Appendix of *Test-Time Linear Out-of-Distribution Detection*

1. Structure of Appendix

In this supplementary we will

- Provide necessary lemmas for the proof.
- Give details of the omitted proof in the main paper.
- Present detailed experiment settings.
- Make further analyses of our experiments, including tests of linear relation, qualitative examples, experiment under different OOD rates and so on.
- List the full experiment results on every OOD datasets.

2. Necessary Lemmas

Definition 1. A random variable with X expectation μ is Sub-Gaussian with parameter ν^2 means:

$$\log(\mathbb{E}[\exp(\lambda(X-\mu))]) \leq \frac{\lambda^2\nu^2}{2} \tag{1}$$

Lemma 2. *Suppose $\boldsymbol{\epsilon}$ is a n-dimensional vector of i.i.d zero mean sub-Gaussian variables with parameter ν^2. For any vector w, $w^\top\boldsymbol{\epsilon}$ is sub-Gaussian of parameter $\|w\|_2^2\nu^2$*

Proof. By independence, we have

$$\begin{aligned}\mathbb{E}[\exp(\lambda w^\top\boldsymbol{\epsilon})] &= \mathbb{E}[\exp(\lambda w_1\epsilon_1 + \cdots + \lambda w_n\epsilon_n)]\\ &= \mathbb{E}[\exp(\lambda w_1\epsilon_1)\cdots\exp(\lambda w_n\epsilon_n)]\\ &= \mathbb{E}[\exp(\lambda w_1\epsilon_1)]\cdots\mathbb{E}[\exp(\lambda w_n\epsilon_n)]\\ &\leq \exp(\lambda^2(w_1^2+\cdots+w_n^2)\nu^2/2)\\ &= \exp(\lambda^2\|w\|_2^2\nu^2/2)\end{aligned} \tag{2}$$

□

Lemma 3. *Suppose $\boldsymbol{\epsilon}$ is a zero-mean n-dimensional vector of sub-Gaussian of parameter ν^2(no need to be i.i.d), we have:*

$$\mathbb{P}(\max_i \epsilon_i > t) \leq n\exp(-\frac{t^2}{2\nu^2}) \tag{3}$$

Proof.

$$\begin{aligned}\mathbb{P}(\max_i \epsilon_i > t) &= \mathbb{P}(\lambda\max_i \epsilon_i > \lambda t)\\ &= \mathbb{P}(\exp(\lambda\max_i \epsilon_i) > \exp(\lambda t))\\ &\leq \frac{\mathbb{E}[\exp(\lambda\max_i \epsilon_i)]}{\exp(\lambda t)}\\ &\leq \frac{\sum_{i=1}^n \mathbb{E}[\exp(\lambda\epsilon_i)]}{\exp(\lambda t)}\\ &\leq n\exp(\nu^2\lambda^2/2 - \lambda t)\end{aligned} \tag{4}$$

Set $\lambda = \frac{t}{\nu^2}$ we finish the proof. □

3. Proof of Theorem

We slightly **generalized** Theorem. 1 in main paper. Let $s_i, \mathbf{z}_i$ and ϵ_i denote the predicted OOD score, extracted feature and prediction error of the i^{th} example, we have $s_i = \mathbf{z}_i^\top\beta + \epsilon_i$, where ϵ_i are independent and identically distributed (i.i.d.) among different samples. We further assume that:
(C1: Ground-Truth Ranking) The ground-truth score can perfectly rank two set A, B with a margin $t > 0$, which means $\mathbf{z}_i^\top\beta > \mathbf{z}_j^\top\beta + t$ for all $i \in A, j \in B$. Where $A \cap B = \emptyset$ and $A \cup B = \Omega$, Ω is the set of all samples.
(C2: Sub-Gaussian Error) The error ϵ follow a sub-Gaussian distribution $\text{SubGau}(0, \nu^2)$ with zero mean and parameters ν^2, where ν is not too larege.

Denote $\hat{\beta}$ as the fitted β of linear regression. We have

Theorem 4 (Ranking Recovery of RTL)**.** *Denote r as the rank of the feature matrix $\mathbf{Z}$ and n as the sample number. If the ratio of the margin to error strength t/ν follow:*

$$t/\nu > min\{2\sqrt{2log(n/\delta)}, 2\sqrt{2r\log(2r/\delta)}\},$$

with at least probability $1-\delta$, $\mathbf{z}_i^\top\hat{\beta}$ rank all samples from A over samples from B.

Proof. Due to assumption C1, we have $min_{i\in A}\mathbf{z}_i^\top\beta > max_{j\in B}\mathbf{z}_j^\top\beta + t$. Without loss of generality, we can assume $min_{i\in A}\mathbf{z}_i^\top\beta > t/2$ and $max_{j\in B}\mathbf{z}_j^\top\beta < -t/2$, which is equivelant to $\mathbf{z}_i^\top\beta > t/2$ for all $i \in A$ and $\mathbf{z}_j^\top\beta < -t/2$ for all $j \in B$.

Write in the form of a matrix, we have $\boldsymbol{S} = \boldsymbol{Z}\beta + \boldsymbol{\epsilon}$, where $\boldsymbol{Z} \in \mathbb{R}^{n\times d}$,$\boldsymbol{S},\boldsymbol{\epsilon} \in \mathbb{R}^{n\times 1}$,$\boldsymbol{Z},\boldsymbol{S}$ and $\boldsymbol{\epsilon}$ as the features, base OOD detector's scores and prediction errors of examples. Here n is the total sample number while d is the feature dimension. By linear regression, we have

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{S} = \left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top(\boldsymbol{Z}\beta+\boldsymbol{\epsilon}) \\ &= \left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{Z}\beta + \left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{\epsilon},\end{aligned} \tag{5}$$

and

$$\begin{aligned}\boldsymbol{Z}\hat{\boldsymbol{\beta}} &= \boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{Z}\beta + \boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{\epsilon} \\ &= \boldsymbol{Z}\beta + \boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{\epsilon},\end{aligned} \tag{6}$$

Where $\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger$ denote the Moore-Penrose Generalized Inverse of $\boldsymbol{Z}^\top\boldsymbol{Z}$ and the last equality is due to $\boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{Z} = \boldsymbol{Z}$, which we will prove later. If $\|\boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{\epsilon}\|_\infty < t/2$, due to the assumption C1, we can conclude that $\boldsymbol{Z}\hat{\boldsymbol{\beta}}$ rank all samples from A over samples from B. Then we divide the proof into two parts:

Part I, $t/\nu > 2\sqrt{2log(n/\delta)}$

In fact we can only deal with a single side inequality to keep the ranking,

$$\max_i\{-\text{sign}(\mathbf{z}_i^\top\beta)[\boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{\epsilon}]_i\} < t/2, \tag{7}$$

where $[\cdot]_i$ denote the i-th element of vector and sign denote signum function that returns the sign of a real number.

First, we check that $-\text{sign}(\mathbf{z}_i^\top\beta)[\boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{\epsilon}]_i$ is a sub-Gaussian variable with paramter ν^2. Recall SVD decomposition, we have $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, where $\boldsymbol{U} \in \mathbb{R}^{n\times r}$,$\boldsymbol{V} \in \mathbb{R}^{d\times r}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$. $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I_r}, \boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{I_r}$, where r is the rank of the feature matrix $\boldsymbol{Z}$. Then we have

$$\begin{aligned}\boldsymbol{Z}^\top\boldsymbol{Z} &= \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^\top \\ \left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger &= \boldsymbol{V}\boldsymbol{\Sigma}^{-2}\boldsymbol{V}^\top,\end{aligned} \tag{8}$$

Further we have

$$\begin{aligned}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{Z} &= \boldsymbol{V}\boldsymbol{\Sigma}^{-2}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^\top \\ &= \boldsymbol{V}\boldsymbol{V}^\top\end{aligned} \tag{9}$$

$$\begin{aligned}\boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{Z} &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{V}^\top \\ &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{Z}\end{aligned} \tag{10}$$

$$\begin{aligned}\boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^{-2}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^\top \\ &= \boldsymbol{U}\boldsymbol{U}^\top\end{aligned} \tag{11}$$

Therefore

$$-\text{sign}(\mathbf{z}_i^\top\beta)[\boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{\epsilon}]_i = -\text{sign}(\mathbf{z}_i^\top\beta)u_i^\top\boldsymbol{U}^\top\boldsymbol{\epsilon}, \tag{12}$$

where $u_i^\top$ denote the i-th row of $\boldsymbol{U}$.

Calculate the 2-norm we have:

$$\begin{aligned}\|-\text{sign}(\mathbf{z}_i^\top\beta)u_i^\top\boldsymbol{U}^\top\|_2^2 &= \|u_i^\top\boldsymbol{U}^\top\|_2^2 = u_i^\top\boldsymbol{U}^\top\boldsymbol{U}u_i \\ &= u_i^\top u_i \le 1,\end{aligned} \tag{13}$$

the last equality is due to the property of incomplete SVD decomposition. By lemma 2 $-\text{sign}(x_i^\top\beta)u_i^\top\boldsymbol{U}^\top\boldsymbol{\epsilon}$ is a sub-Gaussian variable with paramter ν^2. Direct applying lemma 3 we have:

$$\mathbb{P}(\max_i\{-\text{sign}(\mathbf{z}_i^\top\beta)u_i^\top\boldsymbol{U}^\top\boldsymbol{\epsilon}\} \ge t/2) \le n\exp(-\frac{t^2}{8\nu^2}). \tag{14}$$

When $t/\nu > 2\sqrt{2log(n/\delta)}$ we get:

$$\mathbb{P}(\max_i\{-\text{sign}(\mathbf{z}_i^\top\beta)u_i^\top\boldsymbol{U}^\top\boldsymbol{\epsilon}\} \ge t/2) \le \delta. \tag{15}$$

Part II, $t/\nu > 2\sqrt{2rlog(2r/\delta)}$

Using the definition of operator norm, we get

$$\|\boldsymbol{Z}\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^\dagger \boldsymbol{Z}^\top\boldsymbol{\epsilon}\|_\infty = \|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty \le \|\boldsymbol{U}\|_\infty\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty \tag{16}$$

where $\|\boldsymbol{U}\|_\infty$ is the ∞-operator norm of $\boldsymbol{U}$ and $\|\boldsymbol{U}^\top\epsilon\|_\infty$ is the vector ∞-norm. Due to some basic linear algebra we know that $\|\boldsymbol{U}\|_\infty$ is the maxium ℓ_1 norm of the row $\max_i\|u_i^\top\|_\infty$. Using the mean-inequality we have

$$\begin{aligned}\|u_i^\top\|_\infty &= \sum_{j=1}^r |u_{ij}| = r\frac{\sum_{j=1}^r |u_{ij}|}{r} \\ &\le r\sqrt{\frac{\sum_{j=1}^r |u_{ij}|^2}{r}} \le \sqrt{r},\end{aligned} \tag{17}$$

for every i, thus $\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty \le \sqrt{r}\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty$. Therefore

$$\mathbb{P}(\sqrt{r}\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty < t/2) {\le} \mathbb{P}(\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty < t/2) \tag{18}$$

and

$$\mathbb{P}(\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty \ge t/2) \le \mathbb{P}(\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty \ge \frac{t}{2\sqrt{r}}) \tag{19}$$

It is easy to see $[\boldsymbol{U}^\top\boldsymbol{\epsilon}]_i$ is a sub-Gaussian vector with paramter ν^2, and $\boldsymbol{U}^\top\boldsymbol{\epsilon}$ is a r-dimensional random vector, we have

$$\begin{aligned}\mathbb{P}(\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty \ge t/2) &\le \mathbb{P}(\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty \ge \frac{t}{2\sqrt{r}}) \\ &\le 2r\exp(-\frac{t^2}{8\nu^2 r})\end{aligned} \tag{20}$$

when $t/\nu > 2\sqrt{2rlog(2r/\delta)}$,

$$\mathbb{P}(\|\boldsymbol{U}\boldsymbol{U}\boldsymbol{\epsilon}\|_\infty \ge t/2) \le \delta$$

Therefore conclude the proof. □

When $A = \{\text{In-distribution Samples}\}$, and $V = \{\text{Out-of-distribution Samples}\}$, this is just Theorem. 1 of main paper. Note that our theorem not only applies to ranking satisfy the ground-truth ranking assumption for in-distribution and out-of-distribution data, but also applies to the case when the ranking are nearly ground-truth ranking. The ranking equality of $\mathbf{z}^\top\beta$ puts a upper bound on the efficacy RTL, but with high probability, the $\mathbf{z}^\top\hat{\beta}$ produced by our RTL can perform as good $\mathbf{z}^\top\beta$ and outperform the raw OOD score $s = \mathbf{z}^\top\beta + \epsilon$.

We further prove another lemma which describe the tail distribution of mixture of two Sub-Gaussian Variables.

Lemma 5. *The mixture of two Sub-Gaussian of SubGau*$(0, a^2)$ *and SubGau*$(0, b^2)$ *is still a Guassian Mixture with parameter* $max\{a^2, b^2\}$

Proof.

$$\begin{aligned}
&\mathbb{E}[\exp(\lambda(\pi_1 \mathrm{SubGau}(0, a^2) + \pi_2 \mathrm{SubGau}(0, b^2)))] = \\
&\pi_1 \mathbb{E}[\exp(\lambda(\mathrm{SubGau}(0, a^2)))] \\
&+ \pi_2 \mathbb{E}[\exp(\lambda(\mathrm{SubGau}(0, b^2)))] \\
&\leq \pi_1 \exp(\frac{\lambda^2 a^2}{2}) + \pi_2 \exp(\frac{\lambda^2 b^2}{2}) \leq \exp(\frac{\lambda^2 \max\{a^2, b^2\}}{2})
\end{aligned} \tag{21}$$

□

Since the tail distribution of mixture of two Sub-Gaussian variables is mainly decided by the variable with larger parameter, directly applying RTL will be difficult when errors follow $\pi_1 \mathrm{SubGau}(0, \nu_1^2) + \pi_2 \mathrm{SubGau}(0, \nu_2^2)$, where $\pi_1 + \pi_2 = 1$, π_2 is very small but $\nu_2 \gg \nu_1$. To deal with this problem, RTL++ introduced subset selection to filter large error predictions.

4. Experiments Setting

Algorithm 1: Subset selection of RTL++

```
Input: features $\mathbf{z}_i$ and OOD score $s_i$, $1 \leq i \leq n$,
Normalize $\mathbf{z}_i$ to unit Euclidean norm
Apply dimensionality reduction on $\mathbf{z}_i$ to $d \ll n$
Stack $\mathbf{z}_i$ and $s_i$ by rows to $\mathbf{Z}$ and $\mathbf{S}$
Calculate projection $\tilde{\mathbf{Z}} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^\dagger\mathbf{Z}^\top$ and
  $\tilde{\mathbf{S}} = \tilde{\mathbf{Z}}\mathbf{S}$
Solving Lasso $\hat{\gamma} = \mathrm{argmin}_\gamma \frac{1}{2}\|\tilde{\mathbf{S}} - \tilde{\mathbf{Z}}\gamma\|_2^2 + \lambda\|\gamma\|_1$
Select a subset $\hat{\mathbf{Z}}$ with the lowest $p\%$ of $|\hat{\gamma}_i|$.
return $\hat{\mathbf{Z}}$
```

Algorithm 2: Online RTL

```
Input: an OOD score function $S : \mathcal{X} \to \mathbb{R}$, batch
  size $b$, numbers of batch $m$
Initialize: $\mathbf{A} \leftarrow 0$, $\mathbf{b} \leftarrow 0$
for $k = 1, \cdots, m$ do
    Sample $b$ images $\boldsymbol{x}_i, 1 \leq i \leq b$;
    Calculate the features of all images $\mathbf{z}_i = g(\boldsymbol{x}_i)$;
    Calculate the OOD score of all images
      $\hat{s}_i = S(\boldsymbol{x}_i)$;
    Stack $\mathbf{z}_i$ and $\hat{s}_i$ by rows to $\mathbf{Z}$ and $\mathbf{S}$ ;
    $\mathbf{A} \leftarrow \mathbf{A} + \mathbf{Z}^\top\mathbf{Z}$;
    $\mathbf{b} \leftarrow \mathbf{b} + \mathbf{Z}^\top\mathbf{S}$;
    $\hat{\beta} = (\mathbf{A})^\dagger\mathbf{b}$;
    Compute calibrated OOD score of the present
      batch $s_i = \mathbf{z}_i^T\hat{\beta}$;
end
Using $s_i$ as the OOD score.
```

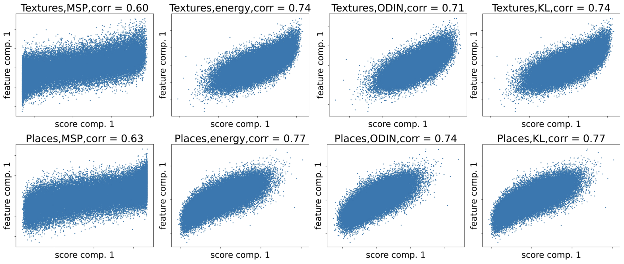

Figure 1. More Visualization of Canonical-Correlation Analysis of ImageNet and two OOD datasets' features(Places and Texture) and OOD scores.

For all results except for changing the percentile of chosen data in RTL, we set the regularization $\lambda = 1e{-}5$ for CIFAR and $\lambda = 1e - 7$ for ImageNet. For CIFAR, we choose percentile $p = 80\%$. For ImageNet we choose $p = 90\%$ for MSP and ODIN, $p = 95\%$ for energy and KL. For experiments on CIFAR-10 and CIFAR-100, we directly use the feature of the penultimate layer of Wide ResNet-40. On ImageNet-1k, we further apply principal component analysis to the output of penultimate layer of ResNetv2-101, to reduce the dimension of feature to 32 in the subset selection procedure of RTL++ in Alg. 1. For GMM model on CIFAR-10 and CIFAR-100, we use Gaussian mixtures of 10 and 100 components respectively, and each component has its own general covariance matrix. For GMM model on ImageNet-1k, a Gaussian Mixture of 1000 components with diagonal covariance matrix is used. The log-likelihood of each sample is used as an OOD score. For local outlier factor model, we set number of neighbors to 20. For Isolation Forest, the number of base estimators is set to 100.

5. Further Experiments Analysis

In this section, we carry out additional experiments of RTL to prove the efficacy of our methods.

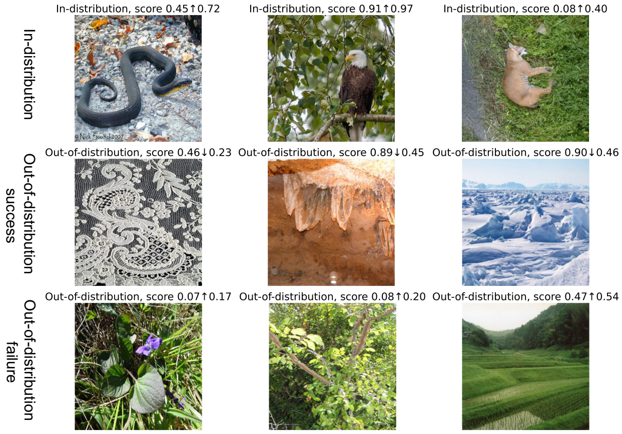
In-distribution
Out-of-distribution success
Out-of-distribution failure
In-distribution, score 0.45↑0.72
In-distribution, score 0.91↑0.97
In-distribution, score 0.08↑0.40
Out-of-distribution, score 0.46↓0.23
Out-of-distribution, score 0.89↓0.45
Out-of-distribution, score 0.90↓0.46
Out-of-distribution, score 0.07↑0.17
Out-of-distribution, score 0.08↑0.20
Out-of-distribution, score 0.47↑0.54
© Nick Scobel 2007

Figure 2. Qualitative examples

Test of Linear Relation between features and scores. Visualization of Canonical-Correlation Analysis of various OOD detection algorithms and benchmark datasets, including Textures and Places, are displayed in Fig 1. We test the linear relationship between features and score in Tab. 1 with ImageNet-1k as in-distribution dataset, based on the significance test of the first canonical correlation. Our test is built on Wilks' Λ, and use the F-approximation of Λ as our statistics. For details of the test, please refer to chapter 11 of [2]. The tests show that the first canonical correlation is significant, which imply the linear relation between features and scores.

Qualitative examples. We display some qualitative examples in Fig. 2. Note that, as Eq. 1 in our main paper, larger score means more ID (In-Distribution). We draw some examples with scores before/after RTL in each title. In the first row, RTL can boost the scores of in-distribution data by training on the entire dataset. In the second row, it tends to correctly decrease scores for OOD images that have little semantic similarity with ID data. However, in the third row, some OOD instances such as plants or scenes that have similar backgrounds to ID data may lead RTL to falsely increase their scores and result in failure.

Revision Under different out-of-distribution rates. On CIFAR-10 and CIFAR-100, we tried to change the rate of out-of-distribution data for OOD detection and apply our RTL. The rate of OOD data varies from 10% to 90% in Tab. 2, with spacing 10%. We report ΔFPR and ΔAUROC, where $\Delta L := (L \text{ with RTL}) - (L \text{ without RTL})$. We fix the total number of data at 5000. Every experiment is repeated 10 times and the average is used as the final results. We find that using MSP or energy as base OOD detectors, our linear revision succeeds in bringing some improvements even with 60% to 80% of OOD data. When the rate of OOD data gets too large, our proposed linear revision may instead worsen the results. This is because when there is a too small number of in-distribution, it is hard to capture the rich information of in-distribution data. However, this is still astonishing because our methods still work when OOD data account for the most part of the observations. Such scenarios violate the assumption of classical outlier detection methods. We argue that this is because we utilized the information of pre-trained classifiers. Even if in-distribution data compose of a smaller part of dataset during OOD inference, OOD detectors tend to produce larger scores for in-distribution data and lower scores for OOD data, and linear regression will keep this tendency.

Revision Under different sample number Still on CIFAR, we keep the rate of in-distribution data and out-of-distribution data at 5:1, then change the number of in-distribution from 10000 to 10 in Tab. 3. It is worth noting that we repeat more rounds when the total sample number is small to get more accurate estimations. Let's denote m as the number of in-distribution data fed in the test time. We correspondingly repeat $\frac{10^5}{m}$ rounds. We find that when number of test data is too small (around the dimension of feature, note that the penultimate layer of Wide ResNet-40 of CIFAR is 128-dimensional), our algorithms will produce no revision because it fits the OOD score perfectly. When the number of data grows, our linear revision may first decease AUROC. However, our algorithm consistently improves the results when there are more than 500 in-distribution samples.

Extend the results with multiple OOD datasets. Because all the above experiments are carried out on single datasets then we take average over all datasets, we wonder if our methods work when multi out-of-distribution domains are simultaneously given. We carry out experiments on large-scale OOD detection, fix the ImageNet-1k validation set as in-distribution. The OOD data are all the combinations of i**N**aturalist, **P**laces, **S**UN and **T**exturs datasets (shorted as **NPST**). The results are visualized in Fig. 3. We find that RTL, as a linear calibration, can improve results on all combination of the given four datasets.

Comparison with ReAct. There are some recently proposed methods to improve OOD detection. The most relevant among these methods is the post-processing method ReAct [3], which rectifies the OOD scores based on the internal feature activation patterns. In Tab. 4, our RTL outperforms ReAct when using both MSP and ODIN as base OOD detectors. More interestingly, ReAct and RTL are complementary and can sometimes achieve better performance when combined.

Results on Noises. In many cases, detecting random noise is more challenging than detecting data from a disjoint dataset. We test two kinds of random noise: uniform noise on $[0, 1]$ and gaussian noise with $\mu = 0.5$ truncated in $[0, 1]$. Although noise based OOD is challenging for ordinary post hoc OOD detector, it is not a problem for RTL as shown in

OOD Dataset	Method	Wilks' Λ	Statistics	df_1	df_2	p-value < 0.01
iNaturalist	MSP	0.5333	24.76	2048	57951	✓
	energy	0.3816	45.86	2048	57951	✓
	ODIN	0.4039	41.77	2048	57951	✓
	KL	0.3816	45.86	2048	57951	✓
Places	MSP	0.6074	18.29	2048	57951	✓
	energy	0.4106	40.63	2048	57951	✓
	ODIN	0.4459	35.17	2048	57951	✓
	KL	0.4106	40.63	2048	57951	✓
SUN	MSP	0.5905	19.63	2048	57951	✓
	energy	0.3895	44.34	2048	57951	✓
	ODIN	0.4264	38.06	2048	57951	✓
	KL	0.3895	44.34	2048	57951	✓
Textures	MSP	0.6373	14.89	2048	53591	✓
	energy	0.4579	30.98	2048	53591	✓
	ODIN	0.4897	27.27	2048	53591	✓
	KL	0.4579	30.98	2048	53591	✓

Table 1. Test of Linear Relationship based Wilks' Λ, where statistics is the F-approximation of Wilks' Λ and df_1 and df_2 is the degree of freedom for F-distribution. ImageNet-1k is used as in-distribution dataset

Dataset	CIFAR-10				CIFAR-100			
Method	MSP		Energy		MSP		Energy	
OOD Percent	ΔFPR↓	ΔAUROC↑	ΔFPR↓	ΔAUROC↑	ΔFPR↓	ΔAUROC↑	ΔFPR↓	ΔAUROC↑
0.1	-38.58	6.08	-12.26	2.33	-23.09	7.66	-8.93	-0.34
0.2	-36.85	5.55	-14.57	3.11	-27.33	8.28	-13.86	2.04
0.3	-34.71	4.79	-15.02	3.28	-28.39	8.16	-16.57	3.02
0.4	-33.06	4.09	-15.35	3.35	-28.06	7.64	-18.19	3.59
0.5	-30.89	3.36	-15.15	3.33	-26.57	6.94	-19.10	3.90
0.6	-28.53	2.59	-14.76	3.20	-24.78	6.11	-19.66	3.98
0.7	-25.89	1.71	-14.28	3.00	-22.32	4.95	-20.11	3.82
0.8	-22.21	0.50	-12.67	2.56	-18.13	3.15	-18.84	3.09
0.9	-14.89	-1.66	-9.80	1.36	-9.92	-0.62	-13.84	0.63

Table 2. Results under different OOD rates

Tab. 5.

Results on Natural Adversarial Examples. The domain shift between in and out-of-distribution data is relatively large for ImageNet OOD detection benchmark. Recently, natural adversarial examples ImageNet-O [1] have been introduced as a hard out-of-distribution dataset. We test of our RTL on ImageNet-O with ImageNet as in-distribution data in Tab. 6. Though our RTL could not achieve improvement as significant as results on the ordinary ImageNet benchmark, our method can help the base OOD detectors to make more precise predictions.

Dataset		CIFAR-10				CIFAR-100			
Method		MSP		Energy		MSP		Energy	
In number	OOD number	ΔFPR↓	ΔAUROC↑	ΔFPR↓	ΔAUROC↑	ΔFPR↓	ΔAUROC↑	ΔFPR↓	ΔAUROC↑
10000	2000	-38.87	6.11	-14.70	3.04	-28.52	8.79	-12.84	1.46
5000	1000	-37.95	5.88	-14.62	3.00	-27.32	8.45	-12.59	1.40
2000	400	-35.34	5.23	-13.92	2.91	-23.95	7.43	-11.41	1.19
1000	200	-31.70	4.17	-12.87	2.71	-19.17	6.02	-9.74	0.91
500	100	-25.23	2.22	-11.08	2.35	-12.83	3.84	-7.33	0.48
200	40	-11.30	-2.43	-6.00	1.26	-3.19	0.17	-2.59	-0.10
100	20	0.04	-4.90	0.34	-0.08	0	0	0	0
50	10	0	0	0	0	0	0	0	0
20	4	0	0	0	0	0	0	0	0
10	2	0	0	0	0	0	0	0	0

Table 3. Results under different total numbers

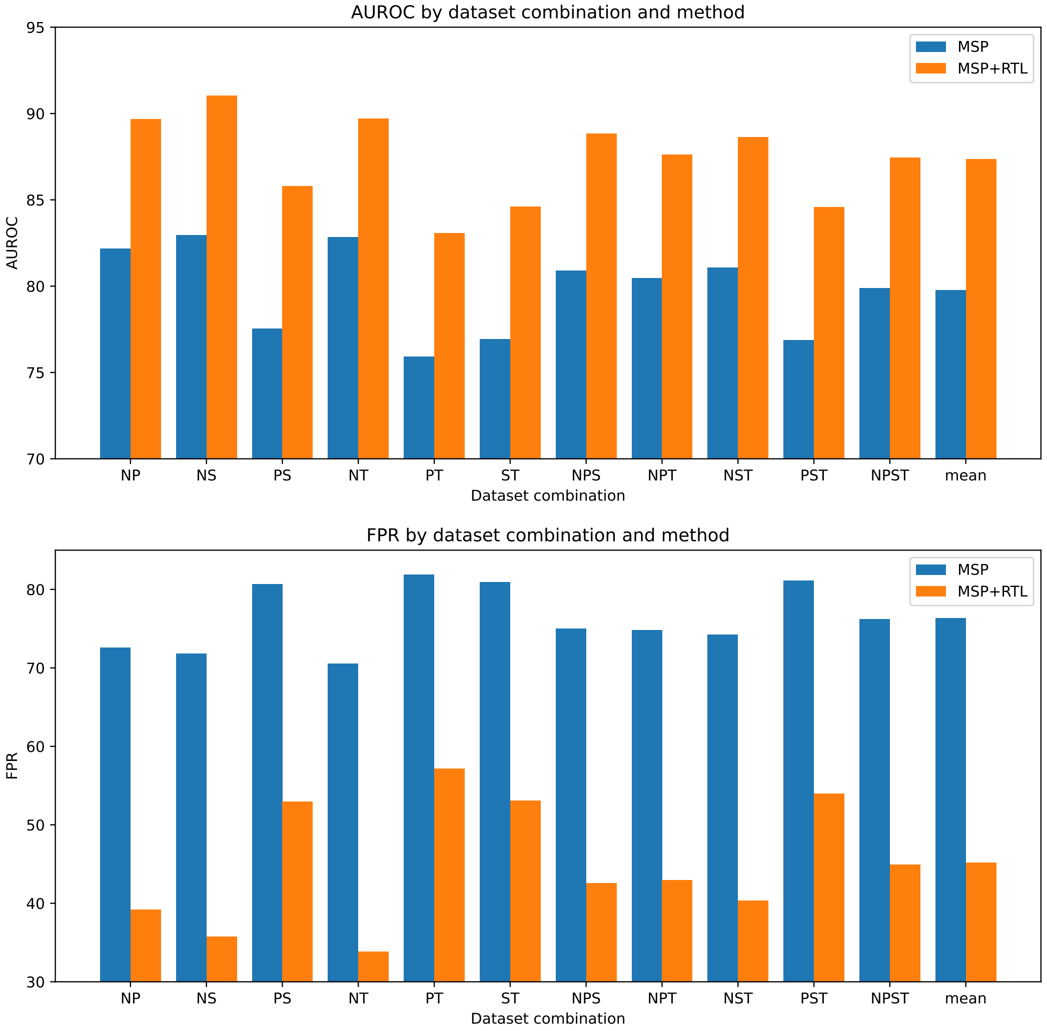

Figure 3. Results of Combination of Datasets

Method	ReAct	RTL	FPR↓	AUROC↑
MSP	no	no	76.98	79.28
	no	yes	**46.84**	86.71
	yes	no	70.19	81.73
	yes	yes	67.83	81.98
ODIN	no	no	72.99	82.56
	no	yes	53.28	85.68
	yes	no	63.64	84.49
	yes	yes	48.85	**87.97**

Table 4. ImageNet-1k with ReAct.

Method	Noise	FPR↓	AUROC↑
MSP	Uniform	99.60	89.13
	Gaussian	99.77	88.08
MSP+RTL	Uniform	0.01	99.50
	Gaussian	0.01	99.34
ODIN	Uniform	99.14	90.12
	Gaussian	99.56	89.15
ODIN+RTL	Uniform	7.05	97.29
	Gaussian	15.38	96.73

Table 5. ImageNet-1k with Noise as OOD data.

	FPR ↓	AUROC↑	AUPR ↑
MSP	94.30	61.69	89.25
MSP+RTL	87.20	67.33	90.55
Energy	94.70	67.86	91.40
Energy+RTL	86.50	72.69	92.55
ODIN	94.90	67.31	91.23
ODIN+RTL	86.45	72.27	92.41
KL	94.70	67.86	91.40
KL+RTL	86.50	72.69	92.55

Table 6. ImageNet-1k with ImageNet-O as OOD data.

6. ImageNet results with AUPR

Dataset	iNaturalist			SUN		
metrics	FPR95	AUROC	AUPR	FPR95	AUROC	AUPR
MSP	63.69	87.59	97.23	79.98	78.34	94.45
+RTL	21.03	94.98	98.71	50.68	87.14	96.61
+RTL++	18.76	95.60	98.90	48.40	88.70	97.17
Energy	64.91	88.48	97.58	65.33	85.32	96.57
+RTL	45.48	91.04	97.96	52.06	88.68	97.38
+RTL++	41.57	92.03	98.22	49.84	89.32	97.61
KL	64.91	88.48	97.58	65.32	85.31	96.57
+RTL	45.48	91.04	97.96	52.06	88.68	97.38
+RTL++	41.57	92.03	98.22	49.84	89.32	97.61
ODIN	62.69	89.36	97.76	71.67	83.92	96.26
+RTL	35.27	92.87	98.35	51.59	88.40	97.25
+RTL++	36.10	92.78	98.33	51.87	88.23	97.21
GradNorm	50.03	90.33	97.83	46.48	89.03	97.29
GMM	87.90	68.43	90.26	89.99	63.29	88.72
LOF	95.16	51.57	83.87	94.89	52.27	84.04
IF	88.58	61.60	87.92	90.12	57.85	86.09
Dataset	Places			Textures		
metrics	FPR95	AUROC	AUPR	FPR95	AUROC	AUPR
MSP	81.44	76.76	94.15	82.73	74.45	95.65
+RTL	57.22	84.48	95.94	58.48	80.24	96.30
+RTL++	56.72	85.32	96.21	59.98	79.91	96.25
Energy	73.02	81.37	95.49	80.87	75.79	96.05
+RTL	62.68	84.35	96.14	69.49	75.39	95.47
+RTL++	62.37	84.05	96.03	70.44	76.52	96.00
KL	73.02	81.37	95.49	80.87	75.79	96.05
+RTL	62.68	84.35	96.14	69.49	75.39	95.47
+RTL++	62.37	84.05	96.03	70.44	76.52	96.00
ODIN	76.27	80.67	95.35	81.31	76.30	96.12
+RTL	60.71	84.44	96.11	66.72	76.78	95.70
+RTL++	61.35	84.28	96.07	67.06	76.58	95.64
GradNorm	60.86	84.82	96.26	61.42	81.07	96.96
GMM	96.85	52.83	84.54	95.37	35.34	83.83
LOF	93.05	56.37	85.64	82.02	65.39	92.77
IF	93.45	50.24	83.02	54.34	87.76	98.27

Table 7. Detailed Experiments on ImageNet-1k with AUPR

7. Detailed Results On CIFAR-10

Dataset	Textures			SVHN			Places365		
Method	FPR	AUROC	AUPR	FPR	AUROC	AUPR	FPR	AUROC	AUPR
MSP	59.50	88.37	97.16	48.98	91.86	98.26	60.32	88.08	97.08
+RTL++	27.35	91.89	97.33	11.24	95.91	98.46	31.49	93.22	98.32
+RTL	20.83	94.56	98.43	10.47	96.71	98.96	29.86	93.46	98.40
Energy	52.33	85.36	95.48	35.49	91.15	97.72	40.16	89.75	97.25
+RTL++	34.74	88.52	95.97	10.28	97.38	99.32	35.82	91.33	97.76
+RTL	36.75	87.23	95.44	14.64	95.73	98.80	37.55	90.76	97.64
Odin	49.62	84.57	95.14	32.88	92.11	98.03	57.14	84.23	95.73
+RTL++	53.33	77.01	91.57	27.00	88.96	95.87	55.31	84.24	95.71
+RTL	60.74	72.08	89.53	27.18	89.31	96.10	63.96	80.08	94.40
KL	52.34	85.36	95.48	35.49	91.15	97.72	40.17	89.75	97.25
+RTL++	34.69	88.53	95.98	10.29	97.39	99.32	35.74	91.33	97.76
+RTL	36.74	87.23	95.44	14.67	95.74	98.80	37.55	90.75	97.63
GradNorm	73.59	57.90	83.12	59.49	70.21	89.45	78.38	60.51	86.99
GMM	60.24	83.07	95.80	98.01	23.62	70.12	80.85	79.14	94.97
IF	99.35	21.64	70.19	78.71	67.98	89.67	91.94	55.09	85.27
LOF	96.90	45.94	80.20	92.86	57.39	86.12	98.62	52.92	84.58
Dataset	LSUN-C			LSUN-R			iSUN		
Method	FPR	AUROC	AUPR	FPR	AUROC	AUPR	FPR	AUROC	AUPR
MSP	30.95	95.63	99.13	52.23	91.49	98.17	56.24	89.80	97.73
+RTL++	0.50	99.83	99.96	2.08	99.51	99.89	8.33	98.17	99.57
+RTL	0.18	99.92	99.98	3.73	99.19	99.82	8.75	98.19	99.59
Energy	8.31	98.34	99.65	27.75	94.15	98.65	33.84	92.51	98.23
+RTL++	0.56	99.81	99.94	5.79	98.81	99.74	9.69	97.98	99.55
+RTL	1.44	99.61	99.89	7.05	98.58	99.69	10.62	97.87	99.54
Odin	15.90	96.98	99.33	26.63	94.58	98.77	32.45	93.29	98.48
+RTL++	39.79	84.25	94.56	17.82	94.81	98.53	21.46	94.14	98.40
+RTL	37.13	85.77	95.23	19.65	94.32	98.41	18.96	95.08	98.69
KL	8.31	98.34	99.65	27.75	94.15	98.65	33.84	92.51	98.23
+RTL++	0.56	99.81	99.94	5.81	98.81	99.74	9.67	97.98	99.55
+RTL	1.42	99.61	99.89	7.02	98.58	99.69	10.62	97.87	99.54
GradNorm	12.07	96.85	99.18	65.27	73.38	92.06	70.27	71.07	91.41
GMM	96.12	46.31	80.17	95.84	58.68	89.29	95.10	58.88	89.14
IF	59.47	82.96	95.11	71.85	76.06	93.43	78.44	71.09	91.75
LOF	97.62	51.69	83.89	94.21	65.72	90.49	94.64	65.04	90.08

Table 8. Details of Experiments On CIFAR-10

8. Detailed Results On CIFAR-100

Dataset	Textures			SVHN			Places365		
Method	FPR	AUROC	AUPR	FPR	AUROC	AUPR	FPR	AUROC	AUPR
MSP	83.55	73.67	93.07	84.00	71.44	92.89	82.30	74.03	93.30
+RTL++	57.25	80.62	93.02	44.29	84.18	94.93	76.63	76.16	93.53
+RTL	64.26	78.55	92.87	56.75	78.72	93.42	75.19	76.47	93.59
Energy	79.32	76.40	93.70	85.43	73.96	93.62	80.11	75.80	93.59
+RTL++	69.14	74.68	91.39	63.51	77.91	93.67	77.56	75.87	93.51
+RTL	70.08	73.51	91.29	64.85	78.73	94.23	80.46	73.55	92.84
Odin	79.57	73.43	92.81	87.73	65.43	90.95	87.14	72.00	92.67
+RTL++	69.01	76.09	92.60	77.75	68.83	90.10	86.30	70.30	91.79
+RTL	65.44	74.61	92.13	67.50	69.81	89.50	88.29	68.33	91.21
KL	79.32	76.40	93.70	85.43	73.96	93.62	80.11	75.80	93.59
+RTL++	69.14	74.68	91.39	63.51	77.91	93.67	77.56	75.87	93.51
+RTL	70.08	73.51	91.29	64.85	78.73	94.23	80.46	73.55	92.84
GradNorm	87.48	60.41	87.46	97.30	55.00	86.83	96.95	53.65	85.83
GMM	92.18	63.26	89.34	99.09	62.07	91.35	86.33	73.32	92.46
IF	95.44	49.11	82.42	79.60	72.67	92.35	87.76	62.94	88.68
LOF	98.47	42.35	80.32	98.08	44.98	83.38	99.26	38.93	80.28

Dataset	LSUN-C			LSUN-R			iSUN		
Method	FPR	AUROC	AUPR	FPR	AUROC	AUPR	FPR	AUROC	AUPR
MSP	66.00	83.85	96.35	82.48	75.39	94.08	82.93	75.62	94.10
+RTL++	20.12	96.09	99.14	33.99	92.43	98.21	30.10	92.88	98.22
+RTL	19.68	96.18	99.15	48.24	88.13	97.11	45.66	88.94	97.28
Energy	35.78	93.46	98.60	79.14	79.42	95.04	81.00	78.99	94.92
+RTL++	25.53	94.94	98.86	56.23	85.66	96.46	56.42	85.53	96.40
+RTL	27.81	93.83	98.51	59.68	83.83	95.98	60.90	83.43	95.85
Odin	57.38	87.05	97.07	69.27	82.48	95.81	66.22	83.01	95.85
+RTL++	29.64	94.12	98.67	23.89	93.72	98.32	23.78	93.36	98.10
+RTL	34.32	92.95	98.37	28.50	91.66	97.67	27.18	91.79	97.57
KL	35.78	93.46	98.60	79.14	79.42	95.04	81.00	78.99	94.92
+RTL++	25.53	94.94	98.86	56.23	85.66	96.46	56.42	85.53	96.40
+RTL	27.80	93.83	98.51	59.68	83.83	95.98	60.90	83.43	95.85
GradNorm	39.47	92.12	98.22	99.06	40.06	80.11	99.06	44.09	82.10
GMM	92.69	68.75	91.34	96.17	78.01	95.34	97.93	74.34	94.19
IF	40.69	90.68	97.72	90.66	60.69	88.15	91.32	60.79	88.44
LOF	97.37	36.64	77.84	98.02	49.36	85.27	98.16	47.67	84.65

Table 9. Details of Experiments On CIFAR-100

References

[1] Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. Natural adversarial examples. In *CVPR*, 2021. 5

[2] Alfin C Rencher. Methods of multivariate analysis. new york: A john wiley & sons. *Inc. Publication*, 2002. 4

[3] Yiyou Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. *Advances in Neural Information Processing Systems*, 34:144–157, 2021. 4