Domain Adaptation for Eye Segmentation

Yiru Shen, Oleg Komogortsev, and Sachin S. Talathi

[1] Facebook Reality Labs
[2] yirus@fb.com; ok1@fb.com; stalathi@fb.com

Abstract. Domain adaptation (DA) has been widely investigated as a framework to alleviate the laborious task of data annotation for image segmentation. Most DA investigations operate under the unsupervised domain adaptation (UDA) setting, where the modeler has access to a large cohort of source domain labeled data and target domain data with no annotations. UDA techniques exhibit poor performance when the domain gap, i.e., the distribution overlap between the data in source and target domain is large. We hypothesize that the DA performance gap can be improved with the availability of a small subset of labeled target domain data. In this paper, we systematically investigate the impact of varying amounts of labeled target domain data on the performance gap for DA. We specifically focus on the problem of segmenting eye-regions from eye images collected using two different head mounted display systems. Source domain is comprised of 12,759 eye images with annotations and target domain is comprised of 4,629 images with varying amounts of annotations. Experiments are performed to compare the impact on DA performance gap under three schemes: unsupervised (UDA), supervised (SDA) and semi-supervised (SSDA) domain adaptation. We evaluate these schemes by measuring the mean intersection-over-union (mIoU) metric. Using only 200 samples of labeled target data under SDA and SSDA schemes, we show an improvement in mIoU of 5.4% and 6.6% respectively, over mIoU of 81.7% under UDA. By using all available labeled target data, models trained under SSDA achieve a competitive mIoU score of 89.8%. Overall, we conclude that availability of a small subset of target domain data with annotations can substantially improve DA performance.

Keywords: Domain Adaptation; Eye Segmentation

1 Introduction

Semantic segmentation is an important problem in computer vision where the objective is to assign labels to each pixel in an image. In recent years, supervised learning methods using convolutional neural networks (CNNs) have enabled significant improvements in the development of models for semantic segmentation [1, 15, 24]. However, supervised training of CNNs require a large amount of images with pixel-wise annotations, which, if done manually is time consuming, non-scalable and label inefficient. To ease the problem of pixel-wise image annotation, unsupervised domain adaptation (UDA) methods have been widely studied. UDA methods attempt to align the distribution of features from a different but related, source domain data, which also contains annotations, to target domain data, where no annotations are available [5, 31].

UDA predominantly uses adversarial training framework to train models that retain good performance for semantic segmentation on source domain data, while at the same time attempting to reduce the discrepancy in feature distribution in both domains. In doing so the hope is that the trained model can semantically segment the target domain data even in the absence of any labeled data. However, UDA have been shown to fail to learn discriminative class boundaries for target domain data. Furthermore, UDA also fails to train models that can generalize well on the target domain data [25, 27, 37]. In addition, the setting under which UDA training is appropriate, may not be suitable for real world scenarios where we also have annotations for a subset of target domain data. In this scenario, the following questions are more relevant:

– Given a fixed annotation budget, how much data need to be labeled in order to reach a reasonable performance on the target domain?
– How to optimally utilize the unlabeled data in target domain?

We focus on answering these questions in the eye-region semantic segmentation research area with a long-term goal of increasing the quality of estimated eye gaze data. Obtained eye tracking quality is critical to many applications in VR/AR such as foveated rendering, intent inference, health assessment, and direct gaze interaction. Specifically, in this work, as a first step toward this goal, we conduct experiments to systematically investigate the impact on segmentation performance for varying levels of target data annotations. Three domain adaptation frameworks are investigated: unsupervised domain adaptation (UDA), where large numbers of target images without target annotations are available [5]; supervised domain adaptation (SDA), where a small number of labeled target images are available [27] and semi-supervised domain adaptation (SSDA), where in addition to large numbers of target images without annotations, a small number of target images with annotations are available [28]. We conduct a series of experiments to train semantic segmentation models using all available labeled source domain data while varying the number of labeled target domain data, which are randomly selected as function of DA framework used for training. Specifically, under UDA, we do not use any of the available labeled target domain data and for SDA, we do not use any of the available unlabeled target domain data.

We investigate a single semantic segmentation model architecture which is trained using one of aforementioned DA frameworks. The model consists of two sub-networks: a segmentation network to predict probability map of an input eye image; and a fully convolutional discriminator network that differentiates the probability map in target domain from those in the source domain. We adopt adversarial training based strategy to train the segmentation network so as to fool the discriminator network by producing distribution of probability maps, which are invariant to the change in domain. The intuition for using adversarial training as a mechanism to adapt the probability map for target domain data from the probability map of the source domain data is based on the fact that both the source domain eye images and the target domain eye images exhibit geometric and spatial structural similarities, for example, the ellipse contour of eyes are always maintained [30].

For the training of semantic segmentation model under SSDA, a semi-supervised loss is designed by generating pseudo-labels for eye regions in the unlabeled target images. Specifically, the discriminator network generates confidence score for each pixel of eye image belonging to the source or the target domain. Only the probability map of regions in the image that are identified as close to source domain are treated as pseudo-labels and trained in the segmentation network via a masked cross entropy loss [10]. This semi-supervised loss enables the model to produce probability map of unlabeled target images close to the distribution of source domain images. Note that although both our method and [10] utilize the signals from discriminator to conduct semi-supervised loss, study in [10] only considers leveraging labeled data and unlabeled data in a single domain. Different from [10], the scope of SSDA in our study is to align the data distributions between source and target domain while at the same time further boost the performance for semantic segmentation by conducting semi-supervise learning in the target domain.

Our contributions are summarized as follows:

– A practical framework to train models under DA by optimally utilizing limited images with annotations and the available unlabeled images in the target domain.
– A systematic comparison of UDA, SDA and SSDA performance as function of the number of available images (and labels) in the target domain.

2 Related Work

2.1 Semantic Segmentation of Eye Regions

Segmentation of periocular regions, including pupil, iris, sclera provide comprehensive information on eyes and is critical for gaze estimation [35]. A majority of studies for eye segmentation have focused on segmenting single trait of eye regions, i.e. sclera, iris, etc. Iris segmentation has been investigated for iris recognition that can be used in personal identification and verification [3, 26]. In [14], ATT-UNet was proposed to learn discriminative features to separate iris and non-iris pixels. Sclera segmentation is usually considered as a preprocessing step for sclera-based recognition in applications of human identification [16, 21, 39]. In [34], ScleraSegNet was proposed to achieve competitive performance for sclera segmentation by using channel-wise attention mechanism. However, limited works have focused on semantic segmentation of all eye regions, i.e. pixel-wise classification due to the lack of availability of large-scale datasets with labeled eye images [4, 17, 18]. Recently, the Open Eye Dataset (OpenEDS) has been released to facilitate development of models for semantic segmentation of all eye regions [7]. Study in [7] trained a modified SegNet model [1] by introducing a multiplicative skip connection between the last layer of the encoder and the decoder network, to classify pupil, iris, sclera and background for each pixel of a given eye image. More recently, few studies have investigated models to segment all eye regions using the OpenEDS datasets, primarily focusing on low computational complexity models [2, 11, 23].

2.2 Domain Adaptation for Semantic Segmentation

Domain adaptation (DA) has been applied to develop semantic segmentation models to ease the problem of data annotations, by aligning the feature or output pixel-wise class distributions between the source and the target images [5,6,9,25]. Three different schemes of DA have been widely studied: UDA [8, 30], SDA [20,28,33], and SSDA [10,36]. For these schemes, adversarial learning has become a prominent approach in semantic segmentation, where the critical step is to train a domain classifier or discriminator to differentiate whether the representation is from source or target domain [6, 9, 32]. Due to the assumption of non-availability of supervision from target domain annotations, many recent work has focused on UDA by aligning distributions between source and target domain in feature space [38], or learning discriminative representations in output space [30]. However, UDA can fail to learn discriminative class boundaries on target domains [25]. SDA is applied when limited labeled target images are available. For example, researchers in [28] proposed a hierarchical adaptation method to align feature distributions between abundant labeled synthetic images in source domain and insufficient labeled images in real-world scenario. In SSDA, in addition to a small amount of labelled target images, a larger amount unlabeled target images is available. However, most of the SSDA is applied in applications of classification and object detection [25, 29, 37], SSDA for semantic segmentation has not been fully explored. In this paper, we revisit this task and compare SSDA with SDA and UDA under varying amount of training images in target domain.

3 Method

3.1 Overview

Fig. 1 provides an overview of our proposed model. Our goal is to perform DA from all available labeled images in the source domain, $\mathcal{X}_s$, to produce annotations for unlabeled images in the target domain, $\mathcal{X}_t$, by leveraging the limited number of labeled images available in the target domain. We aim to investigate how the performance on segmentation of unlabeled target domain images is impacted by varying amounts of labeled target domain images.

Our proposed model consists of two sub-networks: a segmentation network $\boldsymbol{SS}$ and a discriminator network $\boldsymbol{D}$. For segmentation network, we consider the mSegNet architecture for eye segmentation [7], which takes images of dimension $H \times W$ as input and outputs the probability map of dimension $H \times W \times K$,

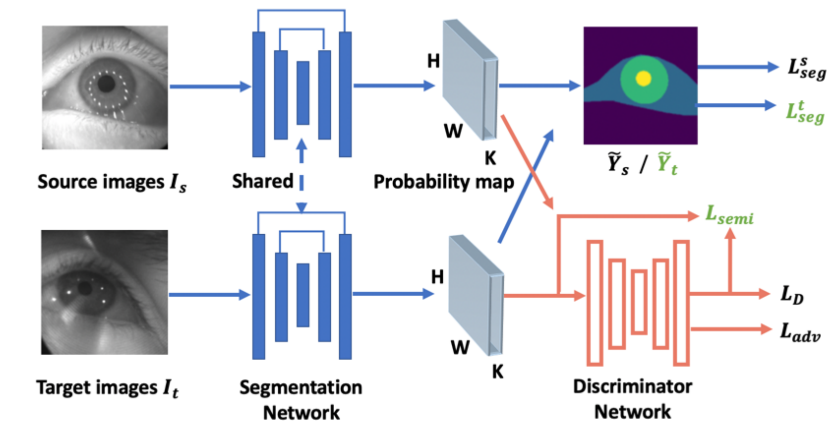

Fig. 1: Architecture overview. Source and target images are passed to segmentation network to produce probability map. Segmentation loss L_{seg} is computed given the probability map. A discriminator loss L_D is calculated to train a discriminator to differentiate whether each pixel is from source or target. An adversarial loss L_{adv} is calculated on target probability map to fool the discriminator. A semi-supervised loss L_{semi} is calculated to boost the training process if additional unlabeled target images are available. Blue arrows indicate forward operation of the segmentation network, while orange arrows indicate that of discriminator. L_{semi} in green is applied when additional unlabeled target images are available, while L^t_{seg} is applied when there are labeled target images.

where K is the number of semantic categories. For discriminator a.k.a. domain classifier, we design a fully convolutional network, which takes probability maps from segmentation network, and outputs a confidence map of dimension $H \times W$. This map indicates how confident the discriminator about each pixel in the input image belonging to source or target domain.

During training, all labeled images in source domain $\mathcal{X}_s$ are used, and we vary the number of labeled images available for training in the target domain $\mathcal{X}_t$. We denote pairs of images with the per-pixel annotations in $\mathcal{X}_s$ as $\{(I_s, Y_s)\}$; pairs of images with per-pixel annotations and the unlabeled images in $\mathcal{X}_t$ as $\{(I_t, Y_t)\}$, $\{I_{t,u}\}$, respectively.

Formally, the inputs for three frameworks for domain adaptation are:

- Unsupervised domain adaptation: $\mathcal{X}_s = \{(I_s, Y_s)\}$; $\mathcal{X}_t = \{I_{t,u}\}$
- Supervised domain adaptation: $\mathcal{X}_s = \{(I_s, Y_s)\}$; $\mathcal{X}_t = \{(I_t, Y_t)\}$; $|Y_s| \gg |Y_t|$
- Semi-supervised domain adaptation: $\mathcal{X}_s = \{(I_s, Y_s)\}$; $X_t = \{(I_t, Y_t); I_{t,u}\}$; $|Y_s| \gg |Y_t|$, $|I_{t,u}| \gg |I_t|$

3.2 Loss Functions

Segmentation Loss Segmentation loss is the cross-entropy loss that compares probability map with the ground truth annotations.

$$L^{d}_{seg} = -\sum_{h,w}\sum_{k\in\mathcal{K}} Y^{(h,w,k)}_{d}\log(\boldsymbol{SS}(I_d)^{(h,w,k)}), \qquad d \in \{\mathcal{X}_s, \mathcal{X}_t\} \tag{1}$$

Adversarial Loss Adversarial loss is used to train the segmentation network to fool the discriminator, in order to make the distribution of target probability map close to source domain.

$$L_{adv} = -\sum_{h,w}\log(\boldsymbol{D}(\boldsymbol{SS}(I_t))^{(h,w)}) \tag{2}$$

Discriminator Loss Discriminator is trained to predict the domain label of each pixel in the probability map. We use $Y_t = 0$ for the source domain and $Y_t = 1$ for the target domain.

$$L_D = -\sum_{h,w}(1-Y_t)\log(1-\boldsymbol{D}(\boldsymbol{SS}(I_s))^{(h,w)}) + Y_t\log(\boldsymbol{D}(\boldsymbol{SS}(I_t))^{(h,w)}) \tag{3}$$

Semi-supervised Loss Semi-supervised loss is used to enhance the training process in a self-taught manner using unlabeled images in the target domain [10,13]. The intuition is that the discriminator generates confidence score for each pixel in the images, and those pixels considered close to source domain could be used as additional information to train the segmentation network. Concretely, we highlight regions in target images where the confidence score is above a threshold, T and treat the probability maps from those regions as the pseudo labels. Therefore, the semi-supervised loss is defined as:

$$L_{semi} = -\sum_{h,w}\sum_{k\in\mathcal{K}} [\![\boldsymbol{D}(\boldsymbol{SS}(I_{t,u}))^{(h,w)} > T]\!] \cdot Y^{(h,w,k)}_{t,u}\log(\boldsymbol{SS}(I_{t,u})^{(h,w,k)}) \tag{4}$$

where $[\![\cdot]\!]$ is the indicator function, and $Y^{(h,w,k)}_{t,u} = 1$ if $k = \mathrm{argmax}_k \boldsymbol{SS}(I_{t,u})^{(h,w,k)}$.

3.3 Objective Functions

In this section, we provide the objective functions for three schemes in domain adaptation. Note that the objective function for discriminator is consistent across unsupervised, supervised and semi-supervised domain adaptation. Hence we follow Equation 3 to train the discriminator and focus more on objective functions of semantic segmentation network $\boldsymbol{SS}$ below.

Unsupervised Domain Adaptation Only unlabeled images in $\mathcal{X}_t$ are available. The goal is to minimize the per-pixel classifier loss in the source domain. Simultaneously, we want to make the distributions of probability maps in the target domain to closely match the probability map of images in the source domain by training segmentation network and discriminator network in a min-max game. The idea is for the segmentation network to be able to take advantage of labeled source domain images for per-pixel classification of target domain images. Besides, semi-supervised loss in Equation 4 is used given the unlabeled target images are available. The objective functions to train the ***SS*** network under UDA can be written as:

$$L_{UDA}^{\boldsymbol{SS}} = L_{seg}^{s} + \lambda_{adv} L_{adv} + \lambda_{semi} L_{semi} \tag{5}$$

where λ_{semi} is the weight for the pseudo annotations generated from the discriminator.

Supervised Domain Adaptation Only a small number of labeled images in $\mathcal{X}_t$ are available. Segmentation network is trained to minimized the per-pixel cross-entropy loss in both domains. The objective functions to train the ***SS*** network under SDA can be written as:

$$L_{SDA}^{\boldsymbol{SS}} = L_{seg}^{s} + L_{seg}^{t} + \lambda_{adv} L_{adv} \tag{6}$$

Semi-supervised Domain Adaptation A majority of unlabeled images are available in $\mathcal{X}_t$, in addition to a small number of labeled images. The goal is to take advantage of the unlabeled images in the target domain to generate pseudo annotations and further improve performance for segmentation of target domain images. As such, the semi-supervised loss is calculated as in Equation 4. The overall objective function to train the ***SS*** network under SSDA can be written as:

$$L_{SSDA}^{\boldsymbol{SS}} = L_{seg}^{s} + L_{seg}^{t} + \lambda_{semi} L_{semi} + \lambda_{adv} L_{adv} \tag{7}$$

3.4 Network Architecture

Segmentation Network The mSegNet network is a modified version of SegNet architecture, wherein a multiplicative skip connection between the last layer of the encoder and the decoder network is introduced, to estimate the probability maps for eye regions, including pupil, iris, sclera and background [7]. The network consists a 7-layer encoder module and a 7-layer decoder module, made up of convolutional layers with kernel-size 3×3 and stride of 1. The number of convolution channels in the encoder are 64, 64, 128, 128, 256, 256, 256, while transposed convolutional layer is used in the decoder, which takes the number of channels of 256, 256, 128, 128, 64, 64 and 4.

Discriminator Discriminator is a fully convolutional network, and we follow the setting in [30] to contain 5 convolution layers, each with kernel 4×4 and stride of 2. The number of channels are 64, 128, 256, 512 and 1. Each convolution layer is followed by a leaky-ReLU non-linearity with parameter 0.2 except for the last layer. An up-sampling layer is added to the last convolutional layer to resize the output to the dimension of input.

4 Experiments and Discussion

We begin by introducing the eye data sets in the source and the target domain. We then investigate how segmentation performance as measured by using the mean intersection-over-union (mIoU) metric [15] is affected by varying the amount of training data in the target domain. Next, we run t-SNE analysis [19] to provide visualizations of the learned probability maps to illustrate the improvements for segmentation of eye images in the target domain using the proposed methods.

4.1 Datasets

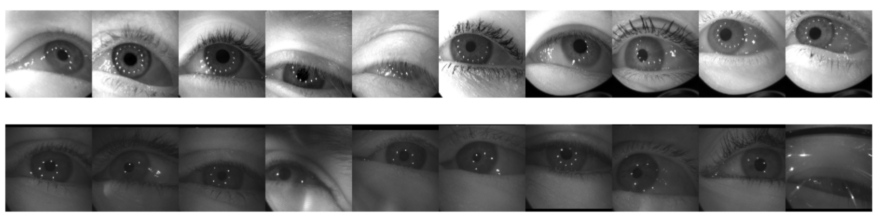

Fig. 2: Top to bottom: examples of images from source and target domain.

Source Domain OpenEDS is a large scale data set of eye-images captured using a virtual reality (VR) head mounted display mounted with two synchronized eye facing cameras recorded at a frame rate of 200 Hz under controlled illumination [7]. This data set was compiled from video capture of the eye-region collected from 152 participants and consists of 12,759 images with pixel-level annotations for key eye-regions: iris, pupil and sclera, at a resolution of 400×640 pixels. Following [7], we use the training set of 8,916 images with the corresponding annotations in OpenEDS as the source domain data.

Target Domain Target domain data was collected using a commercial VR head mounted display that was modified with two eye facing cameras. The data set contains 4,629 eye images at a resolution of 400×640 pixels and 4 semantic categories are provided with pixel-level labels (pupil, iris, sclera and background). For all the experiments we split the data set into images of 1,532, 500 and 2,597 as training, validation and test set. The results of proposed domain adaptation scheme are reported on the test set. Fig. 2 shows some examples from the source and target domains respectively.

4.2 Training Details

We use PyTorch to implement the proposed model [22]. ADAM optimizer [12] is used to train both segmentation network and the discriminator network, with initial learning rate of 0.0001, momentum 0.9 and 0.999 respectively, which is decreased using a polynomial decay function with power of 0.9 [30]. We scale the images to a resolution of 184×184 and train the models for 200 epochs with batch size 8. A parameter search is conducted to find the optimal settings of λ_{adv}, λ_{semi} and T. As a result, we set λ_{adv} as 0.001, λ_{semi} as 0.1, T as 0.8. Algorithm 1 demonstrates the process of training models with varying amount of target domain data.

Algorithm 1: Training Protocol

Input Source domain: $\mathcal{X}_s = \{(I_s, Y_s)\}$;
Target domain (UDA): $\mathcal{X}_t = \{I_{t,u}\}$ or
Target domain (SDA): $\mathcal{X}_t = \{(I_t, Y_t)\}$ or
Target domain (SSDA): $\mathcal{X}_t = \{(I_t, Y_t); I_{t,u}\}$;
N: # unlabeled images selected per iteration;
M: max iterations;
Model Segmentation network SS; Discriminator D;
for *iteration* $= 1$ *to* M **do**
  Initialize SS with parameters that are pretrained on $\mathcal{X}_s$
  UDA Training SS and D with ($\mathcal{X}_s$, $\mathcal{X}_t$) via Eq. 5
  SDA Training SS and D with ($\mathcal{X}_s$, $\mathcal{X}_t$) via Eq. 6
  SSDA Training SS and D with ($\mathcal{X}_s$, $\mathcal{X}_t$) via Eq. 7
  Random select N unlabeled images $I_N \in \mathcal{X}_t$ where:
  UDA $\mathcal{X}_t \leftarrow I_{t,u} \cup I_N$
  SDA $\mathcal{X}_t \leftarrow (I_t, Y_t) \cup (I_N, Y_N)$, Y_N is corresponding labels of I_N
  SSDA $\mathcal{X}_t \leftarrow ((I_t, Y_t) \cup (I_N, Y_N);\ I_{t,u} \setminus I_N)$
end

4.3 Results

Performance with Varying Amount of Training Data In Fig. 3, we chart the segmentation performance, as measured using mIoU, for the three DA frameworks as function of the number of target domain images available in the training data. We note that the mIoU for models trained under all three frameworks increase and reach a plateau as more images in the target domain are available, the extent of increase is much smaller for model trained under UDA. We note that, with only 200 labeled images in the target domain, mIoU is improved by 5.4% (from 81.7% in UDA to 87.1% in SDA) and 6.6% (from 81.7% UDA to 88.3% SSDA), which affirms our hypothesis that even with limited number of labeled target domain images, domain-specific information can boost the performance for segmentation of images in the target domain.

We note that SSDA consistently outperforms SDA, since SSDA is able to leverage the most available information, including images with annotations as well as all the available unlabeled images. As can be seen from equations 6 and 7, the gain in the performance for SSDA relative to SDA is brought about by the addition of adversarial loss as well as the semi-supervised loss from unlabeled target images. As shown in Algorithm 1, however, as more target images with annotations are provided to train the model, the amount of unlabeled target images are decreased. Therefore, the performance gain brought by the adversarial and semi-supervised loss from unlabeled target images is diminishing. This explains the smaller performance gap between SDA and SSDA as more labeled target images are available during training. We finally note that SSDA is able to achieve a competitive mIoU of 89.5% with only 800 labeled target-domain images (53.2% of total data). It is encouraging to observe that the proposed SSDA can efficiently take advantage of insufficient labels while still producing competitive performance on segmentation task.

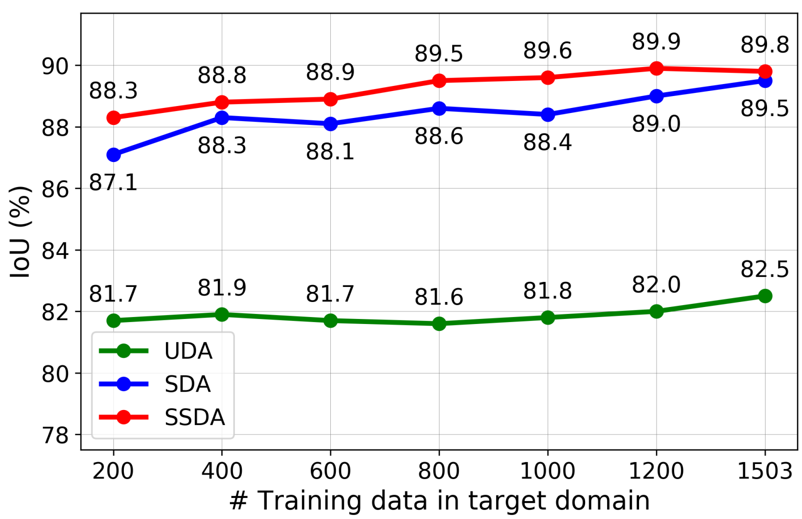

Fig. 3: Performance of three domain adaptation frameworks (unsupervised, supervised and semi-supervised domain adaptation) trained with varying amounts of target images.

Performance with All Training Data In Table 1, we summarize the performance for all the three frameworks for the case when all training images and the corresponding annotations in target domain are used. Two baseline models are being compared: 1) **Source Only:** we train the segmentation model on source domain data. 2) **Target finetuning:** we train the segmentation model on source domain data and then finetuned the segmentation model using the available labeled target domain data. By comparing UDA with Source only, the adversarial loss directly brings an improvement of 15.2%, from 66.3% of baseline to 81.5% for UDA. With additional labeled target domain data available, by comparing SDA and SSDA with target finetuning, mIoU is increased by 0.2% and 0.5%, respectively. Comparing all the domain adaptation based models, given the context of additional labels, mIoU is increased by 8.3%, from 81.5% for UDA to 89.8% for SSDA. Not only does SSDA consistently performs better than SDA over varying number of labeled target domain training images, SSDA also achieved slightly better performance metrics than SDA (0.3%) and target finetuning (0.5%) when all target domain labeled images are available.

Table 1: Results of models trained with all target images (# images = 1,503).

Models	Overall	Pupil	Iris	Sclera	Background
Source only	66.3	60.7	66.4	48.8	89.5
Target finetuning	89.3	89.6	90.6	79.0	97.5
UDA	82.5	82.3	84.0	64.0	94.1
SDA	89.5	89.8	90.7	78.8	97.8
SSDA	89.8	90.0	90.3	79.3	97.9

Visualization of Segmentation Results Fig. 4 shows segmentation results from models trained under different frameworks. Note that the ground truth labels in the target domain are missing fine-grained boundary, i.e. detailed boundaries of sclera. Domain adaptation is able to transfer the knowledge from source domain with high quality labels to target domain where both quality and quantity are limited, which improves segmenting detailed geometry such as sclera in the target images.

Visualization of Feature Clustering Fig. 5 shows the T-SNE [19] of the embeddings of the per-pixel probability maps of training data in source and target domain. The "perplexity" parameter for T-SNE was set as 20. Qualitative results show that with domain adaptation applied, the distributions of source and target domain probability maps overlap closely.

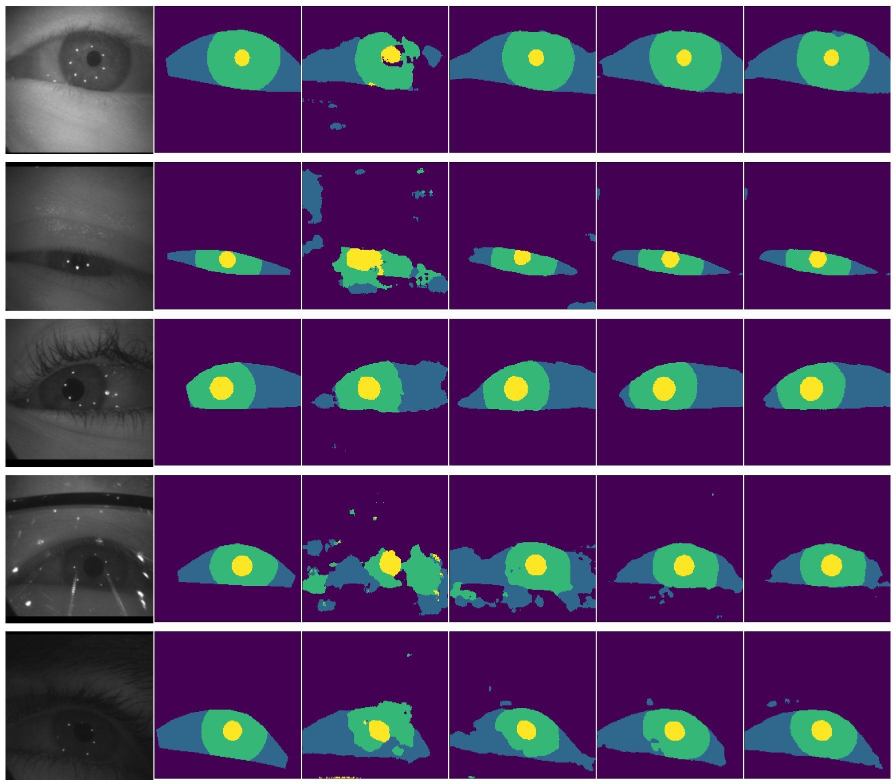

Fig. 4: Examples and segmentation results from different models. From left to right: target images, ground truth, source only, predictions of UDA, SDA and SSDA. From top to bottom: regular-opened eyes, half-opened eyes, long eyelashes, eyeglasses, dim light.

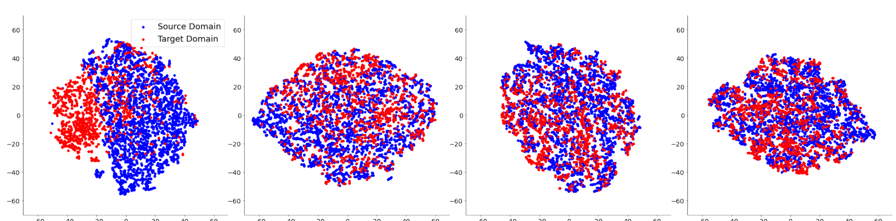

Fig. 5: T-SNE visualization. Blue: source domain. Red: target domain.

5 Conclusions

This paper considers the problem of domain adaptation for eye segmentation, under the setting where a small amount of labeled images and a majority of unlabeled images in the target domain are available. We systematically compare the performance of semantic segmentation models trained under unsupervised (UDA), supervised (SDA) and semi-supervised domain adaptation (SSDA), as a function of the number of labeled target domain training images. Results show that SSDA is able to improve mIoU from 81.7% of UDA to 88.3% when only 200 labeled target domain images are available for training. Furthermore, with 50% of labeled target domain images being used, we are able to achieve a competitive mIoU of 89.5% using SSDA. In conclusion, results demonstrate the benefit to annotate a small number of target domain images to effectively perform domain adaptation. We hope that presented approaches would be useful for eye tracking practitioners that employ segmentation of periocular regions in their eye tracking pipelines to improve data quality of obtained gaze positional data, which is important for providing the broadest possible field of eye tracking-driven applications.

In our future work we plan to explore what aspects of eye tracking data quality (e.g., spatial accuracy, spatial precision, spatial resolution, data loss, etc.) are affected by segmentation of periocular regions and to what degree.

Acknowledgment

We thank our colleagues Jixu Chen and Kapil Krishnakumar for providing target domain data, Robert Cavin, Abhishek Sharma and Elias Guestrin for providing constructive and thoughtful feedback.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence **39**(12), 2481–2495 (2017)
2. Boutros, F., Damer, N., Kirchbuchner, F., Kuijper, A.: Eye-mms: Miniature multi-scale segmentation network of key eye-regions in embedded applications. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 0–0 (2019)
3. Bowyer, K.W., Burge, M.J.: Handbook of iris recognition. Springer (2016)
4. Das, A., Pal, U., Ferrer, M.A., Blumenstein, M., Štepec, D., Rot, P., Emeršič, Ž., Peer, P., Štruc, V., Kumar, S.A., et al.: Sserbc 2017: Sclera segmentation and eye recognition benchmarking competition. In: 2017 IEEE International Joint Conference on Biometrics (IJCB). pp. 742–747. IEEE (2017)
5. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: Bach, F., Blei, D. (eds.) Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 37, pp. 1180–1189. PMLR, Lille, France (07–09 Jul 2015)
6. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. The Journal of Machine Learning Research **17**(1), 2096–2030 (2016)
7. Garbin, S.J., Komogortsev, O., Cavin, R., Hughes, G., Shen, Y., Schuetz, I., Talathi, S.S.: Dataset for eye tracking on a virtual reality platform. In: Symposium on Eye Tracking Research and Applications. pp. 1–10 (2020)
8. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A.A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: ICML. vol. 80, pp. 1994–2003. PMLR (2018)
9. Hoffman, J., Wang, D., Yu, F., Darrell, T.: Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. arXiv preprint arXiv:1612.02649 (2016)
10. Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y., Yang, M.H.: Adversarial learning for semi-supervised semantic segmentation. In: BMVC. p. 65. BMVA Press (2018)
11. Kim, S.H., Lee, G.S., Yang, H.J., et al.: Eye semantic segmentation with a lightweight model. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3694–3697. IEEE (2019)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015), http://arxiv.org/abs/1412.6980
13. Li, Y., Yuan, L., Vasconcelos, N.: Bidirectional learning for domain adaptation of semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6936–6945 (2019)
14. Lian, S., Luo, Z., Zhong, Z., Lin, X., Su, S., Li, S.: Attention guided u-net for accurate iris segmentation. Journal of Visual Communication and Image Representation **56**, 296–304 (2018)
15. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
16. Lucio, D.R., Laroca, R., Severo, E., Britto, A.S., Menotti, D.: Fully convolutional networks and generative adversarial networks applied to sclera segmentation. In: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). pp. 1–7. IEEE (2018)

17. Luo, B., Shen, J., Cheng, S., Wang, Y., Pantic, M.: Shape constrained network for eye segmentation in the wild. In: The IEEE Winter Conference on Applications of Computer Vision. pp. 1952–1960 (2020)
18. Luo, B., Shen, J., Wang, Y., Pantic, M.: The ibug eye segmentation dataset. In: 2018 Imperial College Computing Student Workshop (ICCSW 2018). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2019)
19. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(Nov), 2579–2605 (2008)
20. Motiian, S., Jones, Q., Iranmanesh, S., Doretto, G.: Few-shot adversarial domain adaptation. In: Advances in Neural Information Processing Systems. pp. 6670–6680 (2017)
21. Naqvi, R.A., Loh, W.K.: Sclera-net: Accurate sclera segmentation in various sensor images based on residual encoder and decoder network. IEEE Access **7**, 98208–98227 (2019)
22. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
23. Perry, J., Fernandez, A.: Minenet: A dilated cnn for semantic segmentation of eye features. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 0–0 (2019)
24. 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)
25. Saito, K., Kim, D., Sclaroff, S., Darrell, T., Saenko, K.: Semi-supervised domain adaptation via minimax entropy. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 8050–8058 (2019)
26. Sankowski, W., Grabowski, K., Napieralska, M., Zubert, M., Napieralski, A.: Reliable algorithm for iris segmentation in eye image. Image and vision computing **28**(2), 231–237 (2010)
27. Shu, Y., Cao, Z., Long, M., Wang, J.: Transferable curriculum for weakly-supervised domain adaptation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 4951–4958 (2019)
28. Sun, R., Zhu, X., Wu, C., Huang, C., Shi, J., Ma, L.: Not all areas are equal: Transfer learning for semantic segmentation via hierarchical region selection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4360–4369 (2019)
29. Tang, Y., Wang, J., Gao, B., Dellandréa, E., Gaizauskas, R., Chen, L.: Large scale semi-supervised object detection using visual and semantic knowledge transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2119–2128 (2016)
30. 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)
31. Tsai, Y.H., Sohn, K., Schulter, S., Chandraker, M.: Domain adaptation for structured output via discriminative patch representations. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1456–1465 (2019)
32. 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)

33. Valindria, V.V., Lavdas, I., Bai, W., Kamnitsas, K., Aboagye, E.O., Rockall, A.G., Rueckert, D., Glocker, B.: Domain adaptation for mri organ segmentation using reverse classification accuracy. arXiv preprint arXiv:1806.00363 (2018)
34. Wang, C., He, Y., Liu, Y., He, Z., He, R., Sun, Z.: Sclerasegnet: an improved u-net model with attention for accurate sclera segmentation. In: IAPR International Conference on Biometrics. vol. 1 (2019)
35. Wang, J.G., Sung, E., Venkateswarlu, R.: Estimating the eye gaze from one eye. Computer Vision and Image Understanding **98**(1), 83–103 (2005)
36. Wang, M., Deng, W.: Deep visual domain adaptation: A survey. Neurocomputing **312**, 135–153 (2018)
37. Yao, T., Pan, Y., Ngo, C.W., Li, H., Mei, T.: Semi-supervised domain adaptation with subspace learning for visual recognition. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 2142–2150 (2015)
38. 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)
39. Zhou, Z., Du, E.Y., Thomas, N.L., Delp, E.J.: A new human identification method: Sclera recognition. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans **42**(3), 571–583 (2011)