Supplementary Material for Enhancing the Power of OOD Detection via Sample-Aware Model Selection

1. Proof of Theorem 1

Theorem. Suppose that we have access to a pre-trained model zoo denoted by $\mathcal{M} = \{\phi_1, \phi_2, \ldots, \phi_m\}$ and let the target TPR level be $1-\alpha$ with $\alpha \leq 0.5$. If the test input $\mathbf{x}^*$ is an ID sample that $\mathbf{x}^* \sim P_{\mathbf{x}}$ and $\mathbf{s}_j^* = S(\mathbf{x}^*; \phi_j)$ is independent of $\mathbf{s}_{j'}^* = S(\mathbf{x}^*; \phi_{j'})$ for $\forall j \neq j'$, then Algorithm 1 can identify $\mathbf{x}^*$ as an ID sample with probability not less than $1-\alpha$.

Sketch of proof. According to Section 3.2,

$$p_j = \mathbb{P}\big(S(\mathbf{x}; \phi_j) \leq \mathbf{s}_j^* \big| \mathbf{x} \sim P_{\mathbf{x}}\big) = F(\mathbf{s}_j^*; \phi_j) \sim U[0,1],$$

where $\phi_j \in \mathcal{M}$. Therefore, the density function of p_j is given by

$$f_{p_j}(x) = \begin{cases} 1 & x \in [0,1]; \\ 0 & \text{otherwise.} \end{cases}$$

Then, the joint probability density of the ordered p-values $p_{(1)}, p_{(2)}, ..., p_{(m)}$ is

$$f_{p_{(1)},p_{(2)},...,p_{(m)}}(x_1, x_2, ..., x_m) = m! \prod_{j=1}^{m} f_{p_j}(x_j) = m!$$

Write $\mathcal{E} = \big\{\forall\, 1 \leq j \leq m,\, p_{(j)} \geq \frac{j}{m}\alpha\big\}$. Then we have

$$\begin{aligned}\mathbb{P}(\mathcal{E}|\mathbf{x}^* \sim P_{\mathbf{x}}) &= \int_{\frac{m}{m}\alpha}^{1} \cdots \int_{\frac{1}{m}\alpha}^{1} f_{p_{(1)},\cdots,p_{(m)}}(x_1, \cdots, x_m)\mathrm{d}x_1 \cdots \mathrm{d}x_m \\ &= m!(1-\frac{1}{m}\alpha)(1-\frac{2}{m}\alpha)...(1-\frac{m}{m}\alpha).\end{aligned}$$

Next, we prove that for any $m \geq 1$ and $\alpha \leq 0.5$,

$$m!(1-\frac{1}{m}\alpha)(1-\frac{2}{m}\alpha)...(1-\frac{m}{m}\alpha) \geq 1-\alpha \tag{1}$$

It is easy to see that Eq.(1) holds when $m = 1$. Suppose Eq.(1) holds for $m = m_0$. Then for $m = m_0 + 1$, we have

$$\begin{aligned}&(m_0+1)!(1-\frac{\alpha}{m_0+1})(1-\frac{2\alpha}{m_0+1})...(1-\frac{m_0+1}{m_0+1}\alpha) \\ \geq &(m_0+1)(1-\alpha)(1-\frac{m_0+1}{m_0+1}\alpha) \geq 1-\alpha,\end{aligned}$$

which implies that Eq.(1) also holds for $m = m_0 + 1$. Hence, the proof is finished. □

2. Proof of Proposition 2

Proposition. Assuming an OOD sample $\mathbf{x}^* \sim Q$, we consider a fixed proportion π of pre-trained models capable of recognizing $\mathbf{x}^*$ as an OOD sample. We further assume for any $0 \leq u \leq 1$, $\mathbb{P}\big(p_j \leq u | \phi_j \in \mathcal{A}(\mathbf{x}^*; \mathcal{M})\big) = G(u)$, where

Table 1-S. The result of model selection.

		Ground Truth		
		Inactive	Active	
Select	Inactive	U	T	
	Active	V	S	k
		m_0	m_1	m

Table 2-S. Supplement to Table 2 in the body of the paper: Compare ZODE-KNN detector with single-model KNN detectors. The ID dataset is CIFAR10. All values are percentages. ↓ indicates smaller values are better and vice versa.

Method		**OOD Dataset**											
		SVHN		**LSUN**		**iSUN**		**Texture**		**Places365**		**Average**	
	TPR	FPR↓	AUC↑	FPR↓	AUC	FPR↓	AUC↑	FPR↓	AUC	FPR↓	AUC↑	FPR↓	AUC↑
ResNet18	95.00	27.97	95.49	18.50	96.84	24.68	95.52	26.74	94.97	47.95	90.02	29.17	94.57
ResNet18*	95.00	2.42	99.52	1.78	99.48	20.06	96.74	8.09	98.57	22.82	95.32	11.03	97.93
ResNet34	95.00	26.53	95.85	10.22	98.39	29.45	95.15	31.65	94.53	36.59	92.75	26.89	95.33
ResNet50	95.00	17.31	97.40	7.10	98.83	17.32	97.26	20.85	96.59	41.35	91.61	20.79	96.34
ResNet101	95.00	25.73	96.12	6.65	98.90	19.84	96.80	18.42	96.89	40.57	92.15	22.24	96.17
ResNet152	95.00	34.96	94.98	7.22	98.88	22.30	96.66	20.76	96.60	38.57	92.36	24.76	95.90
DenseNet	95.00	10.22	98.18	7.90	98.60	10.87	97.94	20.78	96.25	50.14	88.92	19.98	95.98
SwinV2-B256	95.00	14.09	98.01	24.98	96.15	61.61	89.82	0.04	100.00	1.14	99.71	20.37	96.74
SwinV2-B384	95.00	28.23	96.48	40.62	94.63	57.80	90.35	0.04	100.00	1.12	99.70	25.56	96.23
SwinV2-L256	95.00	7.87	98.68	14.59	97.35	20.50	96.69	0.02	100.00	0.71	99.82	8.74	98.51
ZODE-KNN(sub-zoo)	94.96	2.12	99.43	1.50	99.61	5.48	98.70	0.16	99.88	9.91	97.99	3.83	99.12
ZODE-KNN	95.14	0.11	99.89	2.18	99.49	5.79	98.62	0.00	100.00	0.00	100	1.62	99.60

$\mathcal{A}(\mathbf{x}^*;\mathcal{M})$ refers to the set of active models that classify $\mathbf{x}^*$ as OOD, i.e., $\mathcal{A}(\mathbf{x}^*;\mathcal{M}) = \{\phi : \phi \in \mathcal{M}, G(\mathbf{x}^*;\phi) = \text{OOD}\}$, and $G(u)$ is a distribution different from the uniform distribution $U[0,1]$ and satisfies $(1-\pi) + \pi G^{'}(0) > \frac{1}{\alpha}$. Then, as the number of pre-trained models approaches infinity, ZODE demonstrates the capability to identify OOD samples with a high probability.

Sketch of proof. Referring to Section 3.2, we have $\mathbb{P}(p_j \le u|\phi_j \notin \mathcal{A}(\mathbf{x}^*;\mathcal{M})) = u$ for all $0 \le u \le 1$. Hence, the p-values of $\mathbf{x}^*$ follow a mixture model with distribution function

$$F(u) = (1-\pi)u + \pi G(u).$$

Suppose Table 1-S presents the model selection result using Algorithm 1. We note that Algorithm 1 successfully detects $\mathbf{x}^*$ as an OOD sample if $S \ge 1$. Therefore we consider the expectation of S and check

$$\mathbb{E}\big(\frac{S}{m_1}\big) = \mathbb{E}\big(\frac{k \times \frac{S}{k}}{m_1}\big) = \mathbb{E}\big(\frac{\frac{k}{m} \times (1-\frac{V}{k})}{\frac{m_1}{m}}\big).$$

According to Chi [2], $\frac{k}{m}$ converges to a positive value $p_*(\alpha, F)$ as $m \to \infty$ when $F^{'}(0) > \frac{1}{\alpha}$ and $F^{'}(0) < +\infty$, which serves as the limit of the proportion of selected pre-trained models. By Theorem 1 and its lemma in Benjamini and Hochberg [1], we have $\mathbb{E}\big(1-\frac{V}{k}\big) \ge 1-\frac{m_0}{m}\alpha = 1-(1-\pi)\alpha$. If m is sufficiently large, then $\mathbb{E}\big(\frac{S}{m_1}\big) \ge \frac{p_*(\alpha,F)(1-(1-\pi)\alpha)}{\pi} \ge \frac{1}{\pi m} = \frac{1}{m_1}$. Since $\mathbb{P}(S>0) = 1-\mathbb{P}(S-E(S) \le -E(S)) \ge 1-\mathbb{P}(|S-E(S)| \ge E(S)) \ge 1-\frac{\sigma(S)}{(E(S))^2}$, if m is sufficiently large and the variance of S, denoted as $\sigma(S)$, is small enough, there exists a high probability that S will be greater than or equal to 1.

3. Supplement to CIFAR experiments

In the CIFAR experiment, we constructed a large model zoo (FPR/AUC: 1.62/99.60) that incorporated more than 20 models, including prominent models such as ResNet, DenseNet, WideResNet, ResNext50 and SwinV2. From these, we handpicked a sub-zoo with superior performance (FPR/AUC: 3.8/99.1) which includes 7 models displayed in Table 2.

Making special mention, our analysis incorporated three structurally different SwinV2 models with multicore architecture. We have compiled detailed results of these SwinV2 models and added the SwinV2 models to our model zoo to provide pooled performance metrics. In Table 2-S, we list some of the single-model results of experiments on CIFAR10, as well as the integrated results of the large model zoo we constructed, and list the results of the sub-zoo we selected (i.e., Table 2 in the text) for reference.

Compared with the Resnet and Densenet models, the Swinv2 model performs particularly well on the data sets Texture and Places365. It complements other models in the model zoo to supplement image features. Compared with the sub zoo with 7 models, the large model including Swinv2 model zoo observed a significant improvement in FPR/AUC (from 3.83/99.12 to 1.62/99.60).

References

[1] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995. 2

[2] Zhiyi Chi. On the performance of fdr control: constraints and a partial solution. *The Annals of Statistics*, 35(4):1409–1431, 2007. 2