OODD: Test-time Out-of-Distribution Detection with Dynamic Dictionary

Supplementary Material

A. Experiments

A.1. More Detailed Results

In Table 8, we provide more detailed experimental results on the CIFAR and ImageNet benchmarks across various types of outliers (C-Out, T-Out, D-Out, None).

A.2. Discussion of Computational Cost

In Table 4, we present the computational overhead of our method. Unlike the naive KNN detection, our approach relies solely on dot products after normalization, significantly reducing overhead, especially with parallel computation. This is equivalent to adding a single linear layer, introducing an additional matrix multiplication with complexity $\mathcal{O}(d \cdot l)$ per sample, where d is the feature dimension, and l denotes the priority queue size. Updating the priority queue has a complexity of $\mathcal{O}(\log l)$. Since $l \leq 2048$ in all experiments, the extra overhead when integrating our method with others (e.g., energy-based or maxlogits) remains negligible.

B. Theoretical Analysis

Setup. The decision function for OOD detection using KNN [31] is given by:

$$G(\mathbf{z}^*; k) = \mathbf{1}\{-r_k(\mathbf{z}^*) \geq \lambda\}, \tag{6}$$

where $r_k(\mathbf{z}^*) = \left\|\mathbf{z}^* - \mathbf{z}_{(k)}\right\|_2$ represents the Euclidean distance to the k-th nearest neighbor, λ is the threshold and $\mathbf{1}\{\cdot\}$ is the indicator function. According to Bayes' rule, the probability of $\mathbf{z_i}$ being ID data is:

$$\hat{p}(g_i = 1 \mid \mathbf{z}_i) = \frac{(1-\varepsilon)\hat{p}_{in}(\mathbf{z}_i)}{(1-\varepsilon)\hat{p}_{in}(\mathbf{z}_i) + \varepsilon \hat{p}_{out}(\mathbf{z}_i)}. \tag{7}$$

where $p(g_i = 0) = \varepsilon$, and the components are defined as: $\hat{p}_{in}(\mathbf{z}_i) = \frac{k}{c_b n (r_k(\mathbf{z}_i))^{m-1}}$,where c_b is constant, $\hat{p}_{\text{out}}(\mathbf{z}_i) = \hat{c}_0 \mathbf{1}\left\{\hat{p}_{\text{in}}(\mathbf{z}_i;) < \frac{\beta\varepsilon\hat{c}_0}{(1-\beta)(1-\varepsilon)}\right\}$. If the OOD dictionary contains $\tilde{\varepsilon}$ proportion of OOD samples out of l total samples, the calibrated probability functions are defined as: $\tilde{p}_{in}(\mathbf{z}_i) = \hat{p}_{in}(\mathbf{z}_i) - \sigma_{in}\hat{p}_{in}(\mathbf{z}_i)$, $\tilde{p}_{out}(\mathbf{z}_i) = \hat{p}_{out}(\mathbf{z}_i) - \sigma_{out}\hat{p}_{out}(\mathbf{z}_i)$, where $\sigma_{in} = \eta_1(1-\tilde{\varepsilon})\tilde{p}_{cal}(\mathbf{z}_i)$, $\sigma_{out} = \eta_2\tilde{\varepsilon}\tilde{p}_{cal}(\mathbf{z}_i)$, $\tilde{p}_{cal}(\mathbf{z}_i) = \frac{\tilde{k}}{c_b l (\tilde{r}_k(\mathbf{z}_i))^{m-1}}$, and $\tilde{r}(\mathbf{z}_i)$ is the distance obtained via KNN OOD detection using the OOD dictionary. η_1, η_2, $\tilde{k}$ are constants, and $\tilde{\varepsilon}$ is bounded by a function related to η_1 and η_2.

Theorem 1 *Given the setup above, if* $\hat{p}_{out}(\mathbf{z}_i) = \hat{c}_0\mathbf{1}\left\{\hat{p}_{in}(\mathbf{z}_i;) < \frac{\beta\varepsilon\hat{c}_0}{(1-\beta)(1-\varepsilon)}\right\}$, $\tilde{\varepsilon} > \frac{\eta_1}{\eta_1+\eta_2}$ *and* $\lambda = -\sqrt[m-1]{\frac{(1-\beta)(1-\varepsilon)k}{\beta\varepsilon c_b n \hat{c}_0}}$, *then there exists* $\delta > 0$ *such that:*

$$\mathbf{1}\{-r_k(\mathbf{z}_i) \geq \lambda\} = \mathbf{1}\{\tilde{p}(g_i = 1 \mid \mathbf{z}_i) \geq \beta + \delta\} \tag{8}$$

Lemma 1 *With the setup specified above and the premise of Theorem 1, the decision boundary can be rewritten as:*

$$\mathbf{1}\{-r_k(\mathbf{z}_i) \geq \lambda\} = \mathbf{1}\{\hat{p}(g_i = 1 \mid \mathbf{z}_i) \geq \beta\}$$
$$= \mathbf{1}\left\{\frac{k(1-\varepsilon)}{k(1-\varepsilon) + \varepsilon c_b n \hat{p}_{\text{out}}(\mathbf{z}_i)(r_k(\mathbf{z}_i))^{m-1}} \geq \beta\right\}. \tag{9}$$

The proof of Lemma 1 is given in [31].

Proof. The calibrated probability of $\mathbf{z_i}$ being ID data is:

$$\tilde{p}(g_i = 1 \mid \mathbf{z}_i) = \frac{(1-\varepsilon)\tilde{p}_{in}(\mathbf{z}_i)}{(1-\varepsilon)\tilde{p}_{in}(\mathbf{z}_i) + \varepsilon\tilde{p}_{out}(\mathbf{z}_i)}$$
$$= \frac{(1-\varepsilon)\hat{p}_{in}(\mathbf{z}_i)}{(1-\varepsilon)\hat{p}_{in}(\mathbf{z}_i) + \frac{1-\sigma_{out}}{1-\sigma_{in}}\varepsilon\hat{p}_{out}(\mathbf{z}_i)}. \tag{10}$$

Since $\tilde{\varepsilon} > \frac{\eta_1}{\eta_1+\eta_2}$, using $\frac{1-\sigma_{out}}{1-\sigma_{in}} = \frac{1-\eta_2\tilde{\varepsilon}\tilde{p}_{cal}}{1-\eta_1(1-\tilde{\varepsilon})\tilde{p}_{cal}} < \frac{1-\eta_2\frac{\eta_1}{\eta_1+\eta_2}\tilde{p}_{cal}}{1-\eta_1(1-\frac{\eta_1}{\eta_1+\eta_2})\tilde{p}_{cal}} = 1$, and combining with Lemma 1, we obtain:

$$\tilde{p}(g_i = 1 \mid \mathbf{z}_i) > \frac{(1-\varepsilon)\hat{p}_{in}(\mathbf{z}_i)}{(1-\varepsilon)\hat{p}_{in}(\mathbf{z}_i) + \varepsilon\hat{p}_{out}}$$
$$= \frac{k(1-\varepsilon)}{k(1-\varepsilon) + \varepsilon c_b n \hat{p}_{\text{out}}(\mathbf{z}_i)(r_k(\mathbf{z}_i))^{m-1}} \geq \beta. \tag{11}$$

Finally, there exists $\delta > 0$ such that:

$$\mathbf{1}\{-r_k(\mathbf{z}_i) \geq \lambda\} = \mathbf{1}\{\tilde{p}(g_i = 1 \mid \mathbf{z}_i) \geq \beta + \delta\}$$

Table 8. Detailed results of proposed approach performance with different outlier types across datasets. CIFAR-10/100 uses D-Out from Tiny-ImageNet 597, whereas ImageNet-200 uses D-Out from ImageNet-800 [38].

		C-Out		T-Out		D-Out		None	
ID dataset	OOD dataset	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑
CIFAR-10	CIFAR-100	40.66	89.76	44.70	89.58	47.24	89.38	43.14	88.98
	Tiny-ImageNet	31.37	92.16	31.23	92.56	31.43	92.58	31.62	91.89
	Near OOD	36.01	90.96	37.97	91.07	39.34	90.98	37.38	90.44
	MNIST	8.92	98.01	7.26	98.35	7.49	98.24	8.16	98.15
	SVHN	13.31	96.73	10.97	97.35	14.38	96.57	12.79	96.92
	Textures	20.16	95.04	18.61	95.33	21.14	94.87	23.64	94.20
	Places365	27.35	93.32	25.80	93.92	25.96	93.90	27.12	93.55
	Far OOD	17.44	95.77	15.66	96.24	17.24	95.89	17.93	95.70
CIFAR-100	CIFAR-10	58.60	81.46	59.12	81.51	61.22	80.90	61.09	81.06
	Tiny-ImageNet	55.36	82.73	55.32	82.78	54.39	82.96	55.39	82.74
	Near OOD	56.98	82.10	57.22	82.15	57.81	81.93	58.24	81.90
	MNIST	3.21	99.37	2.22	99.53	2.48	99.48	2.51	99.48
	SVHN	9.44	98.54	6.70	98.90	7.61	98.78	9.49	98.55
	Textures	35.75	91.47	34.37	92.29	34.84	91.92	34.40	92.07
	Places365	50.55	85.14	50.51	85.35	51.24	85.09	50.92	85.20
	Far OOD	24.74	93.64	23.45	94.02	24.04	93.82	24.33	93.83
ImageNet-200	SSB-hard	64.62	81.76	59.02	84.41	64.94	81.54	64.21	81.66
	NINCO	42.82	89.72	38.26	91.17	42.74	89.76	43.49	89.71
	Near OOD	53.70	85.74	48.64	87.79	53.84	85.65	53.85	85.69
	iNaturalist	9.14	98.54	3.32	99.27	10.48	98.29	9.73	98.48
	Textures	25.06	95.24	21.02	96.21	23.98	95.46	25.44	95.23
	OpenImage-O	28.44	92.67	26.13	94.10	29.04	92.34	29.22	92.63
	Far OOD	20.89	95.47	16.83	96.53	21.17	95.36	21.47	95.45
ImageNet-1K	SSB-hard	80.35	65.90	77.59	67.73	- -	- -	80.36	66.00
	NINCO	62.48	79.41	59.80	81.13	- -	- -	63.06	79.51
	Near OOD	71.41	72.66	68.70	74.43	- -	- -	71.71	72.75
	iNaturalist	21.54	95.75	13.14	97.51	- -	- -	24.30	95.03
	Textures	21.76	96.18	20.62	96.49	- -	- -	21.65	96.25
	OpenIamge-O	39.61	89.60	36.85	91.25	- -	- -	40.40	89.12
	Far OOD	27.63	93.84	23.54	95.08	- -	- -	28.79	93.47