

Duo-SegNet: Adversarial Dual-Views for Semi-Supervised Medical Image Segmentation

Himashi Peiris¹, Zhaolin Chen^{1,2}, Gary Egan², Mehrtash Harandi¹

¹ Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, Australia.

² Monash Biomedical Imaging (MBI), Monash University, Melbourne, Australia.
 {Edirisinghe.Peiris, Zhaolin.Chen, Gary.Egan,
 Mehrtash.Harandi}@monash.edu

Abstract. Segmentation of images is a long-standing challenge in medical AI. This is mainly due to the fact that training a neural network to perform image segmentation requires a significant number of pixel-level annotated data, which is often unavailable. To address this issue, we propose a semi-supervised image segmentation technique based on the concept of multi-view learning. In contrast to the previous art, we introduce an adversarial form of dual-view training and employ a critic to formulate the learning problem in multi-view training as a min-max problem. Thorough quantitative and qualitative evaluations on several datasets, indicate that our proposed method outperforms state-of-the-art medical image segmentation algorithms consistently and comfortably. The code is publicly available at <https://github.com/himashi92/Duo-SegNet>.

Keywords: Deep Learning · Semi-Supervised Learning · Medical Image Segmentation · Multi-View Learning · Adversarial Learning.

1 Introduction

In this paper, we propose a semi-supervised technique based on the concept of multi-view learning [18] to segment medical images. Accurate segmentation of medical images is a key-step in developing Computer-Aided Diagnosis (CAD) and automating various clinical tasks such as image-guided interventions.

The prevailing idea in medical image segmentation is to employ an encoder-decoder structure (*e.g.*, UNet [16] and its variants [20,12]) and formulate the problem as a dense classification/regression problem, depending on the type of input and the desired output. Most of the existing algorithms are breeze through supervised setting when sizable, annotated datasets are available. However, annotating large-scale datasets for image segmentation is challenging and expensive. On one hand, most medical image modalities are hefty in size (*e.g.*, 3D volumes as in MRI and CT) and hence annotation is extremely laborious. On the other hand, annotating medical images requires expert knowledge and cannot be crowd-sourced. Add to this the fact that medical images often contain low contrast slices and ambiguous regions, which in turn makes annotation very

difficult. While attaining large annotated datasets is challenging, unlabeled data comes (almost) for free and is abundant.

Multi-view learning makes use of multiple distinct views of data and benefits from the resulting relationships to achieve accurate models. The principle of consensus [18] states that by minimizing the disagreement on multiple distinct hypotheses, the error of each hypothesis will be minimized. To be specific, suppose h^1 and h^2 are two distinct hypothesis defined on a distribution. Under some mild assumptions and as shown in [5]

$$P(h^1 \neq h^2) \geq \max \{P_{\text{err}}(h^1), P_{\text{err}}(h^2)\} .$$

That is, reducing the disagreement of the two hypotheses minimizes the error of each hypothesis. The consensus principle provides an efficient way with strong theoretical properties to benefit from unlabeled data as shown for example in the celebrated work of Blum and Mitchell [4]. Except a handful of studies [14,13], multi-view learning has been mostly studied under the classification regime [8,7]. This raises a natural question, *is multi-view learning beneficial when it comes to segmentation?* and *if yes, how it must be formulated?* Our work takes a step in this direction and provides a way to cultivate unlabeled data for image segmentation. In particular, we propose a dual-view UNet model and equip it with a critic network. Each UNet provides a view of the data distribution and will be trained by minimizing a supervised loss on the labeled data and a disagreement loss over the unlabeled data. The critic is used to facilitate two objectives, **1.** to ensure that the output of UNets resembles the ground-truth segmentation masks (in the ideal case, the critic is not able to distinguish whether its input is a ground-truth mask or a prediction mask from the UNets), and **2.** to identify confident parts of a prediction mask to enforce agreement across the views. We should stress that unlike classification where (dis)agreement can be formulated readily between predictions, in segmentation, we face a dense prediction problem, meaning a prediction mask includes tens of thousands, if not more, predictions at pixel-level. A naive treatment of (dis)agreement loss could lead to inferior results as simply the networks can overfit to agree on the background, which is the dominant part of the prediction mask in many cases. Therefore, to train the Duo-SegNet, we formulate the learning as a min-max problem by allowing a critic to stand in as a quantitative subjective referee. Thorough empirical evaluations demonstrate that our method performs well both qualitatively and quantitatively, utilizing small fraction of labeled data.

In short, we have made the following contributions in this work, **1.** We propose a dual-view learning scheme for semi-supervised medical image segmentation. **2.** We make use of a critic to regularize the training and identify confident parts of a prediction mask towards learning from unlabelled data.

2 Methodology

We start this section by providing an overview of the Duo-SegNet (see Fig. 1 for a conceptual illustration). Let $\mathcal{X} = \{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_m, \mathbf{Y}_m)\}$ be a labeled

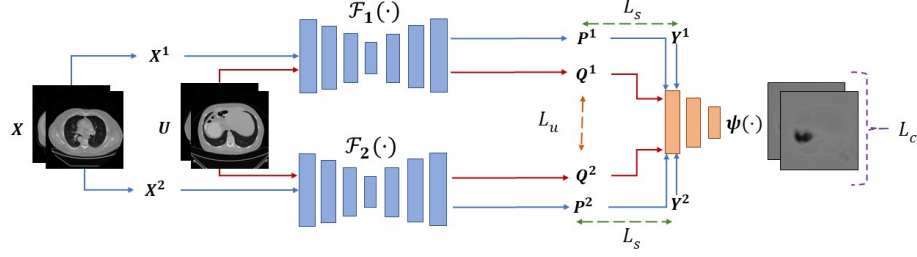


Fig. 1: Proposed Adversarial Dual View Network. $\{\mathcal{F}_i(\cdot)\}_{i=1}^2$ and $\psi(\cdot)$ denote Segmentation networks and Critic network. Here, Critic criticizes between prediction masks and the ground truth masks to perform the min-max game.

set, where each pair $(\mathbf{X}_i, \mathbf{Y}_i)$ consists of an image $\mathbf{X}_i \in \mathbb{R}^{C \times H \times W}$ and its associated ground-truth mask $\mathbf{Y}_i \in \{0, 1\}^{H \times W}$. Furthermore, let $\mathcal{U} = \{\mathbf{U}_i\}_{i=1}^n$, $\mathbf{U}_i \in \mathbb{R}^{C \times H \times W}$ be a set of n unlabelled images with $n \gg m$. The primary objective is to learn a segmentation model from $\mathcal{D} = \mathcal{X} \cup \mathcal{U}$.

The Duo-SegNet includes two basic modules, a dual view segmentation networks and a critic, shown by \mathcal{F}_1 , \mathcal{F}_2 , and ψ respectively in Fig. 1. Each leg in the dual-view network (*i.e.*, \mathcal{F}_1 and \mathcal{F}_2) benefits from an encoder-decoder design, and is structured as a UNet [16]. In this work, by incorporating a critic, we implicitly enforce segmentation networks to create predictions that are more similar to the desired masks holistically. The parameters of the model are the parameters of the dual-view network θ_1 , θ_2 and the critic network θ_C . We collectively show the parameters of the dual-view network (*i.e.*, θ_1 , θ_2) by Θ to avoid cluttering the equations. To train the model, we propose optimizing the following min-max problem:

$$\min_{\Theta} \max_{\theta_C} \mathcal{L}(\Theta; \mathcal{D}). \quad (1)$$

The familiar min-max problem in (1) encourages the dual-view segmentation networks to yield segmentation masks that look like realistic ones by deceiving the critic. Similar to standard multi-view training, our approach adapted dual view training to train two segmentation models collaboratively. We propose to train dual view segmentation networks by minimizing a multitask loss function consists of three loss terms:

$$\mathcal{L}(\Theta; \mathcal{D}) := \lambda_s \mathcal{L}_s(\Theta; \mathcal{X}) + \lambda_u \mathcal{L}_u(\Theta; \mathcal{U}) + \lambda_c \mathcal{L}_c(\Theta; \theta_C; \mathcal{D}), \quad (2)$$

where \mathcal{L}_s , \mathcal{L}_u , and \mathcal{L}_c denote the supervised, the unsupervised and the critic loss respectively. Furthermore, $\lambda_s, \lambda_u, \lambda_c > 0$ are hyper-parameters of the algorithm, controlling the contribution of each loss term. We note that the supervised and unsupervised loss are only dependent on the dual-view networks while the critic loss is defined based on the parameters of the whole model. In practice, we find that tuning the hyper-parameters of the network is not difficult at all and

the Duo-SegNet works robustly as long as these parameters are defined in a reasonable range. For example, in all our experiments in this paper, we set $\lambda_s = 1.0$, $\lambda_u = 0.3$ and $\lambda_c = 0.2$. To better understand Eq. (2), we start with the supervised loss. This loss makes use of the labeled data and can be defined as a cross-entropy (*i.e.*, Eq. (3)) or dice loss (*i.e.*, Eq. (4)) or even a linear combination of both, which is a common practice in segmentation [16].

$$\mathcal{L}_{ce}(\theta_i; \mathcal{X}) = \mathbb{E}_{(\mathbf{X}, \mathbf{Y}) \sim \mathcal{X}} \left[\left\langle \mathbf{Y}, \log(\mathcal{F}_i(\mathbf{X})) \right\rangle \right] \quad (3)$$

$$\mathcal{L}_{dice}(\theta_i; \mathcal{X}) = 1 - \mathbb{E}_{(\mathbf{X}, \mathbf{Y}) \sim \mathcal{X}} \left[\frac{2 \langle \mathbf{Y}, \mathcal{F}_i(\mathbf{X}) \rangle}{\|\mathbf{Y}\|_1 + \|\mathcal{F}_i(\mathbf{X})\|_1} \right], \quad (4)$$

where we use $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j} \mathbf{A}[i,j] \mathbf{B}[i,j]$ and $\|\mathbf{A}\|_1 = \sum_{i,j} |\mathbf{A}[i,j]|$.

We define the unsupervised loss as a means to realize the principle of consensus as;

$$\mathcal{L}_u(\Theta; \mathcal{U}) := \mathbb{E}_{\mathbf{U} \sim \mathcal{U}} \left[\left\langle \mathcal{F}_1(\mathbf{U}), \log(\mathcal{F}_2(\mathbf{U})) \right\rangle + \left\langle \mathcal{F}_2(\mathbf{U}), \log(\mathcal{F}_1(\mathbf{U})) \right\rangle \right]. \quad (5)$$

We identify Eq.(5) as a symmetric form of the cross-entropy loss. In essence, the unsupervised loss \mathcal{L}_u acts as an agreement loss during dual-view training. The idea is that the two segmentation networks should generate similar segmentation masks for unlabeled data. In the proposed method, diversity among two views are retrieved based on segmentation model predictions. Unlike in [13,14] where adversarial examples for a model are used to teach other model in the ensemble, we make use of a critic to approach inter diversity among segmentation models for both labeled and unlabeled distributions. When using the labeled data, two segmentation networks are supervised by both the supervised loss with the ground truth and the adversarial loss. We denote the functionality of the critic by $\Psi : [0, 1]^{H \times W} \rightarrow [0, 1]^{H \times W}$ and define the normalized loss of critic for labeled prediction distribution as:

$$\begin{aligned} \mathcal{L}_{adv1}(\Theta; \theta_C; \mathcal{X}) := & \mathbb{E}_{(\mathbf{X}, \mathbf{Y}) \sim \mathcal{X}} \left[- \sum_{a \in H} \sum_{b \in W} \left\{ (1 - \eta) \log \left(\psi(\mathbf{Y})[a, b] \right) \right. \right. \\ & \left. \left. + \eta \log \left(1 - \psi(\mathcal{F}_i(\mathbf{X}))[a, b] \right) \right\} \right], \end{aligned} \quad (6)$$

where $\eta = 0$ if the sample is generated by the segmentation network, and $\eta = 1$ if the sample is drawn from the ground truth labels. For unlabeled data, it is obvious that we cannot apply any supervised loss since there is no ground truth annotation available. However, adversarial loss can be applied as it only requires the knowledge about whether the mask is from the ground-truth labels or generated by the segmentation networks. The adversarial loss for the distribution of unlabeled predictions is defined as:

$$\mathcal{L}_{adv2}(\Theta; \theta_C; \mathcal{U}) := \mathbb{E}_{\mathbf{U} \sim \mathcal{U}} \left[- \sum_{a \in H} \sum_{b \in W} \left\{ \log \left(1 - \psi(\mathcal{F}_i(\mathbf{U}))[a, b] \right) \right\} \right], \quad (7)$$

Remark 1. The basis of having a critic is that a well trained critic can produce pixel-wise uncertainty map/confidence map which imposes a higher-order consistency measure of prediction and ground truth. So, it infers the pixels where the prediction masks are close enough to the ground truth distribution.

Therefore, the critic loss is defined as:

$$\mathcal{L}_c(\Theta; \theta_C; \mathcal{D}) = \mathcal{L}_{\text{adv1}}(\Theta; \theta_C; \mathcal{X}) + \mathcal{L}_{\text{adv2}}(\Theta; \theta_C; \mathcal{U}) . \quad (8)$$

With this aggregated adversarial loss, we co-train two segmentation networks to fool the critic by maximizing the confidence of the predicted segmentation being generated from the ground truth distribution. To train the critic, we use labeled prediction distribution along with given ground truth segmentation masks and calculate the normalized loss as in Eq.(6). The overall min-max optimization process is summarized in Algorithm 1.

Algorithm 1: Duo-SegNet (*training*)

Input: Define Segmentation networks $\{\mathcal{F}_i(\cdot)\}_{i=1}^2$, critic $\psi(\cdot)$, batch size \mathcal{B} , maximum epoch E_{max} , number of steps k_s and k_c for segmentation networks and critic, Labeled images $\mathcal{X} = \{(X_1, Y_1), \dots, (X_m, Y_m)\}$, Unlabeled images $\mathcal{U} = \{U_1, \dots, U_n\}$ and two labeled sets $\mathcal{X}^1; \mathcal{X}^2 \subset \mathcal{X}$;

Output: Network Parameters $\{\theta_i\}_{i=1}^2$ and θ_C ;

Initialize Network Parameters $\{\theta_i\}_{i=1}^2$ and θ_C ;

```

for  $epoch = 1, \dots, E_{max}$  do
  for  $batch = 1, \dots, \mathcal{B}$  do
    for  $k_s$  steps do
      Generate predictions for labeled data  $\mathcal{F}_1(x)$  for all  $x_i \in \mathcal{X}^1$ ,  $\mathcal{F}_2(x)$  for all  $x_i \in \mathcal{X}^2$  and for unlabeled data  $\mathcal{F}_1(x)$  and  $\mathcal{F}_2(x)$  for all  $U_i \in \mathcal{U}$ ;
      Generate confidence maps for all predictions using  $\psi(\cdot)$ ;
      Let  $\mathcal{L} = \mathcal{L}_s + \mathcal{L}_u + \mathcal{L}_c$ , as defined in Equations. (2) - (8);
      Update  $\{\theta_i\}_{i=1}^2$  by descending its stochastic gradient on  $\mathcal{L}$ ;
    end
    for  $k_c$  steps do
      Generate confidence maps for all labeled predictions and ground truth masks using  $\psi(\cdot)$ ;
      Let  $\mathcal{L}_c = \mathcal{L}_{adv1}$ , as defined in 6;
      Update  $\theta_C$  by ascending its stochastic gradient on  $\mathcal{L}_c$ ;
    end
  end
end

```

3 Related Work

We begin by briefly discussing alternative approaches to semi-supervised learning, with a focus on the most relevant ones to our approach, namely pseudo

labelling [9], mean teacher model [17], Virtual Adversarial Training (VAT) [11], recently published deep co-training [13]. Our goal here is to discuss the resemblance and differences between our proposed approach and some of the methods that have been adopted for semi-supervised image segmentation.

In Pseudo Labelling, as the name implies, the model uses the predicted labels of the unlabeled data and treat it as ground-truth labels for training. The drawback here is, sometimes incorrect pseudo labels may diminish the generalization performance and weaken the training of deep neural networks. In contrast, Duo-SegNet by incorporating a critic, increases the tolerance of these incorrect pseudo labels which stabilizes the generalization performance. Similar to Duo-SegNet, the mean teacher approach benefits from two neural networks, namely the teacher and the student networks. While the student model is trained in a stochastic manner, the parameters of the teacher model are updated slowly by a form of moving averaging of the student’s parameters. This, in turn, results in better robustness to prediction error as one could hope that averaging attenuates the effect of noisy gradient (as a result of incorrect pseudo labels). Unlike mean teacher model, Duo-SegNet simultaneously train both networks and models can learn from one another during training. VAT can be understood as an effective regularization method which optimizes generalization power of a model for unlabeled examples. This is achieved by generating adversarial perturbations to the input of the model, followed by making the model robust to the adversarial perturbations. In contrast to VAT, Duo-SegNet makes use of a critic to judge if predictions are from the same or different distribution compared to labeled examples. With this, segmentation networks are encouraged to generate similar predictive distribution for both labeled and unlabeled data. Similar to our work, in co-training two models are alternately trained on distinct views, while learning from each other is encouraged. Recently, Peng *et al.* introduced a deep co-training method for semi-supervised image segmentation task [13] based on the approach presented by Qiao *et al.* for image recognition task in [14]. In this approach diversity among views are achieved via adversarial examples following VAT [11]. We note that adversarial examples, from a theoretical point of view, cannot guarantee diversity, especially when unlabeled data is considered. That is, a wrong prediction can intensify even more once adversarial examples are constructed from it. In contrast, inspired by Generative Adversarial Networks (GANs) [6] and GAN based medical imaging applications including medical image segmentation [10], reconstruction [15] and domain adaptation [19], our proposed method graciously encloses the min-max formulation in dual-view learning for segmenting medical images where high-confidence predictions for unlabeled data are leveraged, which is simple and effective.

4 Experiments

Implementation Details: The proposed model is developed in PyTorch [3]. Training was done from scratch without using any pre-trained model weights. For training of segmentation network and critic, we use SGD optimizer (LR=1e-

02) and RMSProp optimizer(LR=5e-05), respectively. We divide the original dataset into training (80%) and test set (20%). Experiments were conducted for 5%, 20% and 50% of labeled training sets.

Datasets: We use three medical image datasets for model evaluation covering three medical image modalities : 670 Fluorescence Microscopy (FM) images from Nuclei [1], 20 MRI volumes from Heart [2] and 41 CT volumes from Spleen [2]. For our experiments, 2D images are obtained by slicing the high-resolution MRI and CT volumes for Heart (2271 slices) and Spleen (3650 slices) datasets. Each slice is then resized to a resolution of 256×256 .

Competing Methods and Evaluation Metrics: We compare our proposed method with Mean Teacher [17], Pseudo Labelling [9], VAT [11], Deep Co-training [13] and fully supervised U-Net [16]. For all baselines, we follow the same configurations as for our method. All approaches are evaluated using : 1) Dice Sørensen coefficient (DSC) and 2) Mean Absolute Error (MAE).

Performance Comparison: The qualitative results for the proposed and competing methods are shown in Fig. 3. The quantitative results comparison of the proposed method to the four state-of-the-art methods are shown in Table 1. The results reveal that the proposed method comfortably outperform other studied methods for smaller fractions of annotated data (*e.g.* Spleen 5%). The gap between the Duo-SegNet and other competitors decreases on the Nuclei dataset, when the amount of labeled data increases. That said, we can still observe a significant improvement on the Heart and Spleen dataset. The proposed network can produce both accurate prediction masks and confidence maps representing which regions of the prediction distribution are close to the ground truth label distribution. This is useful when training unlabeled data. Fig. 2 shows the visual analysis of confidence maps.

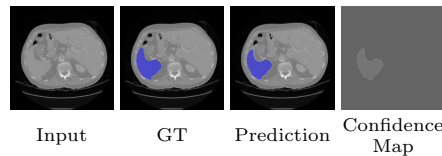


Fig. 2: Visual Analysis of Confidence Map generated by the Critic during training

Ablation Study: We also perform ablation studies to show the effectiveness of adding a critic in dual-view learning in semi-supervised setting and the importance of dual view network structure. In our algorithm, we benefit from unlabeled data via (1) criss-cross exchange of confident regions, (2) improving the critic

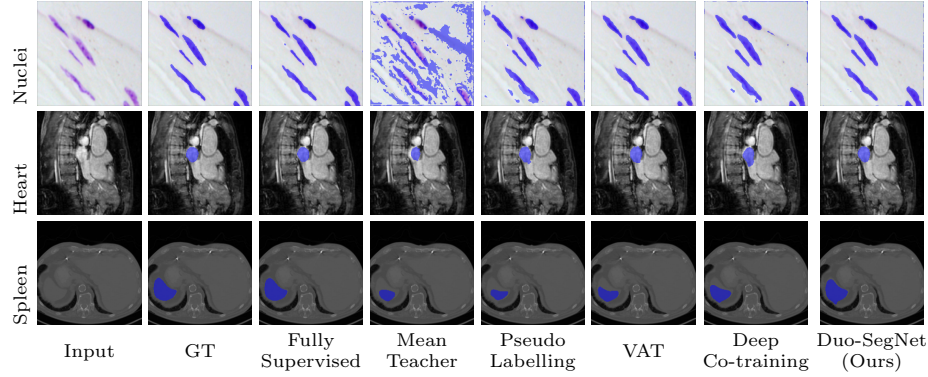


Fig. 3: **Visual comparison of our method with state-of-the-art models.** Segmentation results are shown for 5% of labeled training data.

Table 1: Comparison with state-of-the-art methods.

Dataset	Method	DSC			MAE		
Nuclei	Fully Supervised	91.36			2.25		
		$l_a = 5\%$	$l_a = 20\%$	$l_a = 50\%$	$l_a = 5\%$	$l_a = 20\%$	$l_a = 50\%$
	Mean Teacher	83.78	84.92	87.99	4.78	4.30	3.36
	Pseudo Labeling	60.90	72.46	85.91	8.40	6.37	3.84
	VAT	85.24	86.43	88.45	4.09	3.77	3.26
	Deep Co-training	85.83	87.15	89.20	4.08	3.80	3.08
	Duo-SegNet	87.14	87.83	89.28	3.57	3.43	3.03
Heart	Fully Supervised	97.17			0.02		
		$l_a = 5\%$	$l_a = 20\%$	$l_a = 50\%$	$l_a = 5\%$	$l_a = 20\%$	$l_a = 50\%$
	Mean Teacher	71.00	87.59	93.43	0.22	0.09	0.05
	Pseudo Labeling	65.92	79.86	80.75	0.20	0.13	0.13
	VAT	85.33	91.60	94.83	0.11	0.06	0.04
	Deep Co-training	85.96	91.54	94.55	0.10	0.06	0.04
	Duo-SegNet	86.79	93.21	95.56	0.10	0.05	0.03
Spleen	Fully Supervised	97.89			0.02		
		$l_a = 5\%$	$l_a = 20\%$	$l_a = 50\%$	$l_a = 5\%$	$l_a = 20\%$	$l_a = 50\%$
	Mean Teacher	75.44	90.76	92.98	0.20	0.08	0.06
	Pseudo Labeling	67.70	68.81	84.81	0.24	0.21	0.12
	VAT	78.31	91.37	94.34	0.19	0.07	0.05
	Deep Co-training	79.16	89.65	94.90	0.16	0.09	0.05
	Duo-SegNet	88.02	92.19	96.03	0.10	0.07	0.03

which in essence minimizes an upper-bound of error. To justify this, we conducted an additional experiment without unlabeled data. It can be seen that there is an impact in the performance of segmentation model for varying values for λ_u . For our experiments we choose λ_u in the range of 0.3 to 0.4. All experiments in Table 2 are conducted for spleen dataset with 5% of annotated data.

Table 2: Ablation Study

(a) Network Structure Analysis.

Experiment	DSC	MAE
Duo-SegNet	88.02	0.10
w/o Critic	77.69	0.19
w/o Unlabeled Data	76.67	0.17
One Segmentation Network	82.44	0.16

(b) Hyper-parameter Analysis for λ_u .

λ_u	0.1	0.2	0.3	0.4	0.5
DSC	83.58	85.62	88.02	87.14	78.89
MAE	0.15	0.12	0.10	0.11	0.20

5 Conclusion

We proposed an adversarial dual-view learning approach for semi-supervised medical image segmentation and demonstrated its effectiveness on publicly available three medical datasets. Our extensive experiments showed that employing a min-max paradigm into multi-view learning scheme sharpens boundaries between different regions in prediction masks and yield a performance close to full-supervision with limited annotations. The dual view training can still be improved by self-tuning mechanisms, which will be considered in our future works.

References

1. 2018 data science bowl. <https://www.kaggle.com/c/data-science-bowl-2018> (accessed: 2021-02-14)
2. Medical segmentation decathlon. <http://medicaldecathlon.com/> (accessed: 2021-02-14)
3. Pytorch. <https://pytorch.org/> (accessed: 2021-07-05)
4. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proc. of the eleventh annual conference on Computational learning theory. pp. 92–100 (1998)
5. Dasgupta, S., Littman, M.L., McAllester, D.: Pac generalization bounds for co-training. In: Proc. Advances in Neural Information Processing Systems (NeurIPS). pp. 375–382 (2002)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
7. Kiritchenko, S., Matwin, S.: Email classification with co-training. In: Proceedings of the 2001 conference of the Centre for Advanced Studies on Collaborative research. p. 8. Citeseer (2001)
8. Kumar, A., Daumé, H.: A co-training approach for multi-view spectral clustering. In: Proceedings of the 28th international conference on machine learning (ICML-11). pp. 393–400 (2011)
9. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks

10. Mahmood, F., Borders, D., Chen, R., McKay, G.N., Salimian, K.J., Baras, A., Durr, N.J.: Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Trans. on Medical Imaging* (2019)
11. Miyato, T., Maeda, S.i., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **41**(8), 1979–1993 (2018)
12. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* (2018)
13. Peng, J., Estrada, G., Pedersoli, M., Desrosiers, C.: Deep co-training for semi-supervised image segmentation. *Pattern Recognition* **107**, 107269 (2020)
14. Qiao, S., Shen, W., Zhang, Z., Wang, B., Yuille, A.: Deep co-training for semi-supervised image recognition. In: *Proc. European Conference on Computer Vision (ECCV)*. pp. 135–152 (2018)
15. Quan, T.M., Nguyen-Duc, T., Jeong, W.K.: Compressed sensing mri reconstruction with cyclic loss in generative adversarial networks. *arXiv preprint arXiv:1709.00753* (2017)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Proc. Int. Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. pp. 234–241. Springer (2015)
17. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: *Proc. Advances in Neural Information Processing Systems (NeurIPS)*. pp. 1195–1204 (2017)
18. Xu, C., Tao, D., Xu, C.: A survey on multi-view learning. *arXiv preprint arXiv:1304.5634* (2013)
19. Zhang, Y., Miao, S., Mansi, T., Liao, R.: Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 599–607. Springer (2018)
20. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11. Springer (2018)