

1 Thanks for all the reviewers' positive feedback and constructive suggestions. We will revise our paper accordingly.

2 **To Reviewer 1**

3 **Q1:** Reason why the method hasn't been tested with newer architectures than DeepLab v2.

4 **A1:** Due to the historical reasons, most previous methods are tested with DeepLab v2. To facilitate a fair comparison,
5 we chose DeepLab v2 as our base architecture. And we will supplement more results based on recent state-of-the-art
6 architectures in our revised version, to setup new baselines for future research.

7 **Q2:** The selection of source pixels to define pixel level association. Are you using all as candidates?

8 **A2:** Yes, we used all the source pixels of a certain image as candidates, other than the ones out of the label set. We will
9 revise this part and make it clearer.

10 **Q3:** At which resolution is the pixel level association between source and target pixel computed?

11 **A3:** The resolution for building the associations is at 92×92 , which is around $1/8$ times as the input resolution $730 \times$
12 730 . We will add this detail in our revised manuscript.

13 **Q4:** The way to integrate your method together with a self-training method.

14 **A4:** There are multiple ways to integrate our method with the self-training. The most conventional way is to use the
15 self-training as a second stage training, after achieving the adaptation model using our method. Another possible way,
16 as discussed by the reviewer, is to integrate the pseudo-labeling into the feature matching. Specifically, an additional
17 cycle-association loss, based on the association starting and ending at the *target* pixel, could be imposed. We remain
18 this as a future work.

19 **Q5:** More related work and more implementation details.

20 **A5:** Thanks. We will cite the suggested papers and add more implementation details in our revised manuscript.

21 **To Reviewer 2**

22 **Q1:** How deep in a network the features should be extracted from. Is multi-layer feature association helpful?

23 **A1:** 1) For the feature extractor, as we discussed in Line 145, we applied pixel-level cycle association to the last-layer
24 feature maps of the backbone. And through back-propagation, the features in previous layers can also be adapted. 2)
25 We have attempted to impose the cycle association loss on the multiple layers of feature maps. Empirically we found
26 that it brings minor improvement, but introduces much additional computation cost. And as most previous works didn't
27 align multi-layer feature maps, we chose to only associate the last-layer features to facilitate a relatively fair comparison.
28 3) However, we believe it remains an open question and valuable to further investigate in future.

29 **Q2:** How often do the cycle-consistent associations find a associated pixel in the target domain of the same, correct
30 label? Is it difficult to find matches with classes that are not often represented?

31 **A2:** 1) Among all the cycle-consistent associations, the average correct ratio is above 70% during training. 2) In
32 terms of the cycle association (Sec. 3.2), rarely-presented classes are relatively harder to match (around 10% can be
33 associated). However, our proposed diffusion module (Sec. 3.3) encourages a diverse set of pixels to be adapted, and
34 enables the adaptation of rarely-presented classes via other similar associated classes. Empirically we found that our
35 method works better on the rarely-presented classes, than the source-only baseline and previous state-of-the-art. For
36 example, on the task SYNTHIA \rightarrow Cityscapes, our method achieves 31.7% mean IoU over the classes with rarest pixels
37 (*i.e.* motorcycle, rider, traffic light, bus, and bicycle), outperforming the source-only baseline (24.0%) and previous
38 state-of-the-art (28.2%) by a large margin.

39 **Q3:** More connections to previous works which are out of the domain adaption area.

40 **A3:** Thanks. We will cite the suggested related papers and add more discussions about the connections to them.

41 **To Reviewer 4**

42 **Q1:** The sensitivity of the model to the hyper-parameter settings.

43 **A1:** We have discussed the selection of α in Line 134-135. And we have shown the sensitivity of our model to the other
44 hyper-parameters $\beta_1, \beta_2, \beta_3$ in the supplementary material due to the space limit. We found that our method is robust to
45 the choices of hyper-parameters and achieves consistent superior performance compared to previous state-of-the-arts.

46 **Q2:** More discussions about previous work.

47 **A2:** Thanks. We will add more discussions about how our method differs from existing works in our revision.