Supplementary: Learning Classifiers of Prototypes and Reciprocal Points for Universal Domain Adaptation

Sungsu Hur Inkyu Shin Kwanyong Park Sanghyun Woo In So Kweon
KAIST

Appendices

In this supplementary material, we provide,

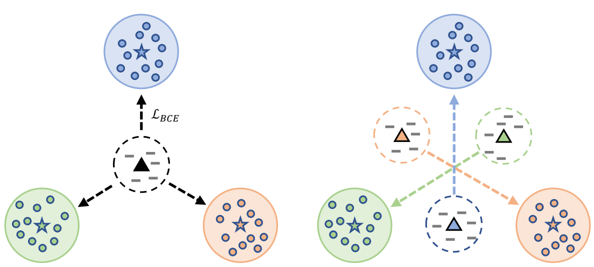
A.1) Pseudo Code of the Proposed Algorithm

A.2) How does Multiple Reciprocal Helps?

A.3) Hyperparameter and Loss Analysis

1. Pseudo Code of the Proposed Algorithm

In this section, we describe the entire training scheme in detail with pseudo code Algo 1.



(a) Single reciprocal (b) Multiple reciprocal

Figure 1: Illustration of single and multiple reciprocal points.

2. How does Multiple Reciprocal Help?

2.1. Overview

CPR shows robust high performance as it uses multiple reciprocal points to represent features from unknown classes. To show the effectiveness of the multiple reciprocal points, we conduct an experiment where the only one reciprocal point is used as [1] compares the single classifier placeholder and multiple classifier placeholders. Each reciprocal point is originally diverted from samples of corresponding class, but for the single reciprocal point setting, the point is trained to be far from all the known source samples by minimizing binary cross entropy loss as shown in Fig. 1a. Therefore, loss function for source data is changed as follows:

$$\mathcal{L}_{src} = \mathcal{L}_{CE_p} + \frac{1}{K}\mathcal{L}_{BCE} + \lambda(\mathcal{L}_o + \mathcal{L}_{split}) \quad (1)$$

where K is the total number of known classes.

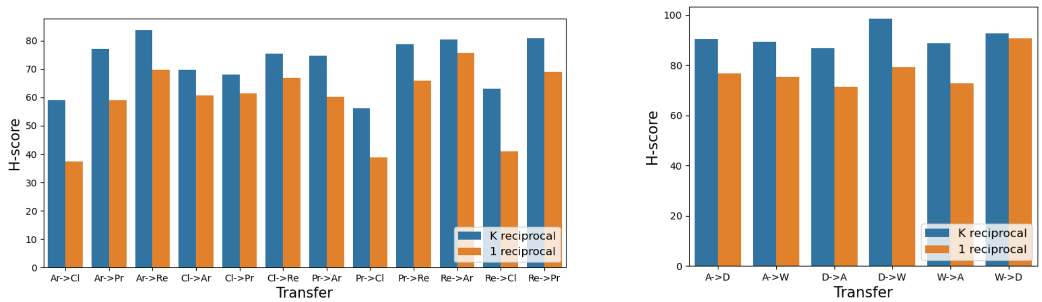
Algorithm 1 Training Pipeline of CPR

Input: source dataset: D_s, target dataset: D_t, feature extractor: g, prototype and reciprocal classifiers: h_p and h_r, two dynamic thresholds ρ_c, ρ_o, warm-up iterations i_w.

for $i, [x^s, x^t]$ in $[D_s, D_t]$ **do**
  /* Network Prediction */
  $x^t_w, x^t_s \leftarrow augment(x^t)$
  Predict p_c, p_p, p_r for x^s, x^t_w, x^t_s
  /* Sample Selection */
  $B_c, B_o = Select(x^t; Nearest)$ /* Sec. 3.3 */
  if $i \geq i_w$ **then**
    /* Multiple Criteria in Sec. 3.3.2 */
    $\hat{B}_c = Select(B_c; Consistency, Threshold)$
    $\hat{B}_o = Select(B_o; Consistency, Threshold)$
  end if
  /* Loss Calculation*/
  /* eq. (9) */
  $\mathcal{L}_{src} = \mathcal{L}_{CE_p} + \mathcal{L}_{CE_r} + \lambda(\mathcal{L}_o + \mathcal{L}_{split})$
  if $i \leq i_{warmup}$ **then**
    $\mathcal{L}_{trg} = \mathcal{L}_{kl}$ /* eq. (8) */
  else
    $\mathcal{L}_{trg} = \mathcal{L}_{kl} + \mathcal{L}_{ent} + \lambda\mathcal{L}_o$ /* eq. (14) */
  end if
  Update parameters of g, h_p, h_r using
  $\mathcal{L}_{all} = \mathcal{L}_{src} + \mathcal{L}_{trg}$
  /* Threshold Update */
  /* eq. (11) & (12) */
  Update ρ_c and ρ_o using B_c and B_o, respectively
end for

Table 1: Ablation study of $\mathcal{L}_{C_R}$ in the adaptation phase. All experiments are conducted on OfficeHome for OSDA.

Method	OfficeHome (25/0/40)												
	Ar→ Cl	Ar→ Pr	Ar→ Re	Cl→ Ar	Cl→ Pr	Cl→ Re	Pr→ Ar	Pr→ Cl	Pr→ Re	Re→ Ar	Re→ Cl	Re→ Pr	Avg
CPR w/o $\mathcal{L}_{CE_r}$	54.6	64.8	74.3	66.0	64.5	64.1	64.5	55.4	72.8	71.6	59.7	73.2	65.5
CPR	57.1	67.2	75.7	64.9	66.8	65.6	64.5	57.3	73.8	71.0	60.9	74.4	66.6

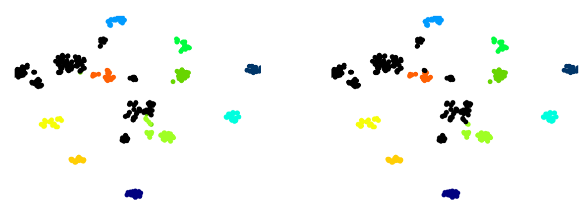


(a) OfficeHome UniDA (b) Office OSDA

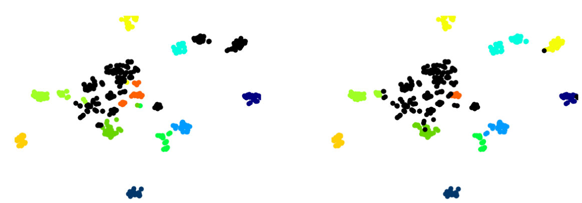
Figure 2: Comparison with single and multiple reciprocal points.

Table 2: Ablation study on α

α	OfficeHome(OSDA)	OfficeHome(UniDA)
0.9	65.8	71.5
0.99	**66.6**	**72.3**
0.999	65.6	71.6



(a) GT CPR (b) Pred CPR



(c) GT w/ single (d) Pred w/ single

Figure 3: Feature visualization on D2W in Office OSDA. Black plots are unknown samples, others are known samples. (a),(b): GT and prediction of the model trained with multiple reciprocal points. (c),(d): GT and prediction of the model trained with single reciprocal point.

2.2. Results

As shown in Fig. 2, we train the model with single reciprocal point to compare with the original CPR under Office OSDA and Office UniDA settings. For the both settings, the performance of all experiments decrease when the model is trained with the single reciprocal point. It implies single reciprocal point is not sufficient for the model to detect unknown samples, which is also shown in Fig. 3. Unknown target features are also distributed across feature space and clustered according to the semantic information as known features. To detect dispersed unknown features, multiple reciprocal points are more desirable to represent various semantic information of unknown features than single reciprocal point.

3. Hyperparameter and Loss Analysis

In the main paper, we quantitatively show the importance of $\mathcal{L}_{split}$ in the ablation study. Moreover, we conduct ablation study of $\mathcal{L}_{CE_r}$ in the adaptation phase. We also further study the sensitivity of a hyperparameter, i.e., α for this framework as shown in Tab 2

Ablation study of $\mathcal{L}_{CE_r}$ in the adaptation phase. We conduct an experiment on OfficeHome under OSDA setting and compare with the CPR. As shown in Table 1, in most scenarios, the performance decreases if $\mathcal{L}_{CE_r}$ is not minimized in the adaptation phase. Although reciprocal classifier is aligned to the target features with $\mathcal{L}_{ent}$, it can lose the semantic knowledge learned from source domain if $\mathcal{L}_{CE_r}$ is not minimized in the adaptation phase. In that sense, minimizing $\mathcal{L}_{CE_r}$ might regularize the model to possess the knowledge.

Analysis of α. We also conduct experiments on OSDA and UniDA of OfficeHome dataset to show CPR is robust to the selection of α. As shown in Tab 2, CPR is robust to the varying α. Compared to the original CPR, other α values seem to have inferior performance. Hence, we choose 0.99 for α.

References

[1] Da-Wei Zhou, Han-Jia Ye, and De-Chuan Zhan. Learning placeholders for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2021.