We thank all the reviewers for their valuable comments, efforts, and time. As the reviewers highlight, we propose a simple (**R1**,**R2**,**R4**) yet effective (**R1**,**R2**) OOD detection method, supported by strong experimental supports (**R1**,**R2**,**R3**,**R4**) and with a clear presentation (**R1**,**R2**,**R3**,**R4**). We respond to each comment one-by-one in what follows.

Common Responses

[R1/R4] Choice of the shifting transformations. Our principle is to choose shifting transformation that generates the most OOD-like samples. Here, to measure *OOD-ness* of a transformation, we use AUROC between in-distribution vs. transformed samples using vanilla SimCLR (as in line 235-246). Via this concrete selection scheme, we choose "rotation" (Rot.) for CIFAR/ImageNet (see Table 4). We remark the scheme can be used for any real-world image datasets. For example, under Describable Texture Dataset (DTD), Rot. is no longer a shifting transformation (as in line 247-257): the above tables show that (a) "Gaussian noise" (Noise) has higher OOD-ness than Rot. on DTD, and (b) our method (CSI) significantly benefit from using Noise instead of Rot in this case.

(a) OOD-ness		(b) AUROC: DTD (in) vs. Textile (out)		
Rot.	Noise	SimCLR	CSI (Rot.)	CSI (Noise)
50.6	**75.7**	70.3	65.9	**80.1**

[R1/R3] Variance over multiple runs. As suggested by **R1** and **R3**, we will include the variance of our results in the final draft, *e.g.*, the right table presents the mean and standard deviation of the mean AUROC on one-class CIFAR-10 (averaged over 10 classes) over 5 runs.

AUROC on one-class CIFAR-10		
Rot+Trans [1]	GOAD [2]	CSI (ours)
89.79$\pm$0.13	85.06$\pm$0.03	94.27$\pm$0.05

[R1/R4] Related work. Following suggestions of **R1** and **R4**, we will add more discussions on related work, particularly on the connection of our method to the existing self-supervised approaches [1,2], *e.g.*, auxiliary classifier (line 122).

Individual Responses

[R1] Fairness of comparison. We put our best efforts to fairly compare our method with the prior works, especially [1,2], *e.g.*, we already have provided *both* the reported values from [1,2] and results from our re-implementations in Table 1. The only difference we aware of the experimental details is the use of different architecture, *i.e.*, ResNet-18, for our method, but we have compensated this issue by using the same architecture in our re-implementations. Somewhat remarkably, our results on this smaller architecture outperform both results of prior methods with significant margins.

[R1] Motivation on the norm in the scoring function. Intuitively speaking, the contrastive loss we minimize enforces the norm to increase, as it is an easier way to maximize cosine similarity (see Appendix I for more details). One of our key findings is that using the norm of the projection outputs as a score function is a surprisingly strong baseline, compared to other reasonable choices. We refer more discussions on other design choices in Appendix E, F, and G. For better presentation, we will add the relevant discussion in the main text following your suggestion.

[R2] Editorial comments. We will revise our manuscript by clarifying all of the following points: (*i*) line 108-118: We will remark that Eq. (3) also uses the transformations for positive samples (SimCLR augmentations), which were implicitly defined in Eq. (2). (*ii*) line 116-118: We will add the navigation: linear evaluation paragraph in Section 3. (*iii*) line 207: We will rephrase the expression; we stated prior methods often fail, since they show 50% of AUROC (*i.e.*, random guess) under the Place-365 dataset. (*iv*) line 241-246: We will fix it to Table 5.

[R3] Novelty. We believe the key novelty of our method belongs to mainly in two aspects: (*i*) We design a surprisingly effective OOD score functions applicable to any contrastive features, an emerging paradigm for representation learning, based on an extensive empirical justification. (*ii*) We report a novel observation that some existing input transformation techniques (*e.g.*, rotation), *i.e.*, *shifting* transformations, could further improve the contrastive features in terms of OOD detection, under our newly proposed contrastive training scheme. As also highlighted by **R2**, we believe our work could provide novel insights for both representation learning and OOD detection.

[R3] Limited to the visual modality. Following the suggestion, we will revise our paper title to emphasize the visual domain we are focusing on in our experiments. Nevertheless, in principle, our key idea of contrasting transformed inputs is applicable to other domains, *e.g.*, using Gaussian noise applicable for any data types, including temporal data.

[R3] Why augmentation helps OOD detection? We claim some augmentations that largely shift data-distribution would behave like OOD, and contrasting them could help to learn a better discriminative features between in- vs. out-of-distribution (line 42-50,114-118). This hypothesis is validated by an extensive empirical study (line 235-246).

[R3] Other comments. (*i*) Pre-train or scratch?: We train all models from scratch (both ours and others). (*ii*) Threshold: One can set the threshold from training data statistics, controlling the margin for precision/recall trade-off.

[R4] Classification accuracy. Although classification is not our primary focus, we found that our method achieves comparable classification accuracy with SimCLR throughout our experiments, *e.g.*, our method achieves test accuracy of 90.19%, while SimCLR shows 90.48%, under the linear evaluation protocol on CIFAR-10 (see line 260-262).

[1] Hendrycks et al. Using self-supervised learning can improve model robustness and uncertainty. *NeurIPS* 2019.
[2] Bergman and Hoshen. Classification-based anomaly detection for general data. *ICLR* 2020.