Introduction

Recently, FixMatch [1] utilizes the confidence-based threshold to select more accurate pseudo-labels and proves the superiority of this technique.

↓

We try to ask — *is the confidence-based threshold really necessary for pseudo-labeling?*

↓

We consider only using distribution alignment (DA) to improve the pseudo-labels without additional hyperparameters.

↓

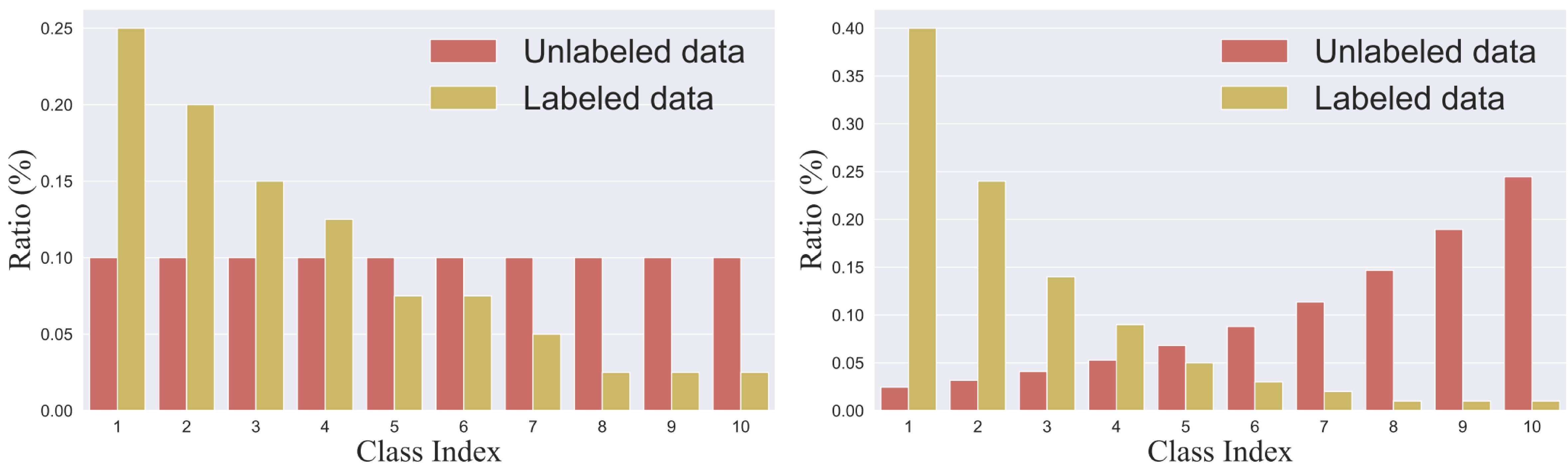
However, original DA is based on a strong assumption: "**labeled data and unlabeled data share the same distribution**".

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

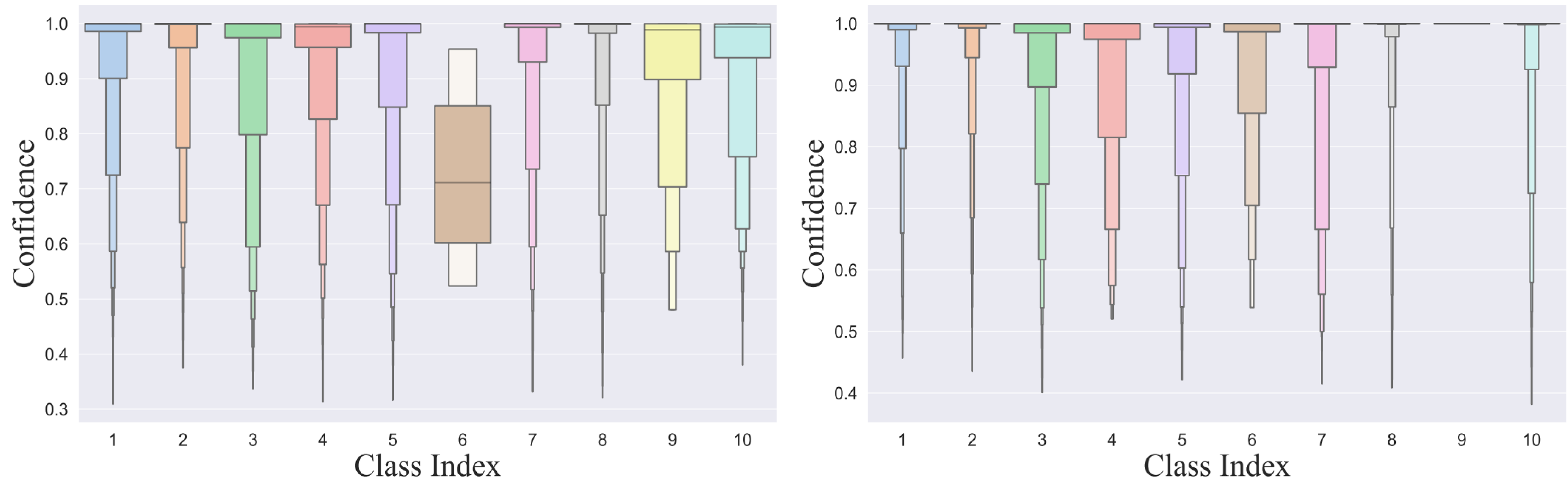
Motivation

The original distribution alignment technique fails in the SSL with mismatched distribution, while the confidence threshold is difficult to set.

Explore a more general distribution alignment technique to address the challenges of mismatched distributions.

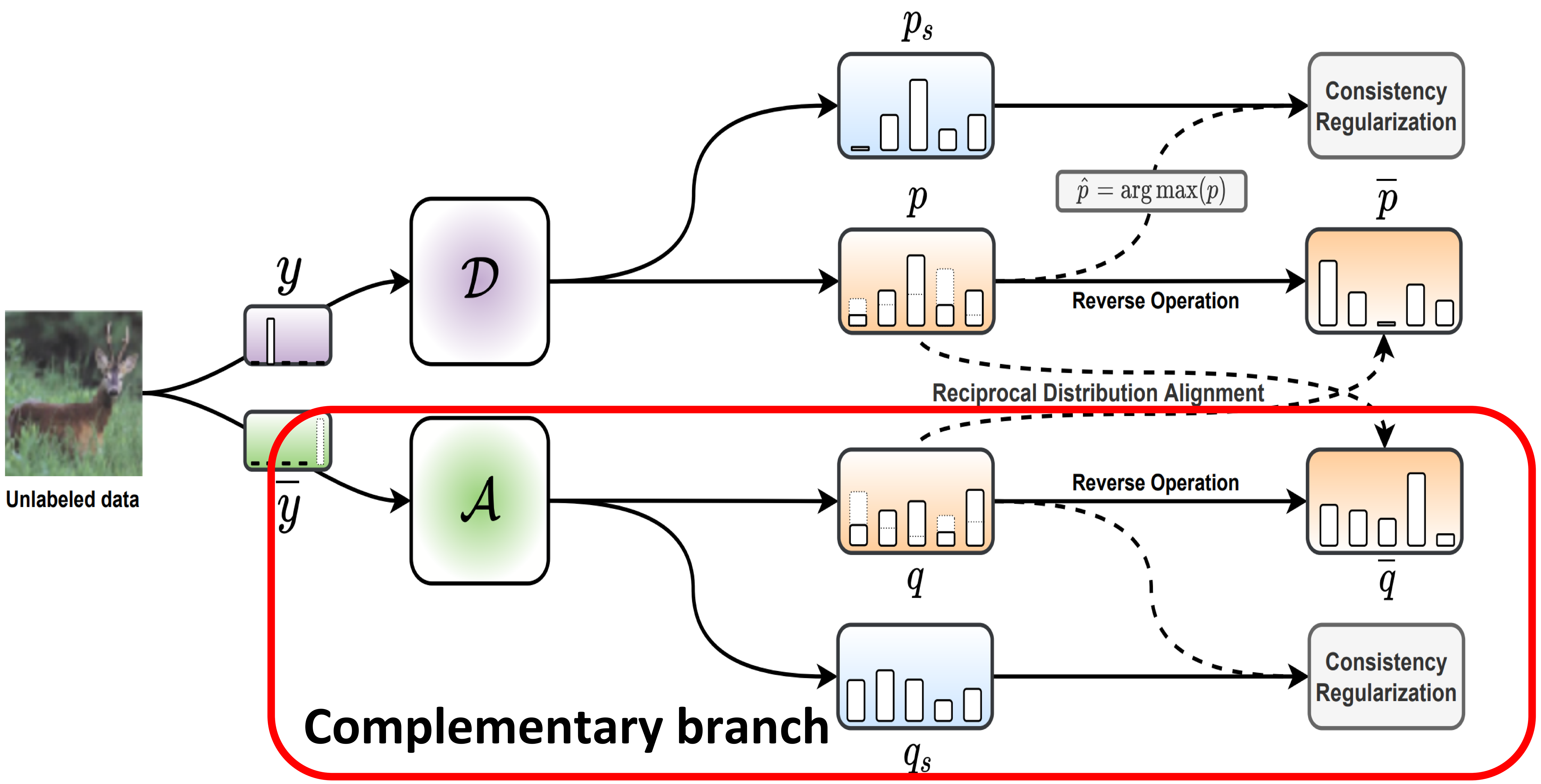


Mismatched distributions



Confidence threshold is difficult to set

A new distribution alignment technique based on **pseudo-label** and **complementary label** distribution is proposed to improve pseudo-label quality

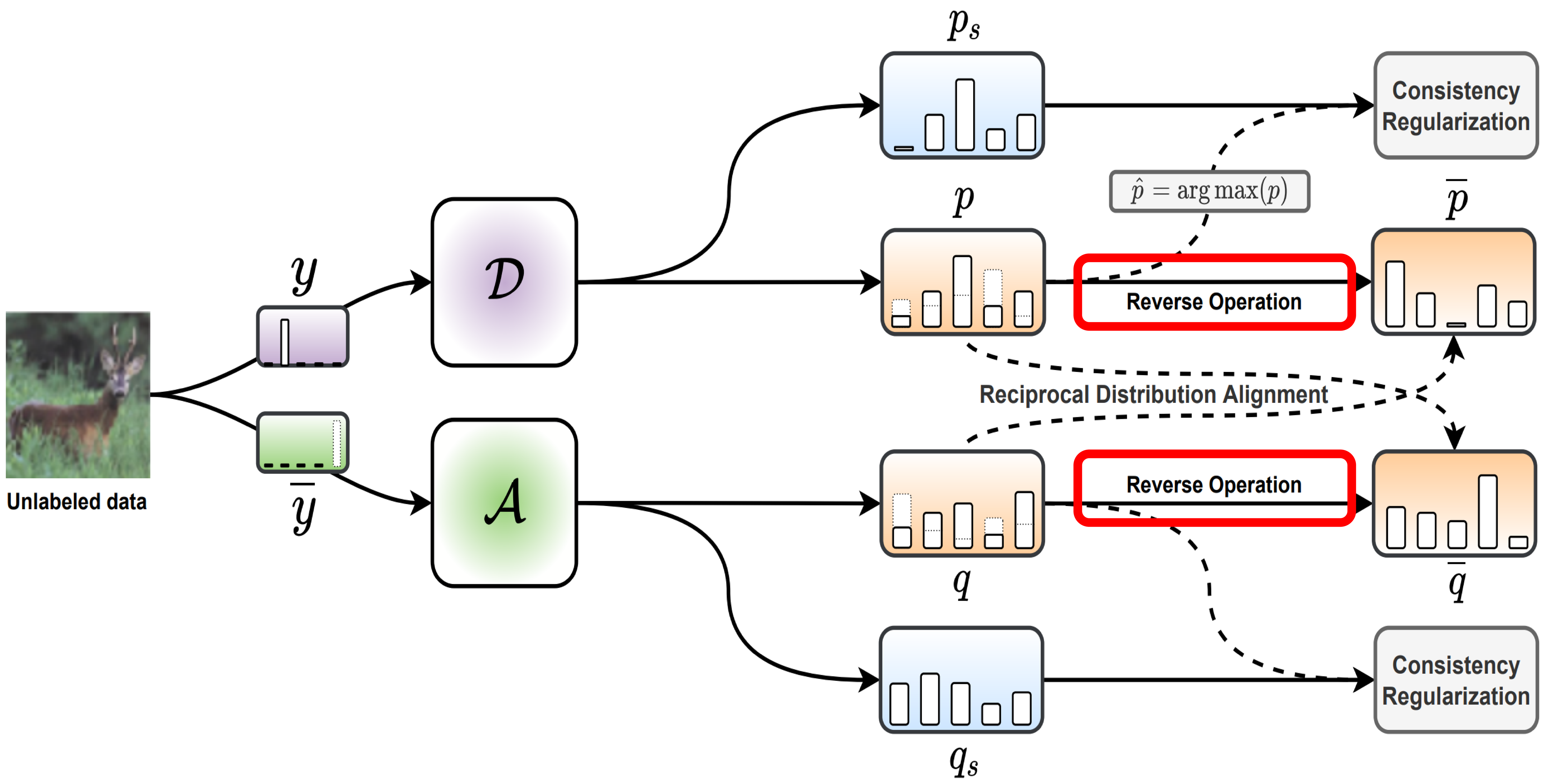


$$\mathcal{I}(y;x) = \mathcal{H}(\mathbb{E}_x[P(y|x)]) - \mathbb{E}_x[\mathcal{H}(P(y|x))]$$

Improve pseudo-labeling by maximizing input-output mutual information [2]

How to maximize input-output mutual information?

[2] Berthelot et al.“**Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring**”. ***ICLR***, *2020*.



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

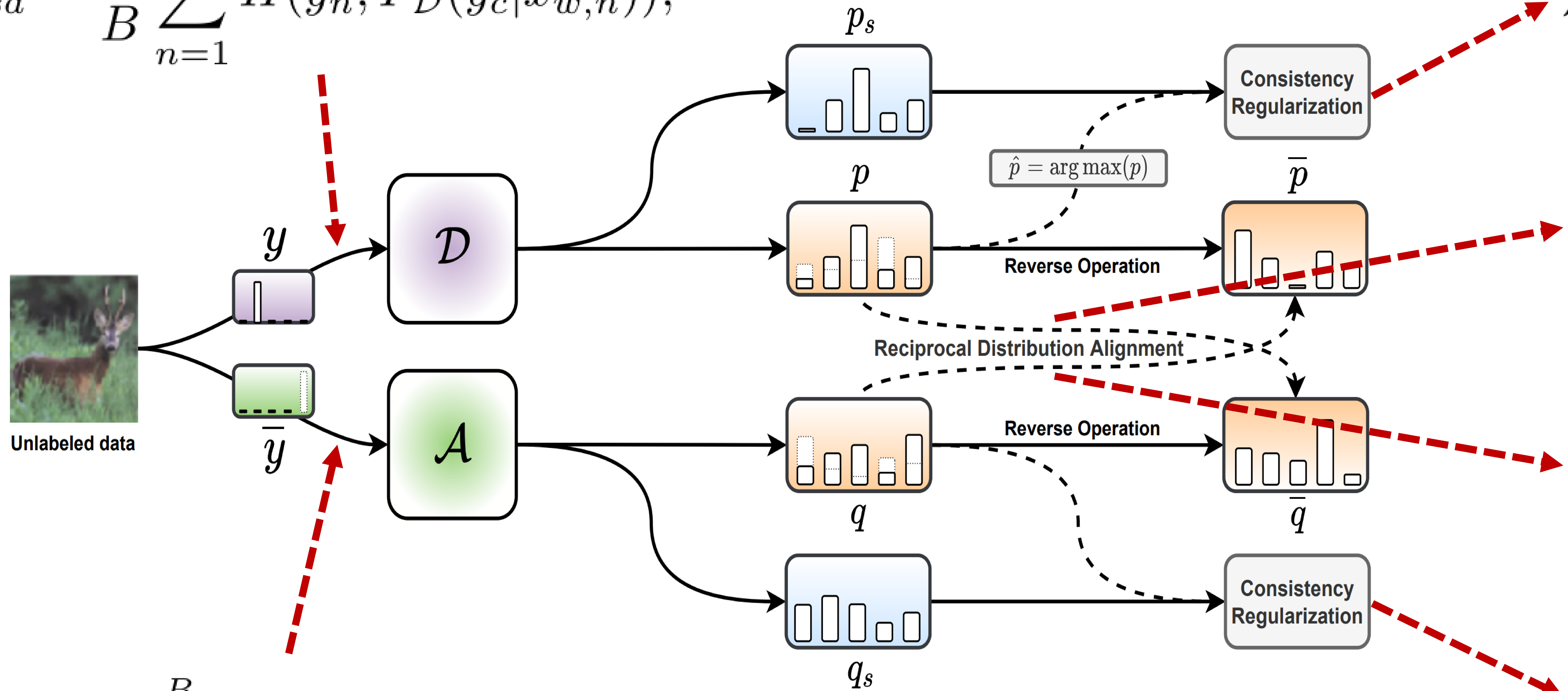
- Generate complementary labels from labeled data to train Auxiliary Classifier $\mathcal{A}$
- Reverse the pseudo-labels and complementary labels output by the Default Classifier and Auxiliary Classifier respectively (**Reverse Operation**)
- Align the pseudo-label distribution from Default Classifier to the reversed distribution of complementary labels from Auxiliary Classifier, while aligning the complementary label distribution to the reversed distribution of pseudo-labels

RDA: Reciprocal Distribution Alignment for Robust Semi-supervised Learning

NANJING UNIVERSITY

$$\mathcal{L}_{sd} = \frac{1}{B}\sum_{n=1}^{B} H(y_n, P_{\mathcal{D}}(y_c|x_{w,n})),$$

$$\mathcal{L}_{cd} = \frac{1}{\mu B}\sum_{n=1}^{\mu B} H(\hat{p}_n, p_{s,n}),$$



$$\tilde{p} = \mathrm{Norm}(p \times \frac{\Psi(\overline{q})}{\Psi(p)}),$$

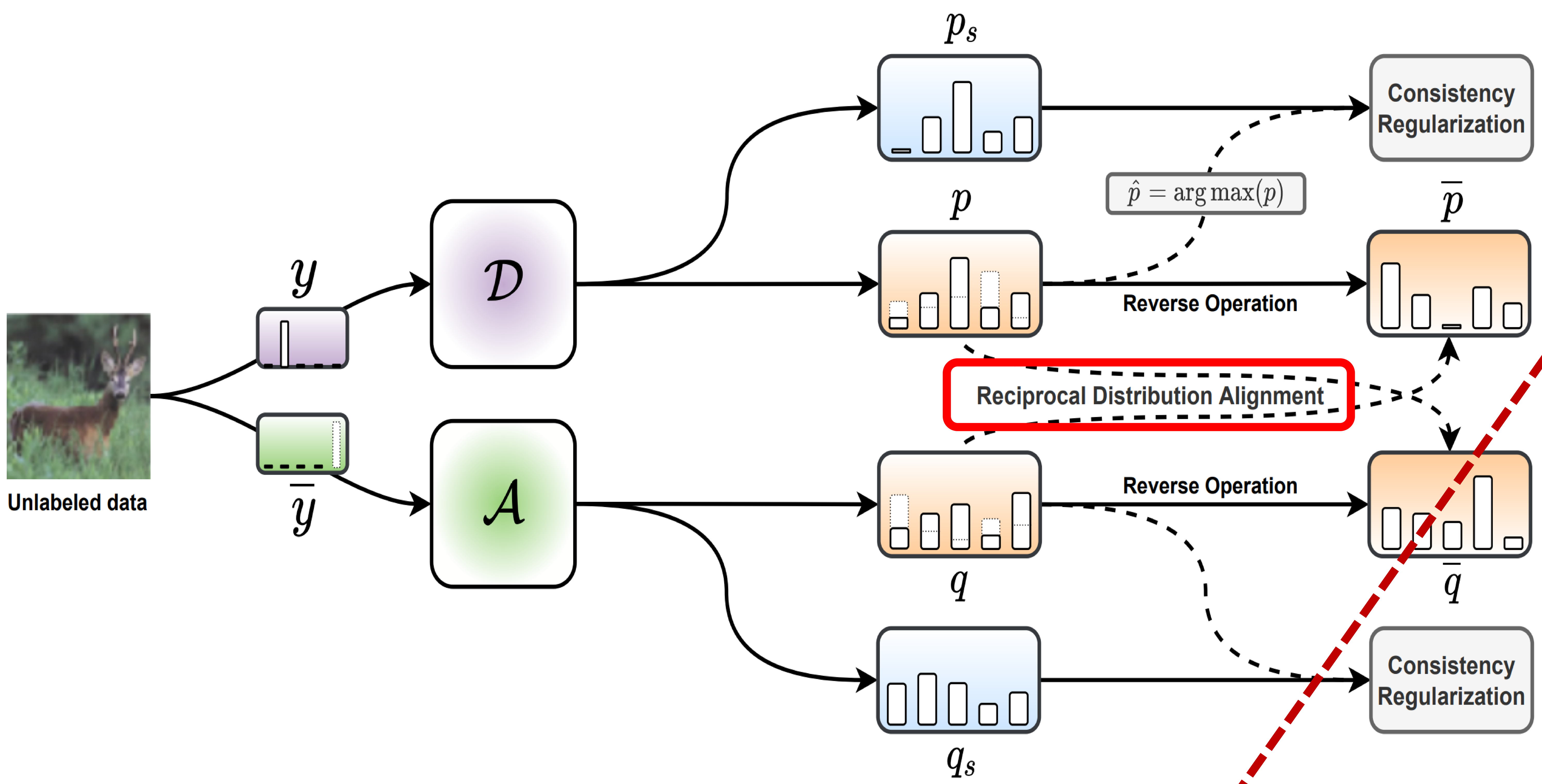
$$\tilde{q} = \mathrm{Norm}(q \times \frac{\Psi(\overline{p})}{\Psi(q)}),$$

$$\mathcal{L}_{ca} = \frac{1}{\mu B}\sum_{n=1}^{\mu B} H(q_n, q_{s,n}),$$

$$\mathcal{L}_{sa} = \frac{1}{B}\sum_{n=1}^{B} H(\overline{y}_n, P_{\mathcal{A}}(y_c|x_{w,n})),$$

$$\mathcal{L} = \mathcal{L}_{sd} + \lambda_a\mathcal{L}_{sa} + \lambda_{cd}\mathcal{L}_{cd} + \lambda_{ca}\mathcal{L}_{ca},$$

RDA: Reciprocal Distribution Alignment for Robust Semi-supervised Learning



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

$$\max_{\mathcal{D},\mathcal{A}} h(\mathcal{D},\mathcal{A}) = \mathcal{H}[\mathbb{E}_u(p)] + \mathcal{H}[\mathbb{E}_u(q)] \Longrightarrow \mathcal{H}[\mathbb{E}_u(p)] + \mathcal{H}[\mathbb{E}_u(q)] \leq \mathcal{H}[\mathbb{E}_u(\overline{p})] + \mathcal{H}[\mathbb{E}_u(\overline{q})].$$

We show theoretically that the input-output mutual information can be maximized by *reciprocal distribution alignment*

◆ Conventional SSL setting： CIFAR10, mini-ImageNet, STL-10

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**

In the conventional SSL setting, where the labeled and unlabeled data have the same distribution and are uniformly distributed, RDA achieves superior performance.

◆ Mismatched distribution scenarios： CIFAR-10/100, mini-ImageNet

- The labeled data is imbalanced, the unlabeled data is balanced
- The labeled data is balanced, the unlabeled data is imbalanced
- The labeled data and unlabeled data are imbalanced and mismatched

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**

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**

Method	CIFAR-10 (γ_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**

In the mismatched scenario, RDA still achieves a superior performance advantage.