We thank the reviewers for their valuable suggestions to improve the draft. We address the concerns below.

To all the reviewers.

a) *"Novelty".* Thanks for providing the citations to self-supervised approaches. Although most of them focus on solving specific challenges (that are different from MSDA), they do deserve a discussion, and we will include one in the revision. Our prime contribution is in the form of insights that lead to a simple design, which makes our work different. For *e.g.*, we find that classifier agreement (A) can hint at the migration of target samples during adaptation (Sec. 4.2c, 4.2e). Likewise, we show that even under category-shift (Sec. 4.2b), the model implicitly aligns only the shared classes across domains (Suppl. Sec. 2), which is relatively less explored in MSDA. Outperforming the state-of-the-art is not the objective here, rather, we wish to broaden the perspective of what deep models can achieve under domain-shift, to promote future research in this area. We will emphasize more on this aspect in the revision.

Reviewer 1:

b) *"Two stage training".* Note, after the warm-start, we do have alternating source/target batches (see L8, L10 in Algo. 1, and implementation in Suppl.). The warm-start helps achieve reliable pseudo-labels before introducing target instances.

c) *"Justification of w".* We empirically verify in Sec. 4.2f that w roughly correlates with the accuracy of pseudo-labels.

d) *"Controlling the noisy pseudo-label".* While the warm-start stage (Sec. 3.1), and the strategy in L139-L150 are intended to subdue the noise in pseudo-labels, one could also consider a subset of $\mathcal{D}_t{}'$ having the most confident target samples to further reduce the noise (see Fig. 6b) since correct labels are often predicted with high confidence [33,12,38].

Reviewer 2:

e) *"Comparison with MFSAN".* MFSAN employs auxiliary loss functions at the feature level and at the output level for adaptation. In contrast, we harness implicit alignment exhibited by deep networks. While MFSAN is applicable in the closed-set scenario, SImpAl can be readily applied under category-shift (Sec. 4.2b, Suppl. Sec. 2). Further, MFSAN selects model based on best target performance, while ours is based on the convergence of agreement rate (A).

Reviewer 3:

f) *"Single Classifier".* The reviewer's observation is correct. One can train a single classifier to exploit implicit alignment. However, there would be no way to measure classifier agreement (a) which has multiple benefits. Particularly, we find that classifier agreement indicates more accurate pseudo-labels (Fig. 5b), while also providing a convenient way to monitor the adaptation process (Fig. 5a). Since, multiple classifier heads enable the measurement of classifier agreement as a by-product (at no additional effort), we choose this approach. Note, while there is no upper limit to the number of classifier heads, employing too many heads delays the convergence of the agreement rate A (due to a greater probability of disagreement), thereby requiring more iterations. Hence we fix the number of heads to be n_d.

g) *"Threshold based self-training / DCTN".* Threshold based self-labeling has been carried out in DCTN [45]. However, as demonstrated in DCTN (see Fig. R1 below for reference), employing a threshold based pseudo-labeling alone results in performance degradation (blue curve) in such a method, where domain-specific classifiers are learned. Thus, an auxiliary adversarial alignment loss *is a requirement* in DCTN, to align the source and the target domains for adaptation. In contrast, SImpAl is motivated from the strong inductive bias of deep models to implicitly align the latent features under supervision (Fig. 2). The approach is based on enforcing classifier consistency (in contrast to learning domain-specific classifiers), which allows adaptation without requiring an explicit alignment loss.

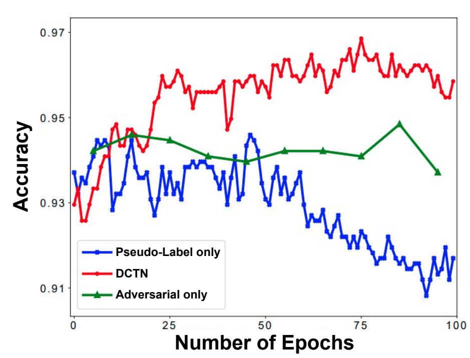

Figure R1: We show Fig. 4a of [45] (DCTN) here for reference (target accuracy vs. training progression). The blue curve corresponds to employing a pseudo-label only algorithm in DCTN (that learns domain-specific classifiers). The performance is observed to deteriorate. See Sec. 5.4 in DCTN [45] for discussion.

Reviewer 4:

h) *"DCTN comparison".* Please refer to point **(g)** above, and the discussion on category-shift in Sec. 4.2b and Suppl. L22-L26.

i) *"Larger domain-gap".* The method also works under large domain gaps, as evaluated on the DomainNet dataset (see Table 1E, and Suppl. Table 1). Further, we demonstrate the alignment of latent features across domains which are vastly different (Quickdraw and Real-world images; see images in Suppl. Fig. 4). Specifically, we are able to perform cross-domain image retrieval (Suppl. Sec. 3). This tool is bundled with the Suppl. code, along with a demonstration video (`image_retrieval_demo.mp4`).