RDA: Reciprocal Distribution Alignment for Robust Semi-supervised Learning

Yue Duan[1], Lei Qi[2], Lei Wang[3], Luping Zhou[4], and Yinghuan Shi[1]

[1] *Nanjing University, China* [2] *Southeast University, China* [3] *University of Wollongong, Australia* [4] *University of Sydney, Australia*

Introduction

Lately, confidence-based pseudo-labeling [1,2] and distribution alignment [2,3,4] have been introduced to SSL, boosting performance to a new height. FixMatch [1] utilizes the confidence-based threshold to select more accurate pseudo-labels and proves the superiority of this technique. Despite this threshold preventing the model from risk of noisy pseudo-labels, since the learning difficulties of different classes are different, a fixed threshold is not a "silver bullet" for all scenarios of SSL. We try to ask — *is the confidence-based threshold really necessary for pseudo-labeling*? Motivated by this, we rethink pseudo-labeling in a hyperparameter-free way while noticing that **distribution alignment (DA)** has been introduced to SSL [2,4]. we consider only using DA to improve the pseudo-labels without additional hyperparameters, i.e., DA is enough for pseudo-labeling. However, original DA is based on a strong assumption: "**labeled data and unlabeled data share the same distribution**". Thus, we have the following motivations:

Motivation

◆ **Improve pseudo-label quality by using only distribution alignment without introducing confidence threshold.**

◆ **Distribution alignment fails in mismatched distribution scenarios.**

◆ **Explore a more general distribution alignment to meet the challenge of mismatched distributions.**

Method
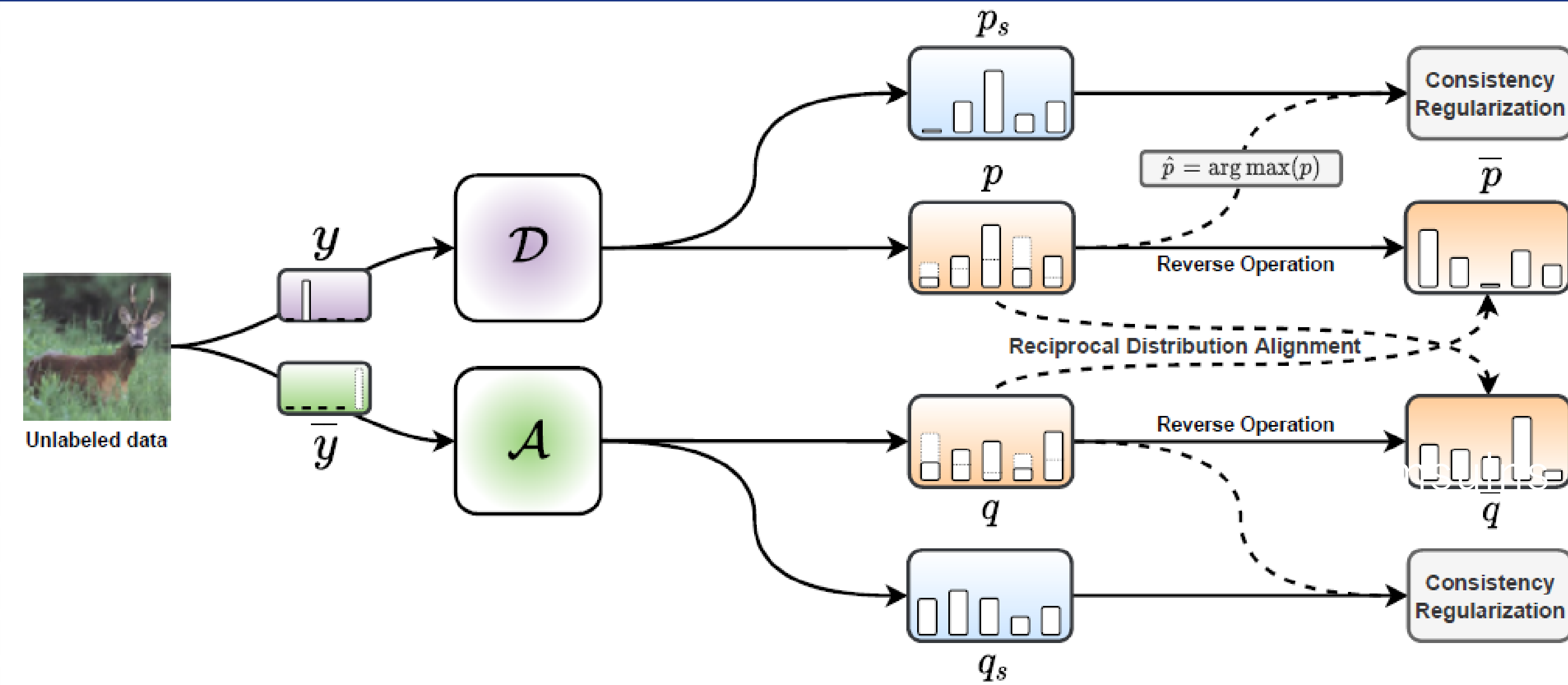
Reciprocal Distribution Alignment (RDA)

We use ground-truth label y and complementary label $\bar{y}$ (selected randomly from classes excluding ground-truth label) of labeled data to train *Default Classifier* $\mathcal{D}$ and *Auxiliary Classifier* $\mathcal{A}$, respectively. Given an unlabeled sample u, $\mathcal{D}$ predicts pseudo-label p and $\mathcal{A}$ predicts complementary label q for its weakly-augmented version. **RDA** is applied on p and q by reciprocally scaling each other to the distributions of their reversed versions obtained by *Reverse Operation* (**Proposition 1**). We then enforce consistency regularization on the aligned pseudo-label and complementary label against corresponding predictions for strongly-augmented u, i.e., p_s (from $\mathcal{D}$) and q_s (from $\mathcal{A}$).

Theoretical Analysis

Proposition 1 (*Reverse Operation*). In the case of using $\mathcal{A}$ to predict pseudo-labels, we have $\bar{q} = Norm(q)$, where $Norm(x)$ is the normalized operation.

Theorem 1. For pseudo-label p and the reversed pseudo-label p obtained by *Reverse Operation*, we show that the entropy of $\bar{p}$ is larger than that of p: $\mathcal{H}(\bar{p}) \geq \mathcal{H}(p)$, where $\mathcal{H}(\cdot)$ refers to the entropy.

Experiments

Method	CIFAR-10				mini-ImageNet	STL-10
	20 labels	40 labels	80 labels	100 labels	1000 labels	1000 labels
MixMatch*	27.84±10.63	51.90±11.76	80.79±1.28	-	-	38.02±8.29
AlphaMatch[†]	-	91.35±3.38			-	-
FixMatch	84.97±10.37	89.18±1.54	91.99±0.71	93.14±0.76	39.03±0.66	65.38±0.42*
CoMatch	88.43±7.22	93.21±1.55	94.08±0.31	**94.55±0.27**	43.72±0.58	79.80±0.38*
RDA	**92.03±2.01**	**94.13±1.22**	**94.24±0.42**	94.35±0.25	**46.91±1.16**	**82.63±0.54**

Tab. 1 Accuracy (%) in the conventional matched SSL setting.

Method	CIFAR-10				CIFAR-100		mini-ImageNet	
	40 labels		100 labels		400 labels	1000 labels	1000 labels	
	$N_0 = 10$	20	40	80	40	80	40	80
FixMatch	85.72±0.93	76.53±3.03	93.01±0.72	71.57±1.88	25.66±0.46	40.22±1.00	36.20±0.36	28.33±0.41
FixMatch w. DA	71.23±1.25	47.85±1.99	56.78±1.28	34.18±0.86	22.66±1.53	31.06±0.51	33.87±0.40	23.53±0.72
CoMatch	60.27±3.22	39.48±2.20	52.82±2.03	26.91±0.75	23.97±0.62	28.35±1.20	30.24±1.37	21.47±0.86
RDA	**92.57±0.53**	**81.78±6.44**	**94.23±0.36**	**79.00±2.67**	**30.86±0.78**	**41.29±0.43**	**42.73±0.84**	**36.73±1.01**

Tab. 2 Accuracy (%) in the mismatched scenario with imbalanced C_x and balanced C_u.

Method	CIFAR-10				mini-ImageNet
	40 labels, $N_0 = 10$		100 labels, $N_0 = 40$		1000 labels, $N_0 = 40$
	$\gamma = 2$	5	5	10	10
FixMatch	74.97±5.80	64.62±6.13	58.72±3.61	57.49±4.56	21.40±0.53
RDA	**88.58±4.05**	**79.90±2.80**	**79.33±1.37**	**70.93±2.91**	**25.99±0.19**

Tab. 3 Accuracy (%) with mismatched and imbalanced C_x, C_u.

Method	CIFAR-10 ($\gamma_l = 100$)				STL-10 ($\gamma_l \neq \gamma_u$)	
	$\gamma_u = 1$	50	150	100 (reversed)	$\gamma_l = 10$	20
FixMatch	68.90±1.95	73.90±0.25	69.60±0.60	65.50±0.05	72.90±0.09	63.40±0.21
DARP	85.40±0.55	77.30±0.17	72.90±0.24	74.90±0.51	77.80±0.33	69.90±0.40
RDA	**93.35±0.24**	**79.77±0.06**	**74.48±0.24**	**79.25±0.52**	**87.21±0.44**	**83.21±0.52**

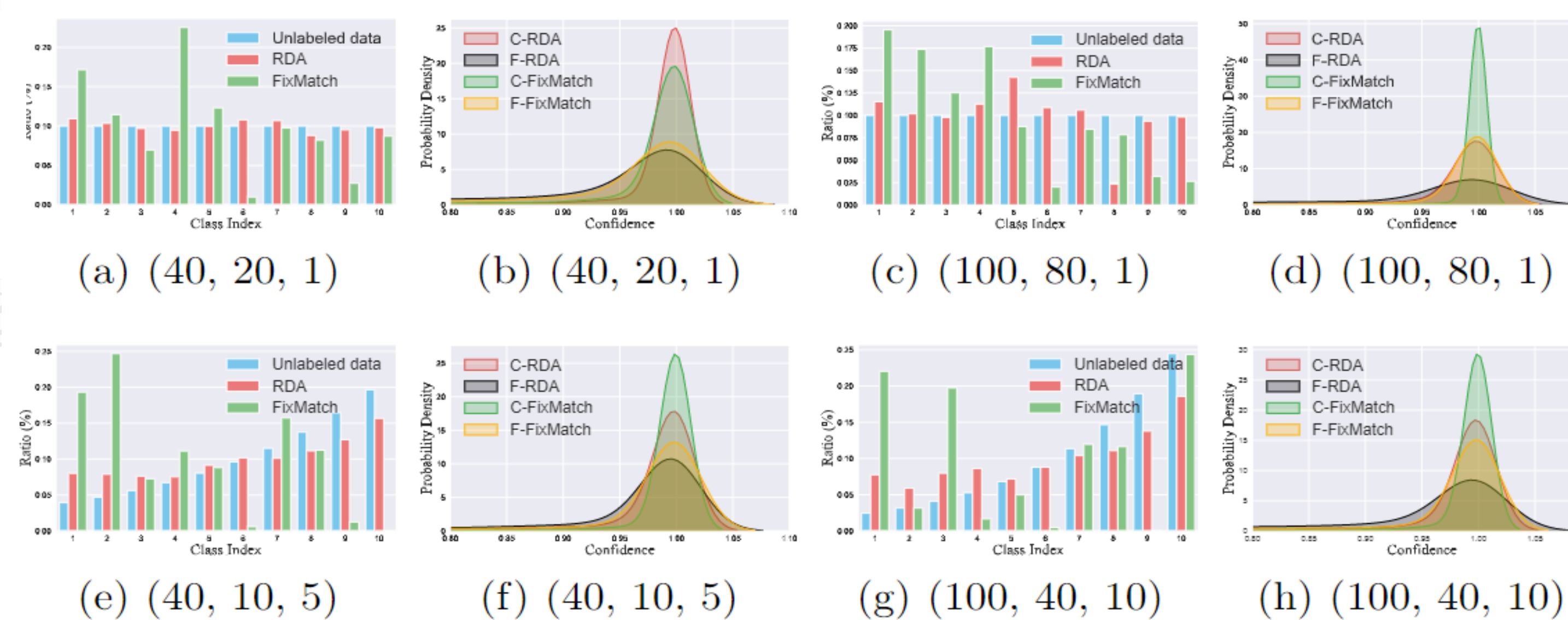Tab. 4 Accuracy (%) under DARP's protocol.

Fig. 1 (x, y, z): (labels, N_0, γ). (a), (c), (e) and (g): Distribution of labels. (b), (d), (f) and (h): Probability density of confidence scores.

Conclusion

First, we describe a scenario that has not been discussed extensively by recently-proposed SSL work: **mismatched distributions**. Second, we improve distribution alignment by proposed **RDA** so that this technique could be applied into mismatched scenario safely. Then we show *RDA results in a form of maximizing the input-out mutual information without any prior information*. Finally, we demonstrate that our method outperforms existing baselines significantly under various scenarios.

References

[1] Sohn, K., Berthelot, D., Li, C.L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., Raffel, C.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In: NeurIPS (2020)

[2] Li, J., Xiong, C., Hoi, S.C.: Comatch: Semi-supervised learning with contrastive graph regularization. In: ICCV (2021)

[3] Bridle, J.S., Heading, A.J., MacKay, D.J.: Unsupervised classifiers, mutual information and "phantom targets". In: NeurIPS (1992)

[4] Berthelot, D., Carlini, N., Cubuk, E.D., Kurakin, A., Sohn, K., Zhang, H., Raffel, C.: Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring. In: ICLR (2020)