). Specifically, when adapting from Warwick-QU to GlandVision, the averaged accuracy and IoU are 0.88 ± 0.0083 (Mean ± SD) and 0.68 ± 0.0021 respectively. On the other hand, the averaged accuracy and IoU are 0.76 ± 0.0105 and 0.57 ± 0.0108 respectively adapting from GlandVision to Warwick-QU. Moreover, Fig. 3 presents qualitative results of two example im-

Table 1. Comparison with state-of-the-art methods for semantic segmentation on GlandVision adapting from Warwick-QU and vice versa.

method	Warwick-QU → GlandVision		GlandVision → Warwick-QU	
	Acc.	IoU	Acc.	IoU
CycleGAN [13]	0.84	0.60	0.74	0.54
CyCADA [3]	0.84	0.62	0.73	0.54
AdapSeg [9]	0.81	0.67	0.72	0.52
DAPNet-NA	0.80	0.58	0.73	0.50
DAPNet-IA	0.85	0.60	0.75	0.55
DAPNet-FA	0.83	0.63	0.74	0.53
DAPNet	**0.88**	**0.68**	**0.76**	**0.57**

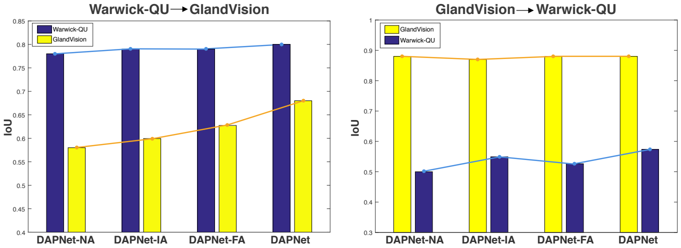


Fig. 4. Performance comparison of different variants of our proposed model in terms of IoU measurements. The trained models are applied to both the source and target domain images for test. The segmentation performance for the source domain maintains at a high level while the performance of the target domain is boosted.

ages for each of the domain adaptation case. Both CycleGAN and CyCADA can successfully detect the gland structures, but the predicted masks contain irregular spot noise. AdaptSeg with only image-level adaptation can hardly segment the gland boundaries clearly. Our proposed DAPNet produces significantly better predictions with accurate layout.

We further conduct ablation study to demonstrate the necessity of the two domain adaptation components of our model. In particular, we compare DAPNet with its three variants, the model trained without domain adaptation modules (DAPNet-NA), only image-level adaptation module (DAPNet-IA) and only feature-level adaptation module (DAPNet-FA). As shown in Table 1, we observe that the performance of the DAPNet-NA drops significantly due to the domain shift and the best results are achieved with DAPNet. It is clear that the two adaptation components can effectively alleviate the discrepancy between two domains. We also show that domain adaptation modules can boosts the segmentation performance on target domain without affecting the results on source domain (see Fig. 4).

4 Conclusions

In this paper, we study the unsupervised domain adaptive segmentation task for histopathology images. We have proposed a dual adaptive pyramid network with two domain adaptation components by adversarial training on both image and feature levels. The model is trained without target domain labels and the test procedure works as normal segmentation networks. Experimental results show that the proposed DAPNet can effectively boost the performance on unlabelled target datasets, and outperform other state-of-the-art approaches.

References

1. Fu, H., Qiu, G., Ilyas, M., Shu, J.: Glandvision: A novel polar space random field model for glandular biological structure detection. In: Proceedings of the British Machine Vision Conference. pp. 42.1–42.12. BMVA Press (2012)
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
3. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: International Conference on Machine Learning. pp. 1994–2003 (2018)
4. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
5. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)
6. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2794–2802 (2017)
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
8. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., et al.: Gland segmentation in colon histology images: The glas challenge contest. Medical image analysis **35**, 489–502 (2017)
9. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7472–7481 (2018)
10. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7167–7176 (2017)
11. Zhang, Y., Qiu, Z., Yao, T., Liu, D., Mei, T.: Fully convolutional adaptation networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6810–6818 (2018)
12. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2881–2890 (2017)
13. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232 (2017)