A. Additional Experiments

We demonstrate additional experiments to provide further insights and understanding of the proposed method.

A.1. ImageNet-100

We evaluate different contrastive learning strategies on a larger dataset, *i.e.,* ImageNet-100, which is a subset of ImageNet with randomly selected 100 classes. To make a fair comparison between methods, we fixed the class prior hyperparameter $\tau^+ = 0.01$ for all methods following [1]. The results in Table 1 compare the performance of linear probing of different methods with ResNet-18 backbone on ImageNet-100. We observe that the proposed hard negative sampling method significantly improved the worst group accuracy while maintaining a competitive average accuracy compared to DCL and HCL. This validates our claim that hard negative mining with relative difficulty leveraging bias-amplifying counterpart is effective in larger datasets.

	Acc	Acc (worst)
DCL [1]	70.22	8.16
HCL [2]	66.08	20.41
WCL	69.01	42.77

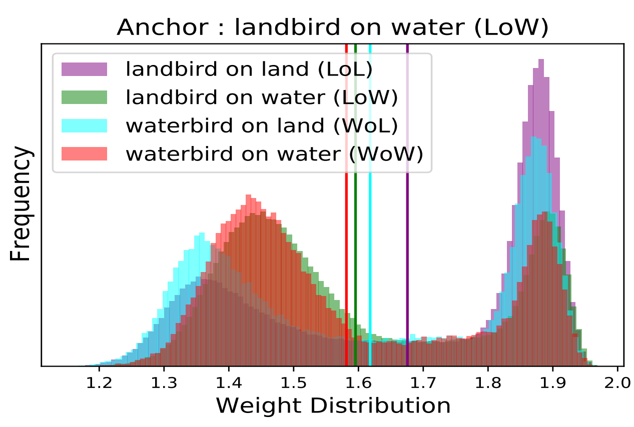
Table 1. Experimental results on ImageNet-100.

A.2. Weight analysis

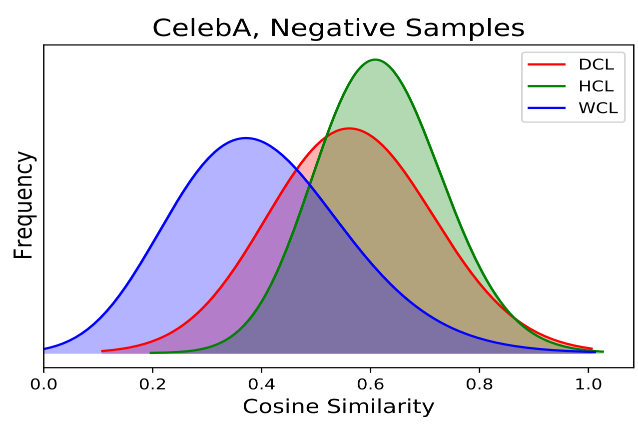
In this subsection, we study quantitative analysis of the proposed reweighing scheme to justify the connection between how contrastive learning model can benefit from it.

[Quantitative analysis of the weight] In Figure 1a, we plot the weight distributions and their mean (vertical line) for different classes when the anchor is "landbird on water (*LoW*)" in CUB dataset. It shows that landbird on land (*LoL*) has the largest peak with high weights and the largest mean among all classes, indicating that *LoL* is the hard negative class. Conversely, waterbird on water (*WoW*) is the easy negative group since it has a low average weight and a small peak at higher weights. Regarding **false negatives**, we observe that they have relatively low weights (*LoW*), suggesting WCL puts less importance on negative samples with the same label as the anchor.

[Cosine similarity analysis] Similar to Robinson *et al.* [2], we study the cosine similarity of the negative samples compared to baselines to support the qualitative result with t-SNE visualization. We plot the cosine similarity distribution of 200K random samples from the CelebA test set in Figure 1b. It shows that the majority of negative samples by WCL have significantly lower cosine similarity than others, indicating WCL effectively contrasts negative samples with the proposed relative difficulty.



(a) Weight distributions and their mean on CUB dataset.



(b) Cosine similarity of negative samples in CelebA.

References

[1] Ching-Yao Chuang, Joshua Robinson, Lin Yen-Chen, Antonio Torralba, and Stefanie Jegelka. Debiased contrastive learning. *arXiv preprint arXiv:2007.00224*, 2020. 1

[2] Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. Contrastive learning with hard negative samples. *arXiv preprint arXiv:2010.04592*, 2020. 1