RDA : Reciprocal Distribution Alignment for Robust SSL (ECCV 2022)

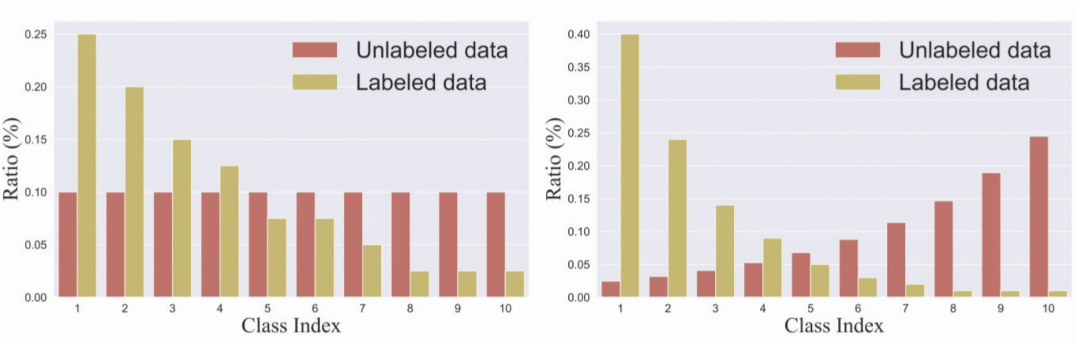
confidence threshold → fixed threshold ✗

dynamically adjust complicated ✗

hyperparameter-free

distribution alignment (DA) → " labeled and unlabeled share the same distribution "

(DA scales the predictions on unlabeled data by prior information about labeled data)



Mismatched distribution

↓

relax the assumption about the class distribution of unlabeled

RDA

maximize the mutual information between the prediction and input data

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

entropy prediction input data

maximize

Eq(1)

Labeled $\mathcal{X} = \{(x_b, y_b)\}_{b=1}^{B}$, Unlabeled $\mathcal{U} = \{(u_b)\}_{b=1}^{\mu B}$ in a batch

$y \in \mathcal{Y} = \{1, \ldots, n\}$: ground-truth label of x

$\bar{y} \in \mathcal{Y} \setminus \{y\}$: complementary label of x (randomly selected)

u_w / u_s : weakly/strongly augmented image for the same unlabeled u

$$\left.\begin{aligned} p &= P_D(y_c|u_w) \\ p_s &= P_D(y_c|u_s) \end{aligned}\right\} \text{pseudo-label} \qquad \left.\begin{aligned} q &= P_A(y_c|u_w) \\ q_s &= P_A(y_c|u_s) \end{aligned}\right\} \text{complementary label}$$

predict pseudo-label

minimize Cross-entropy

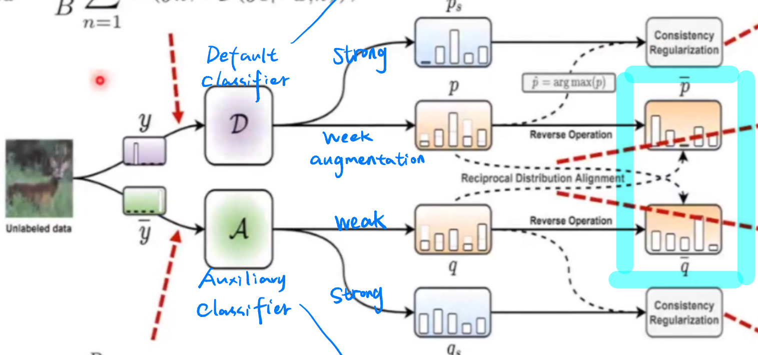
for $\mathcal{X}$

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

for $\mathcal{U}$

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

hard



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

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

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

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

soft

predict complementary label

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

$$\mathrm{DA} \rightarrow \max_{D} \mathcal{H}[\mathbb{E}_u(P_D(y_c|u_w))] \ \& \ \max_{A} \mathcal{H}[\mathbb{E}_u(P_A(y_c|u_w))]$$

$$\downarrow$$

Uniform prediction X

[2] shows making one distribution approach to another can achieve the purpose of maximizing Eq. (1)

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

Mismatched $\longrightarrow$ labeled distribution cannot be directly used.

distribution of class prediction

use $\mathbb{E}_u(p)$, $\mathbb{E}_u(q)$ build a reciprocal alignment

complementary

$\bar{q} = \mathrm{Norm}(\mathbb{1} - q)$, $\mathrm{Norm}(x)$ is the normalized operation $x_i' = x_i / \sum_{j=1}^{n} x_j$

$i \in (1, \cdots, n)$

$$\left.\begin{array}{l} \mathbb{E}_u(p) \longrightarrow \mathbb{E}_u(\bar{q}) \\ \mathbb{E}_u(q) \longrightarrow \mathbb{E}_u(\bar{p}) \end{array}\right\} \Rightarrow \begin{array}{l} \boxed{\breve{p}} = \mathrm{Norm}\left(p \times \dfrac{\phi(\bar{q})}{\phi(p)}\right) \\ \boxed{\breve{q}} = \mathrm{Norm}\left(q \times \dfrac{\phi(\bar{p})}{\phi(q)}\right) \end{array}$$

$\phi(\cdot)$: moving average over last 128 batches

aligned probability distribution

Finally, just replace $\hat{p}_n$ with $\hat{\breve{p}}_n$ in $\mathcal{L}_{cd}$

q_n with $\breve{q}_n$ in $\mathcal{L}_{ca}$