Reviewer 22E1: Experimental results show the superiority compared with SimCLR in terms of the Top-k NN precision (Table 1-5) and unsupervised classification accuracy (Table 2). However, there is a lack of evaluation on the supervised accuracy with a simple classifier, such as a linear layer. Since the model is iteratively trained to minimize the contrastive loss, I think the Top-k NN precision could be improved in the nature. Therefore, the evaluation on supervised accuracy becomes more important to justify the effect of the propose method.
Reviewer 2254: Introducing the dataset's intra-class variation in a contrastive batch is quite interesting for self-supervised learning. However, there are some limitations:
1. Figure 1 is not clear enough, for example, where "Y" is from?
2. How many samples are in the batch of "Y"? However negative sample in Equation
3. The comparison is not enough, there are many SOTA self-supervised methods are not covered.

We thank the reviewers for their insightful critiques and constructive reviews. We will reflect the necessary changes in the future version. We address some reviewer comments below.

Reviewer **22E1** raised **concerns regarding the KNN classifier and linear evaluation**. It is important to note that the contrastive batch approach minimizes distance in the projection space (projection head's output), and the Top-k precision is computed on the representation space (CNN backbone's output). Also, in paper [1], Zheltonozhskii et al. clarify that the KNN classifier is often used as an alternative to the Linear Evaluation metric. In addition, paper [2] finds that while linear classification directly evaluates the learned representation, it cannot predict performance on downstream tasks. We address this by including a downstream unsupervised classification metric for evaluation [1], which has the advantage of assessing the ability of the representations to capture meaningful structures in the dataset beyond simple linear separability.

Reviewer **2254** raised concerns in two main areas:
Clarity regarding 'Y' in Figure 1: We follow the template used by Chen et al. in the baseline paper [3], representing a foundational architectural diagram for a single sample 'X' from the training split. The caption accompanying Figure 1 specifies that 'Y' originates from the same pseudo-class as the sample 'X.' Subsequently, the definition of 'Y' is intrinsically linked to a singular chosen sample X and is not explicitly defined within a batch setting. In our paper, Algorithm 1 explains the procedure behind batch training.
Comparison with other SOTA methods: Our paper's methodological focus is to introduce a novel approach that can be easily integrated with existing SOTA methods of contrastive learning-based self-supervised training for improvement. Our choice of SimCLR [3] as a baseline method is intentional; it serves as a foundational and widely recognized framework in this domain. It allows us to illustrate the simplicity and effectiveness of integrating our proposed approach into established methodologies. Therefore, comparing the proposed approach with the baseline self-supervised setup, i.e., SimCLR, is sufficient.
Finally, we express our gratitude to the technical committee for diligently reviewing our response.

References
[1]. Zheltonozhskii, Evgenii, et al. "Self-supervised learning for large-scale unsupervised image clustering." *arXiv preprint arXiv:2008.10312*(2020).
[2]. Resnick, Cinjon, et al. "Probing the state of the art: A critical look at visual representation evaluation." *arXiv preprint arXiv:1912.00215*(2019).

[3]. Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *International conference on machine learning*. PMLR, 2020.