

Out-of-Distribution Detection through Transformation Rectified Activation

Maozhen Li, Yu Liang, Man Qi and Guanjun Liu

Abstract—Rectified activation is a highly effective post-hoc method for out-of-distribution (OOD) detection, designed to enhance the robustness of Deep Neural Networks (DNNs). It operates by truncating high neuron activations to prevent overconfident predictions on unfamiliar data. However, this approach has two key limitations: it only truncates outlier activations without optimizing the post-truncation relationship between in-distribution (ID) and OOD data, and its performance is often unstable due to the manual selection of the rectification threshold. In this paper, we introduce Transformation Rectified Activation (Trans-ReAct), a novel method that addresses these issues. Trans-ReAct leverages two core innovations: (1) it applies mathematical transformations to strategically amplify the activation differentiation between ID and OOD data, and (2) it introduces a dynamic threshold anchor that automatically adapts the rectification threshold to the ID data, ensuring stable performance. Evaluated on three benchmark datasets, Trans-ReAct significantly outperforms state-of-the-art methods, reducing the false positive rate at 95% true positive rate (FPR95) by up to 9.76% and increasing the Area Under the ROC Curve (AUROC) by up to 2.04%.

Index Terms—Out-of-distribution detection, rectified activation, neural network robustness, transformation rectified activation.

I. INTRODUCTION

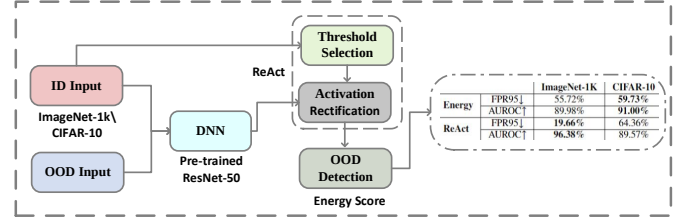
WITH the rethinking of overconfident cognition of deep neural networks (DNNs), the past few years have witnessed a growing interest in developing out-of-distribution (OOD) detection methods to improve model robustness of DNNs [1]. OOD detection aims to recognize the test samples that are not consistent with an in-distribution (ID) data while shall not compromise the classification performance on the ID data.

Several works have been proposed for OOD detection including outlier exposure [2], Dirichlet prior network [3] and Bayesian model [4]. These works mainly do so by adding a set of extra OOD samples or making changes to DNN model structures. It should be pointed out that DNNs [5]–[7] are becoming more diversified and complex which makes it hard to deploy these OOD detection works. In addition to traditional methods, graph-based approaches for anomaly

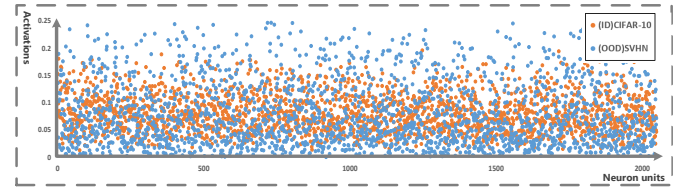
Maozhen Li is with the College of Computer, Shandong Xiehe University Jinan City, 250109, China. He is also with the Department of Electronic and Electrical Engineering, Brunel University of London, UK(Maozhen.Li@brunel.ac.uk).

Yu Liang and Guanjun Liu are with the Key Laboratory of Ministry of Education on Embedded System and Service Computing and also with the Department of Computer Science and Technology, Tongji University, Shanghai 201804, China (micopac001@tongji.edu.cn; liuguanjun@tongji.edu.cn).

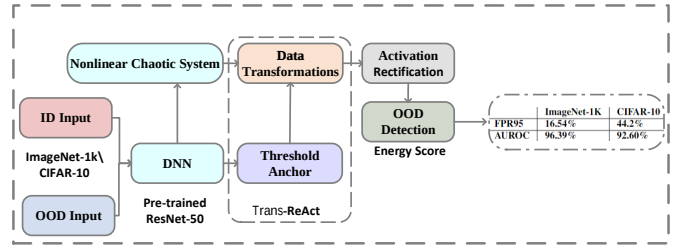
Man Qi is with the College of Computer, Shandong Xiehe University Jinan City, 250109, China. She is also with the School of Engineering, Technology and Design, Canterbury Christ Church University, UK (Man.Qi@canterbury.ac.uk).



(a) Pipeline of ReAct



(b) The activations distribution of ReAct on CIFAR-10 benchmark



(c) Pipeline of Trans-ReAct

Fig. 1: Plots showing (a) the operational pipelines of ReAct and Trans-ReAct on various datasets, (b) the activation density distribution of ReAct on the CIFAR-10 benchmark, where the x-axis represents activation values and the y-axis represents activation density, revealing significant overlap between ID and OOD distributions, and (c) the improved separation achieved by the Trans-ReAct pipeline. The key takeaway is the visual confirmation that our method (c) creates a clearer decision boundary compared to the baseline (b).

and OOD detection have gained traction in social network analysis. For instance, Liu et al. [8] proposed a dual-attention mechanism combined with a graph convolutional network for rumor detection, effectively handling abnormal patterns in complex social networks. Similarly, He et al. [9] utilized a graph convolutional network-based method for rumor blocking, demonstrating the potential of graph neural networks for OOD-like detection tasks in social networks. Furthermore, Guo et al. [10] introduced a biased random walk embedding method for community detection, which indirectly addresses

OOD scenarios by detecting anomalous nodes in attributed networks. These methods highlight the interdisciplinary nature of OOD detection and its applicability across domains.

In addition to traditional methods, graph-based approaches for anomaly and OOD detection have gained traction in social network analysis. For instance, Liu et al. [8] proposed a dual-attention mechanism combined with a graph convolutional network for rumor detection, effectively handling abnormal patterns in complex social networks. Similarly, He et al. [9] utilized a graph convolutional network-based method for rumor blocking, demonstrating the potential of graph neural networks for OOD-like detection tasks in social networks. Furthermore, Guo et al. [10] introduced a biased random walk embedding method for community detection, which indirectly addresses OOD scenarios by detecting anomalous nodes in attributed networks. These methods highlight the interdisciplinary nature of OOD detection and its applicability across domains, including high-performance heterogeneous computing platforms [11], large-scale cloud computing frameworks [12], and industrial IoT-based intelligent systems [13].

Post-hoc OOD detection is a low-cost but effective method that requires no modifications to the training process of a DNN. Rectified activation, as an optimization method, has emerged into one of the most effective approaches to post-hoc OOD detection, and ReAct [14] represents the latest work. Conceptually, rectified activation is a technique that truncates or caps high neuron activations to mitigate over-confident predictions on unfamiliar, OOD inputs. Essentially, it truncates an OOD and ID data into two individual parts based on a rectified activation threshold.

However, rectified activation has two major limitations. On the one hand, it only truncates the outlier activations of most OOD data but fails to further consider the distributional relationship between OOD and ID data after truncation. Specially, rectified activation truncates neurons in the penultimate layer of a pre-trained DNN model using a rectification threshold value c to reduce abnormal neuron activations on an OOD data, but rectified activation does not perform a further optimization on the truncated neurons. On the other hand, rectified activation suffers from unstable performance due to randomness in selection of a rectified activation threshold. In the case of ReAct, it selects c based on the p -th percentile of ID activations. For example, when $p = 90\%$, it indicates that 90 percent of the ID activations are less than the threshold c . It is worth noting that the threshold c selected for one ID data may have a negative impact on OOD detection when applied to another ID data. As shown in Fig. 1(a), ReAct performs well on ImageNet-1K [15] when applying $p = 90\%$ ($c = 1$), but it performs worse than the Energy baseline approach [16] when applying the randomly selected $p = 90\%$ ($c = 0.25$) on CIFAR-10 [17]. It is further observed that the unstable performance of ReAct on CIFAR-10 is due to its weak differentiation between ID and OOD data using a small threshold value ($c = 0.25$). As shown in Fig. 1(b), ID and OOD data are mixed in a limited space which makes it difficult to differentiate one from the other.

To address the two major limitations of rectified activation, we present Trans-ReAct, a rectified activation method

for post-hoc OOD detection through transformations. Here, activation transformation refers to the strategic application of mathematical functions to neuron activation values, designed to amplify the subtle yet critical differences between ID and OOD data patterns. Trans-ReAct introduces a threshold anchor into a pre-trained DNN. Similar OOD and ID data will generate significant differences in the model's activations after undergoing transformations, enabling further separation of OOD data from a vast amount of data. Considering the negative impact of randomness in selection of a rectification threshold value, Trans-ReAct dynamically adjusts a threshold value that is adaptive to an ID data, leading to a stable performance on both ImageNet-1K and CIFAR-10 datasets as shown in Fig. 1(c). This also saves the effort of manually tuning hyperparameters repetitively as incurred in the ReAct approach.

Specifically, the main contributions of the paper are summarized as follows:

- (1) Trans-ReAct utilizes transformations to enlarge the activation differentiation between ID and OOD data, even when they initially show similar activation values. By applying distinct transformations that specifically target the activation gap between ID and OOD data within preset limits, Trans-ReAct successfully decreases the false positive rate (FPR).
- (2) Trans-ReAct introduces a threshold anchor to dynamically adjust a rectification threshold value that is adaptive to an ID data. Trans-ReAct detects a potentially small rectification threshold during the threshold selection process, and adjusts it to the threshold anchor point. In this way, a stable performance on OOD detection can always be achieved. This also relieves the need of manually tuning hyperparameters.
- (3) Trans-ReAct is extensively evaluated on three benchmark datasets - ImageNet-1K, CIFAR-10 and CIFAR-100 [18], and achieves the state-of-the-art performance on OOD detection. Compared with the previous best results, Trans-ReAct reduces FPR95 on average by 8.95% on ImageNet-1K, 9.76% on CIFAR-10 and 8.42% on CIFAR-100. In the meantime, Trans-ReAct increases AUROC on average by 2.04% on ImageNet-1K, 1.69% on CIFAR-10, and 2.1% on CIFAR-100.

The remainder of this work is organized as follows. Section II reviews related work on post-hoc OOD detection methods. Section III presents the methodology of design and implementation of Trans-ReAct including a theoretic analysis. Section IV conducts comprehensive experiments and validates the performance of Trans-ReAct in comparison with the state-of-the-art results. Section V concludes the paper.

II. RELATED WORK

This section gives a review on related background and post-hoc OOD methods. To improve the readability of the paper, we maintain consistency in the formal expressions.

A. Post-hoc OOD Detection

Let $\mathcal{X}$ (resp. $\mathcal{Y}$) be an input (resp. output) space and let $\mathcal{X}_{ID}$ (resp. $\mathcal{X}_{OOD}$) be an ID (resp. OOD) dataset. Let $f : \mathcal{X} \xrightarrow{t} \mathcal{Y}$

be an input space to an output space through a sequence of hidden layers t ($t \in (0, n)$) in a DNN f , and $n \in \mathbb{N}^+$ is the total number layers in f . The goal of OOD detection is to define a decision function Det such that for a given test input $x \in \mathcal{X}$:

$$Det(x, f, n) = \begin{cases} 0 & \text{if } x \sim \mathcal{X}_{OOD} \\ 1 & \text{if } x \sim \mathcal{X}_{ID} \end{cases} \quad (1)$$

Post-hoc OOD detection does not need to introduce additional OOD data or change the structure of an original DNN model during the training process. This type of methods only needs to obtain the output of a pre-trained model during the OOD detection process which extremely reduces computing resources. We highlight two major lines of post-hoc OOD detection methods - score-based OOD detection and OOD detection optimization.

B. Score-based OOD Detection

Methods in this category employs score functions for OOD detection.

1) *MSP*: MSP [19] builds on maximum softmax probabilities which was the first work on post-hoc OOD detection. Compared with erroneously classified and OOD examples, correctly classified ID examples tend to have higher SoftMax probabilities.

2) *ODIN*: ODIN [20] is a grad-based method that uses temperature scaling and input perturbation to improve the discrimination between ID and OOD data. When the input is disturbed, an ID sample will have a larger score change than an OOD sample based on gradients. Besides, since a sufficiently large temperature threshold has a strong smoothing effect, ODIN transforms the softmax score back into a logit space to further distinguishes ID from OOD.

3) *Energy score*: To alleviate the defect of the SoftMax confidence score mechanism, a logit-based method (Energy score [16]) solved the overconfidence issue of ODIN in the form of log-sum-exp based on the energy score while achieving a better performance than ODIN for both ID and OOD data.

4) *KNN*: The KNN-based OOD detection method [21] focuses on the effectiveness of nonparametric nearest neighbour distance in OOD detection. Unlike previous works, the KNN-based OOD detection method does not impose any distribution assumptions and therefore provides greater flexibility and universality. This method shows the effectiveness of nearest neighbour-based OOD detection on multiple benchmarks and establishes superior performance.

5) *Gradient*: Due to the magnitude of gradients being higher for in-distribution (ID) data than that for OOD data, GradNorm [22] directly employs the vector norm of gradients where the gradients are obtained by the KL divergence between the softmax output and a uniform probability distribution, making it informative for OOD detection.

In addition to score-based post-hoc OOD methods, [23] introduces a label-based method in which the testing OOD samples should be relatively far away from the centroids of ID classes. In this work, the minimum Mahalanobis distance is used for all class centroids on OOD detection.

C. OOD Detection Optimization

Methods in this category normally work with score-based approaches for OOD detection. Rectified activation [14] and neurons sparsification [24] are the two mainstream methods of this category.

1) *ReAct*: ReAct provides a simple activation rectification strategy. It displays highly distinctive signature patterns for OOD distributions and generalizes effectively to OOD detection scores. ReAct can work with several commonly used OOD scoring functions derived from the model output including the probability, logit and grad based methods.

2) *DICE*: The core idea of DICE [24] is based on neuron sparsification, and it selectively uses the most salient weights to derive the output for OOD detection. By pruning away noisy signals, DICE reduces the output variance for OOD data.

The advantages of Trans-ReAct over the ReAct and DICE are reflected in its incorporations of transformations and a threshold anchor in achieving a stable performance on OOD detection.

III. METHODOLOGY

In this section, we first introduce the concept of transformations to enlarge the differentiation between ID and OOD data, and then we present the design of Trans-ReAct on OOD detection.

A. Transformation Enlarged OOD Differentiation

Transformations refer to the application of distinct mathematical operations or functions to modify the activation values of a neural network. These transformations are designed to accentuate the inherent differences between ID and OOD data, making them more distinguishable.

Essentially, rectified activation truncates OOD and ID activations into two separate parts based on a threshold. A crucial aspect of optimizing OOD detection is to enlarge the differentiation between ID and OOD data, especially when their activation values are initially similar.

The motivation for using transformations, specifically a power function ($f(x) = x^\theta$), stems from its ability to non-linearly rescale the activation space. Since OOD inputs typically produce lower activation values than ID inputs after rectification, a power function with $\theta > 1$ will disproportionately suppress these already low OOD activations (e.g., $0.2^2 = 0.04$) while having a less severe effect on higher ID activations (e.g., $0.8^2 = 0.64$). This selectively increases the relative gap between the two distributions, making them easier to separate with a simple scoring function.

For OOD detection, distinguishing between ID and OOD data can be challenging due to their close proximity in activation space, which may lead to misclassification by the classifier. By applying transformations to the activation values of ID and OOD data differentiation between ID and OOD data can be enlarged.

Let the output space $\mathcal{Y}_{ID}$ of an ID data in a DNN correspond to the set G, and $\mathcal{Y}_{OOD}$ correspond to the set H. The condition where ID and OOD data is distinguishable can be expressed as:

$$|d(\mathcal{Y}_{ID}(g); \mathcal{Y}_{OOD}(h))| < \kappa \quad (2)$$

where

- $\mathcal{Y}(\cdot)$ represents the activation score of an output space.
- $d(\mathcal{Y}_{ID}(g); \mathcal{Y}_{OOD}(h))$ represents the differentiation between g and h .
- κ represents a positive value

By applying a transformation η to the output space $\mathcal{Y}$, the activation differentiation between ID and OOD data can be expressed as:

$$|d(\mathcal{Y}_{ID}(g, \eta); \mathcal{Y}_{OOD}(h, \eta))| > \kappa \quad (3)$$

It should be noted that the OOD detection task is not confined to a single neuronal activation or a single input data point. For multiple neurons, individual transformations η can still amplify the disparities between ID and OOD data, and choosing an appropriate transformation further enhances the effectiveness on OOD detection. We provide a comprehensive analysis and evidence in Section D.

B. Trans-ReAct for OOD Detection

Rectified activation illustrates that anomaly activation values of neuron units can produce overconfident predictions on OOD data. Rectified activation truncates activations to remove the noises in differentiation of ID from OOD data. Considering a pre-trained DNN, the output of category j after rectified activation in the penultimate layer of a DNN can be represented as:

$$f_j^{ReAct}(x) = \sum_{i=1}^m W_j^i [\min(f_{n-1}^i(x), c)] + b \quad (4)$$

where

- $j \in (0, o]$, o is the number of classes in the output layer n .
- m is the number of units in the penultimate layer, $i \in (0, m]$.
- c is a truncation threshold value.
- $\min(f_{n-1}^i(x), c)$ truncates activations larger than c to limit the effect of noises on OOD data.
- W_j^i is the weight of j in corresponding to i .
- b is a bias vector.

The core elements of Trans-ReAct (i.e. transformations and threshold anchor selection) are further operations on top of ReAct which relies on Eq. (4). Trans-ReAct takes the output in the penultimate layer of a pre-train DNN as an initial condition and performs transformation $\eta(\cdot)$ to enlarge the activation differentiation between ID and OOD data.

Before introducing a transformation, it is worth highlighting the following points:

- Rectified activation limits the activation value of each neuron below the threshold c .
- In the penultimate layer of a DNN, the activation values of ID data are usually higher than that of OOD data except for abnormal OOD activation units after rectification activation.

- For two data points with a significant difference, the difference after a transformation could become larger.

Based on the three points above, we can explore that the threshold c limits the anomalous activation of the OOD data to prevent it from becoming larger through an individual transformation. And after a transformation of both ID and OOD data, the overall differentiation can be larger from a macro perspective. Therefore, we can choose a transformation function from a macro perspective to ensure that all the activation units shall have positive changes after performing a transformation with an aim to achieve better detection results.

The selection of a transformation function needs to be considered from the following two aspects:

- To avoid substantial alterations to the initial distribution of ID data, we shall utilise the mean of the ID distribution as a reference point throughout the entire transformation procedure. Furthermore, this mean reference point shall remain consistent even after several transformations have been carried out.
- After a transformation, the relative activation difference between ID and OOD data shall become larger and meet the sparsity condition of the data.

As a result, we use Multiple and Power functions as transformations in this work. To deal with randomness in selection of c , Trans-ReAct introduces a threshold anchor α which dynamically adjusts the corresponding c that is adaptive to an ID data. A key innovation of Trans-ReAct is this dynamic threshold anchor, α , which adapts the rectification process to the specific characteristics of the in-distribution (ID) data. Unlike the static threshold c used in ReAct, which can be suboptimal if the activation scale varies, our dynamic anchor enhances stability. The selection process is as follows: first, we determine the base threshold c by analyzing the ID training data's activation distribution (e.g., the 90th percentile). Second, α is set according to the rule: $\alpha = 1$ when threshold $c < 1$, and $\alpha = c$ when threshold $c \geq 1$. This rule normalizes the activation range; if activations are low ($c < 1$), it maps them to a consistent $[0, 1]$ space where the subsequent power function is most effective. If activations are already in a good range ($c \geq 1$), we preserve this scale. This adaptive approach ensures consistent performance across different models and mitigates the instability seen with static thresholds.

This adjustment effectively rescales low activation values, thereby enlarging the differentiation potential for both ID and OOD data when the initial threshold c is below the anchor α .

After performing the first transformation (from c to the threshold anchor), we use the Power function to further widen the differentiation gap between ID and OOD data. Since most of the OOD data is distributed in the low activation region, this transformation generates a larger change on the OOD data in comparison with the ID data. Fig. 2 shows an example using ImageNet. Based on the threshold c ($c=1$) obtained by the training data of ImageNet, we randomly extracted 10 sets of data from the test set of ImageNet and OOD dataset respectively. As shown in Fig. 2, the differentiation between the ID and OOD data is enlarged after a transformation of the Power function.

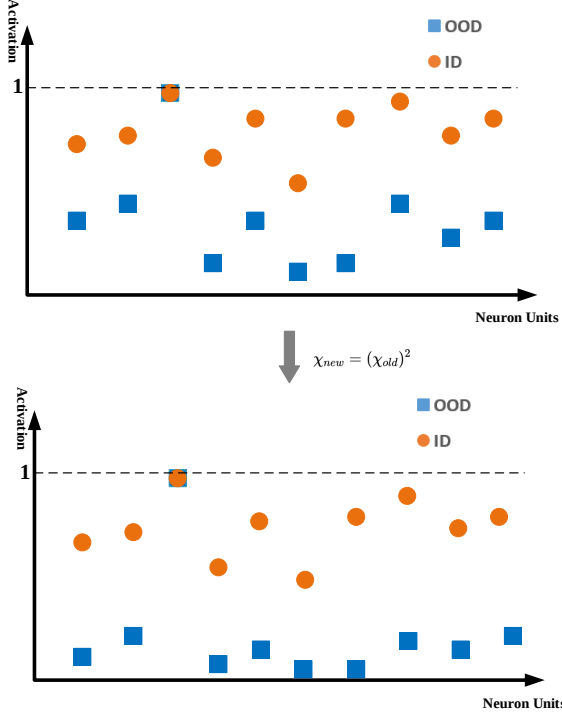


Fig. 2: An illustration of enlarging differentiation between ID (blue) and OOD (orange) data. The x-axis represents sample data points and the y-axis shows their activation values before and after transformation. The power transformation significantly increases the separation between the two distributions.

To maximize the effect of a transformation, we set $\alpha=1$ when threshold $c < 1$, and $\alpha=c$ when threshold $c \geq 1$. We design a sparsity condition for selection of a Power function to ensure that OOD detection can be optimized after executing the Power function.

$$\theta \rightarrow \begin{cases} \theta < 1 & \text{if } Act > 0 \ \& \ \sum_{i=1}^m \frac{\min(f_{n-1}^i(x), c)^\theta}{\min(f_{n-1}^i(x), c)} < 1 \\ \theta \geq 1 & \text{other} \end{cases}$$

where

$$Act = \sum_{i=1}^m \left(\frac{\min(\frac{\alpha}{c} f_{n-1}^i(x_{id}), \alpha)^\theta}{\min(\frac{\alpha}{c} f_{n-1}^i(x_{ood}), \alpha)^\theta} - \frac{\min(\frac{\alpha}{c} f_{n-1}^i(x_{id}), \alpha)}{\min(\frac{\alpha}{c} f_{n-1}^i(x_{ood}), \alpha)} \right) \quad (5)$$

- Act is the differentiation between the ID and OOD data.
- $\sum_{i=1}^m \frac{\min(f_{n-1}^i(x), c)}{\min(f_{n-1}^i(x), c)}$ is the sparse ratio.

We further perform another transformation to ensure that the transformed ID data distribution is close to the original one while the activation values of the transformed OOD become lower. Fig. 3 visualizes the operation process of Trans-ReAct using a threshold anchor α and a transformation in comparison with ReAct using a threshold value c . As shown in Fig. 3(a), Trans-ReAct first performs a transformation from threshold c to threshold anchor α after executing Eq. (6). The

activation values of the neurons obtain through transformation of $x_{new} = \frac{\alpha}{c} \cdot x_{old}$. Fig. 3(b) shows the distributions of ID and OOD data between 0 and threshold anchor α . Fig. 3(c) shows the activation values of ID and OOD data following the transformation function. Finally, we perform the last transformation function of $x_{new} = \frac{c}{\alpha} \cdot x_{old}$ to bring the ID data to its original distribution and make the activation values of the OOD data lower. As shown in Fig. 3(d), the ID data does not incur significant change, while the OOD data is mostly closed to 0.

Combining a threshold anchor α and a chaotic transformation $[\cdot]^\theta$, the model output of a category j after Trans-ReAct can be represented by Eq. (6).

$$f_j^{Trans-ReAct}(x) = \sum_{i=1}^m W_j^i \left[\min\left(\frac{c}{\alpha} \left[\frac{\alpha}{c} f_{n-1}^i(x) \right]^\theta, \alpha\right) \right] + b \quad (6)$$

C. Multi-layer Integration

In addition to the penultimate layer, Trans-ReAct can be used in other layers. The feature information extracted from multiple layers can also be integrated into the penultimate layer in Trans-ReAct to enhance the effect on OOD detection. For this purpose, we design two different integration strategies for improving OOD detection performance with multi-layer rectified activations methods.

The first integration strategy is a post-hoc-based multi-stage strategy (PhMS). For the activation values of neurons on different layers in a pre-trained model, we extract their activation values and the corresponding OOD detection results separately and finally give an integrated result.

The second integration strategy is an all-in-one strategy based on integration within the model (AioIS). For the neuronal activation values on different layers of a pre-trained model, we integrate the activation information to the penultimate layer based on the context association relationship of neurons in different layers and generate results.

D. Theoretical Analysis

In ReAct, it has been theoretically analyzed that the performance on OOD detection mainly depends on the differentiation of activation values between ID and OOD data. Following the theoretical analysis of ReAct, we further enlarge the differentiation of activation values between ID and OOD data.

Let

- $(z_1, \dots, z_m) =: Z$ ($z_i \in (0, c]$) be the activations in the penultimate layer.
- Z_{id} and Z_{ood} be the ID and OOD data respectively.
- $\hat{Z}$ and $\bar{Z}$ be the ReAct and Trans-ReAct activations in the penultimate layer respectively.
- D_{ReAct} and D_{Trans} be the differentiation expectations of $\hat{Z}$ and $\bar{Z}$ respectively, a priori assumption of $(\hat{Z}_{id} - \hat{Z}_{ood}) > 0$ that fits most cases is considered.

For simplicity, we set the parameters α to 1, θ to 2 and c to 1 in the analysis following the experimental settings of

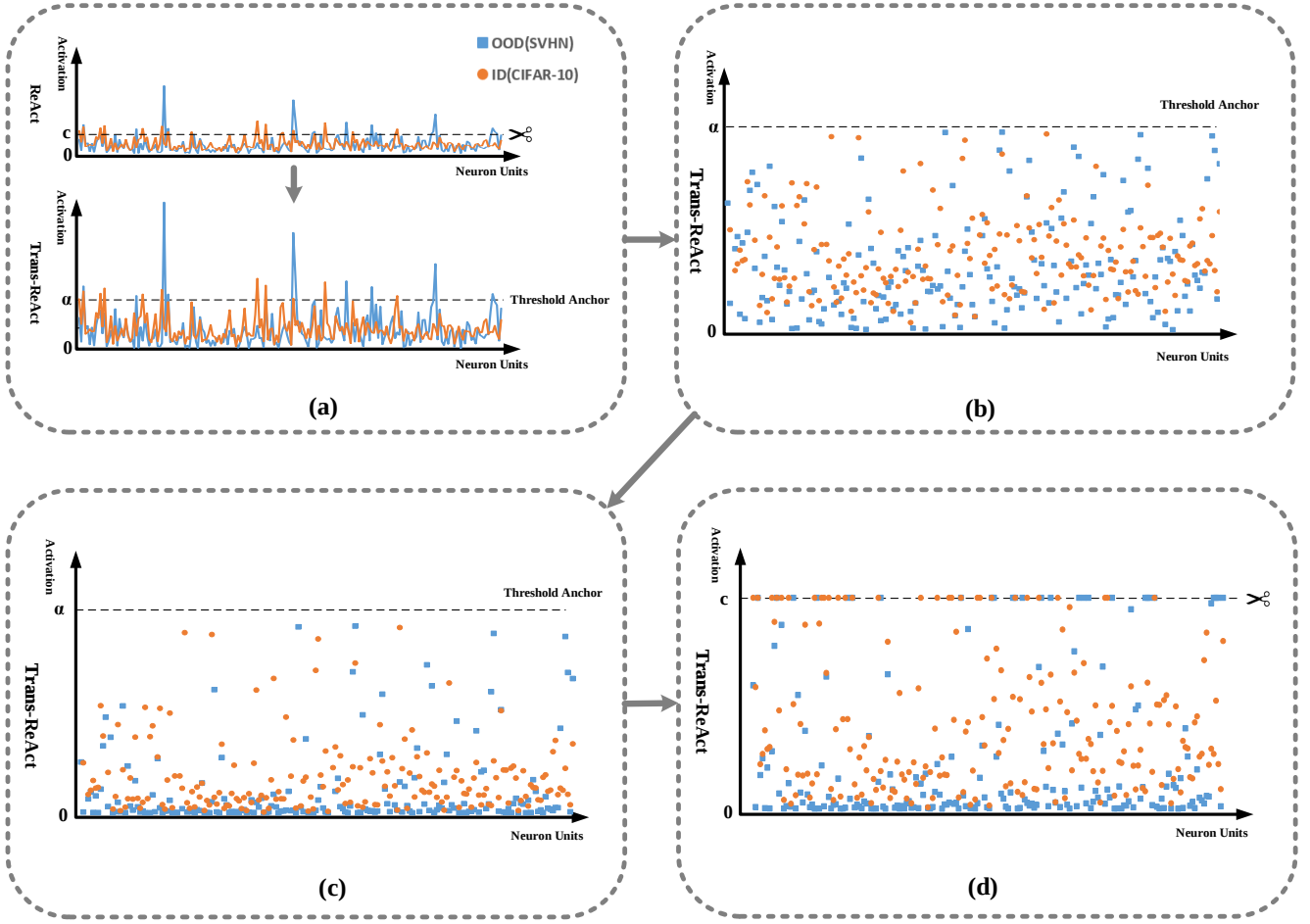


Fig. 3: Plots showing (a) truncation schemes in ReAct and Trans-ReAct respectively, (b) the distributions of ID and OOD data between 0 and threshold anchor α , (c) the distributions of ID and OOD data after a transformation, and (d) the final distributions of ID and OOD data through a regression process.

Trans-ReAct on ImageNet-1K. As a result, Trans-ReAct must satisfy $\tau \geq 0$ as shown in Eq. (7).

IV. EXPERIMENTAL RESULTS

This section validates the performance of Trans-ReAct through a set of experiments. We first present the main results of Trans-ReAct on a number of datasets and models, we then present the ablation experiments by analysing the impact of α and θ on the performance of Trans-ReAct.

$$\begin{aligned}
 \tau &= D_{Trans-ReAct} - D_{ReAct} \\
 &= \mathbb{E} \left(\frac{\bar{Z}_{id} - \bar{Z}_{ood}}{\bar{Z}_{id}} \right) - \mathbb{E} \left(\frac{\hat{Z}_{id} - \hat{Z}_{ood}}{\hat{Z}_{id}} \right) \\
 &= \frac{\sum_{i=1}^m \min_i(z_{id}^2, 1) - \sum_{i=1}^m \min_i(z_{ood}^2, 1)}{\sum_{i=1}^m \min_i(z_{id}^2, 1)} \\
 &\quad - \frac{\sum_{i=1}^m \min_i(z_{id}, 1) - \sum_{i=1}^m \min_i(z_{ood}, 1)}{\sum_{i=1}^m \min_i(z_{id}, 1)} \\
 &= \sum_{i=1}^m \left(\frac{\min_i(z_{ood}, 1)}{\min_i(z_{id}, 1)} - \frac{\min_i(z_{ood}^2, 1)}{\min_i(z_{id}^2, 1)} \right) \\
 &= \sum_{i=1}^m \left(\frac{\min_i(z_{ood}, 1)}{\min_i(z_{id}, 1)} - \frac{\min_i(z_{ood}, 1)^2}{\min_i(z_{id}, 1)^2} \right) \\
 &= \sum_{i=1}^m \frac{\min_i(z_{ood}, 1)}{\min_i(z_{id}, 1)^2} \sum_{i=1}^m (\min_i(z_{id}, 1) - \min_i(z_{ood}, 1)) \\
 &= \sum_{i=1}^m \frac{\min_i(z_{ood}, 1)}{\min_i(z_{id}, 1)^2} (\hat{Z}_{id} - \hat{Z}_{ood})
 \end{aligned} \tag{7}$$

A. Experimental Settings

We evaluated Trans-ReAct on three ID data benchmarks - ImageNet-1K, CIFAR-10 and CIFAR-100.

For OOD data under the CIFAR benchmarks, we validated 6 common benchmark datasets - Textures [25], SVHN [26], Places365 [27], LSUN-Crop [28], LSUN-Resize [28], and iSUN [29]. For OOD data under the ImageNet benchmark, we evaluated Trans-ReAct on Places365 [27], Textures [25], iNaturalist [30] and SUN [31] with non-overlapping classes. The evaluations spanned a range of domains including fine-grained images, scene images and textural images.

Consistent with [14], Trans-ReAct was mainly validated on two aspects - false-positive rates (FPR) and the receiver operating characteristic curve (AUROC) [1], [14]. FPR95 measures the probability of misclassifying OOD data as ID when the true positive rate (the rate of correctly identifying ID data) is held at 95%; a lower FPR95 is better. AUROC represents the likelihood of the model ranking a random ID sample higher than a random OOD sample, providing a comprehensive measure of separability across all thresholds; a higher AUROC is better. The FPR threshold is typically set to 95% in order to correctly classify most of the ID data. All the experiments were conducted using the PyTorch framework [32] in Python 3.8 and a single NVIDIA Tesla V100 GPU (32GB RAM).

B. Evaluations on ImageNet benchmark

1) The Experiment Results of Trans-ReAct on ResNet:

Compared with MSP, ODIN, Mahalanobis, KNN, GradNorm, and Energy, Trans-ReAct achieves the best performance across the 4 OOD datasets on the pre-trained ResNet-50, as shown in Table I. Using the Energy baseline, Trans-ReAct reduces FPR95 by 3.12% and 20.4%, and increases AUROC by 0.55% and 4.51% on iNaturalist and Textures, respectively. These results highlight its ability to handle challenging OOD detection scenarios effectively.

The experimental results of Trans-ReAct on ResNet-50 reveal several important insights:

First, Trans-ReAct consistently outperforms baseline methods such as MSP, ODIN, and Energy in terms of both FPR95 and AUROC, demonstrating its robustness across different scoring functions and OOD detection tasks.

Second, Trans-ReAct achieves significant improvements on challenging datasets like Textures and iNaturalist. For example, it achieves a 20.4% reduction in FPR95 and a 4.51% increase in AUROC on the Textures dataset compared to ReAct with Energy, highlighting its strength in distinguishing fine-grained and textural data.

Third, the method shows stable performance across diverse OOD datasets, including fine-grained images (iNaturalist), scene images (SUN, Places), and textural images (Textures). This robustness ensures strong generalization in real-world scenarios with diverse and unpredictable OOD characteristics.

Fourth, integrating Trans-ReAct with advanced baselines such as DICE further enhances its performance. For instance, combining it with the MSP+DICE baseline achieves an FPR95 of 30.88% and an AUROC of 94.31% on iNaturalist, while the

Energy+DICE combination achieves even better results, with an FPR95 of 15.86% and an AUROC of 96.90%.

Finally, Trans-ReAct maintains computational efficiency by leveraging the pre-trained ResNet-50 architecture without requiring additional costly computations or retraining. This makes it a scalable and practical solution for large-scale OOD detection tasks.

2) The Experiment Results of Trans-ReAct on MobileNet:

To evaluate the performance of Trans-ReAct on a lightweight model, we conducted experiments using a pre-trained MobileNet_v2 [33]. Trans-ReAct was compared with two representative OOD detection optimization methods, ReAct and DICE, using four score-based methods: MSP, Energy, ODIN, and Mahalanobis distance. Both Trans-ReAct and ReAct are compatible with MSP, ODIN, and Energy, while DICE works with MSP and Energy, as shown in Table II.

Compared to ReAct using the Energy baseline on MobileNet_v2, Trans-ReAct reduces FPR95 by 3.71%, 1.04%, and 11.95%, and increases AUROC by 1.47%, 0.84%, and 2.78% on SUN, Places, and Textures datasets, respectively. On average, Trans-ReAct reduces FPR95 by 4.2% and increases AUROC by 1.27% compared to ReAct, while achieving an average reduction of 3.02% in FPR95 and a 1.04% increase in AUROC compared to DICE.

Additionally, Trans-ReAct demonstrates consistent improvements over ReAct and DICE across the four datasets using all applicable baseline methods (MSP, ODIN, and Energy). Notably, when combined with DICE, Trans-ReAct achieves further performance gains. For instance, with the MSP+DICE baseline, Trans-ReAct reduces FPR95 to 60.64% and increases AUROC to 85.23%, while with the Energy+DICE baseline, it achieves an FPR95 of 31.00% and an AUROC of 92.73%.

These results indicate that Trans-ReAct effectively enhances OOD detection even in lightweight architectures like MobileNet_v2, demonstrating its versatility and scalability across different model types and dataset complexities.

3) The Experiment Results of Trans-ReAct on RepVGG:

To further verify the effectiveness of Trans-ReAct, experiments were conducted on RepVGG [34], a state-of-the-art BN-based model, as shown in Table III. Trans-ReAct was compared with ReAct using three baseline methods (MSP, ODIN, and Energy) for OOD detection. Two p -values (90% and 95%) were applied to ReAct as these settings yielded its best performance.

On RepVGG, Trans-ReAct achieved the best average performance across the four datasets, with improvements of 1.2% in FPR95 and 0.27% in AUROC compared to ReAct. Notably, Trans-ReAct consistently outperformed ReAct across most configurations and datasets. For example, under the Energy baseline with $p=95\%$, Trans-ReAct reduced FPR95 to 27.16% and increased AUROC to 94.03%, compared to ReAct's FPR95 of 28.36% and AUROC of 93.77%. Trans-ReAct also showed stable performance across both p -values, further demonstrating its robustness.

However, it is worth noting that on the SUN dataset under the ODIN baseline, ReAct achieved slightly better FPR95 (37.26%) compared to Trans-ReAct (37.31%). Despite this minor discrepancy, Trans-ReAct performed consistently better

TABLE I: The performance of Trans-ReAct on ResNet-50 (%).

Methods	OOD Datasets								Average	
	iNaturalist		SUN		Places		Textures		FPR95↓	AUROC ↑
	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC		
MSP (prob)	52.53	88.41	68.59	81.73	71.56	80.69	66.24	80.47	64.73	82.83
ODIN (grad)	47.66	89.66	60.15	84.59	67.89	81.78	50.23	85.62	56.48	85.41
Mahalanobis (label)	97.00	52.65	98.50	42.41	98.40	41.79	55.80	85.01	87.43	55.47
Energy (logit)	55.72	89.95	59.26	85.89	64.92	82.86	53.72	85.99	58.41	86.17
KNN (label)	59.00	86.47	68.82	80.72	76.28	75.76	11.77	97.07	53.97	85.01
Gradient (grad)	50.03	90.33	46.48	89.03	60.86	84.82	61.42	81.07	54.70	86.31
ReAct(MSP)	45.57	91.90	56.99	87.36	62.26	85.31	57.98	86.93	55.70	87.88
Trans-ReAct(MSP)	33.05	93.93	51.13	88.89	58.34	86.58	44.49	90.28	46.75	89.92
DICE(MSP)	52.52	89.17	66.91	82.34	69.29	81.40	69.13	79.75	64.46	83.17
ReAct(MSP+DICE)	44.38	91.30	51.51	86.92	58.06	84.77	67.71	82.38	55.42	86.34
Trans-ReAct(MSP+DICE)	30.88	94.31	48.27	89.22	55.94	86.94	48.62	89.57	45.93	90.01
ReAct(ODIN)	31.96	93.80	40.51	90.92	49.19	88.37	40.09	90.81	40.44	90.97
Trans-ReAct(ODIN)	24.51	95.43	34.88	92.57	44.21	89.70	25.71	94.57	32.33	93.07
ReAct(Energy)	19.66	96.38	23.41	94.45	33.18	91.96	46.52	90.31	30.69	93.27
Trans-ReAct(Energy)	16.54	96.93	23.40	94.45	32.75	92.03	25.76	94.82	24.61	94.56
DICE(Energy)	26.58	94.51	36.21	90.97	47.55	87.74	32.88	90.48	35.68	90.92
ReAct(Energy+DICE)	19.66	96.12	26.43	93.84	38.03	90.59	29.15	92.66	28.32	93.30
Trans-ReAct(Energy+DICE)	15.86	96.90	25.05	94.21	36.15	91.41	27.20	93.67	26.06	94.05

TABLE II: The performance of Trans-ReAct on MobileNet_V2 (%).

Methods	OOD Datasets								Average	
	iNaturalist		SUN		Places		Textures		FPR95↓	AUROC ↑
	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC		
MSP (prob)	64.29	85.32	77.02	77.10	79.23	76.27	73.51	77.30	73.51	79.00
ODIN (grad)	55.39	87.62	54.07	85.88	57.36	84.71	49.96	85.03	54.20	85.81
Mahalanobis (label)	62.11	81.00	47.82	86.33	52.09	83.63	92.38	33.06	63.60	71.01
Energy (logit)	59.50	88.91	62.65	84.50	69.37	81.19	58.05	85.03	62.39	84.91
ReAct(MSP)	63.85	87.64	76.28	78.75	78.24	77.87	65.21	83.41	70.90	81.92
Trans-ReAct(MSP)	54.00	89.47	68.71	81.80	73.41	80.10	54.18	86.78	62.58	84.54
DICE(MSP)	71.74	83.88	71.77	79.90	75.34	78.88	71.91	79.62	72.69	80.57
ReAct(MSP+DICE)	61.03	87.27	69.49	80.14	73.28	78.68	64.91	82.86	67.18	82.24
Trans-ReAct(MSP+DICE)	54.13	89.62	64.91	82.89	70.64	80.92	52.89	87.48	60.64	85.23
ReAct(ODIN)	40.38	92.60	61.19	85.13	67.49	82.90	37.70	91.42	51.69	88.01
Trans-ReAct(ODIN)	39.84	92.76	55.44	87.34	64.55	84.39	28.83	94.04	47.16	89.63
ReAct(Energy)	43.24	92.70	52.67	87.23	60.06	84.08	40.30	90.98	49.07	88.75
Trans-ReAct(Energy)	43.15	92.71	48.96	88.70	59.02	84.92	28.35	93.76	44.87	90.02
DICE(Energy)	49.17	90.22	44.67	89.81	58.96	85.20	38.75	90.69	47.89	88.98
ReAct(Energy+DICE)	32.57	93.55	31.20	92.81	46.72	88.01	16.21	96.24	31.67	92.65
Trans-ReAct(Energy+DICE)	32.05	93.55	30.58	92.95	45.87	88.00	15.51	96.41	31.00	92.73

on other datasets and metrics, particularly on challenging datasets such as Textures and iNaturalist.

These results demonstrate the effectiveness of Trans-ReAct on state-of-the-art architectures like RepVGG, showcasing its capability to enhance OOD detection performance with minimal parameter adjustments. The consistent improvements across diverse datasets and baseline methods further establish

Trans-ReAct as a reliable and robust solution for OOD detection.

4) *The Experiment Results of Trans-ReAct on BiT:* Going beyond Batch Normalization, we evaluated the performance of Trans-ReAct on BiT [35], which uses GroupNorm [36]. The results are presented in Table IV.

TABLE III: The performance of Trans-ReAct on RepVGG (%).

Methods	OOD Datasets								Average	
	iNaturalist		SUN		Places		Textures		FPR95↓	AUROC ↑
	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC		
MSP	54.12	87.99	67.85	82.07	72.03	80.95	70.94	78.44	66.23	82.36
ReAct(MSP)	40.95	91.99	57.92	86.49	65.31	83.72	36.77	92.16	50.24	88.59
Trans-ReAct(MSP)	40.20	92.50	57.07	87.09	64.79	84.12	35.23	92.72	49.32	89.11
ODIN	40.29	92.51	46.59	88.85	54.43	87.05	48.55	86.21	47.46	88.66
ReAct(ODIN)	25.82	95.38	37.26	91.98	49.96	88.33	13.92	97.36	31.74	93.26
Trans-ReAct(ODIN)	25.44	95.42	37.31	91.99	49.40	88.38	13.33	97.47	31.37	93.32
Energy	49.16	91.30	53.13	87.65	62.35	85.09	58.35	83.67	55.75	86.93
ReAct(Energy, p=90%)	34.23	93.63	40.26	91.03	52.47	87.65	27.87	93.06	38.71	91.34
Trans-ReAct(Energy, p=90%)	31.46	94.55	39.26	91.55	53.24	87.68	22.54	95.22	36.62	92.25
ReAct(Energy, p=95%)	20.75	96.22	32.67	92.72	45.38	89.09	14.65	97.06	28.36	93.77
Trans-ReAct(Energy, p=95%)	19.18	96.45	30.08	93.20	43.16	89.69	16.21	96.77	27.16	94.03

TABLE IV: The performance of Trans-ReAct on BiT (%).

Methods	OOD Datasets								Average	
	iNaturalist		SUN		Places		Textures		FPR95↓	AUROC ↑
	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC		
MSP	52.10	88.52	67.83	82.02	70.51	81.41	66.03	81.49	64.12	83.36
ReAct(MSP)	46.64	89.76	64.41	83.31	67.10	82.56	58.51	85.87	59.17	85.38
Trans-ReAct(MSP)	44.36	91.18	61.96	84.91	64.55	83.85	55.98	87.62	56.71	86.89
ODIN	36.69	93.59	51.94	89.36	58.62	87.39	50.71	88.43	49.49	89.69
ReAct(ODIN)	27.66	94.83	46.79	89.91	53.82	87.65	36.81	92.98	41.27	91.34
Trans-ReAct(ODIN)	26.51	94.98	46.14	89.91	53.19	87.72	33.79	93.50	39.90	91.53
Energy	42.40	92.96	52.02	89.31	57.90	87.18	50.69	88.65	50.75	89.52
ReAct(Energy, p=90%)	20.84	96.02	37.69	91.78	47.70	89.10	31.35	93.67	34.39	92.64
Trans-ReAct(Energy, p=90%)	19.85	96.22	35.83	91.98	45.58	89.86	22.87	95.59	31.03	93.41
ReAct(Energy, p=95%)	27.61	95.12	41.04	91.23	48.43	88.88	37.82	93.25	38.72	92.12
Trans-ReAct(Energy, p=95%)	21.49	96.04	35.15	92.36	42.62	90.01	36.47	93.93	33.93	92.93

As shown in Table IV, Trans-ReAct achieves the best performance across all four datasets, outperforming ReAct on average by 3.36% in FPR95 and 0.77% in AUROC. Notably, under the Energy baseline with $p=90\%$, Trans-ReAct reduces FPR95 to 31.03% and increases AUROC to 93.41%, compared to ReAct's FPR95 of 34.39% and AUROC of 92.64%. Similarly, under the ODIN baseline, Trans-ReAct consistently outperforms ReAct with an average FPR95 reduction to 39.90% and AUROC improvement to 91.53%.

These results demonstrate the robustness and adaptability of Trans-ReAct to alternative normalization methods like GroupNorm. The consistent improvements across different datasets and baseline methods further establish Trans-ReAct as an effective approach for OOD detection in state-of-the-art architectures.

5) *The Experiment Results on Object Detection* : Trans-ReAct can be used for computer vision tasks. For object detection, using BDD100K as ID data and MS-COCO as OOD data, the following results in Table V show that Trans-ReAct also performs best on the ResNet-50 backbone.

TABLE V: A comparison with competitive OOD detection methods on CIFAR-10 (%).

Methods	FPR95 ↓	AUROC ↑	mAP(ID) ↑
Energy	78.36	73.75	31.0
Energy+ReAct	75.83	76.13	30.4
Energy+Trans-ReAct	73.89	78.27	30.7

Overall, Trans-ReAct outperforms ReAct on average by 6.08% on FPR95 and 1.29% on AUROC in ResNet-50, on average by 4.2% on FPR95 and 1.27% on AUROC in MobileNet_v2, on average by 1.2% on FPR95 and 0.27% on AUROC in RepVGG, on average by 3.36% on FPR95 and 0.77% on AUROC in BiT.

C. Evaluations on CIFAR Benchmark

This section evaluates the performance of Trans-ReAct on CIFAR-10 and CIFAR-100 datasets.

TABLE VI: A comparison with competitive OOD detection methods on CIFAR-10 (%).

Models	Methods	OOD Datasets												Average	
		SVHN		LSUN		LSUN_Resize		iSUN		Textures		Places365			
		FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC
ResNet-18	Energy	71.99	86.19	41.81	92.44	33.99	94.57	38.64	93.70	50.27	90.58	50.68	88.97	47.90	91.07
	ReAct	76.01	87.18	21.68	96.29	30.82	95.31	35.37	94.45	51.70	90.56	50.14	89.52	44.32	92.22
	Trans-ReAct	67.91	88.98	11.85	97.99	29.76	95.38	33.47	94.57	47.43	91.48	47.31	90.21	39.62	93.10
ResNet-50	Energy	59.73	91.00	3.90	99.01	25.29	95.84	29.32	95.35	54.65	89.30	39.04	92.70	35.32	93.87
	ReAct	64.36	89.57	6.54	98.67	29.29	95.07	32.75	94.58	50.44	90.39	41.08	91.93	37.41	93.37
	Trans-ReAct	44.20	92.60	1.02	99.49	17.61	96.74	20.74	96.39	46.84	91.79	35.50	93.37	27.65	95.06
DenseNet	Energy	38.88	94.18	3.79	99.18	11.83	97.77	12.32	97.77	56.58	86.37	36.53	93.32	26.66	94.76
	ReAct	36.49	94.65	3.75	99.18	11.41	97.81	11.89	97.78	53.37	89.15	36.38	93.35	25.55	95.32
	Trans-ReAct	29.85	95.55	3.44	99.30	8.98	98.14	9.58	98.07	46.31	91.59	35.99	93.38	22.36	96.01

To ensure a fair comparison, we followed the same training settings as ReAct, using two pre-trained models—ResNet-18 and ResNet-50—based on Yerlan’s code [37]. Specifically, ResNet-18 and ResNet-50 were trained for 100 epochs with an initial learning rate of 0.1, which decayed by a factor of 10 at epochs 50, 75, and 90. We compared Trans-ReAct with ReAct using the Energy scoring function and also validated its performance on the pre-trained DenseNet provided in DICE.

1) *Evaluations on CIFAR-10*: Table VI presents the evaluation results of Trans-ReAct on six OOD datasets: Textures, SVHN, Places365, LSUN-Crop, LSUN-Resize, and iSUN.

From the results, Trans-ReAct consistently outperforms both Energy and ReAct across all datasets. For ResNet-18, Trans-ReAct achieves an average FPR95 reduction of 8.28% and an AUROC improvement of 2.03% compared to Energy, while reducing FPR95 by 4.7% and increasing AUROC by 0.88% compared to ReAct.

For ResNet-50, Trans-ReAct shows even greater improvements. It reduces FPR95 by 7.67% and increases AUROC by 1.19% compared to Energy, and reduces FPR95 by 9.76% and increases AUROC by 1.69% compared to ReAct.

On DenseNet, Trans-ReAct demonstrates similar effectiveness, reducing FPR95 by 4.3% and increasing AUROC by 1.25% compared to Energy, while reducing FPR95 by 3.19% and increasing AUROC by 0.69% compared to ReAct.

These results highlight the robustness and adaptability of Trans-ReAct across different model architectures. Particularly, its performance gains on challenging datasets like Textures and SVHN demonstrate its ability to better distinguish between ID and OOD data compared to existing methods.

2) *Evaluations on CIFAR-100*: Table VII presents the evaluation results of Trans-ReAct on six OOD datasets: Textures, SVHN, Places365, LSUN-Crop, LSUN-Resize, and iSUN.

For ResNet-18, Trans-ReAct consistently outperforms both Energy and ReAct across all datasets. On average, Trans-ReAct achieves a significant reduction of 18.01% (compared to Energy) and 2.16% (compared to ReAct) in FPR95, along with an AUROC improvement of 6.36% (compared to Energy) and 0.36% (compared to ReAct).

For ResNet-50, Trans-ReAct outperforms Energy and ReAct on five datasets, with the exception of LSUN. On average, Trans-ReAct achieves a reduction of 8.65% (compared to

Energy) and 8.42% (compared to ReAct) in FPR95, as well as an AUROC improvement of 5.87% (compared to Energy) and 3.77% (compared to ReAct).

For DenseNet, Trans-ReAct demonstrates consistent superiority over Energy and ReAct on all datasets. It achieves an average reduction of 8.00% (compared to Energy) and 5.23% (compared to ReAct) in FPR95, and an AUROC improvement of 4.94% (compared to Energy) and 2.40% (compared to ReAct).

These results highlight the robustness of Trans-ReAct across various architectures and datasets. Notably, its substantial improvements on FPR95 indicate its effectiveness in reducing false positives, while the consistent AUROC gains demonstrate its enhanced capability to distinguish OOD data from ID data. Trans-ReAct’s performance across multiple architectures further establishes its reliability as a versatile and effective OOD detection method.

These results highlight the robustness and adaptability of Trans-ReAct across different model architectures. As a key takeaway, the consistent and significant reduction in FPR95, such as the 9.76% improvement on ResNet-50 over ReAct, underscores our method’s enhanced ability to distinguish between ID and OOD data compared to existing methods.

D. ID and OOD Differentiation

Table VIII provides a comprehensive comparison of the activation results of both ReAct and Trans-ReAct methods using 5,000 ID data samples (ImageNet-1K) and 5,000 OOD data samples (iNaturalist) in the penultimate layer of a pre-trained ResNet-50 model. Each cell in Table VIII represents the number of activation units within a specific activation range, allowing for a detailed understanding of the distribution of activation values for both ID and OOD data.

Compared to ReAct, Trans-ReAct demonstrates a significant improvement in differentiating ID and OOD data by dynamically adjusting the activation values. Specifically, Trans-ReAct reduces the average activation value of OOD data by 43%, while the reduction for ID data is only 30%. This targeted reduction highlights Trans-ReAct’s ability to selectively suppress the activation of OOD data without overly diminishing the meaningful activations of ID data. Consequently, Trans-ReAct achieves a larger separation in the activation space between ID and OOD data, which is crucial for effective OOD detection.

TABLE VII: A comparison with competitive OOD detection methods on CIFAR-100 (%).

Models	Methods	OOD Datasets												Average	
		SVHN		LSUN		LSUN_resize		iSUN		Textures		Places365			
		FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC
ResNet-18	Energy	84.27	83.84	37.45	94.02	65.86	85.63	69.65	83.98	89.06	72.39	84.33	72.79	71.77	82.11
	ReAct	53.55	91.21	34.49	93.92	52.46	89.04	53.08	88.66	63.51	87.01	78.46	78.18	55.92	88.01
	Trans-ReAct	51.62	91.66	24.76	95.73	52.30	89.04	53.01	88.68	62.70	87.46	78.18	78.26	53.76	88.47
ResNet-50	Energy	74.20	86.57	12.93	97.60	83.67	79.02	85.88	78.46	91.81	69.46	84.42	72.93	72.15	80.67
	ReAct	78.27	86.68	30.84	94.06	82.73	83.52	82.64	83.95	71.95	83.19	85.08	75.21	71.92	84.44
	Trans-ReAct	61.04	89.95	25.41	95.02	74.44	86.36	73.75	86.77	63.12	85.77	83.25	75.35	63.50	86.54
DenseNet	Energy	87.93	81.32	14.65	97.44	70.54	80.16	74.45	78.98	84.18	71.04	80.69	75.77	68.74	80.79
	ReAct	84.15	80.96	25.52	94.94	59.98	87.91	65.19	86.59	78.05	78.66	82.94	70.07	65.97	83.19
	Trans-ReAct	74.95	84.53	11.98	97.75	59.75	88.23	64.44	86.67	73.44	80.58	79.91	76.61	60.74	85.73

A closer examination of the activation distribution reveals that:

- For **ReAct**, the ID data activations are primarily concentrated in higher activation ranges (e.g., 0.35–0.45), while the OOD data activations exhibit a broader spread, with a noticeable presence in both low and moderate activation ranges (e.g., 0.05–0.3). This overlap between ID and OOD data activations in the higher ranges compromises ReAct’s ability to distinguish between the two.
- In contrast, **Trans-ReAct** shifts the OOD data activations towards lower ranges (e.g., 0–0.15), effectively reducing their overlap with the ID data activations. Additionally, Trans-ReAct maintains a strong concentration of ID data activations in the mid-range intervals (e.g., 0.2–0.3), ensuring that the essential features of ID data are preserved.

The total activation values further underscore these differences. For ReAct, the total activation for ID data is 901.9, significantly higher than the OOD data’s 614.55. While this indicates some level of separation, the overlap in distribution limits its effectiveness. On the other hand, Trans-ReAct not only reduces the total activation for both ID (616.65) and OOD data (354.2), but also achieves a sharper contrast between the two distributions, as evidenced by the greater reduction in OOD activations relative to ID activations.

This analysis demonstrates the practical benefits of using Trans-ReAct in real-world scenarios. By compressing OOD activations into lower ranges while preserving key ID characteristics, Trans-ReAct achieves better differentiation and reduces false positive rates. Its dynamic thresholding mechanism eliminates the need for manual tuning, ensuring robust performance across datasets. Additionally, the reduced overall activations lower computational complexity, making it well-suited for resource-constrained environments.

In conclusion, Table VIII highlights that Trans-ReAct effectively separates ID and OOD activations, dynamically enhancing detection performance and outperforming ReAct in precision and robustness.

E. Ablation Study on Threshold Anchor α

To evaluate the effect of the threshold anchor α in Trans-ReAct, we employed Energy as the baseline scoring function in the tests. For ReAct, we set the value of parameter p to 65%, 85%, 90%, and 95%, respectively. As shown in Table

IX, ReAct performs best when p is set to 90% on ImageNet-1K, 95% on CIFAR-10, and 95% on CIFAR-100. However, as discussed earlier, ReAct suffers from the randomness in setting parameter p , necessitating a manual process to determine the optimal rectification threshold for stable OOD detection performance.

In contrast, Trans-ReAct leverages the threshold anchor α to dynamically adjust the rectification threshold, bypassing the need for manually setting parameter p . This dynamic adjustment improves both the robustness and generalization of Trans-ReAct across diverse datasets. As seen in Table IX, Trans-ReAct achieves consistently better performance than ReAct in terms of FPR95 and AUROC across ImageNet-1K, CIFAR-10, and CIFAR-100. For instance, on ImageNet-1K, Trans-ReAct reduces FPR95 from 30.69 (ReAct) to 24.61, while improving AUROC from 93.27 to 94.56.

The results demonstrate that the threshold anchor α is effective in mitigating the instability caused by manual parameter tuning, thus enhancing OOD detection performance across datasets.

F. Ablation Study on Sensitivity to θ

The parameter θ controls the sparsity condition in Trans-ReAct, ensuring that ID activation values are not excessively diluted compared to OOD data. To study the impact of θ , we conducted a series of experiments on ImageNet-1K as the ID dataset and iNaturalist as the OOD dataset. Table X summarizes the results, showing that Trans-ReAct consistently outperforms ReAct across various θ settings.

It can be observed that varying θ results in minor fluctuations in Trans-ReAct’s performance. For example, increasing θ from 1.5 to 1.7 slightly reduces FPR95 from 16.71 to 16.45, while AUROC improves from 96.89 to 96.94. These results indicate that Trans-ReAct is robust to the choice of θ , maintaining high performance across different configurations.

The improved robustness and performance of Trans-ReAct compared to ReAct further validate the effectiveness of the sparsity mechanism in θ . Additionally, the stability of Trans-ReAct eliminates the need for labor-intensive hyperparameter tuning.

G. OOD Detection on Different Layer:

We evaluated the effects of multiple layers on ResNet-50 (ID: CIFAR-10) as shown in Table XI. For conventional

TABLE VIII: Distribution of activation counts in the penultimate layer of a pre-trained ResNet-50 when applying ReAct and Trans-ReAct. Results are shown for 5,000 in-distribution samples (ImageNet-1K) and 5,000 out-of-distribution samples (iNaturalist), with each cell indicating the number of activation units whose values fall within the specified activation range.

Methods	Data	Activations														Total Activations
		0-0.05	0.05-0.1	0.1-0.15	0.15-0.2	0.2-0.25	0.25-0.3	0.3-0.35	0.35-0.4	0.4-0.45	0.45-0.5	0.5-0.55	0.55-0.6	0.6-0.65	0.65-0.7	
ReAct	ID(ImageNet-1K)	0	0	0	4	16	43	200	589	661	351	121	43	14	6	901.9
	OOD(iNaturalist)	0	50	229	275	448	203	347	149	68	98	41	47	24	29	614.55
Trans-ReAct	ID(ImageNet-1K)	0	21	62	164	646	569	387	132	45	51	7	4	0	0	616.65
	OOD(iNaturalist)	262	540	431	246	145	94	83	59	31	20	20	13	12	8	354.2

TABLE IX: The effect of rectification threshold (%).

	Methods	FPR95	AUROC	ID acc.	Rectification
		↓	↑	↑	Threshold c
ImageNet	Energy	58.41	86.17	75.08	-
	ReAct(p=90)	30.69	93.27	74.82	1.00
	ReAct(p=95)	34.31	92.79	75.38	1.50
	ReAct(p=85)	33.48	92.37	74.06	0.84
	ReAct(p=65)	76.62	73.09	68.76	0.50
	Trans-ReAct	27.02	94.19	75.41	$\alpha=c$
	Trans-ReAct	24.61	94.56	75.56	$\alpha=1$
CIFAR-10	Energy	35.32	93.87	93.32	-
	ReAct(p=90)	51.39	89.04	92.27	0.25
	ReAct(p=95)	46.70	90.88	92.66	0.36
	ReAct(p=85)	64.99	83.47	91.38	0.18
	ReAct(p=65)	96.31	62.20	77.46	0.08
	Trans-ReAct	27.65	95.06	93.34	$\alpha=1$
	Trans-ReAct	27.65	95.06	93.34	$\alpha=1$
CIFAR-100	Energy	71.77	82.11	73.54	-
	ReAct(p=90)	59.92	87.37	70.26	0.48
	ReAct(p=95)	55.92	88.01	72.31	0.73
	ReAct(p=85)	70.19	84.57	68.39	0.36
	ReAct(p=65)	88.58	73.71	40.27	0.17
	Trans-ReAct	53.76	88.47	72.82	$\alpha=1$
	Trans-ReAct	53.76	88.47	72.82	$\alpha=1$

TABLE X: The effect of θ on Trans-ReAct performance.

Methods	FPR95	AUROC	ID acc.
	↓	↑	↑
ReAct	19.66	96.38	74.82
Trans-ReAct($\theta=1.5$)	16.71	96.89	75.41
Trans-ReAct($\theta=1.7$)	16.45	96.94	75.51
Trans-ReAct($\theta=2.0$)	16.54	96.93	75.56
Trans-ReAct($\theta=3.0$)	17.60	96.75	75.52

DNNs, the further back the layer fits the information, the stronger the ability to fit information, so the better on OOD detection. For this purpose, we carried out experiments to add multi-layer activation information for OOD detection. It is found in Table XI that the OOD detection after multi-layer accumulation is better than that using a single layer.

In this paper, we use penultimate layer for presentation being consistent with other related papers. To the best of our knowledge, most existing OOD detection methods use the penultimate layer for model feature extraction.

V. CONCLUSION

In this paper, we presented Trans-ReAct, a novel post-hoc method that significantly improves out-of-distribution (OOD) detection. By introducing mathematical transformations and a dynamic threshold anchor, our method effectively enlarges

TABLE XI: A comparison with competitive OOD detection methods on different layers (%), evaluated on ResNet-50 with CIFAR-10 as ID. The integration methods, Penultimate and Hidden Layer Mean Score (PhMS) and Adaptive Integrated Layer Score (AioIS), show superior performance by combining features from multiple layers.

	Method	OOD Datasets			
		SVHN		Textures	
		FPR95	AUROC	FPR95	AUROC
Bottleneck-2	Energy	86.82	81.43	60.80	86.00
	ReAct	79.66	84.92	55.82	87.89
	Trans-ReAct	74.19	86.49	53.51	88.73
Bottleneck-3	Energy	76.01	87.18	51.70	90.56
	ReAct	72.12	87.95	50.21	91.00
	Trans-ReAct	70.74	88.41	50.96	90.71
Integration(PhMS)	Trans-ReAct	44.98	92.62	45.77	91.89
Integration(AioIS)	Trans-ReAct	43.91	92.65	45.12	91.93

the activation gap between in-distribution and OOD data while ensuring stable, robust performance without manual hyperparameter tuning. Experimental results confirm that Trans-ReAct consistently outperforms state-of-the-art methods across multiple benchmarks and model architectures.

The practical impact of this work is particularly relevant for safety-critical systems. The stability and reliability offered by Trans-ReAct are crucial for applications such as autonomous driving, where a vehicle must reliably identify and react to unseen road conditions, or in medical imaging, where a diagnostic model must flag anomalous scans that fall outside its training domain. By reducing false positives and providing more dependable OOD detection, Trans-ReAct offers a path toward more trustworthy and deployable AI systems. Future work will explore the application of these principles to other data modalities and more complex, multi-modal architectures, and will also consider integrating OOD detection strategies into load balancing mechanisms in cloud environments, such as game-theoretic multi-server scheduling [38], to enhance system robustness under varying workloads.

REFERENCES

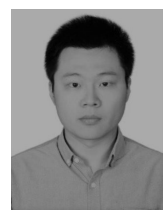
- [1] J. Yang, K. Zhou, Y. Li, and Z. Liu, “Generalized out-of-distribution detection: A survey,” *arXiv preprint arXiv:2110.11334*, 2021.
- [2] D. Hendrycks, M. Mazeika, and T. Dietterich, “Deep anomaly detection with outlier exposure,” in *International Conference on Learning Representations (ICLR)*, 2018.
- [3] J. Nandy, W. Hsu, and M. L. Lee, “Towards maximizing the representation gap between in-domain & out-of-distribution examples,” *Advances in Neural Information Processing Systems (NIPS)*, vol. 33, pp. 9239–9250, 2020.

- [4] K. Osawa, S. Swaroop, M. E. E. Khan, A. Jain, R. Eschenhagen, R. E. Turner, and R. Yokota, "Practical deep learning with bayesian principles," *Advances in neural information processing systems (NIPS)*, vol. 32, 2019.
- [5] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11 976–11 986.
- [6] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, pp. 770–778.
- [8] X. Liu, Z. Zhao, Y. Zhang, C. Liu, and F. Yang, "Social network rumor detection method combining dual-attention mechanism with graph convolutional network," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 5, pp. 1125–1136, 2023.
- [9] Q. He, D. Zhang, X. Wang, L. Ma, Y. Zhao, F. Gao, and M. Huang, "Graph convolutional network-based rumor blocking on social networks," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 6, pp. 1254–1266, 2023.
- [10] K. Guo, Z. Zhao, Z. Yu, W. Guo, R. Lin, Y. Tang, and L. Wu, "Network embedding based on biased random walk for community detection in attributed networks," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 6, pp. 1298–1308, 2023.
- [11] Y. Xu, K. Li, J. Hu, and K. Li, "A genetic algorithm for task scheduling on heterogeneous computing systems using multiple priority queues," *Information Sciences*, vol. 270, pp. 255–287, 2014.
- [12] J. Chen, K. Li, Z. Tang, K. Bilal, S. Yu, C. Weng, and K. Li, "A parallel random forest algorithm for big data in a spark cloud computing environment," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 4, pp. 919–933, 2016.
- [13] B. Pu, K. Li, S. Li, and N. Zhu, "Automatic fetal ultrasound standard plane recognition based on deep learning and iiot," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7771–7780, 2021.
- [14] Y. Sun, C. Guo, and Y. Li, "React: Out-of-distribution detection with rectified activations," *Advances in Neural Information Processing Systems (NIPS)*, vol. 34, pp. 144–157, 2021.
- [15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.
- [16] W. Liu, X. Wang, J. Owens, and Y. Li, "Energy-based out-of-distribution detection," *Advances in Neural Information Processing Systems (NIPS)*, vol. 33, pp. 21 464–21 475, 2020.
- [17] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.
- [18] S. Singla, S. Singla, and S. Feizi, "Improved deterministic l2 robustness on cifar-10 and cifar-100," in *International Conference on Learning Representations (ICLR)*, 2021.
- [19] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *International Conference on Learning Representations (ICLR)*. PMLR, 2017.
- [20] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *International Conference on Learning Representations (ICLR)*, 2018.
- [21] Y. Sun, Y. Ming, X. Zhu, and Y. Li, "Out-of-distribution detection with deep nearest neighbors," in *International Conference on Machine Learning (ICML)*. PMLR, 2022, pp. 20 827–20 840.
- [22] R. Huang, A. Geng, and Y. Li, "On the importance of gradients for detecting distributional shifts in the wild," *Advances in Neural Information Processing Systems*, vol. 34, pp. 677–689, 2021.
- [23] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," *Advances in neural information processing systems (NIPS)*, vol. 31, 2018.
- [24] Y. Sun and Y. Li, "Dice: Leveraging sparsification for out-of-distribution detection," in *European Conference on Computer Vision*, 2022.
- [25] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing textures in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2014, pp. 3606–3613.
- [26] Q. V. Le, J. Ngiam, Z. Chen, D. K. Chia, P. S. Koh, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," in *Proceedings of the 22nd International Conference on Neural Information Processing Systems (NIPS)*, 2011, pp. 2155–2163.
- [27] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 6, pp. 1452–1464, 2017.
- [28] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, "Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop," *arXiv preprint arXiv:1506.03365*, 2015.
- [29] P. Xu, K. A. Ehinger, Y. Zhang, A. Finkelstein, S. R. Kulkarni, and J. Xiao, "Turkergaze: Crowdsourcing saliency with webcam based eye tracking," *arXiv preprint arXiv:1504.06755*, 2015.
- [30] G. Van Horn, O. Mac Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, and S. Belongie, "The inaturalist species classification and detection dataset," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2018, pp. 8769–8778.
- [31] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in *IEEE computer society conference on computer vision and pattern recognition (CVPR)*. IEEE, 2010, pp. 3485–3492.
- [32] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *arXiv preprint arXiv:1912.01703*, 2019.
- [33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2018, pp. 4510–4520.
- [34] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "Repvgg: Making vgg-style convnets great again," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 13 733–13 742.
- [35] A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, and N. Houlsby, "Big transfer (bit): General visual representation learning," in *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 2020, pp. 491–507.
- [36] Y. Wu and K. He, "Group normalization," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.
- [37] Y. Idelbayev, "Proper ResNet implementation for CIFAR10/CIFAR100 in PyTorch," https://github.com/akamaster/pytorch_resnet_cifar10, 2021, accessed: 2021-07-20.
- [38] C. Liu, K. Li, and K. Li, "A game approach to multi-servers load balancing with load-dependent server availability consideration," *IEEE Transactions on Cloud Computing*, vol. 9, no. 1, pp. 1–13, 2018.



He is a Fellow of the British Computer Society (BCS) and the Institute of Engineering and Technology (IET).

Maozhen Li is a Professor in the Department of Electronic and Computer Engineering, Brunel University London, UK. He received the PhD from the Institute of Software, Chinese Academy of Sciences in 1997. His main research interests include high performance computing, big data analytics and intelligent systems with applications to smart grid, smart manufacturing and smart cities. He has over 240 research publications in these areas including 4 books. He has served over 30 IEEE conferences and is on the editorial board of a number of journals.

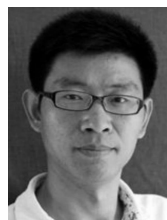


Yu Liang received the PhD degree from the Department of Computer Science and Technology, Tongji University, Shanghai in 2023. He is currently a Research Scientist at Shanghai Satellite Network Research Institute Co., Ltd, China. His research interests include lightweight deep neural networks (DNNs), interpretation methods for DNNs and graph DNNs. He has over 10 research publications in these areas. His review work on local interpretation methods for deep neural networks has received 243 citations since its publication in 2021.



Man Qi is a Reader in Computing at Canterbury Christ Church University. Her research interests are in Cyber Security, Data Intelligence, IoT and HCI. Dr Qi published over 80 research papers including over 30 journal papers and is the editorial board member for 5 international journals. She has been the PhD external examiners for many Universities in Australia and UK. Dr Qi has served as chair/program committee member for around 50 international conferences and been long term reviewer for many international journals. She is Fellow of British Computer

Society (FBCS) and Fellow of Higher Education Academy (FHEA).



Guanjun Liu (M'16, SM'19) received the Ph.D. degree in computer software and theory from Tongji University, Shanghai, China, in 2011. He was a Post-Doctoral Research Fellow with the Singapore University of Technology and Design, Singapore, from 2011 to 2013. He was a Post-Doctoral Research Fellow with the Humboldt University of Berlin, Berlin, Germany, from 2013 to 2014, supported by the Alexander von Humboldt Foundation. He is currently a Professor with the Department of Computer Science and Technology, Tongji University. He has

authored over 100 articles and two books. His research interests include Petri net theory, model checking, machine learning, information security, real-time concurrent system and multi-agent system.