2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)
Venice, Italy, April 8-11, 2019

DOMAIN ADAPTIVE SEGMENTATION IN VOLUME ELECTRON MICROSCOPY IMAGING

Joris Roels[1,3] *Julian Hennies*[4] *Yvan Saeys*[2,3] *Wilfried Philips*[1] *Anna Kreshuk*[4]

[1] Department of Telecommunications and Information Processing,
Ghent University / IMEC, Ghent, Belgium
[2] Department of Applied Mathematics, Computer Science and Statistics,
Ghent University, Ghent, Belgium
[3] Center for Inflammation Research, VIB, Ghent, Belgium
[4] Cell Biology and Biophysics, EMBL, Heidelberg, Germany

ABSTRACT

In the last years, automated segmentation has become a necessary tool for volume electron microscopy (EM) imaging. So far, the best performing techniques have been largely based on fully supervised encoder-decoder CNNs, requiring a substantial amount of annotated images. Domain Adaptation (DA) aims to alleviate the annotation burden by 'adapting' the networks trained on existing groundtruth data (source domain) to work on a different (target) domain with as little additional annotation as possible. Most DA research is focused on the classification task, whereas volume EM segmentation remains rather unexplored. In this work, we extend recently proposed classification DA techniques to an encoder-decoder layout and propose a novel method that adds a reconstruction decoder to the classical encoder-decoder segmentation in order to align source and target encoder features. The method has been validated on the task of segmenting mitochondria in EM volumes. We have performed DA from brain EM images to HeLa cells and from isotropic FIB/SEM volumes to anisotropic TEM volumes. In all cases, the proposed method has outperformed the extended classification DA techniques and the finetuning baseline. An implementation of our work can be found on https://github.com/JorisRoels/domain-adaptive-segmentation.

Index Terms— Electron microscopy, segmentation, domain adaptation

1. INTRODUCTION

Recent developments in volume electron microscopy (EM) have dramatically increased the throughput and simplified the acquisition of large-scale datasets. The problem of segmenting the resulting volumes has also received a lot of attention [1, 2, 3, 4]. For a specific use-case (*e.g.* segmentation of neuron circuits [1, 2, 4] or mitochondria [3]), the state-of-the-art workflows are based on training encoder-decoder networks using large amounts of pixel-level labels. The extracted features are typically data-dependent and high performance on slightly different datasets (*e.g.* different microscope or sample preparation protocol) is therefore not always guaranteed.

Domain adaptation (DA) tackles the problem of building a predictive model for a target dataset with no or very few labels by using a relatively large labeled source dataset. The state-of-the-art in (deep) DA is however largely focused on classification [5, 6, 7] and an extension to segmentation is not straightforward. Recent developments in the field of segmentation show promising results [8, 9], even specifically for EM [10]. However, they regularize only a small fraction of the extracted features or are based on adversarial networks, which are hard to optimize for end-users without significant deep learning expertise. In this work, we introduce a natural extension of classification-based DA techniques to encoder-decoder segmentation networks by regularizing the encoder features. This regularization significantly improves the network performance in the target domain over classical finetuning, at an additional computational cost. Furthermore, we propose a new unsupervised DA method for such networks based on auto-encoder feature alignment [11, 12] which avoids computationally intensive regularization metrics or challenging adversarial network training without sacrificing segmentation performance.

We start with a brief overview of the related work in classification and segmentation DA (section 2). Next, we propose an extension of classification-based DA techniques to the segmentation task in section 3. section 4 describes the new unsupervised auto-encoder DA method in more detail (section 4). Finally, all methods are validated on two mitochondria segmentation use-cases in volume EM data (section 5).

This research has been made possible by the Agency for Flanders Innovation & Entrepreneurship (VLAIO). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research. We would like to thank Anna Steyer and Yannick Schwab (EMBL - Volume Correlative Light and Electron Microscopy) for the provided FIB-SEM HeLa dataset.

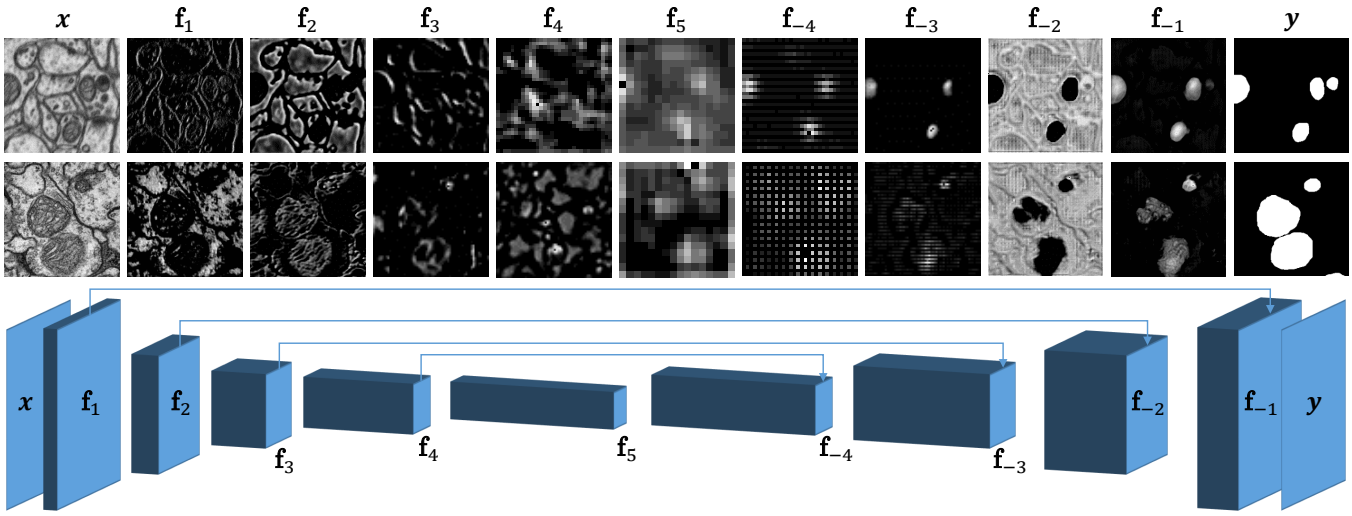

Fig. 1. Encoder-decoder segmentation network architecture with skip connections. The top and bottom row illustrate layer activations extracted from source and target data, respectively, with a network that was trained on the source. The domain shift is especially visible in the encoder features ($\mathbf{f}_i$), whereas the decoder features ($\mathbf{f}_{-i}$) are much closer to an actual segmentation result. This motivates discrepancy regularization on the encoder features.

2. RELATED WORK

Segmentation in volume EM is a semantic segmentation problem where each pixel is to be assigned the appropriate class label. The current state-of-the-art is largely based on extracting features in the encoder through various convolution and pooling stages and returning to a segmentation at the original resolution through the decoder with skip-connections [1, 2, 4].

Most DA approaches are designed for classification and align source and target features by including a domain discrepancy loss. The work of [5] models this discrepancy by means of a distribution similarity metric termed maximum mean discrepancy (MMD). Alternatively, a feature correlation difference (CORAL) is proposed in [6]. In [13], domain classifiers and gradient reversal layers are introduced to align the feature distributions in an adversarial setup (DANN).

The first DA method for semantic segmentation was proposed in [8]. It is based on classical CNN feature extractors where the last feature layer is aligned using an adversarial loss. The recent work of [9] also employs domain confusion for alignment, but additionally normalizes the visual appearance of source and target data. Alternatively, the work of [10] proposes shared decoders combined with MMD regularization on the final decoder activations.

3. DOMAIN ADAPTATION SEGMENTATION

Inspired by [10], we propose an extension of classification-based DA to encoder-decoder segmentation, using the MMD, CORAL and DANN approaches. Unsupervised DA segmentation assumes a labeled source $\mathcal{S} = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1,\dots,n^s}$ of images $\mathbf{x}_i^s \in \mathbb{R}^N$ and pixel-level labels $\mathbf{y}_i^s \in \{0,\dots,C-1\}^N$ and an unlabeled target $\mathcal{T} = \{\mathbf{x}_i^t\}_{i=1,\dots,n^t}$ of images $\mathbf{x}_i^t \in \mathbb{R}^N$ for which the goal is to maximize target segmentation performance. In the semi-supervised setup, there is also a small amount of target labels $\mathbf{y}_i^t \in \{0,\dots,C-1\}^N$ available. For notational convenience, we define $\mathcal{L}_s$ as any segmentation loss (*e.g.* cross entropy), $\hat{\mathbf{y}}^{s/t}$ is the output of the source/target segmentation network, $\mathbf{f}_i^{s/t}$ and $\mathbf{f}_{-i}^{s/t}$ are the final feature activations on level i in respectively the encoder and decoder for source/target (see figure 1).

The discussed classification DA approaches (DANN, CORAL and MMD) include a domain regularization loss $\mathcal{L}_d$ on the source and target features. The aligned features are usually the final activations used for classification. In an encoder-decoder setup, the feature extractor and pixel-wise classifier are not that clearly separated due to the skip connections between encoder and decoder layers. Nevertheless, we denote that the encoder activations largely contain segmentation features, whereas the decoder activations largely serve for segmentation refining and resolution enhancement (see figure 1). Additionally, high-resolution encoder features (*i.e.* the first layers) require less alignment compared to the low-resolution (high-level) encoder features, which should be more domain-invariant. Therefore, we propose to regularize each encoder feature activation level in a weighted fashion, *i.e.*:

$$\mathcal{L} = \mathcal{L}_s(\hat{\mathbf{y}}^s, \mathbf{y}^s) + \sum_i \lambda_i \mathcal{L}_d(\mathbf{f}_i^s, \mathbf{f}_i^t) \qquad (1)$$

where λ_i are regularization parameters and increasing w.r.t. i.

Note that a target segmentation loss can be added to the loss function in equation (1) in the semi-supervised case.

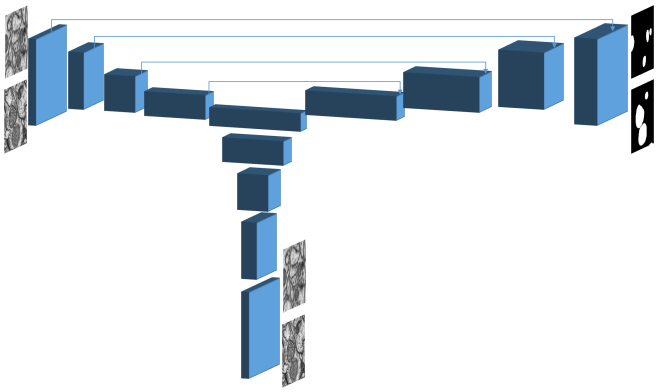

Fig. 2. Proposed unsupervised DA approach: a second decoder is attached to the encoder-decoder segmentation network which reconstructs both the source and target data.

However, we experienced that this limits the capacity of the segmentation network significantly. Therefore, the network is initially trained unsupervised and finetuned with the available target labels in the semi-supervised case.

4. Y-NET

The techniques discussed in the previous section compensate the domain shift between source and target domain by introducing feature (distribution) similarity metrics. They are, however, computationally intensive (*e.g.* MMD computation, correlation matrix computation in CORAL, domain classification in DANN). The work of [11], however, shows that auto-encoders are able to extract generic useful features for classification, whereas the recent work of [12] shows that these architectures are also able to align feature distributions. This motivates our idea of introducing a second decoder to the classical encoder-decoder setup which serves to reconstruct the input data which originates from both source and target domain (see figure 2). The complete architecture is trained end-to-end with the following loss function:

$$\mathcal{L} = \mathcal{L}_s(\hat{\mathbf{y}}^s, \mathbf{y}^s) + \lambda_r^s \mathcal{L}_r(\hat{\mathbf{x}}^s, \mathbf{x}^s) + \lambda_r^t \mathcal{L}_r(\hat{\mathbf{x}}^t, \mathbf{x}^t) \quad (2)$$

where $\mathcal{L}_r$ is a reconstruction loss function (*e.g.* mean-squared error), $\hat{\mathbf{x}}^{s/t}$ are reconstructions of the source/target inputs obtained by the auto-encoding sub-network and $\lambda_r^{s/t}$ are regularization parameters. The network is initially trained in an unsupervised fashion, after which the reconstruction decoder is discarded. Similar as in section 3, the remaining segmentation network is finetuned on the target labels in the semi-supervised case.

5. RESULTS & DISCUSSION

We validate the discussed DA approaches on the problem of mitochondria segmentation in volume EM data. The

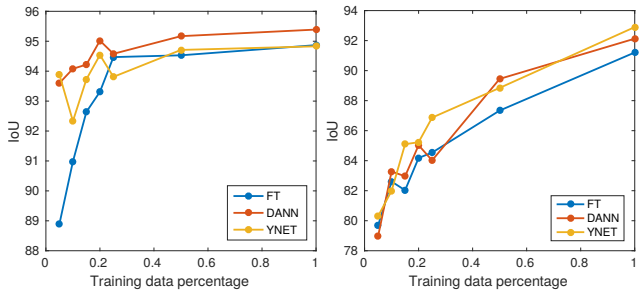

Fig. 3. Segmentation performance on the HeLa (left) and Drosphila dataset (right) after finetuning on various fractions of the target data.

source dataset consists of two annotated $165 \times 1024 \times 768$ FIB-SEM acquisitions (respectively for training and testing) of the CA1 hippocampus region at 5 nm^3 isotropic resolution. We consider two target volumes. The first dataset (HeLa) consists of a $64 \times 512 \times 512$ annotated FIB-SEM block of a HeLa cell at 5 nm lateral and 8 nm axial resolution. The second dataset (Drosophila) [14] is an annotated $20 \times 1024 \times 1024$ serial section Transmission Electron Microscopy (ssTEM) block of the Drosophila melanogaster third instar larva ventral nerve cord at 5 nm lateral and 50 nm axial resolution. Note that mitochondria in the HeLa data are significantly different from those in the source data and that the Drosophila data originates from a different modality and is highly anisotropic, which makes DA particularly challenging. Both target datasets are split along the x axis: 67% and 33% was used for training and testing, respectively.

	FT	MMD	CORAL	DANN	Y-NET
H	8.83	2.33	3.46	11.80	**22.51**
D	28.70	44.96	40.28	**49.90**	49.55

Table 1. Segmentation performance (in terms of IoU) of the discussed DA approaches on the HeLa (H) and Drosophila (D) dataset in the unsupervised setting.

We compare the methods described in sections 3 and 4 to the classical finetuning baseline (FT) which pre-trains the segmentation network on the source and finetunes on the available target labels. Segmentation performance on the target test set is measured by means of the intersection-over-union (IoU). Figure 1 summarizes the unsupervised results for the HeLa and Drosophila dataset. Generally speaking, all the DA approaches significantly outperform the finetuning baseline on the Drosophila data, whereas the domain shift in the HeLa data is too large for MMD and CORAL regularization. For both datasets, DANN is the best performing regularization-based technique. The proposed Y-NET achieves similar to better performance. By finetuning on a fraction of the target labels, we denote that DANN and Y-NET generally outperform the finetuning baseline (figure 3). Figure 4 shows

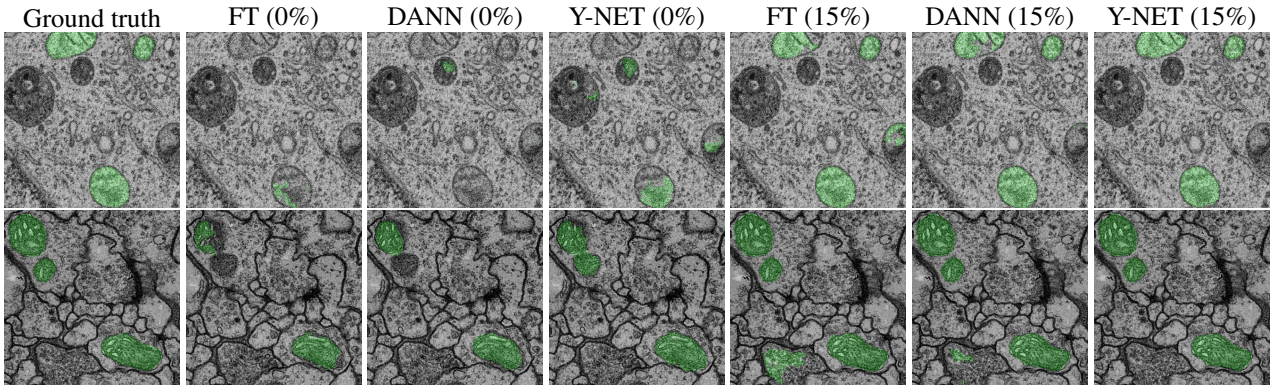

Fig. 4. Qualitative comparison of the finetuning baseline (FT), DANN and Y-NET for the HeLa (top) and Drosophila (bottom) dataset. We illustrate segmentation results in the unsupervised and semi-supervised setting (using 15% of the target labels).

qualitative segmentation results on the HeLa and Drosophila datasets. Both DANN and Y-NET are able to detect large fractions of mitochondria and outperform the finetuning baseline significantly. Note that the Y-NET approach avoids erroneous detections obtained by finetuning, *e.g.* the upper left mitochondria and the lower left structure in the Drosophila data.

6. CONCLUSION

Convolutional neural networks deliver state-of-the-art segmentation results, with the down-side of requiring large amounts of labeled data. Similar shortcomings can be found in all supervised deep learning tasks, but image classification problems have been the target of most domain adaptation work so far. We have demonstrated how the domain adaptation techniques originally proposed for classification can be extend to encoder-decoder segmentation networks. We have also introduced a new DA approach which overcomes the domain shift by training an additional decoder unsupervised on both source and target domains. We believe that the conceptually simple auto-encoding alignment approach will ease the application of CNN-based segmentation in biomedical imaging. In future work, we plan to address unsupervised approaches such as zero-shot learning for segmentation in volume electron microscopy.

7. REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science*, 2015.

[2] H. Chen, X. Qi, J.-z. Cheng, and P.-a. Heng, "Deep Contextual Networks for Neuronal Structure Segmentation," *Proceedings of the 30th Conference on Artificial Intelligence*, 2016.

[3] I. Oztel, G. Yolcu, I. Ersoy, T. White, and F. Bunyak, "Mitochondria Segmentation in Electron Microscopy Volumes using Deep Convolutional Neural Network," *IEEE International Conference on Bioinformatics and Biomedicine*, 2017.

[4] J. Funke, F. D. Tschopp, W. Grisaitis, A. Sheridan, C. Singh, S. Saalfeld, and S. C. Turaga, "Large Scale Image Segmentation with Structured Loss based Deep Learning for Connectome Reconstruction," *Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[5] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, U. Dogan, M. Kloft, F. Orabona, and T. Tommasi, "Domain-Adversarial Training of Neural Networks," *Journal of Machine Learning Research*, 2016.

[6] B. Sun and K. Saenko, "Deep CORAL: Correlation alignment for deep domain adaptation," in *Lecture Notes in Computer Science*, 2016.

[7] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep Transfer Learning with Joint Adaptation Networks," in *International Conference on Machine Learning*, 2017.

[8] J. Hoffman, D. Wang, F. Yu, and T. Darrell, "FCNs in the Wild: Pixel-level Adversarial and Constraint-based Adaptation," *arXiv preprint arXiv:1612.02649*, 2016.

[9] Y. Zhang, Z. Qiu, T. Yao, D. Liu, and T. Mei, "Fully Convolutional Adaptation Networks for Semantic Segmentation," *Conference on Computer Vision and Pattern Recognition*, 2018.

[10] R. Bermudez-Chacon, P. Marquez-Neila, M. Salzmann, and P. Fua, "A domain-adaptive two-stream U-Net for electron microscopy image segmentation," in *International Symposium on Biomedical Imaging*, 2018.

[11] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, "Deep reconstruction-classification networks for unsupervised domain adaptation," in *Lecture Notes in Computer Science*, 2016.

[12] L. Hu, M. Kan, S. Shan, and X. Chen, "Duplex Generative Adversarial Network for Unsupervised Domain Adaptation," *Conference on Computer Vision and Pattern Recognition*, 2018.

[13] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning Transferable Features with Deep Adaptation Networks," in *International Conference on Machine Learning*, 2015.

[14] S. Gerhard, J. Funke, J. Martel, A. Cardona, and R. Fetter, "Segmented anisotropic ssTEM dataset of neural tissue," 2013.