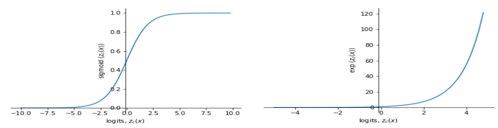
We thank the reviewers for their insightful comments. We first address the major concerns raised by the reviewers, followed by their minor questions/ comments. We shall incorporate their suggestions into the paper.

[R1 & R4] Justification of sigmoid (logistic) function in proposed regularizer (Eq 9, 10). By limiting logits, z_c to values that are (approximately) greater than 5 for in-domain examples, and less than -5 for OOD examples, we would have the desirable sharp uni-modal or multi-modal Dirichlet distributions respectively, maximizing their representation gaps (recall Fig 1; paper). Beyond these values, the cross-entropy loss should be the dominant term in the loss function to improve classification accuracy. The use of sigmoid function in our regularizer satisfies this condition by providing an implicit upper (lower) bounds on the concentration parameters for in-domain (OOD) examples (see Fig(Rebuttal)(a)).

(a) $\frac{1}{K}\sum_c \mathrm{sigmoid} z_c(\boldsymbol{x})$ (b) $\sum_c \exp z_c(\boldsymbol{x})$



Fig(Rebuttal): Growth of regularizers w.r.t logits.

In contrast, using the precision, $\alpha_0 = \sum_c \exp z_c(\boldsymbol{x})$ as the regularizer leads to large logit values for in-domain examples (see Fig(Rebuttal)(b)). However, it makes the cross-entropy loss term negligible (Eq. 9), leading to degrading the in-domain classification accuracy. Further $\exp z_c(\boldsymbol{x})$ is not a symmetric function. Hence, it does not equally constrain the network to produce small fractional concentration parameters (i.e $\alpha_c = \exp z_c(\boldsymbol{x}) \to 0$) for OOD examples, that leads to the desired multi-modal Dirichlet distributions (Fig 1d; paper). Moreover, in practice the choice of $\sum_c \exp z_c(\boldsymbol{x})$ (or $\sum_c z_c(\boldsymbol{x})$) leads the training loss to NaN.

[R2 & R4] "Sometimes have a significant drop" in misclassification detection Table 2. Table 2 presents the AUPR scores for misclassification detection. However, AUPR may not be an ideal metric for comparison, as it greatly depends on the *base rates* i.e no. of misclassifications vs correct predictions (see accuracy vs. AUPR scores in Appendix Table 8) [1]. We instead recommend comparing the AUROC scores in Table 8 (appendix), where we achieve comparable scores with the other non-ensemble based OOD models. Further, DPN is consistent with Bayesian ensemble techniques (Eq 2, without marginalizing $\boldsymbol{\theta}$; Lines [107-115]), which would further improve the misclassification detection task.

Table (Rebuttal) gives the comparison of *root mean square calibration error (RMS)* using the same experimental setup as Hendrycks et al [1]. We achieve similar performances as non-Bayesian OE [1], and better results for C100 and TIM. Our proposed regularizer scales up (or down) the concentration parameters, $\alpha_c = \exp z_c(\boldsymbol{x})$ for in-domain (OOD) examples, without disturbing their relative values i.e $\exp z_c(\boldsymbol{x}) / \sum_c \exp z_c(\boldsymbol{x})$ $\big(= p(\omega_c|\boldsymbol{x}^*, D)$; the predictive categorical$\big)$. Hence, it does not lead to over-confidence for in-distribution examples.

	C10	C100	TIM
Baseline	16.2$\pm$0.0	6.6$\pm$0.3	5.2$\pm$0.0
MCDP	15.7$\pm$0.1	6.7$\pm$0.0	5.3$\pm$0.2
DE	16.1$\pm$NA	6.8$\pm$NA	6.2$\pm$NA
EDL	15.5$\pm$0.1	10.1$\pm$0.4	10.3$\pm$0.4
OE	6.4$\pm$0.4	3.8$\pm$0.1	4.2$\pm$0.1
DPN$_{rev}$	9.2$\pm$0.4	10.4$\pm$0.1	7.2$\pm$0.5
DPN$^+$	**6.3$\pm$0.3**	4.3$\pm$0.0	2.8$\pm$0.3
DPN$^-$	6.5$\pm$0.2	**3.5$\pm$0.1**	**2.7$\pm$0.3**

Table(Rebuttal): Root mean square calibration error

Finally, our proposed maximizes the representation gap between in-domain and OODs to *confidently determine the source of uncertainty* and improves the *OOD detection performance*. Maximizing the gap between in-domain correct predictions and misclassifications is an important and interesting problem for future research.

Reviewer 1: Eq. 12: First term is an expectation on joint dist. $\tilde{P}_T(\boldsymbol{x}, y)$ where $\boldsymbol{x}$ and y are continuous and discrete random variables. Denoting sigmoid (logistic) function as σ (apply $p(x, y) = p(y|x)p(x)$ followed by rearranging):

$$\mathbb{E}_{\tilde{P}_T(\boldsymbol{x},y)}\Big[\sum_{c=1}^{K} \frac{-\lambda_T \sigma(z_c(\boldsymbol{x}))}{K}\Big] = \int_x \Big[\sum_y \Big[\sum_{c=1}^{K} \frac{-\lambda_T \sigma(z_c(\boldsymbol{x}))}{K}\Big] p(y|\boldsymbol{x})\Big] p(x)dx = \mathbb{E}_{P_T(\boldsymbol{x})}\Big[-\frac{\lambda_T}{K}\sum_{k=1}^{K} p(y=\omega_k|\boldsymbol{x})\Big[\sum_{c=1}^{K}\sigma(z_c(\boldsymbol{x}))\Big]\Big]$$

Deep ensemble (DE) results and **In-domain accuracy** are included in Appendix Table 8-12 and Table 8. We have also included the results for **EDL** in Table (Rebuttal) and shall include their remaining results in our main paper.

Reviewer 2: Fig 5: We normalized the scores for better visualization.
Detecting **distributional shift across non-semantic factors** can be useful for in-distribution generalization on specific domains by understanding the limitations of a classifier (Yarin Gal, Ph.D. thesis; 2016).

Reviewer 3: Q 3.1 Appendix A.2 and B.1 provide additional ablation studies and results on the synthetic dataset for Reverse-KL loss to further justify our claims. (Due to space constraints, we could not include them in the main paper.)
Q 3.2 In Table 1 and 2, OE represents the non-Bayesian model (state-of-the-art) by Hendrycks et al [1].
Q 3.3 Our C10, C100 and TIM tasks respectively uses CIFAR-10, CIFAR-100 and TinyImageNet with 10, 100 & 200 classes. Please refer to Appendix Table 7 where we present the experimental setup.

Reviewer 4: Q4: Dataset with highly different characteristics as OOD training set lead to poor OOD detection performance. This is well-studied in (Lee et al., 2018; Hendrycks et al., 2019; Malinin et al., 2019).
Q5: $\sum_c \exp(z_c(\boldsymbol{x}))$ as a uncertainty measure for OE leads to poor OOD detection performances (in most of the cases) as it does not control the absolute values of $\exp z_c(\boldsymbol{x})$ terms.

Reference: *[1] Deep Anomaly Detection with Outlier Exposure (Hendrycks et al., ICLR 2019)*