

Deep Adversarial Networks for Biomedical Image Segmentation Utilizing Unannotated Images

Yizhe Zhang¹(✉), Lin Yang¹, Jianxu Chen¹, Maridel Fredericksen²,
David P. Hughes², and Danny Z. Chen¹

¹ Department of Computer Science and Engineering, University of Notre Dame,
Notre Dame, IN 46556, USA

yzhang29@nd.edu

² Department of Entomology and Department of Biology, Center for Infectious
Disease Dynamics, Pennsylvania State University, University Park, PA 16802, USA

Abstract. Semantic segmentation is a fundamental problem in biomedical image analysis. In biomedical practice, it is often the case that only limited annotated data are available for model training. Unannotated images, on the other hand, are easier to acquire. How to utilize unannotated images for training effective segmentation models is an important issue. In this paper, we propose a new deep adversarial network (DAN) model for biomedical image segmentation, aiming to attain consistently good segmentation results on both annotated and unannotated images. Our model consists of two networks: (1) a segmentation network (SN) to conduct segmentation; (2) an evaluation network (EN) to assess segmentation quality. During training, EN is encouraged to distinguish between segmentation results of unannotated images and annotated ones (by giving them different scores), while SN is encouraged to produce segmentation results of unannotated images such that EN cannot distinguish these from the annotated ones. Through an iterative adversarial training process, because EN is constantly “criticizing” the segmentation results of unannotated images, SN can be trained to produce more and more accurate segmentation for unannotated and unseen samples. Experiments show that our proposed DAN model is effective in utilizing unannotated image data to obtain considerably better segmentation.

1 Introduction

Deep learning models [1,10] have achieved many successes in biomedical image segmentation. To obtain good segmentation performance, a decent amount of (pixel-wise) annotated images is often required to train such models. Due to high costs of pixel-wise annotation and large image sizes in applications

Electronic supplementary material The online version of this chapter (doi:10.1007/978-3-319-66179-7.47) contains supplementary material, which is available to authorized users.

(e.g., 3D image stacks with hundreds of slices, 2D whole-tissue images with hundreds of millions of pixels), it is common that annotation for only a small subset of all image data is available. Thus, when training a deep learning model using annotated images, one may also have a considerable number of unannotated images at hand. Such unannotated images are often from the original data distribution (containing useful information) and are free to use. Hence, a natural question is: How could we utilize unannotated images to benefit and improve segmentation?

Some recent attempts [5, 7] were made to utilize weakly annotated images in natural scene image segmentation. Bounding box (to bound an object of interest) and image level label (to show what objects appear in the images) are two common weak annotation methods for their settings. However, in biomedical image segmentation, there can be numerous more object instances (e.g., cells) than in natural scene images, and drawing bounding box still requires a great deal of effort. Also, there can be much fewer object classes in biomedical images than in natural scene images, and image level labels may be less useful in biomedical settings since almost all the images may contain all the object classes for segmentation (e.g., cells, glands). Thus, it is important to exploit unannotated images as well as annotated images for effective biomedical image segmentation.

Using unannotated data together with annotated data to train a learning model is not new. In [14], it combined an auxiliary unsupervised learning task to help the supervised training of a neural network; the intermediate layers are shared among both the supervised and unsupervised learning tasks. Consequently, its network can be trained for better generality. Using this approach, different choices for unsupervised learning tasks were proposed (e.g., reconstructing the input of the model through an encoding and decoding stage [8], a classification task for transforming the input to specially designed class labels [3]). As pointed out in [9], a key drawback of this approach is that, since unsupervised and supervised learning tasks have different goals, the unsupervised learning part may not always be helpful to the supervised learning part via the shared model parameters. To alleviate this problem, Ladder networks (with skip connections) were used to reduce the burden put on the encoding layers by the unsupervised learning part [9]. Despite this, the inherent problem of having different goals for supervised and unsupervised learning tasks was still not well resolved.

It would be ideal to use both annotated and unannotated data to serve the same goal (e.g., using both for training a segmentation network, as in our problem). A major difficulty is, since no ground truth is given for unannotated data, back-propagation errors after the forward pass cannot be directly computed for unannotated data. Our key idea is to train a deep neural network to compute approximate errors for unannotated data, using adversarial training [4, 11].

In this paper, we propose a new adversarial training approach, i.e., a deep adversarial network (DAN) model, for producing consistently good segmentation for both annotated and unannotated images. Our DAN model consists of two networks: (1) a segmentation network (SN) to conduct segmentation; (2) an evaluation network (EN) to assess the quality of SN's segmentation. During training,

EN is encouraged to distinguish between segmentation results of unannotated and annotated samples by giving them different scores, while SN is encouraged to produce segmentation results of unannotated images such that EN cannot distinguish these from the annotated ones. Through an iterative adversarial training process, because EN is constantly “criticizing” the segmentation of unannotated images using its learned feature mappings (describing what good segmentation looks like), SN can be trained to produce more and more accurate segmentation for unannotated and unseen samples. Our method is inspired by [4,6]. Different from [6], our adversarial networks are designed to utilize unannotated images.

Experiments using the 2015 MICCAI Gland Challenge dataset [13] and a 3D fungus segmentation dataset show that our DAN model is effective in utilizing unannotated image data to obtain segmentation of considerably better quality.

2 Method

This section describes our adversarial training model utilizing unannotated data, and discusses a key issue: How to construct the input for the evaluation network.

2.1 Adversarial Networks Using Unannotated Data

There are two networks in our DAN model: a segmentation network SN and an evaluation network EN. SN takes an input image I and produces segmentation probability maps for I . EN takes the segmentation probability maps and the corresponding input image I , and determines a score indicating the quality of the segmentation: 1 (for good quality) or 0 (for not good quality).

During the model training, EN is encouraged to give high scores (1) for segmentation of annotated images and low scores (0) for segmentation of unannotated images. SN is trained using annotated images and is also encouraged to produce segmentation results of unannotated images such that EN might give them high scores. Below we describe the details of our adversarial training model.

Given M annotated training images X_m , their corresponding segmentation ground truth Y_m , and N unannotated images U_n , we define the loss function as

$$\ell(\theta_S, \theta_E) = \sum_{m=1}^M \ell_{mce}(S(X_m), Y_m) - \lambda \left[\sum_{m=1}^M \ell_{bce}(E(S(X_m), X_m), 1) + \sum_{n=1}^N \ell_{bce}(E(S(U_n), U_n), 0) \right]$$

where θ_S and θ_E are the parameters of the two networks SN and EN respectively, ℓ_{mce} is the multi-class cross-entropy loss, and ℓ_{bce} is the binary-class cross-entropy loss. The first term in the loss function is for the supervised training of SN using annotated images, and the second term forms the adversarial training part. The training process minimizes part of the loss with respect to the parameters θ_S of SN, while maximizing the loss with respect to the parameters θ_E of EN. More specifically, training EN aims to minimize

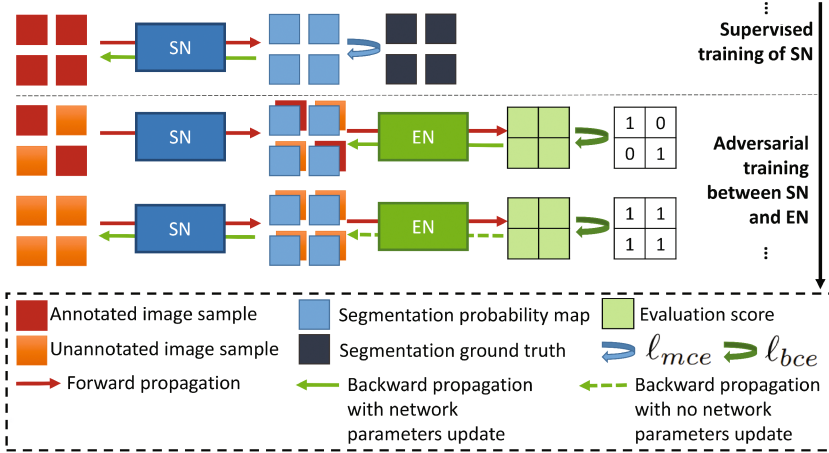


Fig. 1. Illustrating the processes in one iteration of our model training (with the mini-batch size = 4). First, SN is trained using annotated images and their corresponding ground truth images; then, EN is trained to give different scores to segmentations of annotated images and unannotated images; finally, SN is trained to improve the segmentation quality of unannotated images (based on EN’s learned feature mappings).

$$\lambda \left[\sum_{m=1}^M \ell_{bce}(E(S(X_m), X_m), 1) + \sum_{n=1}^N \ell_{bce}(E(S(U_n), U_n), 0) \right]$$

with respect to the parameters θ_E of EN, and training SN aims to minimize

$$\sum_{m=1}^M \ell_{mce}(S(X_m), Y_m) - \lambda \left(\sum_{n=1}^N \ell_{bce}(E(S(U_n), U_n), 0) \right)$$

with respect to the parameters θ_S of SN. As in [4], when updating SN, we replace the term $-\lambda(\sum_{n=1}^N \ell_{bce}(E(S(U_n), U_n), 0))$ by $\lambda(\sum_{n=1}^N \ell_{bce}(E(S(U_n), U_n), 1))$. A standard stochastic gradient descent method can be applied to optimize this loss function. Since the adversarial training part may be less useful prior to the stage when SN can produce reasonably good segmentation for the annotated training images, we set $\lambda = 0.1$ initially, and set $\lambda = 1$ after 30000 iterations. The value of lambda should be small (< 1) before SN can produce decent segmentation results. Too large lambda (e.g. $\lambda = 10$) may cause the training to fail to train a reasonable SN. Figure 1 shows our training process. Figure 2 gives more details of the SN and EN architectures. Our SN largely follows the architecture of DCAN [2], but with no split up-sampling (deconvolution) paths used. Our EN follows the main architecture of the classic VGG16 network [12].

2.2 Constructing the Input of the Evaluation Network

The input information provided to EN is crucial to the whole adversarial training system. A simple form for the input of EN could be just the segmentation probability maps, which allow EN to examine useful morphological properties of the segmented biomedical objects and help assess segmentation quality.

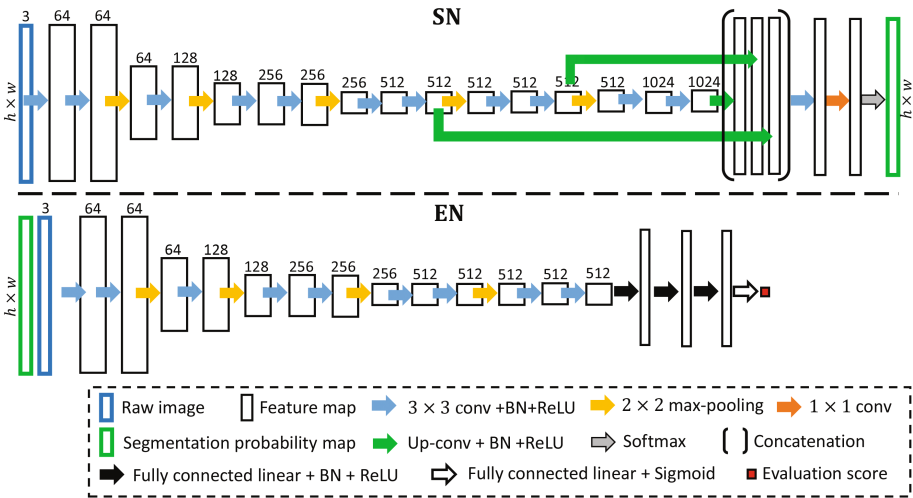


Fig. 2. Architectural details of our segmentation network (SN) and evaluation network (EN).

A more effective way to construct input for EN is to combine segmentation probability maps and the corresponding input image. This allows EN to explore the correlations between the segmentation and input image for evaluating the segmentation quality. However, giving the input image to EN could potentially be problematic since EN might come up with a way to give an evaluation score only based on the appearance of the input image without examining the segmentation probability maps. This would make the whole adversarial training useless with respect to improving the segmentation performance for unannotated images. Below we discuss two main methods for combining the segmentation maps and the input image to construct the input for EN.

Concatenation. Two possible ways to concatenate the segmentation probability map and input image are: directly concatenate them, or transform them to two feature maps and concatenate the feature maps. With either method, since EN has separate model parameters for handling information from the segmentation maps and from the input image, it is possible that only information from the raw image input is utilized for EN’s decision making.

Element-wise multiplication. A good aspect of element-wise multiplication is that it can “force” the segmentation probability maps and input image to mix

at the very initial stage. Thus, all the model parameters are jointly trained using information from both the segmentation and input image. This ensures that the segmentation probability maps are used in EN’s decision making and in the entire adversarial training process. However, since element-wise multiplication essentially performs a pixel-wise gate operation (using the input image) on the segmentation probability maps in both the forward pass from SN to EN and backward propagation from EN to SN, it could happen that lower intensity structures (e.g., cell nuclei, gland borders in H&E stained images) may have very little influence on both the decision making process of EN and the parameter updates of SN. In order to reduce this bias, we use both the input image and its inverted image for mixing with the segmentation probability maps. Suppose I is one channel from the raw image input and P is one probability map produced by SN. We mix them by $I \cdot P$ and $(1 - I) \cdot P$ (two maps obtained). We mix every possible pair of I and P and concatenate all the obtained maps to form the input of EN.

3 Experiments and Results

To evaluate the effectiveness of our DAN model on utilizing unannotated images for segmentation, we test and compare DAN and several related models using two data sets: the 2015 MICCAI Gland Challenge dataset [13] for gland segmentation in H&E stained tissue images (e.g., see the top row of Fig. 3), and an in-house 3D electron microscopy (EM) image dataset for fungus segmentation.

Gland segmentation. This dataset [13] has 85 training images (37 benign (BN), 48 malignant (MT)), 60 testing images (33 BN, 27 MT) in part A, and 20 testing images (4 BN, 16 MT) in part B. As our unannotated training data, we acquired 100 additional H&E stained intestinal images from an in-house dataset (e.g., see the bottom row of Fig. 3).

Table 1 shows the gland segmentation results of our DAN model and several closely related models. For fair comparison, an adversarial training model (SSAN [6]), a semi-supervised learning model (Ladder networks [9]), and our DAN model all use the same segmentation network as SN (the base model).

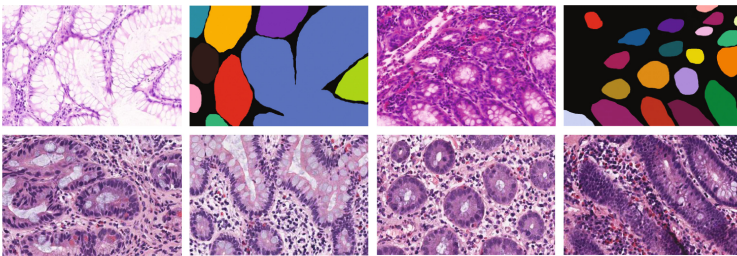


Fig. 3. Top row: Image samples and their corresponding instance-level segmentation in the Gland Challenge dataset. Bottom row: Our unannotated training image samples.

CUMedVision [2] and multichannel models [15, 16] were very recently designed especially for gland segmentation. CUMedVision [2] won the 2015 MICCAI Gland Segmentation Challenge, and the multichannel model in [15] is the best-known model with a sophisticated network structure. As we show, based on a relatively simple segmentation network (SN) and effective use of unannotated images via adversarial training, DAN (using SN) can improve the segmentation performance and give better overall segmentation results than the state-of-the-art methods. Figure 4 gives visual segmentation results of difficult cases in malignant tissues.

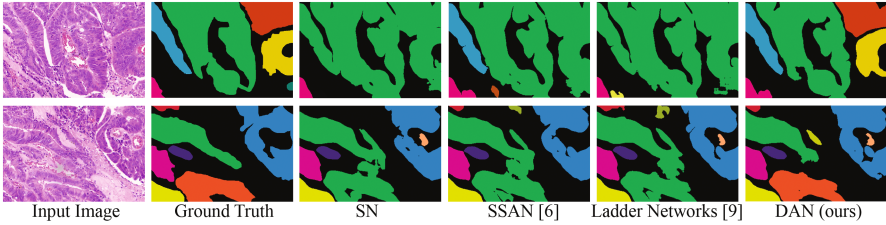


Fig. 4. Instance-level segmentation results on some malignant cases.

Table 1. Summary of the gland segmentation results. SSAN [6] is a latest adversarial network for semantic segmentation, Ladder networks [9] are a state-of-the-art model for semi-supervised learning, CUMedVision [2] won the 2015 MICCAI Gland Segmentation Challenge, and Multichannel2 [15] is the current best published model for gland segmentation on the MICCAI dataset.

Method	# images used		F_1 Score		ObjectDice		ObjectHausdorff	
	Anno.	Unanno.	part A	part B	part A	part B	part A	part B
SN (base model)	85	0	0.9071	0.825	0.898	0.826	48.740	126.479
SSAN [6]	85	0	0.9060	0.836	0.886	0.818	53.393	128.385
Ladder networks [9]	85	100	0.9047	0.833	0.893	0.818	45.418	110.984
CUMedVision [2]	85	0	0.912	0.716	0.897	0.718	45.418	160.347
Multichannel1 [16]	85	0	0.858	0.771	0.888	0.815	54.202	129.930
Multichannel2 [15]	85	0	0.893	0.843	0.908	0.833	44.129	116.821
DAN (ours)	85	100	0.916	0.855	0.903	0.838	45.276	104.982

Fungus segmentation. We also test our DAN model using four 3D EM image stacks (\sim size $1658 \times 1705 \times 100$ each) for fungus segmentation. The 3D EM images are captured from body tissues of ants. In biomedical applications, one may often have only a limited amount of annotated 2D slices for model training for 3D segmentation problems. To model such scenarios, we use only one slice

Table 2. Results for pixel-level fungus segmentation in EM images.

Method	# slices used		Recall	Precision	F_1 Score
	Anno.	Unanno.			
SN (base line)	4	0	0.9020	0.9287	0.9152
SSAN [6]	4	0	0.9067	0.9295	0.9179
Ladder networks [9]	4	40	0.9168	0.9223	0.9195
DAN (ours)	4	40	0.9302	0.9428	0.9364

from each stack to form the annotated images; 10 extra slices are utilized from each stack to form the unannotated images; 20 slices in each stack are marked with ground truth for testing the segmentation performance of different models.

Table 2 shows the pixel-level fungus segmentation results of our model and three closely related models. Our model produces considerably better results.

4 Conclusions

In this paper, we proposed a deep adversarial network that can effectively utilize unannotated image data for training biomedical image segmentation neural networks with better generalization and robustness.

Acknowledgement. This project was supported in part by the National Science Foundation under grant CCF-1640081, and the Nanoelectronics Research Corporation (NERC), a wholly-owned subsidiary of the Semiconductor Research Corporation (SRC), through Extremely Energy Efficient Collective Electronics (EXCEL), an SRC-NRI Nanoelectronics Research Initiative under Research Task ID 2698.005. The research was supported in part by NSF grants CCF-1217906, CNS-1629914, CCF-1617735, and IOS-1558062, and NIH grant R01 GM116927-02.

References

1. Chen, H., Qi, X., Cheng, J.Z., Heng, P.A.: Deep contextual networks for neuronal structure segmentation. In: AAAI, pp. 1167–1173 (2016)
2. Chen, H., Qi, X., Yu, L., Heng, P.A.: DCAN: deep contour-aware networks for accurate gland segmentation. In: CVPR, pp. 2487–2496 (2016)
3. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. In: NIPS, pp. 766–774 (2014)
4. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., et al.: Generative adversarial nets. In: NIPS, pp. 2672–2680 (2014)
5. Hong, S., Noh, H., Han, B.: Decoupled deep neural network for semi-supervised semantic segmentation. In: NIPS, pp. 1495–1503 (2015)
6. Luc, P., Couprie, C., Chintala, S., Verbeek, J.: Semantic segmentation using adversarial networks. arXiv preprint [arXiv:1611.08408](https://arxiv.org/abs/1611.08408) (2016)

7. Papandreou, G., Chen, L.C., Murphy, K., Yuille, A.L.: Weakly and semi-supervised learning of a DCNN for semantic image segmentation. arXiv preprint [arXiv:1502.02734](https://arxiv.org/abs/1502.02734) (2015)
8. Ranzato, M., Szummer, M.: Semi-supervised learning of compact document representations with deep networks. In: ICML, pp. 792–799 (2008)
9. Rasmus, A., Berglund, M., Honkala, M., Valpola, H., Raiko, T.: Semi-supervised learning with Ladder networks. In: NIPS, pp. 3546–3554 (2015)
10. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)
11. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training GANs. In: NIPS, pp. 2226–2234 (2016)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
13. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., et al.: Gland segmentation in colon histology images: the GlaS challenge contest. *Med. Image Anal.* **35**, 489–502 (2017)
14. Sudderth, S.C., Kergosien, Y.L.: Rule-injection hints as a means of improving network performance and learning time. In: Almeida, L.B., Wellekens, C.J. (eds.) EURASIP 1990. LNCS, vol. 412, pp. 120–129. Springer, Heidelberg (1990). doi:[10.1007/3-540-52255-7_33](https://doi.org/10.1007/3-540-52255-7_33)
15. Xu, Y., Li, Y., Liu, M., Wang, Y., Fan, Y., Lai, M., Chang, E.I., et al.: Gland instance segmentation by deep multichannel neural networks. arXiv preprint [arXiv:1607.04889](https://arxiv.org/abs/1607.04889) (2016)
16. Xu, Y., Li, Y., Liu, M., Wang, Y., Lai, M., Chang, E.I.-C.: Gland instance segmentation by deep multichannel side supervision. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 496–504. Springer, Cham (2016). doi:[10.1007/978-3-319-46723-8_57](https://doi.org/10.1007/978-3-319-46723-8_57)