STYLE TRANSFER AND PSEUDO-LABEL FILTERING IMPROVE TRANSFERABILITY IN CELL ORGANELLE SEGMENTATION SCENARIOS

Dmitrii Seletkov[⋆∘1], *Simon Reiß*[⋆†1], *Alexander Freytag*[†], *Constantin Seibold*[‡], *Rainer Stiefelhagen*[⋆]

[⋆]Karlsruhe Institute of Technology [∘] Technical University of Munich [†]Carl Zeiss AG [‡]UK Essen

ABSTRACT

Our community has experienced a lot of progress in analyzing biomedical images driven by semantic segmentation solutions. However, the insufficient ability to adapt to new data distributions limits their applicability. As an example, we observed that cell organelle segmentation models can easily drop by more than 60% in relative accuracy when applied to differently imaged cell data. While bridging this gap is possible by collecting new annotations for new data, it is highly repetitive, inefficient, and expensive. In this work, we evaluate how unsupervised and weakly supervised domain adaptation techniques can help to close this gap more efficiently. We answer the questions of how well domain adaptation techniques perform in cell organelle segmentation and whether easy-to-obtain image-level information gives specific benefits. Based on our findings, we propose StyleFilter: a simple and effective approach that uses image-level labels and leads to an observed improvement in 19.2% absolute DICE over the naive transfer baseline in electron microscopy-based domain adaptation for cell organelle segmentation.

Index Terms— Cell organelle segmentation, domain adaptation, focused ion beam electron microscopy

1. INTRODUCTION AND RELATED WORK

Understanding the semantic meaning of each pixel in an image can help in many applications (*e.g.*, in medical scenarios by quantifying the progression of a disease) or even lead to new scientific insights. Deriving such rich information via manual inspection is often impossible due to the required time investment, but with computational support, it becomes feasible at scale. Automatically deriving such pixel-precise understanding is most commonly done via deep neural networks, which require large amounts of data and annotation during training. However, acquiring such annotations in biomedical domains is challenging since experts have to provide them. This, again, requires experts to spend their time on annotation rather than their actual work.

Even if collecting annotations for training would be feasible, a second challenge occurs from domain shifts when applying a trained model to new data, which has slightly different properties than the training data. As an example, see in Figure 1 how the segmentation quality changes from the training domain (left) to sightly different new domains (middle and right). Domain adaptation (DA) emerged as a technology field to address this challenge. During training, DA exploits annotated source data and additional target data, either without further annotations (unsupervised DA) or with weak or partial labels (weakly- / semi-supervised DA). In this paper, we explore UDA and WDA in the context of multi-class cell organelle segmentation within focused ion beam electron microscopy imagery and look into whether image-level labels are worth their annotation time.

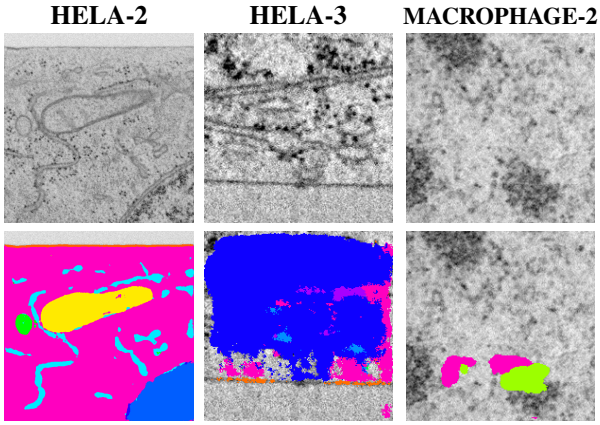

Fig. 1. Different imaging domains in OpenOrganelle and their effects on predictions of a HELA-2-trained model.

The primary subject of investigation for segmentation-related UDA has been traffic scene segmentation, where the domain gap is induced by training on synthetic data and adapting to real driving scenes (*e.g.*, see [1] and references therein). In the biomedical domain, style transfer [2], self-supervised reconstruction of target images [3], or adversarial learning for distinguishing source and target domains [4] were previously investigated. In this work, we analyze how well UDA methods perform on cell organelle segmentation in a challenging multi-class and partial class-overlap setting. On top of that, we investigate WDA and the resulting benefit of additional information at the cost of additional annotation efforts. Paul *et al.* [5] showed that this additional information can greatly improve the segmentation adaptation at manageable costs. Also, WDA has been extensively analyzed for urban scenarios, *e.g.*, with adversarial objectives [5] or multi-curriculum learning [6]. However, we are not aware of any cell-related biomedical WDA approaches.

[1]authors contributed equally to this work

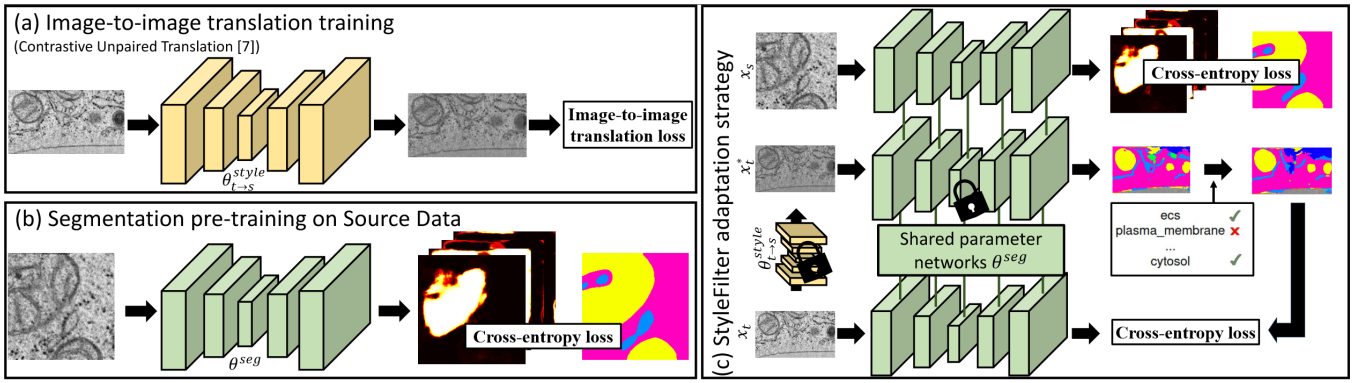

Fig. 2. StyleFilter for weakly supervised domain adaptation consists of three main steps: (a) image-to-image translation training, (b) source segmentation training, and (c) StyleFilter adaptation with PLF. Locks indicate that no back-propagation is applied.

Our contributions summarize to: (1) We investigate domain adaptation in a multi-class cell organelle segmentation scenario with partially overlapping class distribution under a suitable evaluation protocol, (2) we evaluate a variety of adaptation techniques to understand what works and where limitations remain, and (3) we propose StyleFilter as a training strategy for bridging the domain gap in electron microscopy cell imaging, combining style transfer with pseudo-label filtering to exploit image-level information.

2. METHODS AND MATERIALS

Methodology: We focus on UDA and WDA scenarios, *i.e.*, a source domain is represented by training images $x_s^i, i = 1, \ldots, N_S$ and associated pixel-wise annotations m_s^i and a target domain is represented by training images $x_t^j, j = 1, \ldots, N_T$ plus optional image-level labels $l_t^j \in \mathcal{R}^C$. For such scenarios, we introduce StyleFilter, which consists of three main steps: (1) train a style transfer model $\theta_{t\to s}^{style}(\cdot)$, (2) pre-train a segmentation network on the source data with pixel-wise annotations, and (3) adapt this model to the target domain by exploiting $\theta_{t\to s}^{style}(\cdot)$ and image-level labels.

In the first step, we exploit images from source and target domains. We train a style transfer model $\theta_{t\to s}^{style}$ which is able to transform an image x_t^j from the target domain to mimic the style from the source domain: $\theta_{t\to s}^{style}(x_t^j) = \overset{\star}{x}{}_t^j$ where $\overset{\star}{x}{}_t^j$ is the target domain image x_t^j stylized as an image from the source domain. In our experiments, we use the unsupervised image-to-image translation approach Contrastive Unpaired Translation (CUT) [7], displayed in fig. 2 (a), which only requires unlabeled source and target domain images.

The second stage includes training a segmentation network θ^{seg} on source domain image-mask pairs (x_s, m_s) (or alternatively select an already trained model). This training can be a simple training with the cross-entropy loss:

$$L_{CE}(x_s^i, m_s^i) = -\frac{1}{\Omega} \sum_{k,l,c=1}^{H,W,C} m_s^{c,k,l} \cdot \log(\theta^{seg}(x_s^i)^{c,k,l}) \;, \quad (1)$$

where $\theta^{seg}(x_s)$ is the post-softmax segmentation prediction of the segmentation model and the running indices k, l, c going over all pixel-locations $\Omega = HWC$ in the image of size $H \times W$ and the corresponding C individual class predictions.

Now, in the third step, the adaptation to the target domain is realized. Due to the domain gap, applying the trained model θ^{seg} as-is on target data would result in poor prediction quality. Therefore, we apply the style transfer model θ^{style} from step one to target training images x_t^j. We now draw inspiration from [8] where pseudo-labels are generated in a teacher-student setup, but only the student is used for back-propagation based on the pseudo-labels. In our case, the stylized images $\overset{\star}{x}{}_t^j$ are used to generate self-inferred pseudo-labels for every unlabeled image:

$$\mathcal{P}(x_t^j) = \sum_{k,l=0}^{W,H} \arg\max_c \; \theta^{seg}(\overset{\star}{x}{}_t^j)^{c,k,l} \;. \quad (2)$$

Hence, we simply take the maximally activating class prediction at each location as pseudo-ground-truth, which is subsequently used for an additional loss $L_{CE}(x_t^j, \mathcal{P}(x_t^j))$.

In WDA, we have additional information about the presence and absence of classes in the target domain image x_t^j through image-level labels l_t^j. Inspired by the idea of pseudo-label filtering (PLF) from semi-weakly supervised learning in [9, 10], we can refine $\mathcal{P}(x_t^j)$ using l_t^j:

$$\mathcal{P}^{filter}(x_t^j, l_t^j) = \sum_{k,l=0}^{W,H} \arg\max_c \; \theta^{seg}(\overset{\star}{x}{}_t^j)^{c,k,l} \cdot l_t^c \;, \quad (3)$$

where class predictions are multiplied with an indicator variable l_t^c which is zero if class c is absent and one otherwise.

In summary, we obtain the following StyleFilter loss: $L_{SF} = L_{CE}(x_s^i, m_s^i) + L_{CE}(x_t^j, \mathcal{P}^{filter}(\overset{\star}{x}{}_t^j, l_t^j))$, where we assumed for the ease of notation that each batch during training consists of a single sample x_s^i from the source domain and a single sample x_t^j from the target domain. A visualization is also given in Figure 2. At inference time, we use $\mathcal{P}(\overset{\star}{x}_t)$ for unseen images x_t.

Experimental setup: As a basis for our investigation, we choose the OpenOrganelle dataset collection [11] of electron microscopy data containing different cells. Specifically, we chose the HELA-2 cell (2, 321 train images) as the source domain and HELA-3 (1, 634 train images) as well as

Method	ex cell sp	plas mem	mito	vesicle	mvb	lyso	endo ret	nucleus	nuc env	micro tub	cytosol	mean
HELA-2 → HELA-3												
Source-only	77.2 ±31.5	23.1 ±24.1	10.5 ±21.8	0.0 ±0.0	0.7 ±1.0	–	5.2 ±7.8	43.5 ±33.0	0.9 ±2.0	0.0 ±0.0	29.1 ±23.1	19.0 ±9.1
CUT [7]	55.8 ±32.2	26.3 ±22.1	56.9 ±24.4	3.8 ±4.1	2.5 ±2.5	–	45.2 ±13.6	50.3 ±15.8	**13.5 ±13.2**	0.1 ±0.1	60.6 ±20.5	31.5 ±8.7
Y-Net [3]	63.9 ±22.2	0.1 ±0.1	3.6 ±6.2	0.0 ±0.0	0.5 ±1.1	–	1.4 ±2.0	25.6 ±27.7	0.0 ±0.0	0.0 ±0.0	44.7 ±18.5	14.0 ±5.9
Paul *et al.* [5]	51.8 ±35.2	33.1 ±32.6	8.9 ±12.6	0.0 ±0.1	0.0 ±0.1	–	0.2 ±0.4	40.6 ±32.8	0.0 ±0.0	0.0 ±0.0	59.6 ±17.0	19.4 ±8.7
Paul *et al.* \ AdvL [5]	**95.1 ±2.0**	**47.1 ±30.7**	28.0 ±16.8	0.3 ±0.7	5.2 ±6.8	–	9.3 ±7.2	40.9 ±30.4	0.6 ±1.1	0.2 ±0.5	51.0 ±19.2	27.8 ±4.9
Paul *et al.* \ AdvL + CUT [5, 7]	81.3 ±8.1	29.9 ±10.3	56.0 ±27.1	**11.8 ±10.3**	**11.1 ±8.2**	–	43.2 ±14.7	28.6 ±25.7	3.4 ±5.3	**0.5 ±1.0**	69.4 ±13.1	33.5 ±9.3
StyleFilter \ CUT	90.8 ±8.3	2.0 ±4.5	6.7 ±13.1	0.0 ±0.0	0.4 ±0.5	–	11.6 ±16.5	**64.9 ±31.6**	6.4 ±13.2	0.0 ±0.0	73.5 ±13.9	25.6 ±6.2
StyleFilter (Ours)	93.0 ±5.6	29.0 ±25.8	**68.0 ±15.3**	7.6 ±7.6	7.1 ±6.2	–	**47.1 ±9.5**	42.7 ±19.2	10.1 ±11.3	0.0 ±0.0	**77.3 ±14.1**	**38.2 ±8.1**
Target-only	89.7 ±7.4	46.3 ±25.1	84.2 ±10.9	5.8 ±6.1	3.6±3.2	–	43.2 ±23.8	77.1 ±17.3	45.4 ±21.5	0.0 ±0.0	78.1 ±13.5	47.3 ±7.2
HELA-2 → MACROPHAGE-2												
Source-only	**61.9 ±22.9**	0.0 ±0.0	–	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	0.2 ±0.4	1.2 ±1.3	–	–	7.9 ±3.0
CUT [7]	18.2 ±21.0	0.2 ±0.2	–	0.1 ±0.	1.2 ±2.7	0.0 ±0.1	7.9 ±5.6	40.9 ±18.96	32.5 ±11.7	–	–	12.6 ±3.7
Paul *et al.* \ AdvL + CUT [5, 7]	25.3 ±21.5	**1.3 ±2.2**	–	1.2 ±2.1	0.2 ±0.5	0.2 ±0.5	**15.1 ±15.5**	28.1 ±28.0	38.0 ±26.8	–	–	13.7 ±8.6
StyleFilter (Ours)	17.2 ±9.4	0.9 ±0.5	–	**1.5 ±2.3**	**0.4 ±0.5**	**0.9 ±1.3**	10.3 ±12.6	**42.8 ±18.9**	**39.7 ±24.3**	–	–	**14.2 ±4.4**
Target-only	92.2 ±6.5	20.0 ±9.7	–	0.5 ±0.9	1.3 ±2.7	37.4 ±28.7	10.5 ±12.6	36.9 ±15.2	47.2 ±20.9	–	–	30.7 ±5.5

Table 1. Segmentation performance in DICE with standard deviation (**best**, runner up) for domain adaptation from HELA-2 to HELA-3 and MACROPHAGE-2. CUT and Y-Net are UDA techniques; the remaining methods utilize image-level labels. The symbols + and \ refer to either inclusion or exclusion of the corresponding components.

MACROPHAGE-2 ($1,482$ train images) as target domains to quantify the degree of domain shift between them. We follow the evaluation protocol of Reiß *et al.* [9] and use the same merged classes and the first five cross-validation splits for our domain adaptation experiments. The classes of HELA-2 are a superset of the present classes in HELA-3 and MACROPHAGE-2, which results in a partial domain adaptation setting [1]. Therefore, in the evaluations, lyso is excluded for HELA-2 → HELA-3 and mito, micro tub, and cytosol are excluded for HELA-2 → MACROPHAGE-2.

The lower bound of our adaptation scenarios are source-only models, which are trained on source data and evaluated on target domain test sets without any adaptation. An upper bound is fully supervised training using only target training data with pixel-wise annotations and evaluating on target domain test data. The difference in performance between source-only and target-only models indicates the strength of the gap between the source and target domain. We measure segmentation accuracy on the target domain via DICE score, averaged over five cross-validation splits. For all experiments, two 11GB NVIDIA RTX 2080 Ti GPUs are used for training. The baseline Paul *et al.* [5] takes the longest to train, which amounts to a wall-clock training time of close to 2 days.

Implementation details: As segmentation architecture, we use Unet [12] as today's standard model in biomedical image segmentation. During training, we pad all images to the maximum size of 500×500, rotate them by multiples of $90°$, and apply Gaussian noise and color jittering for augmentation. We pre-train on source data for 100 epochs. We use the AdamW optimizer with a momentum term of 0.9, a constant learning rate of 0.0001, and a weight decay of 0.01. We then successively train the adaptation to the target domain for another 100 epochs. Batches of size two with one source and one target image are used. Every five epochs, we evaluate the model on the validation set, select the best validation model after training, and finally evaluate it on the test set.

We compare our approach with the following baselines: **CUT** [7] refers to style transfer from the target domain to source, **Y-Net** [3] refers to a network trained through self-supervised image reconstruction in the target domain and supervised training in source. **Paul *et al.*** refers to the domain adaptation approach as presented in [5] which includes an adversarial learning- (AdvL) and a classification branch. The addition of '\ AdvL' refers to variants of Paul *et al.* that do not use adversarial learning while '\ CUT' indicates that the style transfer component is not applied.

3. RESULTS

Quantitative results: In Table 1, we report the segmentation results of different adaptation strategies on cell organelle segmentation. When adapting from HELA-2 to HELA-3, the domain gap is 28.3% and for HELA-2 to MACROPHAGE-2 22.8%. Simply applying the CUT style transfer to the target domain test images before passing them to the source-only model massively improves the DICE score from 19.0% to 31.5%, showing the efficacy of this adaptation strategy. Interestingly, this simple baseline already outperforms the sophisticated Y-Net, which adds a self-supervised reconstruction loss in the target domain, as well as the full setup of Paul *et al.* even though it utilizes image-level labels in the target domain through a classification branch and adversarial learning. Surprisingly, omitting the adversarial loss from it and only training with the classification branch (Paul *et al.* \ AdvL) improves the results to 27.8%. While it is still inferior to the simple style-transfer baseline CUT. Combining both ideas (Paul *et al.* \ AdvL + CUT) improves the DICE score over CUT's performance to 33.5%. Hence, this improvement of $+2.0\%$ is brought by the added image-level labels. We further observe that inferring a segmentation in the original target domain with a StyleFilter-trained model directly is not successful (StyleFilter \ CUT). In contrast, infer-

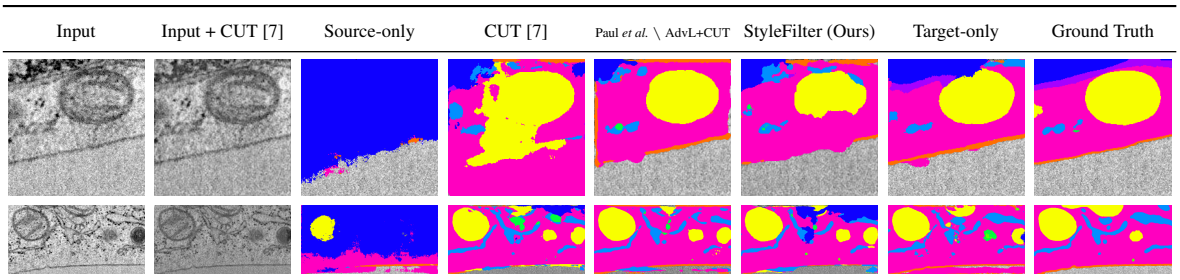

Fig. 3. Qualitative results for HELA-2 $\rightarrow$ HELA-3 domain adaptation. The source-only lower bound fails in segmenting the image; only after applying techniques that utilize style transfer the results noticeably improve. The symbols $+$ and $\setminus$ refer to either inclusion or exclusion of the corresponding components. Best viewed in color and by zooming in.

ring the segmentation on the target to source style-transferred images leads to an increased performance of 38.2%, as StyleFilter was trained with the style transfer module as an integral part. Compared to the best UDA approach, StyleFilter adds $+6.7\%$ in DICE due to the effective utilization of image-level labels, narrowing down the domain gap from 13.8% to 9.1%.

Is style-transfer-based adaptation the definitive way to go for electron microscopy imagery? The HELA-2 and MACROPHAGE-2 datasets are arguably visually more dissimilar to each other. We also observe from quantitative results in the lower part of Table 1 that if the style transfer between the domains is harder to learn, then the benefit for segmentation performance also decreases. Hence, no benefit can be expected from StyleFilter in cases of strongly different domains, which can be viewed as a limitation of the approach.

Qualitative results We finally inspect the drastic effect of the domain gap from HELA-2 $\rightarrow$ HELA-3 and the effect of adaptation in Figure 3, where the source-only baseline fails completely. Interestingly, although the input image and its style-transferred variant look similar to the human eye, they lead to a starkly different segmentation. Clearly, adaptation strategies help to predict a reasonable segmentation of cell organelles.

4. DISCUSSION AND CONCLUSION

Since hardware and sample types vary heavily between microscopy applications, analysis algorithms need to be adaptable between tasks to reuse gathered annotations as efficiently as possible. We therefore investigated how state-of-the-art domain adaptation methods which were developed for street scene applications help in microscopy applications. To our surprise, we found that some methods barely surpass source-only performance, but applying unpaired image translation improved existing methods notably. We finally introduced StyleFilter, a WDA strategy that combined image translation with pseudo-label filtering and found consistent gains in adaption performance. Successfully adapting across visually different domains remains an open challenge.

Acknowledgment This work was supported by funding from Carl Zeiss AG and by funding from the pilot program Core-Informatics of the Helmholtz Association (HGF).

Compliance with Ethical Standards: This research study was conducted retrospectively using imaging data of eukaryotic cell lines which is available in open access at https://openorganelle.janelia.org/. Ethical approval was *not* required as confirmed by the CC BY 4.0 license attached with the open access data.

5. REFERENCES

[1] Csurka et al., "Unsupervised domain adaptation for semantic image segmentation: a comprehensive survey," *ArXiv*, 2021.

[2] Franco-Barranco et al., "Deep learning based domain adaptation for mitochondria segmentation on em volumes," *CMPB*, 2022.

[3] Joris Roels et al., "Domain adaptive segmentation in volume electron microscopy imaging," in *ISBI 2019*.

[4] Haq and Huang, "Adversarial domain adaptation for cell segmentation," in *MIDL*, 2020.

[5] Paul et al., "Domain adaptive semantic segmentation using weak labels," *CoRR*, vol. abs/2007.15176, 2020.

[6] Lv et al., "Weakly-supervised cross-domain road scene segmentation via multi-level curriculum adaptation," *TCSVT*, 2021.

[7] Park et al., "Contrastive learning for unpaired image-to-image translation," *CoRR*, vol. abs/2007.15651, 2020.

[8] Sohn et al., "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," *NeurIPS 2020*.

[9] Reiß et al., "Decoupled semantic prototypes enable learning from diverse annotation types for semi-weakly segmentation in expert-driven domains," in *CVPR 2023*.

[10] Bae et al., "One weird trick to improve your semi-weakly supervised semantic segmentation model," *ArXiv*, 2022.

[11] Heinrich et al., "Whole-cell organelle segmentation in volume electron microscopy," *Nature*, 2021.

[12] Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.