Cross-Adversarial Local Distribution Regularization for Semi-supervised Medical Image Segmentation

Thanh Nguyen-Duc[1(✉)], Trung Le[1], Roland Bammer[1], He Zhao[2], Jianfei Cai[1], and Dinh Phung[1]

[1] Monash University, Melbourne, Australia
{thanh.nguyen4,trunglm,roland.bammer,jianfei.cai,dinh.phung}@monash.edu
[2] CSIRO's Data61, Melbourne, Australia
he.zhao@ieee.org

Abstract. Medical semi-supervised segmentation is a technique where a model is trained to segment objects of interest in medical images with limited annotated data. Existing semi-supervised segmentation methods are usually based on the smoothness assumption. This assumption implies that the model output distributions of two similar data samples are encouraged to be invariant. In other words, the smoothness assumption states that similar samples (e.g., adding small perturbations to an image) should have similar outputs. In this paper, we introduce a novel cross-adversarial local distribution (Cross-ALD) regularization to further enhance the smoothness assumption for semi-supervised medical image segmentation task. We conducted comprehensive experiments that the Cross-ALD archives state-of-the-art performance against many recent methods on the public LA and ACDC datasets.

Keywords: Semi-supervised segmentation · Adversarial local distribution · Adversarial examples · Cross-adversarial local distribution

1 Introduction

Medical image segmentation is a critical task in computer-aided diagnosis and treatment planning. It involves the delineation of anatomical structures or pathological regions in medical images, such as magnetic resonance imaging (MRI) or computed tomography (CT) scans. Accurate and efficient segmentation is essential for various medical applications, including tumor detection, surgical planning, and monitoring disease progression. However, manual medical imaging annotation is time-consuming and expensive because it requires the domain

knowledge from medical experts. Therefore, there is a growing interest in developing semi-supervised learning that leverages both labeled and unlabeled data to improve the performance of image segmentation models [16,27].

Existing semi-supervised segmentation methods exploit smoothness assumption, e.g., the data samples that are closer to each other are more likely to to have the same label. In other words, the smoothness assumption encourages the model to generate invariant outputs under small perturbations. We have seen such perturbations being be added to natural input images at data-level [4,9,14,19,21], feature-level [6,17,23,25], and model-level [8,11,12,24,28]. Among them, virtual adversarial training (VAT) [14] is a well-known one which promotes the smoothness of the local output distribution using adversarial examples. The adversarial examples are near decision boundaries generated by adding adversarial perturbations to natural inputs. However, VAT can only create one adversarial sample in a run, which is often insufficient to completely explore the space of possible perturbations (see Sect. 2.1). In addition, the adversarial examples of VAT can also lie together and lose diversity that significantly reduces the quality of adversarial examples [15,20]. Mixup regularization [29] is a data augmentation method used in deep learning to improve model generalization. The idea behind mixup is to create new training examples by linearly interpolating between pairs of existing examples and their corresponding labels, which has been adopted in [2,3,19] to semi-supervised learning. The work [5] suggests that Mixup improves the smoothness of the neural function by bounding the Lipschitz constant of the gradient function of the neural networks. However, we show that mixing between more informative samples (e.g., adversarial examples near decision boundaries) can lead to a better performance enhancement compared to mixing natural samples (see Sect. 3.3).

In this paper, we propose a novel cross-adversarial local distribution regularization for semi-supervised medical image segmentation for smoothness assumption enhancement[1]. Our contributions are summarized as follows: **1)** To overcome the VAT's drawback, we formulate an adversarial local distribution (ALD) with Dice loss function that covers all possible adversarial examples within a ball constraint. **2)** To enhance smoothness assumption, we propose a novel cross-adversarial local distribution regularization (Cross-ALD) to encourage the smoothness assumption, which is a random mixing between two ALDs. **3)** We also propose a sufficiently approximation for the Cross-ALD by a multiple particle-based search using semantic feature Stein Variational Gradient Decent (SVGDF), an enhancement of the vanilla SVGD [10]. **4)** We conduct comprehensive experiments on ADCD [1] and LA [26] datasets, showing that our Cross-ALD regularization achieves state-of-the-art performance against existing solutions [8,11,12,14,21,22,28].

---

[1] The Cross-ALD implementation in https://github.com/PotatoThanh/Cross-adversarial-local-distribution-regularization.

2   Method

In this section, we begin by reviewing the minimax optimization problem of virtual adversarial training (VAT) [14]. Given an input, we then formulate a novel adversarial local distribution (ALD) with Dice loss, which benefits the medical semi-supervised image segmentation problem specifically. Next, a cross-adversarial local distribution (Cross-ALD) is constructed by randomly combining two ALDs. We approximate the ALD by a particle-based method named semantic feature Stein Variational Gradient Descent (SVGDF). Considering the resolution of medical images are usually high, we enhance the vanilla SVGD [10] from data-level to feature-level, which is named SVGDF. We finally provide our regularization loss for semi-supervised medical image segmentation.

2.1   The Minimax Optimization of VAT

Let $\mathbb{D}_l$ and $\mathbb{D}_{ul}$ be the labeled and unlabeled dataset, respectively, with $P_{\mathbb{D}_l}$ and $P_{\mathbb{D}_{ul}}$ being the corresponding data distribution. Denote $\boldsymbol{x} \in \mathbb{R}^d$ as our d-dimensional input in a space $\boldsymbol{X}$. The labeled image $\boldsymbol{x}_l$ and segmentation ground-truth $\boldsymbol{y}$ are sampled from the labeled dataset $\mathbb{D}_l$ ($\boldsymbol{x}_l, \boldsymbol{y} \sim P_{\mathbb{D}_l}$), and the unlabeled image sampled from $\mathbb{D}_{ul}$ is $\boldsymbol{x} \sim P_{\mathbb{D}_{ul}}$.

Given an input $\boldsymbol{x} \sim P_{\mathbb{D}_{ul}}$ (i.e., the unlabeled data distribution), let us denote the ball constraint around the image $\boldsymbol{x}$ as $C_\epsilon(\boldsymbol{x}) = \{\boldsymbol{x}' \in \boldsymbol{X} : ||\boldsymbol{x}' - \boldsymbol{x}||_p \leq \epsilon\}$, where ϵ is a ball constraint radius with respect to a norm $|| \cdot ||_p$, and $\boldsymbol{x}'$ is an adversarial example[2]. Given that f_θ is our model parameterized by θ, VAT [14] trains the model with the loss of ℓ_{vat} that a minimax optimization problem:

$$\ell_{vat} := \min_\theta \mathbb{E}_{\boldsymbol{x} \sim P_{\mathbb{D}_{ul}}} \Big[\max_{\boldsymbol{x}' \in C_\epsilon(\boldsymbol{x})} D_{\mathrm{KL}}(f_\theta(\boldsymbol{x}'), f_\theta(\boldsymbol{x})) \Big], \tag{1}$$

where D_{KL} is the Kullback-Leibler divergence. The inner *maximization problem* is to find an adversarial example near decision boundaries, while the *minimization problem* enforces the local smoothness of the model. However, VAT is insufficient to explore the set of of all adversarial examples within the constraint C_ϵ because it only find one adversarial example $\boldsymbol{x}'$ given a natural input $\boldsymbol{x}$. Moreover, the works [15,20] show that even solving the *maximization problem* with random initialization, its solutions can also lie together and lose diversity, which significantly reduces the quality of adversarial examples.

2.2   Adversarial Local Distribution

In order to overcome the drawback of VAT, we introduce our proposed adversarial local distribution (ALD) with Dice loss function instead of D_{KL} in [14,15]. ALD forms a set of all adversarial examples $\boldsymbol{x}'$ within the ball constraint given

---

[2] A sample generated by adding perturbations toward the adversarial direction.

an input $\boldsymbol{x}$. Therefore, the distribution can helps to sufficiently explore all possible adversarial examples. The adversarial local distribution $P_\theta(\boldsymbol{x}'|\boldsymbol{x})$ is defined with a ball constraint C_ϵ as follow:

$$P_\theta(\boldsymbol{x}'|\boldsymbol{x}) := \frac{e^{\ell_{Dice}(\boldsymbol{x}',\boldsymbol{x};\theta)}}{\int_{C_\epsilon(\boldsymbol{x})} e^{\ell_{Dice}(\boldsymbol{x}'',\boldsymbol{x};\theta)} d\boldsymbol{x}''} = \frac{e^{\ell_{Dice}(\boldsymbol{x}',\boldsymbol{x};\theta)}}{Z(\boldsymbol{x};\theta)}, \tag{2}$$

where $P_\theta(\cdot|\boldsymbol{x})$ is the conditional local distribution, and $Z(\boldsymbol{x};\theta)$ is a normalization function. The ℓ_{Dice} is the Dice loss function as shown in Eq. 3

$$\ell_{Dice}(\boldsymbol{x}',\boldsymbol{x};\theta) = \frac{1}{C} \sum_{c=1}^{C} [1 - \frac{2||p_\theta(\hat{\boldsymbol{y}}_c|\boldsymbol{x}) \cap p_\theta(\tilde{\boldsymbol{y}}_c|\boldsymbol{x}')||}{||p_\theta(\hat{\boldsymbol{y}}_c|\boldsymbol{x}) + p_\theta(\tilde{\boldsymbol{y}}_c|\boldsymbol{x}')||}], \tag{3}$$

where C is the number of classes. $p_\theta(\hat{\boldsymbol{y}}_c|\boldsymbol{x})$ and $p_\theta(\tilde{\boldsymbol{y}}_c|\boldsymbol{x}')$ are the predictions of input image $\boldsymbol{x}$ and adversarial image $\boldsymbol{x}'$, respectively.

2.3   Cross-Adversarial Distribution Regularization

Given two random samples $\boldsymbol{x}_i, \boldsymbol{x}_j \sim P_\mathbb{D}$ $(i \neq j)$, we define the cross-adversarial distribution (Cross-ALD) denoted $\tilde{P}_\theta$ as shown in Eq. 4

$$\tilde{P}_\theta(\cdot|\boldsymbol{x}_i,\boldsymbol{x}_j) = \gamma P_\theta(\cdot|\boldsymbol{x}_i) + (1-\gamma)P_\theta(\cdot|\boldsymbol{x}_j) \tag{4}$$

where $\gamma \sim \text{Beta}(\alpha,\alpha)$ for $\alpha \in (0,\infty)$, inspired by [29]. The $\tilde{P}_\theta$ is the Cross-ALD distribution, a mixture between the two adversarial local distributions.
Given Eq. 4, we propose the Cross-ALD regularization at two random input images $\boldsymbol{x}_i, \boldsymbol{x}_j \sim P_\mathbb{D}$ $(i \neq j)$ as

$$R(\theta,\boldsymbol{x}_i,\boldsymbol{x}_j) := \mathbb{E}_{\tilde{\boldsymbol{x}}' \sim \tilde{P}_\theta(\cdot|\boldsymbol{x}_i,\boldsymbol{x}_j)}[\log \tilde{P}_\theta(\tilde{\boldsymbol{x}}'|\boldsymbol{x}_i,\boldsymbol{x}_j)] = -H(\tilde{P}_\theta(\cdot|\boldsymbol{x}_i,\boldsymbol{x}_j)), \tag{5}$$

where H indicates the entropy of a given distribution.

When minimizing $R(\theta,\boldsymbol{x}_i,\boldsymbol{x}_j)$ or equivalently $-H(P_\theta(\cdot|\boldsymbol{x}_i,\boldsymbol{x}_j))$ w.r.t. θ, we encourage $P_\theta(\cdot|\boldsymbol{x}_i,\boldsymbol{x}_j)$ to be closer to a uniform distribution. This implies that the outputs of $f(\tilde{\boldsymbol{x}}') = f(\tilde{\boldsymbol{x}}'') = $ a constant c, where $\tilde{\boldsymbol{x}}', \tilde{\boldsymbol{x}}'' \sim \tilde{P}_\theta(\cdot|\boldsymbol{x}_i,\boldsymbol{x}_j)$. In other words, we encourages the invariant model outputs under small perturbations. Therefore, minimizing the Cross-ALD regularization loss leads to an enhancement in the model smoothness. While VAT only enforces local smoothness using one adversarial example, Cross-ALD further encourages smoothness of both local and mixed adversarial distributions to improve the model generalization.

2.4   Multiple Particle-Based Search to Approximate the Cross-ALD Regularization

In Eq. 2, the normalization $Z(\boldsymbol{x};\theta)$ in denominator term is intractable to find. Therefore, we propose a multiple particle-based search method named SVGDF

to sample $\boldsymbol{x}'^{(1)}, \boldsymbol{x}'^{(2)}, \ldots, \boldsymbol{x}'^{(N)} \sim P_\theta(\cdot|\boldsymbol{x})$. N is the number of samples (or *adversarial particles*). SVGDF is used to solve the optimization problem of finding a target distribution $P_\theta(\cdot|\boldsymbol{x})$. SVGDF is a particle-based Bayesian inference algorithm that seeks a set of points (or particles) to approximate the target distribution without explicit parametric assumptions using iterative gradient-based updates. Specifically, a set of adversarial particles $(\boldsymbol{x}'^{(n)})$ is initialized by adding uniform noises, then projected onto the ball C_ϵ. These adversarial particles are then iteratively updated using a closed-form solution (Eq. 6) until reaching termination conditions (, number of iterations).

$$\boldsymbol{x}'^{(n),(l+1)} = \prod_{C_\epsilon} \Big(\boldsymbol{x}'^{(n),(l)} + \tau * \big(\phi(\boldsymbol{x}'^{(n),(l)}) \big) \Big)$$

$$s.t. \ \phi(\boldsymbol{x}') = \frac{1}{N} \sum_{j=1}^{N} [k(\Phi(\boldsymbol{x}'^{(j),(l)}), \Phi(\boldsymbol{x}')) \nabla_{\boldsymbol{x}'^{(j),(l)}} \log P(\boldsymbol{x}'^{(j),(l)}|\boldsymbol{x}) \quad (6)$$

$$+ \nabla_{\boldsymbol{x}^{(j),(l)}} k(\Phi(\boldsymbol{x}'^{(j),(l)}), \Phi(\boldsymbol{x}'))],$$

where $\boldsymbol{x}'^{(n),(l)}$ is a nth adversarial particle at l^{th} iteration ($n \in \{1, 2, ..., N\}$, and $l \in \{1, 2, ..., L\}$ with the maximum number of iteration L). $\prod_{C_\epsilon}$ is projection operator to the C_ϵ constraint. τ is the step size updating. k is the radial basis function (RBF) kernel $k(\boldsymbol{x}', \boldsymbol{x}) = \exp\left\{ \frac{-||\boldsymbol{x}'-\boldsymbol{x}||^2}{2\sigma^2} \right\}$. Φ is a fixed feature extractor (e.g., encoder of U-Net/V-Net). While vanilla SVGD [10] is difficult to capture semantic meaning of high-resolution data because of calculating RBF kernel (k) directly on the data-level, we use the feature extractor Φ as a semantic transformation to further enhance the SVGD algorithm performance for medical imaging. Moreover, the two terms of ϕ in Eq. 6 have different roles: (i) the first one encourages the adversarial particles to move towards the high density areas of $P_\theta(\cdot|\boldsymbol{x})$ and (ii) the second one prevents all the particles from collapsing into the local modes of $P_\theta(\cdot|\boldsymbol{x})$ to enhance diversity (e.g.,pushing the particles away from each other). Please refer to the Cross-ALD Github repository for more details.

SVGDF approximates $P_\theta(\cdot|\boldsymbol{x}_i)$ and $P_\theta(\cdot|\boldsymbol{x}_j)$ in Eq. 4, where $\boldsymbol{x}_i, \boldsymbol{x}_j \sim P_{\mathbb{D}_{ul}}$ ($i \neq j$). We form sets of adversarial particles as $\mathbb{D}_{adv}|\boldsymbol{x}_i = \{ \boldsymbol{x}_i'^{(1)}, \boldsymbol{x}_i'^{(2)}, \ldots, \boldsymbol{x}_i'^{(N)}\}$ and $\mathbb{D}_{adv}|\boldsymbol{x}_j = \{\boldsymbol{x}_j'^{(1)}, \boldsymbol{x}_j'^{(2)}, \ldots, \boldsymbol{x}_j'^{(N)}\}$. The problem (5) can then be relaxed to

$$R(\theta, \boldsymbol{x}_i, \boldsymbol{x}_j) := \mathbb{E}_{\boldsymbol{x}_i'^{(n)} \sim P_{\mathbb{D}_{adv}|\boldsymbol{x}_i}, \boldsymbol{x}_j'^{(m)} \sim P_{\mathbb{D}_{adv}|\boldsymbol{x}_j}} \left[\ell_{Dice}(\tilde{x}', \tilde{x}; \theta) \right]$$

$$s.t. : \tilde{x}' = \gamma \boldsymbol{x}_i'^{(n)} + (1-\gamma)\boldsymbol{x}_j'^{(m)}; \ \tilde{x} = \gamma \boldsymbol{x}_i + (1-\gamma)\boldsymbol{x}_j, \quad (7)$$

where $\gamma \sim \text{Beta}(\alpha, \alpha)$ for $\alpha \in (0, \infty)$.

2.5   Cross-ALD Regularization Loss in Medical Semi-supervised Image Segmentation

In this paper, the overall loss function ℓ_{total} consists of three loss terms. The first term is the dice loss, where labeled image $\boldsymbol{x}_l$ and segmentation ground-truth $\boldsymbol{y}$

are sampled from labeled dataset $\mathbb{D}_l$. The second term is a contrastive learning loss for inter-class separation ℓ_{cs} proposed by [21]. The third term is our Cross-ALD regularization, which is an enhancement of ℓ_{vat} to significantly improve the model performance.

$$\ell_{total} := \min_{\theta} \ \mathbb{E}_{(\boldsymbol{x}_l, \boldsymbol{y}) \sim P_{\mathbb{D}_l}} \left[l_{Dice}(\boldsymbol{x}_l, \boldsymbol{y}; \theta) \right] + \lambda_{cs} \ \mathbb{E}_{\boldsymbol{x}_l \sim P_{\mathbb{D}_l}, \boldsymbol{x} \sim P_{\mathbb{D}_{ul}}} \left[\ell_{cs}(\boldsymbol{x}_l, \boldsymbol{x}) \right]$$
$$+ \lambda_{Cross-ALD} \ \mathbb{E}_{(\boldsymbol{x}_i, \boldsymbol{x}_j) \sim P_{\mathbb{D}_{ul}}} \left[R(\theta, \boldsymbol{x}_i, \boldsymbol{x}_j) \right], \tag{8}$$

where λ_{cs} and $\lambda_{Cross-ALD}$ are the corresponding weights to balance the losses. Note that our implementation is replacing ℓ_{vat} loss with the proposed Cross-AD regularization in SS-Net code repository[3] [21] to reach the state-of-the-art performance.

3   Experiments

In this section, we conduct several comprehensive experiments using the ACDC[4] dataset [1] and the LA[5] dataset [26] for 2D and 3D image segmentation tasks, respectively. For fair comparisons, all experiments are conducted using the identical setting, following [21]. We evaluate our model in challenging semi-supervised scenarios, where only 5% and 10% of the data are labeled and the remaining data in the training set is treated as unlabeled. The Cross-ALD uses the U-Net [18] and V-Net [13] architectures for the ACDC and LA dataset, respectively. We compare the diversity between the adversarial particles generated by our method against vanilla SVGD and VAT with random initialization in Sect. 3.1 . We then illustrate the Cross-AD outperforms other recent methods on ACDC and LA datasets in Sect. 3.2. We show ablation studies in Sect. 3.3. The effect of the number particles to the model performance is studied in the Cross-ALD Github repository.

3.1   Diversity of Adversarial Particle Comparison

Settings. We fixed all the decoder models (U-Net for ACDC and V-Net for LA). We run VAT with random initialization and SVGD multiple times to produce adversarial examples, which we compared to the adversarial particles generated using SVGDF. SVGDF is the proposed algorithm, which leverages feature transformation to capture the semantic meaning of inputs. Φ is the decoder of U-Net in ACDC dataset, while Φ is the decoder of V-Net in LA dataset. We set the same radius ball constraint, updating step, and etc. We randomly pick three images from the datasets to generate adversarial particles. To evaluate their diversity, we report the sum squared error (SSE) between these particles. Higher SSE indicates more diversity, and for each number of particles, we calculate the average of the mean of SSEs.

---

[3] https://github.com/ycwu1997/SS-Net.
[4] https://www.creatis.insa-lyon.fr/Challenge/acdc/databases.html.
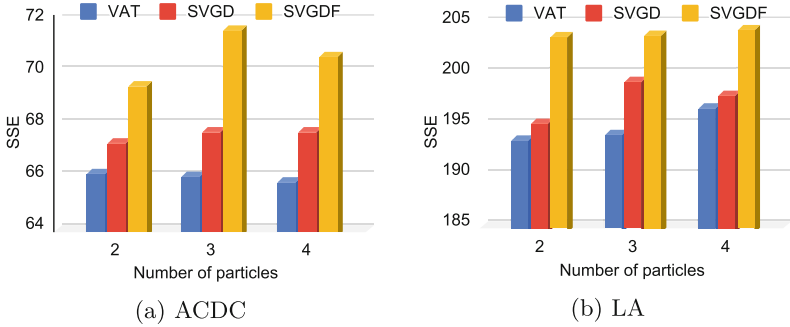[5] http://atriaseg2018.cardiacatlas.org.

Fig. 1. Diversity comparison of our SVGDF, SVGD and VAT with random initialization using sum of square error (SSE) of ACDC and LA datasets.

Table 1. Performance comparisons with six recent methods on ACDC dataset. All results of existing methods are used from [21] for fair comparisons.

Method	# Scans used		Metrics				Complexity	
	Labeled	Unlabeled	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓	Para.(M)	MACs(G)
U-Net	3(5%)	0	47.83	37.01	31.16	12.62	1.81	2.99
U-Net	7(10%)	0	79.41	68.11	9.35	2.7	1.81	2.99
U-Net	70(All)	0	91.44	84.59	4.3	0.99	1.81	2.99
UA-MT [28]	3 (5%)	67(95%)	46.04	35.97	20.08	7.75	1.81	2.99
SASSNet [8]			57.77	46.14	20.05	6.06	1.81	3.02
DTC [11]			56.9	45.67	23.36	7.39	1.81	3.02
URPC [12]			55.87	44.64	13.6	3.74	1.83	3.02
MC-Net [22]			62.85	52.29	7.62	2.33	2.58	5.39
SS-Net [21]			65.82	55.38	6.67	2.28	1.83	2.99
Cross-ALD (**Ours**)			**80.6**	**69.08**	**5.96**	**1.9**	1.83	2.99
UA-MT [28]	7 (10%)	63(90%)	81.65	70.64	6.88	2.02	1.81	2.99
SASSNet [8]			84.5	74.34	5.42	1.86	1.81	3.02
DTC [11]			84.29	73.92	12.81	4.01	1.81	3.02
URPC [12]			83.1	72.41	4.84	1.53	1.83	3.02
MC-Net [22]			86.44	77.04	5.5	1.84	2.58	5.39
SS-Net [21]			86.78	77.67	6.07	**1.4**	1.83	2.99
Cross-ALD (**Ours**)			**87.52**	**78.62**	**4.81**	1.6	1.83	2.99

Results. Note that the advantage of SVGD over VAT is that the former generates diversified adversarial examples because of the second term in Eq. 6 while VAT only creates one example. Moreover, vanilla SVGD is difficult to capture semantic meaning of high-resolution medical imaging because it calculates kernel k on image-level. In Fig. 1, our SVGDF produces the most diverse particles compared to SVGD and VAT with random initialization.

3.2   Performance Evaluation on the ACDC and la Datasets

Settings. We use the metrics of Dice, Jaccard, 95% Hausdorff Distance (95HD), and Average Surface Distance (ASD) to evaluate the results. We compare our

Table 2. Performance comparisons with six recent methods on LA dataset. All results of existing methods are used from [21] for fair comparisons.

Method	# Scans used		Metrics				Complexity	
	Labeled	Unlabeled	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓	Para.(M)	MACs(G)
V-Net	4(5%)	0	52.55	39.6	47.05	9.87	9.44	47.02
V-Net	8(10%)	0	82.74	71.72	13.35	3.26	9.44	47.02
V-Net	80(All)	0	91.47	84.36	5.48	1.51	9.44	47.02
UA-MT [28]	4 (5%)	76(95%)	82.26	70.98	13.71	3.82	9.44	47.02
SASSNet [8]			81.6	69.63	16.16	3.58	9.44	47.05
DTC [11]			81.25	69.33	14.9	3.99	9.44	47.05
URPC [12]			82.48	71.35	14.65	3.65	5.88	69.43
MC-Net [22]			83.59	72.36	14.07	2.7	12.35	95.15
SS-Net [21]			86.33	76.15	9.97	2.31	9.46	47.17
Cross-ALD (**Ours**)			**88.62**	**79.62**	**7.098**	**1.83**	9.46	47.17
UA-MT [28]	8 (10%)	72(90%)	87.79	78.39	8.68	2.12	9.44	47.02
SASSNet [8]			87.54	78.05	9.84	2.59	9.44	47.05
DTC [11]			87.51	78.17	8.23	2.36	9.44	47.05
URPC [12]			86.92	77.03	11.13	2.28	5.88	69.43
MC-Net [22]			87.62	78.25	10.03	1.82	12.35	95.15
SS-Net [21]			88.55	79.62	**7.49**	1.9	9.46	47.17
Cross-ALD (**Ours**)			**89.92**	**81.78**	7.65	**1.546**	9.46	47.17

Cross-ALD to six recent methods including UA-MT [28] (MICCAI'19), SASSNet [8] (MICCAI'20), DTC [11] (AAAI'21) , URPC [12] (MICCAI'21) , MC-Net [22] (MICCAI'21), and SS-Net [21] (MICCAI'22). The loss weights $\lambda_{Cross-ALD}$ and λ_{cs} are set as an iteration dependent warming-up function [7], and number of particles $N = 2$. All experiments are conducted using the identical settings in the Github repository[6] [21] for fair comparisons.

Results. Recall that our Cross-ALD generates diversified adversarial particles using SVGDF compared to vanilla SVGD and VAT, and further enhances smoothness of cross-adversarial local distributions. In Table 1 and 2, the Cross-ALD can significantly outperform other recent methods with only 5%/10% labeled data training based on the four metrics. Especially, our method impressively gains 14.7% and 2.3% Dice score higher than state-of-the-art SS-Net using 5% labeled data of ACDC and LA, respectively. Moreover, the visualized results of Fig. 2 shows Cross-ALD can segment the most organ details compared to other methods.

3.3   Ablation Study

Settings. We use the same network architectures and parameter settings in Sect. 3.2, and train the models with 5% labeled training data of ACDC and LA. We illustrate that crossing adversarial particles is more beneficial than random

---

[6] https://github.com/ycwu1997/SS-Net.

Table 3. Ablation study on ACDC and LA datasets.

Dataset	Method	# Scans used		Metrics			
		Labeled	Unlabeled	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓
ACDC	U-Net	4(5%)	0	47.83	37.01	31.16	12.62
	RanMixup	4 (5%)	76(95%)	61.78	51.69	8.16	3.44
	VAT			63.87	53.18	7.61	3.38
	VAT + Mixup			66.23	56.37	7.18	2.53
	SVGD			66.53	58.09	6.41	2.4
	SVGDF			73.15	61.71	6.32	2.12
	SVGDF + ℓ_{cs}			74.89	62.61	6.52	2.01
	Cross-ALD (**Ours**)			**80.6**	**69.08**	**5.96**	**1.9**
LA	V-Net	3(5%)	0	52.55	39.6	47.05	9.87
	RanMixup	3 (5%)	67(95%)	79.82	67.44	16.52	5.19
	VAT			82.27	70.46	13.82	3.48
	VAT + Mixup			83.28	71.77	12.8	2.63
	SVGD			84.62	73.6	11.68	2.94
	SVGDF			86.3	76.17	10.01	2.11
	SVGDF + ℓ_{cs}			86.55	76.51	9.41	2.24
	Cross-ALD (**Ours**)			**87.52**	**78.62**	**4.81**	**1.6**
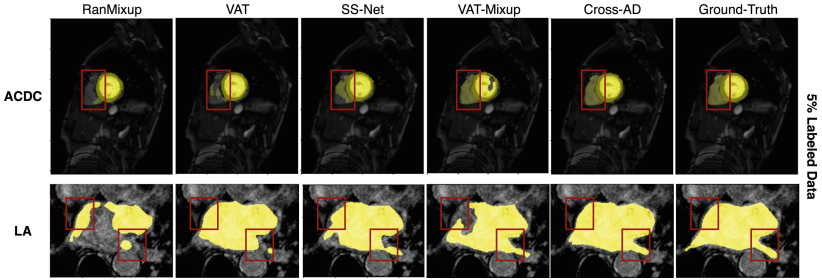
Fig. 2. Visualization results of several semi-supervised segmentation methods with 5% labeled training data and its corresponding ground-truth on ACDC and LA datasets.

mixup between natural inputs (RanMixup [29]) because these particles are near decision boundaries. Recall that our SVGDF is better than VAT and SVGD by producing more diversified adversarial particles. Applying SVGDF's particles and ℓ_{cs} (SVGDF + ℓ_{cs}) to gain the model performance in the semi-supervised segmentation task, while Cross-ALD efficiently enhances smoothness to significantly improve the generalization.

Result. Table 3 shows that mixing adversarial examples from VAT outperform those from RanMixup. While SVGDF + ℓ_{cs} is better than SVGD and VAT, the proposed Cross-ALD achieves the most outstanding performance among comparisons methods. In addition, our method produces more accurate segmentation masks compared to the ground-truth, as shown in Fig. 2.

4   Conclusion

In this paper, we have introduced a novel cross-adversarial local distribution (Cross-ALD) regularization that extends and overcomes drawbacks of VAT and Mixup techniques. In our method, SVGDF is proposed to approximate Cross-ALD, which produces more diverse adversarial particles than vanilla SVGD and VAT with random initialization. We adapt Cross-ALD to semi-supervised medical image segmentation to achieve start-of-the-art performance on the ACDC and LA datasets compared to many recent methods such as VAT [14], UA-MT [28], SASSNet [8], DTC [11], URPC [12] , MC-Net [22], and SS-Net [21].

References

1. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging **37**(11), 2514–2525 (2018)

2. Berthelot, D., et al.: Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. arXiv preprint arXiv:1911.09785 (2019)

3. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: A holistic approach to semi-supervised learning. Adv. Neural Inform. Process. Syst. **32** (2019)

4. French, G., Laine, S., Aila, T., Mackiewicz, M., Finlayson, G.: Semi-supervised semantic segmentation needs strong, varied perturbations. arXiv preprint arXiv:1906.01916 (2019)

5. Gyawali, P., Ghimire, S., Wang, L.: Enhancing mixup-based semi-supervised learningwith explicit lipschitz regularization. In: 2020 IEEE International Conference on Data Mining (ICDM), pp. 1046–1051. IEEE (2020)

6. Lai, X., et al.: Semi-supervised semantic segmentation with directional context-aware consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1205–1214 (2021)

7. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242 (2016)

8. Li, S., Zhang, C., He, X.: Shape-aware semi-supervised 3D semantic segmentation for medical images. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 552–561. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_54

9. Li, X., Yu, L., Chen, H., Fu, C.W., Xing, L., Heng, P.A.: Transformation-consistent self-ensembling model for semisupervised medical image segmentation. IEEE Trans. Neural Netw. Learn. Syst. **32**(2), 523–534 (2020)

10. Liu, Q., Wang, D.: Stein variational gradient descent: A general purpose bayesian inference algorithm. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (eds.) Proceedings of NeurIPS. vol. 29 (2016)

11. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 8801–8809 (2021)
12. Luo, X., et al.: Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In: Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 318–329. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_30
13. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
14. Miyato, T., Maeda, S.i., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE TPAMI **41**(8), 1979–1993 (2018)
15. Nguyen-Duc, T., Le, T., Zhao, H., Cai, J., Phung, D.Q.: Particle-based adversarial local distribution regularization. In: AISTATS, pp. 5212–5224 (2022)
16. Ouali, Y., Hudelot, C., Tami, M.: An overview of deep semi-supervised learning. arXiv preprint arXiv:2006.05278 (2020)
17. Ouali, Y., Hudelot, C., Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12674–12684 (2020)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241. Springer (2015)
19. Sohn, K., et al.: Fixmatch: simplifying semi-supervised learning with consistency and confidence. Adv. Neural. Inf. Process. Syst. **33**, 596–608 (2020)
20. Tashiro, Y., Song, Y., Ermon, S.: Diversity can be transferred: Output diversification for white-and black-box attacks. Proc. NeurIPS **33**, 4536–4548 (2020)
21. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 13435, pp. 34–43. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16443-9_4
22. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 297–306. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_28
23. Wu, Z., Shi, X., Lin, G., Cai, J.: Learning meta-class memory for few-shot semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 517–526 (2021)
24. Xia, Y., et al.: 3D semi-supervised learning with uncertainty-aware multi-view co-training. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3646–3655 (2020)
25. Xie, Y., Zhang, J., Liao, Z., Verjans, J., Shen, C., Xia, Y.: Intra-and inter-pair consistency for semi-supervised gland segmentation. IEEE Trans. Image Process. **31**, 894–905 (2021)
26. Xiong, Z., et al.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Med. Image Anal. **67**, 101832 (2021)
27. Yang, X., Song, Z., King, I., Xu, Z.: A survey on deep semi-supervised learning. IEEE Transactions on Knowledge and Data Engineering (2022)

28. Yu, L., Wang, S., Li, X., Fu, C.-W., Heng, P.-A.: Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 605–613. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_67

29. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)