

1 We sincerely appreciate insightful comments and positive feedback from the reviewers: important problem (R1, R3),
2 simple and effective (R1, R4), and good experiments (R1, R2, R4). We respond to each comment one by one.

3 (R2, R3) Experiments on datasets other than CIFAR-10. In fact, the experimental results on CIFAR-100, STL-10,
4 and SUN-397 datasets are presented in Sections C and D of the supplementary material. In these results, DARP
5 consistently outperforms other baselines. We mention this in Line 148; however, we will make it clear in the final draft.

6 (R1) Imbalanced SSL. We first clarify that SSL algorithms + DARP are the *only* ‘imbalanced’ SSL algorithms in
7 our experiments, which consistently outperform the two types of baselines: re-balancing (RB) and semi-supervised
8 learning (SSL). Here, RB algorithms are designed for imbalanced classification, but they do *not* use the unlabeled
9 data. Conversely, SSL algorithms use the unlabeled data but they do *not* consider the class imbalance. We suspect
10 that R1 uses “imbalanced SSL” for denoting RB algorithms: it is not surprising that performances of RB and SSL are
11 comparable as one of them is not superior to the other in general. We will make this point clear in the final draft.

12 (R1) Derivation/confusion in Algorithm 1. We first note that the formal derivation of Algorithm 1 is presented in
13 Sections A and B of the supplementary material. Here, we use X, Y for denoting auxiliary variables for Algorithm 1
14 while x denotes the input. However, to avoid the confusion, we will substitute X, Y to α, β in the final draft.

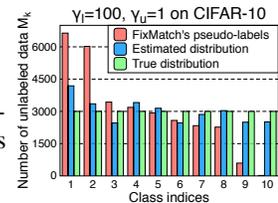
15 (R1) Unclear sentence. Following the R1’s comment, we will clarify “DARP increases...” to “When we apply DARP
16 to the SSL algorithm, the additional running time incurred by DARP is at most 20% of the running time of the vanilla
17 SSL algorithm in our experiments.” in the final draft. We thank R1 for pointing out this.

18 (R2) Discussion for the proposed estimation. We provide the detailed discussion of our estimation scheme in Section
19 F of the supplementary material. The linear relationship of the confusion matrix with the models’ prediction and the
20 true distribution of the unlabeled data is the key of the proposed estimation. We will move this to the main text.

21 (R2) Additional feedback. We thank R2 for careful reading to improve the presentation of our paper. We feel sorry
22 for not providing our response to them due to the space limit. We will clarify all comments in the final draft.

23 (R3) Pseudo-labels for SSL algorithms. As R3 mentioned, most SSL algorithms encourage the original data and the
24 augmented data to have similar predictions, typically by matching the model’s prediction of the original data to some
25 target vector (e.g., the model’s prediction of the augmented data). In our paper, the target vector is denoted by the
26 ‘pseudo-label’, i.e., it is a soft-label and not necessarily one-hot. To clarify the R3’s confusion, we will emphasize our
27 definition of the pseudo-label in the final draft.

28 (R3) True distribution of the unlabeled data. We first clarify that the true distribution
29 of the unlabeled data is NOT necessarily required for DARP. As reported in Table 2 and
30 experimental results on STL-10 in Table 4 of the supplementary material, DARP “with-
31 out the true distribution” is super effective for the imbalanced SSL problem. This is because
32 we estimate the true distribution via our estimation scheme. Our estimation scheme is
33 based on the simple linear relationship (see Section 4.2 and Section F of the supplement-
34 ary material for details) widely used in various problems, e.g., domain adaptation [L⁺18]
35 and noisy labels [H⁺18]. In particular, as illustrated in the right figure, our estimation scheme (blue) effectively
36 approximates the true distribution (green). Namely, DARP can refine pseudo-labels to have distribution close to the true
37 one while the distribution of pseudo-labels generated by FixMatch (red) is highly biased. We will clarify these points
38 and move the detailed discussion of the estimation scheme in the supplementary material (Section F) to the main text.



39 Finally, as mentioned in Line 194, the distribution of the unlabeled data can be also inferred from that of labeled data
40 without such estimation, if both data are sampled from the same distribution (arguably the most practical scenario).

41 [L⁺18] Lipton et al., “Detecting and Correcting for Label Shift with Black Box Predictors”, ICML, 2018.

42 [H⁺18] Hendrycks et al., “Using Trusted Data to Train Deep Networks on Labels Corrupted by Severe Noise”, NeurIPS, 2018.

43 (R4) Computational cost. The computational cost of Algorithm 1 is $O(KM)$ where K, M denote the number of
44 classes and the number of the unlabeled data, respectively. In our experiments, DARP increases at most 20% of the
45 overall training time of the vanilla SSL algorithm. We will add a related discussion in the final draft.

46 (R4) Fail to compare with simple baseline. In fact, DARP has been compared with algorithms that R4 suggested, i.e.,
47 utilizing both re-balancing and the unlabeled data (i.e., SSL), in Section E of the supplementary material. Here, DARP
48 also benefits by re-balancing since it only resolves pseudo-labels’ bias toward the majority classes. We will move this
49 result to the experiment section in the final draft.

50 (R4) Real-world imbalanced dataset. Experimental results on the imbalanced real-world scene categorization dataset
51 (SUN-397) is presented in Section D of the supplementary material. We will clarify this in the final draft.

52 (R4) Related work. In Section 4.3, we compare our method with recent SSL algorithms using the class distribution as
53 prior knowledge and discuss them. Nevertheless, following the R4’s suggestion, we will add a discussion about these
54 algorithms and works in “Semi-Supervised Learning Literature Survey” to the related work section to the final draft.