Dual Energy-Based Model with Open-World Uncertainty Estimation for Out-of-distribution Detection

Supplementary Material

8. The derivation of Equation (20)

Based on Equation (3), we have

$$\begin{aligned}&\partial_\theta D_{\mathrm{KL}}\left(p_{\text{data}}\left(x, y^{+}\right) \| p_\theta(\tilde{x}, y^{+})\right) \\ &= \mathbb{E}_{x, y^{+} \sim p_{\text{data}}\left(x, y^{+}\right)}\left[\partial_\theta E_\theta(x, y^{+})\right] \\ &- \mathbb{E}_{\tilde{x}, y^{+} \sim p_\theta\left(x, y^{+}\right)}\left[\partial_\theta E_\theta(\tilde{x}, y^{+})\right] \end{aligned} \tag{30}$$

Building upon the result presented in Equation (30), we can leverage this gradient to guide the model towards minimizing the discrepancy between the data distribution $p_{\text{data}}\left(x, y^{+}\right)$ and the model distribution $p_\theta(\tilde{x}, y^{+})$. Then we have

$$\mathbb{E}_{x, y^{+} \sim p_{\text{data}}\left(x, y^{+}\right)}\left[E_\theta(x, y^{+})\right] - \mathbb{E}_{\tilde{x}, y^{+} \sim p_\theta\left(x, y^{+}\right)}\left[E_\theta(\tilde{x}, y^{+})\right] \tag{31}$$

$$\begin{aligned}&= \mathbb{E}_{x, y^{+} \sim p_{\text{data}}\left(x, y^{+}\right)}\left[\left(-\log \mathtt{P}_{\mathtt{conf}}(y^{+}|x)\right)\right] \\ &- \mathbb{E}_{\tilde{x}, y^{+} \sim p_\theta\left(x, y^{+}\right)}\left[-\log \mathtt{P}_{\mathtt{conf}}(y^{+}|\tilde{x})\right] \\ &= \mathbb{E}_{x, y^{+} \sim p_{\text{data}}\left(x, y^{+}\right)}\left[-\log \mathtt{P}_{\mathtt{conf}}(y^{+}|x)\right] \\ &+ \mathbb{E}_{\tilde{x}, y^{+} \sim p_\theta\left(x, y^{+}\right)}\left[\log(1 - \sum_{k \neq y^{+}} \mathtt{P}_{\mathtt{conf}}(y^{+}|\tilde{x}))\right] \\ &\leq \mathbb{E}_{x, y^{+} \sim p_{\text{data}}\left(x, y^{+}\right)}\left[\left(-\log \mathtt{P}_{\mathtt{conf}}(y^{+}|x)\right)\right] \\ &+ \mathbb{E}_{\tilde{x}, y^{+} \sim p_\theta\left(x, y^{+}\right)}\left[\left(\log(1 - \mathtt{P}_{\mathtt{conf}}(y^{-}|\tilde{x}))\right)\right] \end{aligned} \tag{32}$$

The inequality in the final step arises from the fact that $\sum_{k \neq y^{+}} \mathtt{P}_{\mathtt{conf}}(y^{+}|\tilde{x}) \geq \mathtt{P}_{\mathtt{conf}}(y^{-}|\tilde{x})$. This derivation provides an upper bound for the difference between the expectations under the data distribution and the model distribution.

9. More Results of DEBO

More results on different hyper-parameters λ_1 and λ_2 in Equation (25), as illustrated in Table 4.

The impact of hyperparameter γ. As shown in Figure 2, we investigate the influence of different hyperparameters γ in Equation (29) on the performance of OOD detection. The experimental results indicate that on the CIFAR10 and CIFAR100 datasets, when γ is fixed at 1, the AUROC performance metric exhibits more excellent performance. On the ImageNet-200 dataset, however, when γ takes a value of 0.3, the OOD detection performance reaches its optimum. This fully demonstrates that the introduction of the hyperparameter γ can significantly enhance the model's adaptability to different data distributions.

Datasets	λ_1	λ_2	FPR95↓	AUROC↑	AUPR-In↑	AUPR-Out↑	ID-ACC↑
CIFAR10	0.1	0.1	15.51	96.07	94.20	96.58	94.67
	0.5	0.5	16.24	95.94	94.55	96.20	94.59
	1.0	1.0	16.69	96.10	93.54	96.65	94.49
	0.1	1.0	14.41	96.67	94.53	97.19	94.86
CIFAR100	0.1	0.1	47.93	85.77	80.94	87.11	75.98
	0.5	0.5	38.13	88.42	83.45	89.45	75.48
	1.0	1.0	33.73	89.76	86.30	89.97	76.26
	0.1	1.0	43.68	87.35	81.72	88.39	75.06
ImageNet-200	0.1	0.1	41.12	88.10	83.24	87.03	86.26
	0.5	0.5	42.70	86.65	82.03	85.00	85.96
	1.0	1.0	42.59	86.99	82.46	86.16	85.27
	0.1	1.0	42.12	87.09	82.77	86.29	85.87

Table 4. More results on different hyper-parameters λ_1 and λ_2 in Equation (25).

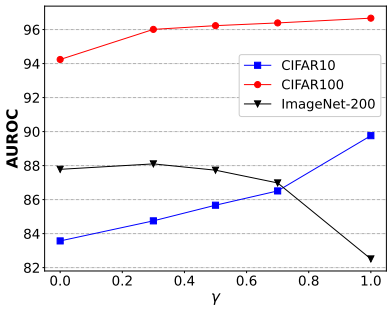

Figure 2. The impact of hyperparameter γ.

Different numbers of negative classes. We further conduct an ablation study to investigate the impact of varying the number of negative classes in our proposed DEBO, as illustrated in Table 5. Notably, when the number of negative classes is set to 1, we deviate from the conditional entropy minimization approach defined by $C_{\mathtt{en}}$ in Equation (24). Instead, we employ a cross-entropy loss function to maximize the confidence of the (K+1)-th class. This modification can be formally expressed as:

$$\mathbb{E}_{\tilde{x} \sim p_{\text{ood}}(x)}\left[-\log \mathtt{P}_{\mathtt{conf}}(K+1|x)\right] \tag{33}$$

The remaining components of the DEBO loss function remain consistent across all configurations. We systematically adjust the negative classes to correspond with the dif-

	Number of negative classes	FPR95↓	AUROC↑	AUPR-In↑	AUPR-Out↑	ID-ACC↑
CIFAR10	1	16.11	96.18	94.22	96.73	93.97
	K	14.41	96.67	94.53	97.19	94.86
	2K	18.10	96.03	93.22	96.68	94.81
CIFAR100	1	45.22	85.27	81.60	85.77	75.48
	K	33.73	89.76	86.30	89.97	76.26
	2K	43.41	86.01	81.31	86.38	74.38
ImageNet-200	1	41.77	87.85	82.92	86.58	86.09
	K	41.12	88.10	83.24	87.03	86.26
	2K	41.44	87.93	83.06	86.79	81.81

Table 5. Ablation study on the number of negative classes in DEBO framework.

ferent class configurations examined in this analysis. As evidenced in Table 5, our results demonstrate that setting the number of negative classes to K yields the best performance.

Different model architectures. In this section, comparative experiments are further conducted using CIFAR100 as the ID training dataset on different model architectures, specifically ResNet-34. As shown in Table 6, compared with the comparison methods, our DEBO consistently demonstrates superior performance.

Methods	FPR95↓	AUROC↑	AUPR-In↑	AUPR-Out↑	ID-ACC↑
MSP	55.76	78.97	73.71	81.03	77.32
EBO	55.65	80.17	73.54	82.39	77.32
CIDER	57.83	79.35	72.66	81.11	N/A
OE	53.13	83.18	76.02	86.05	77.93
EBO (w. D_{aux})	48.47	84.72	77.41	87.14	72.03
UDG	67.03	74.58	67.21	76.81	72.60
MIXOE	66.65	74.60	67.08	77.09	75.51
MCD	54.86	78.40	72.53	79.24	75.68
S_{dec} (ours)	**42.75**	**87.63**	**82.46**	**88.86**	77.17

Table 6. Comparative experiments using different model architectures (ResNet-34) with CIFAR100 as the ID training dataset.

Different OOD auxiliary training data. We further conduct comparative experiments using different auxiliary OOD training datasets with CIFAR100 as the ID training dataset. This OOD dataset consists of data generated by a generative model, as presented in the paper "DREAM-OOD"[6]. As shown in Table 7, compared with the baseline methods, our method consistently demonstrates superior performance.

Methods	FPR95↓	AUROC↑	AUPR-In↑	AUPR-Out↑	ID-ACC↑
OE	42.40	83.81	78.07	87.50	**76.76**
EBO (w. D_{aux})	42.81	83.71	78.12	86.69	74.58
UDG	50.25	82.97	77.01	86.68	65.42
MIXOE	51.09	82.49	76.81	84.53	75.72
MCD	51.61	79.79	74.60	80.55	76.30
S_{dec} (ours)	**38.91**	**86.25**	**81.40**	**88.42**	76.24

Table 7. Comparative experiments using different auxiliary OOD training datasets with CIFAR100 as the ID training dataset.

Methods	ACC.↑	IS↑
JEM [9]	92.9	8.76
Ours	92.34	8.77

Table 8. Results on CIFAR10. The ACC. denotes the classification accuracies and IS is widely used metric for evaluating the performance of generative models.

10. Results of Dual Energy-Based Model

We employ the proposed Dual Energy-Based Model to train a hybrid discriminative-generative model, demonstrating its capability to jointly model both label and data distributions. Table 8 presents the accuracy and generation performance of our Dual Energy-Based Model using a Wide-ResNet 28-10 backbone architecture on the CIFAR-10 dataset. Figure 3 illustrates the qualitative results of the generated data.

It is noteworthy that training this hybrid discriminative-generative model necessitates increased computational resources and time due to the incorporation of SGLD sampling in the training process. Moreover, the primary focus of this paper remains on OOD detection, specifically the Dual Energy-Based Model for OOD Detection (DEBO) method proposed herein.

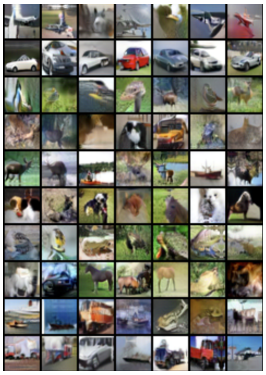

Figure 3. Generated images of our proposed Dual Energy-Based Model.

11. Hardware and Software

All experiments are conducted using PyTorch and Python 3.8, leveraging the computational capabilities of NVIDIA A6000 GPUs.

12. Results on Individual Datasets.

We present comprehensive comparative results of several representative methods on each OOD test dataset, as shown in Table 9, specifically for the CIFAR100 dataset.

Table 9. Detailed comparative experiments using CIFAR100 as the ID training dataset.

Methods	OOD test datasets	FPR95↓	AUROC↑	AUPR-In↑	AUPR-Out↑
MSP	cifar10	63.27	77.10	77.82	74.75
	mnist	51.31	76.83	55.15	94.26
	svhn	61.57	78.58	63.81	89.94
	texture	65.21	75.34	84.15	60.49
	places365	57.61	78.68	60.28	91.34
	LSUN-resize	49.68	82.86	84.18	81.71
	iSUN	50.99	82.02	84.76	78.67
	LSUN-Crop	50.21	80.87	83.17	78.55
EBO	cifar10	64.22	77.42	77.62	74.80
	mnist	51.21	75.63	53.80	93.43
	svhn	53.13	84.12	71.53	92.64
	texture	66.47	75.73	84.07	61.69
	places365	58.26	79.14	60.06	91.37
	LSUN-resize	41.13	87.81	88.49	87.11
	iSUN	44.48	86.49	88.25	84.10
	LSUN-Crop	41.01	84.36	86.79	81.10
OE	cifar10	62.58	75.87	77.70	71.26
	mnist	46.03	85.83	62.54	97.26
	svhn	41.61	88.88	79.41	95.17
	texture	56.76	82.45	88.35	75.03
	places365	58.32	78.88	60.24	91.52
	LSUN-resize	48.44	86.44	86.61	86.86
	iSUN	47.96	86.81	87.96	86.17
	LSUN-Crop	47.56	83.15	85.01	80.75
EBO (w. D_{aux})	cifar10	68.38	73.94	74.46	70.28
	mnist	51.77	89.70	57.40	98.31
	svhn	47.80	84.73	74.40	93.39
	texture	65.06	79.92	85.84	72.44
	places365	64.16	77.12	55.72	90.96
	LSUN-resize	38.82	87.90	89.08	85.61
	iSUN	39.22	88.39	90.14	85.61
	LSUN-Crop	27.72	92.80	93.32	91.75
CIDER	cifar10	81.24	68.06	66.54	65.94
	mnist	65.30	76.30	46.11	95.43
	svhn	19.12	97.06	92.04	99.00
	texture	55.29	84.45	89.71	77.71
	places365	69.00	75.53	52.36	90.74
	LSUN-resize	64.17	77.81	78.15	76.10
	iSUN	69.24	75.48	77.22	72.66
	LSUN	64.51	81.99	80.42	83.61
Ours	cifar10	63.06	75.09	75.98	69.85
	mnist	16.02	96.92	83.87	99.47
	svhn	20.14	94.90	90.61	97.61
	texture	50.08	86.61	91.12	80.24
	places365	54.46	81.16	64.02	92.40
	LSUN-resize	17.39	96.14	96.35	95.49
	iSUN	14.91	96.63	97.11	95.65
	LSUN-Crop	33.82	90.62	91.37	89.04