FIDEI: Hyperparameter-Free Out-of-Distribution Detection

A.Q.M. Sazzad Sayyed[†], Nathaniel D. Bastian[‡] and Francesco Restuccia[†]
[†]Northeastern University [‡]United States Military Academy
Corresponding author: sayyed.a@northeastern.edu

Abstract

We propose a hyperparameter-free Out-of-Distribution (OOD) detection framework named FIDEI, which combines geometric and output-space cues to achieve superior OOD detection performance. FIDEI constructs a classifier-aligned hyperspherical feature space, computes a relative class-alignment score conditional to background alignment, and combines this geometric signal with logit-level confidence in a calibrated way that matches the scale of the two. FIDEI operates post hoc, does not require retraining of the backbone Deep Neural Network (DNN) nor access to auxiliary OOD data for hyperparameter tuning – thus avoiding catastrophic failure upon wrong choice of the hyperparameter. We provide a geometric interpretation showing that the resulting score approximates a cosine-margin criterion conditioned on background deviation, supporting the robustness of FIDEI across near- and far-OOD settings. Extensive experiments on CIFAR and ImageNet benchmarks demonstrate that FIDEI achieves consistently strong and balanced performance across datasets and architectures, with average relative improvement of 10% in terms of False Positive Rate (FPR) on ImageNet200 evaluations, while being more than $9\times$ faster than methods comparable in performance in high-throughput settings. Experiments on edge devices show that FIDEI introduces less than 3% additional latency beyond backbone inference.

1. Introduction

DNNs are usually overconfident with OOD samples like unstructured noise or semantically unrelated images [26]. Such overconfidence decreases their applicability in safety-critical applications [30]. As a result, OOD detection has become an important problem [39]. Post-hoc OOD detection is particularly attractive since it does not require retraining and can work on any given pre-trained DNN [12, 21]. However, the performance of the post-hoc detectors often degrades when test samples semantically differ from the training distribution (e.g., CIFAR-10 vs. CIFAR-100 or fine-grained ImageNet subsets) but still share considerable features [10, 29].

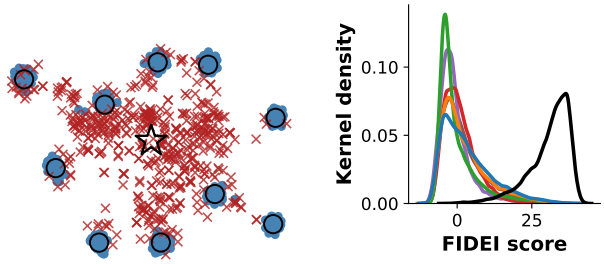

Figure 1. **Left:** t-SNE projection of centred, L2-normalized penultimate features. CIFAR-10 test images (blue circles) cluster around their class prototypes (open black circles), whereas Tiny-ImageNet images (red crosses) drift toward the global background prototype (black star), illustrating the geometry exploited by FIDEI. **Right:** Kernel-density estimates of FIDEI scores for ID (black) and six OOD datasets; far-OOD sets (SVHN, Textures, MNIST, Places365) are separated, while near-OOD CIFAR-100 and Tiny-ImageNet still yield Area Under Receiver Operating Curve (AUROC) ≥ 0.95.

Key conceptual gap of current approaches. A well-trained classifier induces a feature space in which In-Distribution (ID) samples concentrate around respective class centroids [28] with OOD samples being distanced [1], as shown in Figure 1. Post-hoc detectors use Mahalanobis distance [18, 29], distance to k-nearest-neighbors , [32], distance to decision-boundary [20] or energy in null-space [35]. However, existing approaches often incur substantial computational or memory costs, e.g., for K classes and d dimensional features, class-conditional covariance estimation in [29] requires $O(K \cdot d^2)$ storage, or $O(N \cdot d)$ memory or storing samples for K-nearest neighbor search in [32]. Many also ignore how the classifier shapes the feature space [32]. In Section 6, we show that accounting for this structure significantly improves OOD detection.

Our approach. We propose FIDEI[1], a post-hoc OOD detection framework that is hyperparameter-free and lightweight. FIDEI has three main steps: (1) center features using a minimum-norm offset that yields zero logit at the final layer [35]; (2) project features onto the unit hypersphere to prioritize angular information [24]; and (3) score

[1]"Fidei" is a Latin word referring to faith, trust, or confidence.

inputs by nearest-class prototype distance conditioned on distance to a global background prototype, motivated by the stronger attraction of OOD samples to the background (Figure 1) and supported by prior work [20, 29]. We then fuse this score with the energy score [21] via scale-matched combination to incorporate logit confidence [21], producing a unified confidence measure without retraining or covariance estimation and yielding strong and consistent OOD detection (Figure 1, right).

Our contributions are summarized as follows:

• We propose a lightweight and hyperparameter-free OOD detection approach that extends distance-based methods by leveraging classifier-induced feature geometry, directional stability, and conditioning on a background prototype;

• Extensive experiments on OpenOOD benchmarks across CIFAR and ImageNet demonstrate that `FIDEI` achieves strong and consistent performance across architectures, yielding up to a 10% relative improvement in FPR on ImageNet200 and ranking first in $14/24$ CIFAR100 architecture–metric settings, all without hyperparameter tuning. Additional experiments on resource-constrained edge platforms show that `FIDEI` incurs negligible inference-time overhead (approximately 2–3$\%$ beyond backbone inference) and no measurable increase in power consumption, confirming its practicality for real-world deployment.

2. Related Work

OOD detection has been widely studied through confidence estimation, feature geometry, and distributional deviation. Early post-hoc methods use softmax-based confidence [12] or energy scores [21], based on the tendency of ID data to produce more concentrated logits. Later work improves logit geometry via training-time changes, e.g., logit normalization to reduce feature-norm variability [37] and Decoupling MaxLogit (DML) [43], which decomposes outputs into angular and radial components (MaxCosine/-MaxNorm). While effective, these methods require retraining or modifying objectives, limiting their use with pre-trained models.

Another line of work leverages internal model responses. Gradient-based approaches such as Gradient Norm (GradNorm) [3] use loss sensitivity for detection, while activation-based methods like Rectified Activation (ReAct) [31] and Activation Shaping by Binarizing (ASH-B) [8] improve separability by clamping, pruning, or reshaping activations; [38] refines [8] by partially scaling features instead of pruning. Despite simplicity, these methods often rely on global heuristics and dataset-specific hyperparameters and do not explicitly model class-conditional geometry. Distance- and density-based detectors more directly target feature-space structure. Mahalanobis scoring [18] and kNN detectors [32] measure deviation from ID feature statistics, while RMDS [29] improves robustness by subtracting a background distance term. However, estimating second-order statistics or nearest-neighbor structure in high-dimensional spaces can be unstable with limited calibration data and often degrades on near-OOD cases. ViM [35] incorporates classifier-induced geometry by centering features using the affine structure of the final linear layer and scoring energy in a residual subspace, but it remains largely class-agnostic and can miss fine-grained class-conditional cues important for near-OOD detection. More recent approaches improve OOD detection by learning stronger representations (e.g., self-supervised features with Mahalanobis scoring [10], or supervised/uncertainty-aware contrastive learning [33, 36] and hard-negative mining [41]). These methods typically require additional training and provide limited mechanistic interpretability at inference time.

`FIDEI` separates itself since it is purely post-hoc, avoids retraining and covariance estimation, and explicitly leverages classifier-aligned, class-conditional structure via normalized class prototypes and a global background prototype. By casting OOD detection as a relative, background-subtracted comparison in a geometry aligned with the classifier's decision boundaries, `FIDEI` yields a more interpretable OOD decision.

3. Notation and Problem Formulation

We denote a DNN parameterized by weights θ with $\mathcal{F}_\theta : \mathcal{X} \to \mathbb{R}^K$, where K denotes the number of outputs. The network outputs a logit vector $\mathbf{g}(\mathbf{x}) = \mathcal{F}_\theta(\mathbf{x})$ for an input $\mathbf{x} \in \mathcal{X}$. An associated feature representation $\phi(\mathbf{x}) \in \mathbb{R}^d$ can be extracted from the penultimate layer. Let $\mathbf{X}$ denote the input random variable with realization $\mathbf{x}$. We use $\mathcal{P}_{id}$ to represent the ID data distribution and $\mathcal{D}_{id}$ to denote a dataset sampled from $\mathcal{P}_{id}$. Similarly, $\mathcal{P}_{ood}$ and $\mathcal{D}_{ood}$ denote the OOD distribution and dataset, respectively. At test time, samples may originate from either $\mathcal{P}_{id}$ or an unknown $\mathcal{P}_{ood}$.

OOD detection as hypothesis testing. The goal of OOD detection is to decide whether a test input $\mathbf{x}$ was generated from $\mathcal{P}_{id}$ or $\mathcal{P}_{ood}$. To this end, we design a scalar scoring function $S : \mathcal{X} \to \mathbb{R}$ such that the induced score distributions

$$S(\mathbf{x}) \sim \mathcal{P}^S_{id} \quad \text{if } \mathbf{x} \sim \mathcal{P}_{id}, \qquad S(\mathbf{x}) \sim \mathcal{P}^S_{ood} \quad \text{if } \mathbf{x} \sim \mathcal{P}_{ood}, \tag{1}$$

are well separated. A decision is obtained via thresholding:

$$1(\mathbf{x}) = \begin{cases} 1, & \text{if } S(\mathbf{x}) \geq \tau, \\ 0, & \text{if } S(\mathbf{x}) < \tau, \end{cases} \tag{2}$$

where $1(\mathbf{x}) = 1$ indicates an ID prediction and $1(\mathbf{x}) = 0$ indicates an OOD prediction. The threshold τ controls

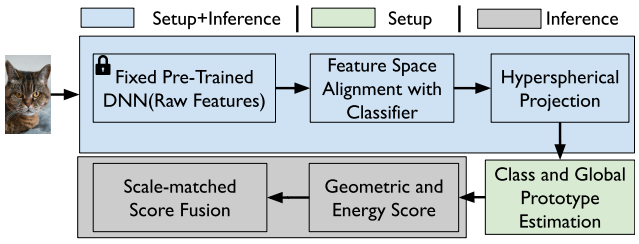

Figure 2. Overview of FIDEI. The complete pipeline consists of setup and inference phases.

the trade-off between True Positive Rate (TPR) and FPR. In the ideal case, $\mathcal{P}_{id}^{S}$ and $\mathcal{P}_{ood}^{S}$ have disjoint support, allowing perfect separation. In practice, performance is evaluated by fixing a target TPR and reporting the corresponding FPR. Without loss of generality, we assume that higher scores correspond to higher confidence of being ID.

4. The FIDEI Framework

As mentioned in Section 1, we compute Euclidean distance to class-mean relative to background mean on a classifier-aligned hyperspherical feature space. We integrate this distance based score with energy score computed from the output logits after scaling. The scale-factor is computed from the training data. The overall pipeline consists of a *setup* phase and an *inference* phase.

4.1. Setup Phase

Feature Space Alignment with Classifier Geometry. Let the pretrained classifier produce logits

$$\mathbf{g}(\mathbf{x}) = W\phi(\mathbf{x}) + \mathbf{b}, \tag{3}$$

where $\phi(\mathbf{x}) \in \mathbb{R}^d$ denotes the penultimate-layer feature, and $(W, \mathbf{b})$ are the parameters of the final linear layer. We compute a classifier-aware centering offset

$$\mathbf{u} = -W^{\dagger}\mathbf{b}, \tag{4}$$

where $W^{\dagger}$ denotes the Moore–Penrose pseudoinverse. This offset corresponds to the minimum-norm feature shift that approximately removes the affine bias of the classifier. In other words, this centers the features at the point where all logits would be zero – aligning the feature space with the decision geometry of the model. As we will show in Section 5, this makes FIDEI more robust for classification [6] and detection [43].

Hyperspherical feature representation. [43] shows that directional information is more stable under distribution shift. We project the features onto unit hypersphere after centering with $\mathbf{u}$. Specifically, for each sample $\mathbf{x}$ in the training set, we compute the centered and normalized feature

$$\hat{\mathbf{z}}(\mathbf{x}) = \frac{\phi(\mathbf{x}) - \mathbf{u}}{\|\phi(\mathbf{x}) - \mathbf{u}\|_2}. \tag{5}$$

Prototype estimation. Using the centered and normalized features $\hat{\mathbf{z}}(\mathbf{x})$, we compute i) class prototypes $\boldsymbol{\mu}_c = \mathbb{E}[\hat{\mathbf{z}}(\mathbf{x}) \mid y = c]$, and ii) a global background prototype $\boldsymbol{\mu} = \mathbb{E}[\hat{\mathbf{z}}(\mathbf{x})]$.

Relative Euclidean geometric score. For each sample in the train set, we define a geometry-based score

$$S_g(\mathbf{x}) = \max_c \left(\|\hat{\mathbf{z}}(\mathbf{x}) - \boldsymbol{\mu}\|_2^2 - \|\hat{\mathbf{z}}(\mathbf{x}) - \boldsymbol{\mu}_c\|_2^2 \right), \tag{6}$$

which measures how much closer a sample is to its nearest class prototype than to the global background mean. This relative formulation reduces sensitivity to global feature shifts and emphasizes class-aligned geometry.

Energy score and scale calibration. We compute the energy score from logits as given by Eqn. 7:

$$E(\mathbf{x}) = \log \sum_{k=1}^{K} \exp(g_k(\mathbf{x})). \tag{7}$$

We note that simply adding the geometric $S_g(\mathbf{x})$ and the energy score $E(\mathbf{x})$ can be detrimental, as the larger of the two can dominate the other one. This will be equivalent to using the larger score only and may even hurt the OOD detection in some cases. To solve this challenge, we compute a scale-factor α as the ratio of the average maximum logit and the average geometric score. Specifically, we set

$$\alpha = \frac{\mathbb{E}_{\mathcal{D}_{\text{id}}}\left[\max_k g_k(\mathbf{x})\right]}{\mathbb{E}_{\mathcal{D}_{\text{id}}}\left[S_g(\mathbf{x})\right] + \varepsilon}, \tag{8}$$

where ε is a small constant for numerical stability. This choice rescales the geometric score so that its magnitude is comparable to that of the logit-based confidence, enabling stable additive fusion without introducing additional hyperparameters.

4.2. Inference Phase

Given a test input $\mathbf{x}$, we compute its feature representation $\phi(\mathbf{x})$ and logits $\mathbf{g}(\mathbf{x})$ using the pretrained network.

Geometric score. We calculate the normalized, centered feature $\hat{\mathbf{z}}(\mathbf{x})$ using Eq. (5). Then we compute the relative geometric score $S_g(\mathbf{x})$ with the centered and normalized features $\hat{z}(x)$ as in Eq. (6).

Energy score. We compute the energy score $E(\mathbf{x})$ from logits using Eq. (7).

Fused OOD score. The final FIDEI confidence score is defined as

$$S(\mathbf{x}) = E(\mathbf{x}) + \alpha\, S_g(\mathbf{x}), \tag{9}$$

where higher values indicate greater confidence that $\mathbf{x}$ originates from the ID distribution. This fused score leverages complementary information: the energy term captures global logit-level confidence, while the geometric term captures class-aligned deviations in feature space. This can also be interpreted as the ID confidence obtained from a virtual-logit similar to [35]. The major difference is the source of the logit – instead of null-space energy estimation, which is sensitive to the number of null-space dimensions, we compute the logit from the distance to the class and background prototypes.

Decision rule. OOD detection is performed by thresholding $S(\mathbf{x})$ as described in Section 3. No retraining or fine-tuning is required, and all additional computation is performed post hoc.

5. Theoretical Analysis

In this section, we provide a theoretical analysis of FIDEI. We first characterize the geometric properties of the relative Euclidean score used by the method, and then show how conditioning on background deviation improves separability between ID and OOD. We reuse the same notations as introduced earlier and provide a definition for any new notation introduced in place. The proofs are provided in Supplementary Sec. A

Relative Euclidean Geometry on Unit Hypersphere

Proposition 1 (Cosine-Margin Equivalence) *For any input $\mathbf{x}$ and class c, assuming that distance to class mean $\boldsymbol{\mu}_c$ is approximately constant across classes, ranking samples by the geometric score $S_g(\mathbf{x})$ in 6 is equivalent to ranking them by $\max_c \hat{\mathbf{z}}(\mathbf{x})^\top(\boldsymbol{\mu}_c - \boldsymbol{\mu})$.*

Proposition 1 shows that S_g approximately implements a directional margin test on the unit hypersphere, measuring how much more aligned a sample is with a class prototype than with the background structure. As a result, S_g emphasizes angular, rather than radial, deviations from the ID manifold. This can be interpreted to be a key reason for the robustness of FIDEI in both near- and far-OOD settings based on the observation of robustness of angular metrics in previous work [43].

Conditioning on Background Deviation One of our key design choices in FIDEI is to evaluate class separation relative to the distance from a global background prototype. We adapt the argument from [20] to show how this improves the performance of FIDEI.

$$x = \|\hat{\mathbf{z}} - \boldsymbol{\mu}\|_2^2, \qquad y = \max_c (x - \|\hat{\mathbf{z}} - \boldsymbol{\mu}_c\|_2^2) \tag{10}$$

Here, x measures deviation from the background feature structure, while y measures class-conditional separation conditioned on that deviation.

Proposition 2 (Conditioning on Background Deviation [20]) *Let $f_{X,Y}$ and $g_{X,Y}$ denote the joint distributions of (X, Y) for ID and OOD, with marginals f_X, g_X and conditionals $f_{Y|X}, g_{Y|X}$. If the background deviation distributions are matched, $f_X = g_X$, then*

$$D_{\mathrm{KL}}(f_Y \,\|\, g_Y) \le \mathbb{E}_{X\sim f_X}\left[D_{\mathrm{KL}}\big(f_{Y|X}(\cdot|X) \,\|\, g_{Y|X}(\cdot|X)\big)\right] \tag{11}$$

Proposition 2 shows that conditioning class-separation statistics on background deviation increases separation between ID and OOD. Unlike prior work that enforces conditioning via explicit normalization or regularization, FIDEI achieves this effect implicitly through relative Euclidean distance.

Fusion with Energy-Based Confidence In addition to the geometric score, FIDEI incorporates the energy score

$$E(\mathbf{x}) = \log \sum_{k=1}^{K} \exp(g_k(\mathbf{x})), \tag{12}$$

which captures global logit-level confidence.

The final FIDEI score is defined as

$$S(\mathbf{x}) = E(\mathbf{x}) + \alpha\, S_g(\mathbf{x}), \tag{13}$$

where $\alpha > 0$ is calibrated on ID data.

Let $\alpha = \frac{\mathbb{E}_{\mathcal{P}_{id}}[\max_k g_k(\mathbf{x})]}{\mathbb{E}_{\mathcal{P}_{id}}[S_g(\mathbf{x})]+\varepsilon}$ with $\varepsilon > 0$. Then, on ID data, $E(\mathbf{x})$ and $\alpha S_g(\mathbf{x})$ have comparable expected magnitudes, preventing either term from dominating the fused score.

Interpretation. The calibrated fusion combines complementary signals: $S_g(\mathbf{x})$ captures class-aligned geometric consistency, while $E(\mathbf{x})$ captures global uncertainty reflected in logits. As a result, the fused score improves robustness to both near-OOD and far-OOD samples.

Taken together, our theoretical analysis shows that FIDEI i) implements an approximate directional, class-aware separation criterion in normalized feature space and ii) implicitly conditions class separation on background deviation, increasing ID/OOD separation. These properties explain the comparable and consistent empirical performance of FIDEI across diverse datasets and architectures.

Table 1. OOD detection performance on CIFAR-10. We report Near- and Far-OOD results aggregated over datasets (FPR@95%TPR ↓, AUROC ↑).

Method	ResNet-18				ViT-Base			
	Near OOD		Far OOD		Near OOD		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
MSP	53.61	87.69	31.22	91.03	10.17	97.10	6.43	97.90
GradNorm	95.26	53.93	89.24	60.50	93.87	75.79	97.51	76.54
React	71.07	86.51	42.58	90.84	7.73	98.25	2.66	99.31
ViM	47.69	88.56	25.68	93.17	**6.65**	**98.55**	**2.07**	99.43
KNN	**34.42**	**90.73**	23.73	93.13	8.16	98.34	2.72	99.38
ASH	88.98	74.20	76.36	78.46	88.29	61.03	86.38	58.92
GEN	57.89	87.80	34.11	91.57	7.40	98.21	3.45	99.10
FDBD	36.62	90.45	23.49	93.20	7.62	98.29	2.89	99.27
SCALE	84.76	82.28	63.46	87.99	7.63	98.17	2.91	99.22
RMDS	42.27	89.57	24.31	92.46	7.50	97.96	3.70	98.87
FIDEI (Ours)	34.73	90.71	**22.81**	**93.39**	7.67	98.40	2.83	**99.32**

Table 2. OOD detection performance on CIFAR-100. We report Near- and Far-OOD results aggregated over datasets (FPR@95%TPR ↓, AUROC ↑).

Method	ResNet-18				ViT-Base			
	Near OOD		Far OOD		Near OOD		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
MSP	54.56	80.38	58.89	77.65	40.08	88.82	48.28	82.19
GradNorm	86.32	69.58	88.92	62.31	83.83	72.14	85.69	67.13
React	56.61	80.63	56.15	79.91	50.78	82.61	48.61	83.04
ViM	62.86	74.88	46.07	83.41	45.47	85.34	43.42	85.75
KNN	**53.58**	**80.98**	**45.98**	83.05	43.28	85.71	41.57	86.36
ASH	66.18	78.33	59.57	81.18	60.58	79.68	61.69	79.59
GEN	56.81	80.45	51.26	81.78	49.51	83.59	47.33	84.17
FDBD	55.06	81.26	54.47	80.60	35.74	92.13	39.28	87.36
SCALE	69.86	73.94	62.15	75.60	38.75	92.15	44.80	85.74
RMDS	60.48	78.95	79.57	70.03	**30.91**	**92.52**	**39.22**	86.11
FIDEI (Ours)	57.63	81.10	54.29	81.24	40.59	92.40	39.75	**93.20**

6. Experimental Analysis

We conduct extensive experiments to evaluate the effectiveness, robustness, and scalability of FIDEI as a post-hoc OOD detector. We focus our evaluation on (i) benchmarking performance against state-of-the-art methods under standard protocols, (ii) validating the geometric principles underlying FIDEI and (iii) analyzing calibration efficiency, stability, and failure modes.

6.1. Experimental Setup

Datasets. As in prior work, we use CIFAR-10 [16], CIFAR-100 [16], ImageNet-200 [5], and ImageNet-1k[5] benchmarks. We follow the OpenOOD [40, 42] protocol for all benchmarks. For CIFAR benchmarks, CIFAR-100 and TinyImageNet [17] are treated as near-OOD, and datasets such as SVHN [25], MNIST[7], Textures[4], and Places365[44] are treated as far-OOD. For ImageNet benchmarks, we use SSB-Hard[34] and NINCO[2], and far-OOD datasets such as iNaturalist[14], Textures, and OpenImage-O[15]. For Imagenet, we also evaluate following the full-spectrum OOD protocol.

Models. We evaluate both convolutional and transformer-based architectures. Specifically, we use ResNet-18 [11] and ViT-B16 [9] for CIFAR benchmarks, and ConvNext-small [23] and ViT-B16 for ImageNet benchmarks. Further

Table 3. OOD detection performance on ImageNet-1K. We report Near- and Far-OOD results aggregated over datasets (FPR@95%TPR ↓, AUROC ↑).

Method	ConvNeXt-S				ViT-B/16			
	Near OOD		Far OOD		Near OOD		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
ViM	70.13	78.66	25.10	92.64	49.73	87.24	22.28	93.71
GEN	60.28	82.58	45.95	90.08	**46.51**	**88.10**	**17.34**	**96.43**
ASH	94.98	62.98	94.56	66.09	82.37	70.66	71.74	75.91
KNN	65.63	78.31	26.81	92.77	65.42	76.47	28.67	93.26
FDBD	66.33	78.84	28.32	92.10	53.78	85.36	20.11	95.24
SCALE	87.92	62.96	89.09	57.97	46.56	88.14	17.37	96.44
FIDEI (Ours)	**57.76**	**82.75**	**25.31**	**93.67**	56.33	83.36	19.83	95.43

Table 4. OOD detection on ImageNet-200 (aggregated Near/-Far OOD). We report Near- and Far-OOD results aggregated over datasets (FPR@95%TPR ↓, AUROC ↑).

Method	ConvNeXt-S				ViT-B/16			
	Near OOD		Far OOD		Near OOD		Far OOD	
	FPR ↓	AUROC ↑	FPR ↓	AUROC ↑	FPR ↓	AUROC ↑	FPR ↓	AUROC ↑
MSP	79.78	78.45	71.86	83.86	57.38	84.20	29.69	93.59
EBO	95.37	57.02	96.99	53.30	49.34	88.76	16.22	96.79
React	78.65	72.45	73.52	77.39	50.86	88.41	16.50	96.71
ViM	39.81	88.83	11.34	95.92	**32.78**	**92.77**	16.33	96.03
KNN	40.48	91.53	8.09	97.32	40.96	90.61	9.51	98.04
ASH	95.94	59.31	94.65	64.79	49.82	88.62	15.74	96.81
GEN	61.80	84.00	45.59	91.06	49.57	88.61	15.58	96.84
FDBD	59.60	82.95	21.23	94.19	46.69	88.90	14.51	96.66
SCALE	90.14	54.34	78.25	59.15	49.74	88.61	15.79	96.80
RMDS	35.87	92.05	8.17	97.17	39.86	91.39	10.66	95.87
FIDEI (Ours)	**34.81**	**93.06**	**7.63**	**97.52**	39.69	91.80	**7.67**	**97.93**

results on DINOV2(ViT-S/14 and ViT-B/14) are provided in the Appendix.

Baselines. We compare FIDEI against a diverse set of post-hoc OOD detectors, including confidence-based methods (MSP[12], Energy[21], GEN[22]), feature-based approaches (ReAct[31], ViM[35]), and distance-based techniques (RMDS[29], kNN[32]), as well as recent strong baselines such as FDBD[20] and Scale[13].

Evaluation Metrics. We report AUROC and FPR@95%TPR. Lower FPR and higher AUROC indicate better performance. All results follow the OpenOOD evaluation protocol.

Hardware and Hyperparameter Specification We have conducted our experiments on Intel(R) Xeon(R) Silver 4410Y CPU with 64 cores and A100 GPU with 80GB of RAM. For the runtime latency analysis on CIFAR10 dataset and Vit-B/16 architecture, we reported the average of three runs, while each run consisted of 10 batches of size 1024 each. We reported the average runtime for a single batch. All DNNs used for experiments are either pre-trained following standard training recipes and kept fixed during OOD detection, or fetched from PyTorch Hub. For experiments on edge devices, we have utilized Raspberry Pi 5 and Jetson Orin Nano.

6.2. Comparison with State-of-the-Art Methods

We report the comparison of FIDEI with the state-of-the-art methods from Table 1-4. We report the average performance in near- and far-OOD setting here. The detailed results per OOD dataset is reported in the appendix.

CIFAR Benchmarks. Tables 1-2 show that across CIFAR-10 and 100 benchmarks and both ResNet-18 and ViT-Base architectures, FIDEI consistently matches or outperforms state-of-the-art OOD detectors in both Near- and Far-OOD regimes. On CIFAR-10 with ResNet-18, FIDEI achieves the best average Far-OOD performance (MNIST, SVHN, Texture, Places365) with FPR 22.81 and AUROC 93.39, and ranks second on Near-OOD with only a 0.31 gap in AUROC from KNN. With ViT-Base, FIDEI remains competitive, reaching among the highest average Far-OOD AUROC (99.32) and staying close to the best Near-OOD methods despite largely saturated performance. On CIFAR-100, FIDEI continues to be strong and stable. With ViT-Base, it attains the highest Far-OOD AUROC (93.20) and the second-best Near-OOD AUROC, showing consistent Far-OOD gains on MNIST, SVHN, Textures, and Places365 where several Near-OOD-focused methods (e.g., SCALE [38], FDBD [20], RMDS [29]) are less stable. Overall, FIDEI provides architecture-agnostic OOD detection on CIFAR-10/100 and is the top AUROC performer among competing methods, including FDBD [20] and RMDS [29], across both ResNet-18 and ViT-Base.

ImageNet Benchmarks. Tables 18 and 19 (Appendix C) show detailed results on ImageNet-200 while Table 4 summarizes them. On ConvNeXt-Small, FIDEI leads across both Near- and Far-OOD, achieving the best Near-OOD average (FPR 34.81, AUROC 93.06) and the strongest Far-OOD aggregate (FPR 7.63, AUROC 97.52). It improves over strong baselines (ViM, RMDS, KNN) on challenging Near-OOD sets (SSB-Hard, NINCO) while also attaining the lowest or near-lowest FPR and highest AUROC on Far-OOD sets (iNaturalist, OpenImage-O, Textures). On ViT-B/16, FIDEI is best in the Far-OOD regime (FPR 7.67, AUROC 97.93), outperforming ViM and RMDS and showing consistent gains on iNaturalist, Textures, and OpenImage-O, where several Near-OOD-optimized methods are more variable. This also applies to ImageNet-1K. From Tables 17 and 16, we can see that FIDEI remains competitive with top methods (excluding weaker baselines such as MSP, EBO, and React). For ConvNeXt, FIDEI achieves AUROC gains of about 1% over ViM in Far-OOD and improves Near-OOD FPR by 2.52% over GEN; no method is uniformly better than FIDEI on ConvNeXt-Small. For ViT-B/16, GEN is best in both regimes, but FIDEI stays close (within ∼1% Far-OOD AUROC). Importantly, GEN can degrade without the right hyperparameter setting (Near-OOD: FPR 57.17, AUROC 82.26; Far-OOD: FPR 29.90, AUROC 88.94), while SCALE is more stable on ViT-B/16 ImageNet-1K but performs poorly on most other architectures/benchmarks. Overall, FIDEI is more reliable than specialized SOTA detectors and does not require hyperparameter tuning. In summary, FIDEI provides strong, architecture-agnostic OOD detection on ImageNet-200 with an improved Near–Far balance.

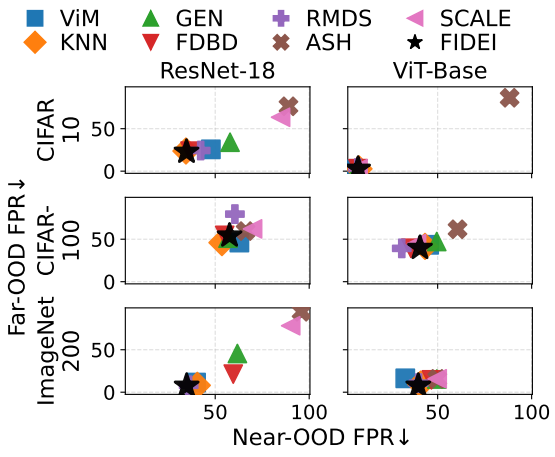

Figure 3. Near–Far OOD trade-off for the baselines across CIFAR-10, CIFAR-100, and ImageNet-200. The closer to the lower left corner a method is, the better its performance.

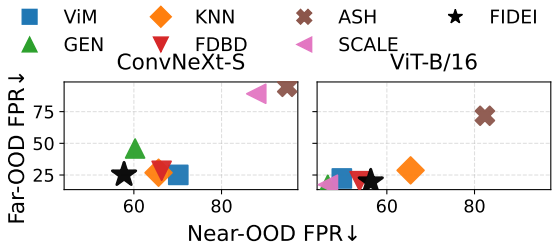

Figure 4. Near–Far OOD trade-off for the baselines across ConvNext-S and ViT-B/16 architectures. FIDEI shows better or at least comparable trade-off performance for both.

6.3. Near-Far Trade-off Evaluation

As shown in Figure 3 and 4, we can observe that FIDEI consistently occupies a favorable region of the Near–Far OOD trade-off space across CIFAR-10, CIFAR-100, ImageNet-200, and ImageNet-1K. This means FIDEI achieves competitive or lower Far-OOD false positive rates without sacrificing Near-OOD performance. In contrast, several strong baselines exhibit a pronounced trade-off: methods such as ASH and SCALE often perform poorly on Far-OOD despite acceptable Near-OOD behavior, while others (e.g., KNN, RMDS) can achieve low Far-OOD FPR at the cost of degraded Near-OOD separation on smaller

Table 5. Comparison of FIDEI with baseline approaches on ImageNet-200(FSOOD) benchmark for ConvNeXt-Small.

	SSB-Hard		NINCO		Near OOD		iNaturalist		Textures		OpenImage-O		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
EBO	90.79	51.44	92.70	50.66	91.75	51.05	90.88	52.67	96.55	51.99	92.09	55.65	93.17	54.74
React	79.88	60.76	85.74	53.11	82.81	56.94	82.63	53.72	95.48	52.02	88.97	50.97	89.03	52.24
ViM	74.48	54.91	73.64	51.75	**74.06**	53.33	66.74	63.17	69.84	61.88	**54.93**	62.49	60.12	61.68
KNN	84.40	56.48	73.30	58.77	78.85	57.62	53.77	64.94	**54.17**	**73.73**	55.92	68.20	54.62	68.96
ASH	**73.63**	**72.46**	84.43	**68.18**	79.03	**70.32**	73.22	65.63	68.86	65.46	53.11	65.31	83.51	65.93
GEN	89.66	52.09	81.21	58.64	85.43	55.37	73.09	66.95	92.48	58.24	60.87	66.53	75.48	68.36
FDBD	81.70	58.49	82.68	52.05	82.19	55.27	63.36	65.89	72.34	62.76	67.04	64.12	67.58	64.26
SCALE	86.41	68.45	89.28	61.64	87.85	65.05	84.19	64.42	90.36	55.15	73.66	56.76	82.73	61.27
RMDS	82.91	54.79	**70.20**	60.82	76.56	57.81	56.69	67.10	71.41	70.87	55.41	69.46	54.26	69.14
FIDEI (Ours)	82.89	68.12	71.29	67.94	77.09	68.03	**49.07**	**83.55**	57.85	71.86	55.58	**75.90**	**54.17**	**77.10**

datasets. FIDEI avoids this collapse, remaining close to the Pareto frontier across architectures and dataset scales. Notably, this behavior is consistent across both CNN and ViT backbones, suggesting that geometry–energy fusion of FIDEI yields a more balanced confidence signal that generalizes beyond dataset- or architecture-specific biases.

6.4. Full Spectrum OOD Benchmark

Table 5 reports OOD detection performance on the ImageNet-200 full-spectrum OOD evaluation benchmark using a ConvNeXt-Small backbone. Overall, FIDEI demonstrates a favorable balance between Near- and Far-OOD detection, outperforming or matching competitive baselines in aggregate despite strong variability across individual methods. On Near-OOD benchmarks (SSB-Hard and NINCO), while some methods such as ASH achieve strong AUROC on specific datasets, they exhibit inconsistent behavior across shifts; in contrast, FIDEI attains competitive Near-OOD performance (AUROC of 68.03) without relying on aggressive feature suppression or scaling. On Far-OOD benchmarks, FIDEI shows clear advantages: it achieves the *best overall Far-OOD AUROC* (77.10) and the *lowest average Far-OOD FPR* (54.17), outperforming ViM, RMDS, KNN, and FDBD. In particular, FIDEI substantially improves detection on iNaturalist and OpenImage-O, where several Near-OOD-optimized methods suffer from elevated false positives. These results indicate that FIDEI scales favorably to large-vocabulary ImageNet settings, providing robust Far-OOD separation while maintaining stable Near-OOD behavior, and avoiding the regime-specific brittleness observed in many prior approaches.

6.5. Ablation Study

Table 6 reports an ablation study on CIFAR-10 with a ResNet-18 backbone, where we remove one component of the proposed OOD detector at a time and evaluate performance on Near- and Far-OOD benchmarks. Removing *hyperspherical normalization* causes the largest degradation, with Near-OOD FPR increasing from 34.7 to 62.9 and AUROC dropping from 90.7 to 87.4, and Far-OOD FPR rising from 22.8 to 37.8, highlighting the importance of suppressing feature-norm variability and enforcing directional geometry. Removing *relative background subtraction* results in a moderate but consistent drop (Near-OOD FPR 34.7 to 38.2; Far-OOD AUROC 93.4 to 91.4), indicating that relative scoring against a global background prototype is critical for discounting shared semantic structure from consideration, particularly in semantically similar scenarios. Removing *classifier-aware centering* leads to a smaller yet consistent degradation (Near-OOD FPR 34.7 to 43.5), suggesting improved alignment with classifier geometry. Overall, while normalization is the single most impactful component, the best performance is achieved only when all three components are combined, confirming that centering, normalization, and background subtraction address complementary geometric failure modes and jointly enforce a classifier-aligned, direction-aware, and relative representation space for robust OOD detection.

Table 6. Ablation study on CIFAR-10 with ResNet-18, reporting Near-OOD and Far-OOD performance. Lower FPR and higher AUROC are better.

Method	Near-OOD		Far-OOD	
	FPR ↓	AUROC ↑	FPR ↓	AUROC ↑
Full (All)	**34.73**	**90.71**	**22.81**	**93.39**
Without Centering	43.54	89.77	24.89	93.19
Without Normalization	62.90	87.40	37.82	92.05
Without BG Subtraction	38.22	89.47	29.19	91.35

Table 7. Inference latency (ms) for different OOD detection methods on ViT-b16 with batch size 1024 for CIFAR10 dataset.

Methods	FIDEI	ViM	KNN
Latency(ms)	70	650	920

Fusion Ablation. Table 8 evaluates the contribution of each component and the sensitivity of the fusion weight. Geometric and energy scores alone provide competitive performance but neither achieves the best Near- and Far-OOD

Table 8. Fusion ablation on CIFAR-10 (ResNet-18).

Method	Near AUROC↑	Far AUROC↑
Geometric only	88.55	89.56
Energy only	86.95	91.80
Unweighted fusion	87.44	92.08
$0.8\alpha^\star$	87.35	92.03
$1.2\alpha^\star$	87.52	92.12
$\alpha^\star$ (Ours)	**90.71**	**93.39**

trade-off. Naive unweighted fusion also underperforms, indicating a scale mismatch between the two scores. In contrast, the proposed calibration consistently yields the best performance. Perturbing the calibrated weight by only $\pm 20\%$ significantly degrades performance, increasing Near-OOD FPR by more than 29%. These results validate the importance of the proposed scale-matching strategy and show that the calibration derived solely from ID data is critical for robust score fusion.

6.6. Latency, Memory, and Power Efficiency

We evaluate the computational and memory overhead of FIDEI relative to baseline OOD detectors. Table 7 shows that FIDEI incurs negligible inference overhead and substantially lower algorithmic cost than competing methods. In particular, in high-throughput GPU settings, FIDEI achieves approximately a $9\times$ reduction in inference-time overhead compared to state-of-the-art ViM and more than a $13\times$ reduction compared to KNN, owing to its lightweight class-prototype–based scoring and avoidance of null-space projections. In terms of memory, FIDEI requires only $O(Kd)$ storage to maintain class and background prototypes, where K denotes the number of classes and d the feature dimensionality. By contrast, KNN-based methods incur memory costs that grow linearly with the calibration set size, while covariance-based approaches require $O(d^2)$ memory for global statistics and $O(Kd^2)$ for class-conditional formulations. This favorable memory profile enables efficient calibration and deployment without storing per-sample features or high-dimensional covariance matrices. We further assess deployment efficiency on resource-constrained edge platforms using ImageNet-1K with ViT-B/16, including a CPU-only Raspberry Pi and a GPU-enabled NVIDIA Jetson Orin Nano. Across both platforms, FIDEI introduces *negligible additional latency* beyond the backbone forward pass. On the Raspberry Pi, all evaluated detectors–FIDEI, ViM, GEN, and KNN–exhibit nearly identical inference latency (approximately 470 ms per sample), indicating that post-hoc OOD scoring is fully amortized by the cost of ViT inference. Similarly, on the Jetson device, FIDEI achieves real-time performance at approximately 15 FPS, with a mean latency of ~ 67 ms per sample, matching the runtime of competing detectors.

These results are consistent with our earlier experiments on high-end accelerators. While FIDEI exhibits substantially lower inference-time overhead than ViM in high-throughput GPU settings (e.g., CIFAR-10 on NVIDIA A100), on edge devices the ViT forward pass dominates overall latency, causing inference-time differences among post-hoc detectors to vanish in practice. Importantly, power measurements on the Jetson platform show that FIDEI incurs no additional energy cost compared to baseline OOD detectors, with average power remaining stable at approximately 10 W (corresponding to ~ 0.67 J per sample). Taken together, these results demonstrate that FIDEI offers a favorable trade-off between runtime, memory efficiency, and deployability, enabling practical OOD-aware inference on both high-performance and edge computing platforms.

6.7. Further Discussions

To take a closer look at the performance of FIDEI, we examine (i) the class-dependent bias term governed by the prototype norm $\|\boldsymbol{\mu}_c\|_2^2$ and (ii) intra-class compactness measured by the variance of squared distances to the class prototype, $\mathrm{Var}(\|\hat{\mathbf{z}} - \boldsymbol{\mu}_c\|_2^2)$ (Table 9). For ResNet-18 on CIFAR-10, $\mathrm{Var}(\|\boldsymbol{\mu}_c\|_2^2)$ is very small, supporting the approximately constant-norm regime assumed in Proposition 1 and hence the cosine-margin interpretation of S_g. However, for ViT and for CIFAR-100, $\mathrm{Var}(\|\boldsymbol{\mu}_c\|_2^2)$ increases by one to two orders of magnitude, indicating a more bias-affected regime in which the exact shifted-margin form (Remark in Proposition 1) is the appropriate interpretation. Consistent with Section 5, both prototype-norm variability and intra-class variance increase from CIFAR-10 to CIFAR-100, and this joint shift correlates with degraded OOD detection performance, suggesting that departures from the low-bias, compact-cluster regime partially explain the performance drop of FIDEI.

Table 9. Empirical prototype-norm variance and intra-class variance across architectures and datasets.

Architecture / Dataset	$\mathrm{Var}(\|\boldsymbol{\mu}_c\|^2)$	$\mathbb{E}_c[\mathrm{Var}(\|f - \boldsymbol{\mu}_c\|^2)]$
ResNet-18 / CIFAR-10	9.7×10^{-5}	$\approx 1.4 \times 10^{-3}$
ViT / CIFAR-10	6.9×10^{-3}	$\approx 1.9 \times 10^{-2}$
ResNet-18 / CIFAR-100	2.8×10^{-3}	$\approx 2.1 \times 10^{-2}$
ViT / CIFAR-100	9.4×10^{-3}	$\approx 2.3 \times 10^{-2}$

7. Conclusion

In this work, we have proposed FIDEI a fully hyperparameter-free and computationally inexpensive approach to detect OOD. FIDEI shows consistency across diverse datasets and architectures. We connect this robustness with its interpretation as an approximate cosine-margin type metric. FIDEI pushes the boundary of distance-based OOD detection in feature space, effectively combining the different cues of OOD for more reliable detection.

8. Acknowledgements

This work was supported in part by NSF under Grants CNS-2312875 and OAC-2530896, DARPA under Cooperative Agreement D25AC00374-00, ONR under Award N00014-23-1-2221, and AFOSR under Award FA9550-23-1-0261.

References

[1] Mouïn Ben Ammar, Nacim Belkhir, Sebastian Popescu, Antoine Manzanera, and Gianni Franchi. NECO: NEural collapse based out-of-distribution detection. In *The Twelfth International Conference on Learning Representations*, 2024. 1

[2] Julian Bitterwolf, Maximilian Mueller, and Matthias Hein. In or out? fixing imagenet out-of-distribution detection evaluation. In *ICML*, 2023. 5

[3] Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *Proceedings of the 35th International Conference on Machine Learning*, pages 794–803. PMLR, 2018. 2

[4] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014. 5

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 5

[6] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 3

[7] Li Deng. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012. 5

[8] Andrija Djurisic, Nebojsa Bozanic, Arjun Ashok, and Rosanne Liu. Extremely simple activation shaping for out-of-distribution detection. In *The Eleventh International Conference on Learning Representations*, 2022. 2

[9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021. 5

[10] Stanislav Fort, Jie Ren, and Balaji Lakshminarayanan. Exploring the limits of out-of-distribution detection. In *NeurIPS*, 2021. 1, 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5

[12] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations*, 2016. 1, 2, 5

[13] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joseph Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. In *International Conference on Machine Learning*, pages 8759–8773. PMLR, 2022. 5

[14] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset, 2018. 5

[15] Ivan Krasin, Tom Duerig, Neil Alldrin, Andreas Veit, Sami Abu-El-Haija, Serge Belongie, David Cai, Zheyun Feng, Vittorio Ferrari, Victor Gomes, Abhinav Gupta, Dhyanesh Narayanan, Chen Sun, Gal Chechik, and Kevin Murphy. Openimages: A public dataset for large-scale multi-label and multi-class image classification. *Dataset available from https://github.com/openimages*, 2016. 5

[16] Alex Krizhevsky et al. Learning multiple layers of features from tiny images. 2009. 5

[17] Ya Le and Xuan S. Yang. Tiny imagenet visual recognition challenge. 2015. 5

[18] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems*, 31, 2018. 1, 2

[19] Jiachen Liang, RuiBing Hou, Minyang Hu, Hong Chang, Shiguang Shan, and Xilin Chen. Revisiting logit distributions for reliable out-of-distribution detection. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 12

[20] Litian Liu and Yao Qin. Fast decision boundary based out-of-distribution detector. *ICML*, 2024. 1, 2, 4, 5, 6, 11

[21] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in neural information processing systems*, 33:21464–21475, 2020. 1, 2, 5

[22] Xixi Liu, Yaroslava Lochman, and Christopher Zach. Gen: Pushing the limits of softmax-based out-of-distribution detection. In *CVPR*, pages 23946–23955, 2023. 5

[23] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022. 5

[24] Yifei Ming, Yiyou Sun, Ousmane Dia, and Yixuan Li. How to Exploit Hyperspherical Embeddings for Out-of-Distribution Detection? In *The Eleventh International Conference on Learning Representations*, 2023. 1

[25] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011. 5

[26] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 427–436, 2015. 1

[27] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 12

[28] Vardan Papyan, X. Y. Han, and David Donoho. Prevalence of neural collapse during the terminal phase of deep learning training. *PNAS*, 117(40):24652–24663, 2020. 1, 11

[29] Jie Ren, Peter J Liu, Emily Fertig, Jasper Snoek, Ryan Poplin, Mark Depristo, Joshua Dillon, and Balaji Lakshminarayanan. A simple fix to mahalanobis distance for improving near-ood detection. In *NeurIPS*, 2021. 1, 2, 5, 6

[30] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Margarethe Schmidt-Erfurth, and Georg Langs. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. In *Information Processing in Medical Imaging*, 2017. 1

[31] 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. 2, 5

[32] Yiyou Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution Detection with Deep Nearest Neighbors. *ICML*, 2022. 1, 2, 5

[33] Sameer Vaze, Bohyung Han, Neil Alldrin, and Deva Ramanan. Open-set recognition: A good closed-set classifier is all you need. In *ICLR*, 2022. 2

[34] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need. In *ICLR*, 2022. 5

[35] Haoqi Wang, Zhizhong Li, Litong Feng, and Wayne Zhang. Vim: Out-of-distribution with virtual-logit matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4921–4930, 2022. 1, 2, 4, 5

[36] Zhen Wang, Xiaoyu Zhang, and Xinyue Zhang. Better safe than sorry: Improving out-of-distribution detection with uncertainty quantification. *arXiv preprint arXiv:2303.12094*, 2023. 2

[37] Hongxin Wei, Renchunzi Xie, Hao Cheng, Lei Feng, Bo An, and Yixuan Li. Mitigating neural network overconfidence with logit normalization. In *International Conference on Machine Learning*, pages 23631–23644. PMLR, 2022. 2

[38] Kai Xu, Rongyu Chen, Gianni Franchi, and Angela Yao. Scaling for training time and post-hoc out-of-distribution detection enhancement. In *ICLR*, 2024. 2, 6

[39] Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *arXiv e-prints*, pages arXiv–2110, 2021. 1

[40] Jingkang Yang, Pengyun Wang, Dejian Zou, Zitang Zhou, Kunyuan Ding, Wenxuan Peng, Haoqi Wang, Guangyao Chen, Bo Li, Yiyou Sun, Xuefeng Du, Kaiyang Zhou, Wayne Zhang, Dan Hendrycks, Yixuan Li, and Ziwei Liu. Openood: Benchmarking generalized out-of-distribution detection. 2022. 5

[41] Ting Yu, Hao Liu, and Zeyu Zhao. Hard negative mining for out-of-distribution detection in contrastive learning. In *CVPR*, 2023. 2

[42] Jingyang Zhang, Jingkang Yang, Pengyun Wang, Haoqi Wang, Yueqian Lin, Haoran Zhang, Yiyou Sun, Xuefeng Du, Yixuan Li, Ziwei Liu, Yiran Chen, and Hai Li. Openood v1.5: Enhanced benchmark for out-of-distribution detection. *arXiv preprint arXiv:2306.09301*, 2023. 5

[43] Zihan Zhang and Xiang Xiang. Decoupling maxlogit for out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3388–3397, 2023. 2, 3, 4

[44] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 5

A. Proofs

In this section, we provide the proofs to the propositions in Section 5 in manuscript. We reuse the same notations as introduced earlier and provide a definition for any new notation introduced in place.

Relative Euclidean Geometry on Unit Hypersphere Proposition 1(Cosine-Margin Equivalence)

For any input $\mathbf{x}$ and class c, assuming that distance to class mean $\boldsymbol{\mu}_c$ is approximately constant across classes, ranking samples by the geometric score $S_g(\mathbf{x})$ in 6 is equivalent to ranking them by $\max_c \hat{\mathbf{z}}(\mathbf{x})^\top(\boldsymbol{\mu}_c - \boldsymbol{\mu})$.

Proof. Fix any sample and write $\hat{\mathbf{z}} = \hat{\mathbf{z}}(x)$. For any class c,

$$\|\hat{\mathbf{z}} - \boldsymbol{\mu}\|_2^2 = \|\hat{\mathbf{z}}\|_2^2 + \|\boldsymbol{\mu}\|_2^2 - 2\hat{\mathbf{z}}^\top\boldsymbol{\mu}, \tag{14}$$

$$\|\hat{\mathbf{z}} - \boldsymbol{\mu}_c\|_2^2 = \|\hat{\mathbf{z}}\|_2^2 + \|\boldsymbol{\mu}_c\|_2^2 - 2\hat{\mathbf{z}}^\top\boldsymbol{\mu}_c. \tag{15}$$

Subtracting yields

$$\|\hat{\mathbf{z}} - \boldsymbol{\mu}\|_2^2 - \|\hat{\mathbf{z}} - \boldsymbol{\mu}_c\|_2^2 = \big(\|\boldsymbol{\mu}\|_2^2 - \|\boldsymbol{\mu}_c\|_2^2\big) + 2\hat{\mathbf{z}}^\top(\boldsymbol{\mu}_c - \boldsymbol{\mu}). \tag{16}$$

Therefore,

$$S_g(x) = \max_c \left[2\hat{\mathbf{z}}^\top(\boldsymbol{\mu}_c - \boldsymbol{\mu}) + \|\boldsymbol{\mu}\|_2^2 - \|\boldsymbol{\mu}_c\|_2^2\right]. \tag{17}$$

If $\|\boldsymbol{\mu}_c\|_2^2$ is constant across c (or varies negligibly relative to the inner-product term), then the additive offset $\|\boldsymbol{\mu}\|_2^2 - \|\boldsymbol{\mu}_c\|_2^2$ is class-independent (or approximately so) and thus does not affect the $\arg\max$ over c. Moreover, multiplying by the positive constant 2 does not change rankings. Hence ranking by $S_g(x)$ is equivalent to ranking by $\max_c \hat{\mathbf{z}}^\top(\boldsymbol{\mu}_c - \boldsymbol{\mu})$. □

Even without the constant-norm approximation, Eq. (16) shows that S_g is a *shifted cosine-margin* type test where each class margin is offset by $-\|\boldsymbol{\mu}_c\|_2^2$. The constant norm term can be justified from the theory of neural collapse[28].

Conditioning on Background Deviation Proposition 2(Conditioning on Background Deviation [20])

Let $f_{X,Y}$ and $g_{X,Y}$ denote the joint distributions of (X, Y) for ID and OOD, with marginals f_X, g_X and conditionals $f_{Y|X}, g_{Y|X}$. If the background deviation distributions are matched, $f_X = g_X$, then

$$D_{\mathrm{KL}}(f_Y \,\|\, g_Y) \;\le\; \mathbb{E}_{X\sim f_X}\left[D_{\mathrm{KL}}\big(f_{Y|X}(\cdot|X) \,\|\, g_{Y|X}(\cdot|X)\big)\right]. \tag{18}$$

Let (X, Y) denote the pair of random variables induced by the scoring construction (in our case, $X = \|\hat{\mathbf{z}} - \boldsymbol{\mu}\|_2^2$ and $Y = \max_c\big(X - \|\hat{\mathbf{z}} - \boldsymbol{\mu}_c\|_2^2\big)$). Let $f_{X,Y}$ and $g_{X,Y}$ denote the joint distributions of (X, Y) under ID and OOD, with marginals f_X, g_X and f_Y, g_Y and conditionals $f_{Y|X}, g_{Y|X}$.

Step 1: Marginalization cannot increase KL (data processing). Consider the mapping $(X, Y) \mapsto Y$ (i.e., it discards X). By the data processing inequality for D_{KL}, applying any measurable map (in particular, marginalization) cannot increase KL divergence. Hence,

$$D_{\mathrm{KL}}(f_Y \,\|\, g_Y) \;\le\; D_{\mathrm{KL}}(f_{X,Y} \,\|\, g_{X,Y}). \tag{19}$$

Step 2: Chain rule for KL on the joint. Using the standard KL chain rule,

$$\begin{aligned}
&D_{\mathrm{KL}}(f_{X,Y} \,\|\, g_{X,Y}) \\
&= \int f_{X,Y}(x,y) \log \frac{f_{X,Y}(x,y)}{g_{X,Y}(x,y)}\, dx\, dy \\
&= \int f_{X,Y}(x,y) \log \frac{f_X(x) f_{Y|X}(y|x)}{g_X(x) g_{Y|X}(y|x)}\, dx\, dy \\
&= \int f_X(x) \log \frac{f_X(x)}{g_X(x)}\, dx + \\
&\quad \int f_X(x) \left[\int f_{Y|X}(y|x) \log \frac{f_{Y|X}(y|x)}{g_{Y|X}(y|x)}\, dy\right] dx \\
&= D_{\mathrm{KL}}(f_X \,\|\, g_X) \;+\; \mathbb{E}_{X\sim f_X}\left[D_{\mathrm{KL}}\big(f_{Y|X} \,\|\, g_{Y|X}\big)\right]
\end{aligned} \tag{20}$$

Step 3: Apply the matched-background assumption $f_X = g_X$. Under the assumption that the background deviation distributions are matched, $f_X = g_X$, we have

$$D_{\mathrm{KL}}(f_X \,\|\, g_X) = 0, \tag{21}$$

and thus (20) reduces to

$$D_{\mathrm{KL}}(f_{X,Y} \,\|\, g_{X,Y}) = \mathbb{E}_{X\sim f_X}\left[D_{\mathrm{KL}}\big(f_{Y|X} \,\|\, g_{Y|X}\big)\right]. \tag{22}$$

Conclusion. Combining (19) and (22) yields

$$D_{\mathrm{KL}}(f_Y \,\|\, g_Y) \;\le\; \mathbb{E}_{X\sim f_X}\left[D_{\mathrm{KL}}\big(f_{Y|X} \,\|\, g_{Y|X}\big)\right]. \tag{23}$$

In words: when the background variable has matched marginals across ID and OOD, conditioning on it cannot reduce, and can strictly increase, the information-theoretic separability between ID and OOD in the task-relevant variable.

In `FIDEI`, $X = \|\hat{\mathbf{z}} - \boldsymbol{\mu}\|_2^2$ captures global/background deviation of the test feature from a global prototype, while $Y = \max_c\big(X - \|\hat{\mathbf{z}} - \boldsymbol{\mu}_c\|_2^2\big)$ measures deviation-specific class separation. The inequality above formalizes that discarding X (i.e., using Y unconditioned on background deviation) can only lose separability, whereas incorporating X via conditioning (or an equivalent relative construction) improves OOD detection in the matched-background regime.

B. OOD Detection in Foundation Models

In order to understand how FIDEI fares for OOD detection with foundation models, we ran FIDEI with DinoV2(ViT-S/14 and ViT-B/14 architectures) for ImageNet-1k benchmark. We include another benchmark LogitGap[19] which is specialized for foundation models. This benchmark shows that even for highly capable foundation models, FIDEI performs competitively with the SOTA approaches. As the DINOV2 DNN are proven to have specialized subspaces [27], it is natural that subspace based approach ViM dominates in this setting. Apart from ViM, FIDEI outperforms or performs competitively with the other OOD detectors e.g., LogitGap, scale, GEN. AUROC of these approaches are within 1% of each other. This shows that FIDEI is reliable for foundation models as well. While FIDEI shows to be weak for near-OOD setting with foundation models, it shows strong performance in far-OOD setting – even outperforming the LogitGap in far-OOD setting.

C. Detailed Per-OOD Dataset Results

We provide the detailed per-OOD datset results in Table 12–Table 19.

D. Reference Implementation of FIDEI (FIDEI)

We provide a reference PyTorch implementation of the proposed FIDEI (FIDEI) post-hoc OOD detector. The code closely follows the formulation described in Section 4, including classifier-aware centering, hyperspherical normalization, relative geometric scoring, and scale-matched fusion with the energy score. This implementation is used for all experiments unless stated otherwise.

Table 10. Comparison of OOD detection methods on ImageNet-1K benchmark for DINOv2 ViT-S/14.

	SSB-Hard		NINCO		Near OOD		iNaturalist		Textures		OpenImage-O		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
GEN	81.91	70.28	52.50	83.21	67.20	76.74	9.19	97.50	34.78	91.40	23.71	93.94	22.56	94.28
KNN	88.82	63.61	76.32	77.19	82.57	70.40	6.66	98.53	48.66	88.99	35.53	92.25	30.28	93.26
LogitGap	81.37	70.99	52.22	83.64	66.80	77.32	9.85	97.35	35.98	91.11	24.01	93.83	23.28	94.10
Scale	81.92	70.90	54.08	83.03	68.00	76.96	8.30	97.80	35.59	91.18	25.35	93.69	23.08	94.22
ViM	**77.41**	**75.19**	**54.01**	**84.27**	**65.71**	**79.73**	**1.62**	**99.42**	**19.97**	**95.66**	**27.46**	**94.06**	**16.35**	**96.38**
FIDEI (Ours)	82.94	69.06	58.41	83.10	70.67	76.08	3.86	99.06	40.76	90.62	25.84	94.21	23.49	94.63

Table 11. Comparison of OOD detection methods on ImageNet-1K benchmark for DINOv2 ViT-B/14.

	SSB-Hard		NINCO		Near OOD		iNaturalist		Textures		OpenImage-O		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
GEN	72.92	78.55	39.18	89.45	56.05	84.00	3.47	99.07	35.48	91.83	16.08	96.33	18.34	95.74
KNN	82.83	71.68	55.97	85.92	69.40	78.80	3.58	99.01	37.38	91.92	22.59	95.31	21.19	95.41
LogitGap	72.56	78.23	40.31	88.96	56.43	83.60	2.55	99.15	33.98	91.95	15.81	96.36	17.05	95.85
Scale	73.06	78.33	40.61	89.06	56.83	83.70	3.45	99.05	34.88	91.90	16.41	96.31	18.25	95.75
ViM	**66.80**	**82.60**	**33.11**	**91.85**	**49.95**	**87.22**	**0.71**	**99.73**	**14.20**	**97.23**	**12.72**	**97.44**	**9.21**	**98.13**
FIDEI (Ours)	78.11	75.84	43.21	89.72	60.66	82.78	2.18	99.36	31.94	92.64	15.34	96.76	16.49	96.25

Table 12. Comparison of FIDEI with baseline approaches on CIFAR10 benchmark for ResNet-18.

	CIFAR100		TIN		Near OOD		MNIST		SVHN		Textures		Places365		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
MSP	59.90	86.73	47.32	88.64	53.61	87.69	19.23	93.95	23.82	91.68	40.16	89.13	41.67	89.35	31.22	91.03
Gradnorm	95.66	52.74	94.87	55.11	95.26	53.93	78.93	70.06	91.01	50.29	97.53	53.71	89.48	67.95	89.24	60.50
React	75.50	85.24	66.63	87.79	71.07	86.51	20.42	94.38	42.28	90.30	67.63	87.27	40.00	91.41	42.58	90.84
ViM	53.60	87.44	41.79	89.67	47.69	88.56	18.06	94.25	**18.29**	**94.49**	**21.88**	**94.76**	44.51	89.15	25.68	93.17
KNN	**37.90**	**89.75**	30.93	91.71	**34.42**	**90.73**	20.61	94.41	20.43	93.01	24.42	93.02	29.44	92.10	23.73	93.13
ASH	89.97	72.63	88.00	75.77	88.98	74.20	67.90	84.07	83.33	70.61	86.95	74.05	67.25	85.09	76.36	78.46
GEN	63.68	86.71	52.10	88.89	57.89	87.80	20.12	95.00	24.07	92.18	46.13	89.36	46.12	89.74	34.11	91.57
FDBD	41.59	89.34	31.66	91.55	36.62	90.45	19.28	94.95	22.83	92.34	24.82	92.86	**27.02**	**92.67**	23.49	93.20
SCALE	86.10	80.57	83.41	84.00	84.76	82.28	35.14	93.19	67.83	86.39	84.92	83.47	65.92	88.89	63.46	87.99
RMDS	48.80	88.38	35.74	90.76	42.27	89.57	18.88	94.20	20.90	92.24	25.87	91.93	31.60	91.48	24.31	92.46
FIDEI (Ours)	38.60	89.61	**30.87**	**91.81**	34.73	90.71	**18.61**	**95.08**	19.89	93.22	23.54	93.12	29.18	92.16	**22.81**	**93.39**

Table 13. Comparison of FIDEI with baseline approaches on CIFAR10 benchmark for ViT-Base.

	CIFAR100		TIN		Near OOD		MNIST		SVHN		Textures		Places365		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
MSP	10.82	96.87	9.51	97.33	10.17	97.10	7.86	97.32	4.54	98.41	4.40	98.56	8.92	97.30	6.43	97.90
Gradnorm	94.41	74.21	93.33	77.37	93.87	75.79	98.66	74.93	94.79	74.93	89.84	77.51	97.99	69.74	97.51	76.54
React	8.97	97.98	6.49	98.52	7.73	98.25	3.29	99.34	**0.94**	99.71	5.39	98.72	2.66	99.31	2.66	99.31
ViM	**7.66**	**98.47**	**5.63**	**98.63**	**6.65**	**98.55**	**2.62**	99.35	2.48	99.36	**1.52**	99.55	**1.59**	**99.58**	**2.07**	99.43
KNN	8.81	98.20	7.51	98.49	8.16	98.34	2.68	99.42	2.55	99.60	4.34	99.05	2.31	99.45	2.72	99.38
ASH	91.23	58.91	85.36	63.15	88.29	61.03	76.44	59.79	69.79	60.52	93.92	51.05	88.24	61.26	86.38	58.92
GEN	8.34	97.93	6.46	98.50	7.40	98.21	5.20	98.61	1.27	99.66	1.24	**99.68**	6.09	98.47	3.45	99.10
FDBD	8.54	98.08	6.69	98.51	7.62	98.29	3.27	99.18	1.98	99.45	1.52	99.55	4.78	98.90	2.89	99.27
SCALE	8.60	97.86	6.67	98.47	7.63	98.17	3.91	98.92	0.94	**99.71**	0.90	99.74	5.89	98.52	2.91	99.22
RMDS	8.57	97.67	6.43	98.26	7.50	97.96	3.82	98.91	2.66	99.10	3.14	98.87	5.17	98.62	3.70	98.87
FIDEI (Ours)	8.40	98.23	6.93	98.57	7.67	98.40	2.93	**99.60**	1.59	99.60	1.49	99.58	5.24	98.81	2.83	**99.32**

Table 14. Comparison of FIDEI with baseline approaches on CIFAR100 benchmark for ResNet-18.

	CIFAR10		TIN		Near OOD		MNIST		SVHN		Textures		Places365		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
MSP	59.10	78.54	50.01	82.22	54.56	80.38	63.47	73.55	55.44	79.37	61.24	78.07	55.40	79.61	58.89	77.65
Gradnorm	84.58	70.33	88.06	68.84	86.32	69.58	85.62	67.60	93.58	52.15	97.85	52.62	78.62	76.88	88.92	62.31
React	61.39	78.64	51.82	82.63	56.61	80.63	63.34	74.10	52.37	82.45	54.04	80.80	54.86	80.16	56.15	79.91
ViM	71.16	71.73	54.57	78.03	62.86	74.88	**46.87**	81.76	44.68	83.95	38.76	**86.60**	53.99	81.34	46.07	**83.41**
KNN	**58.56**	**79.39**	48.60	82.56	**53.58**	**80.98**	60.92	76.24	46.76	**86.11**	40.80	86.00	**51.46**	**83.84**	**45.98**	83.05
ASH	67.56	76.89	64.80	79.78	66.18	78.33	71.73	76.63	49.36	83.83	52.41	84.86	64.78	79.41	59.57	81.18
GEN	61.51	79.11	52.10	81.79	56.81	80.45	65.03	74.70	50.44	84.23	43.23	86.08	56.35	82.13	51.26	81.78
FDBD	60.69	79.24	49.43	83.28	55.06	81.26	60.39	76.81	52.07	80.82	48.17	83.44	57.26	81.34	54.47	80.60
SCALE	76.93	70.70	62.79	77.18	69.86	73.94	69.79	71.57	62.33	75.05	55.31	79.33	61.17	76.46	62.15	75.60
RMDS	70.19	75.14	50.77	82.75	60.48	78.95	57.98	76.86	91.97	64.75	85.61	69.60	82.71	68.89	79.57	70.03
FIDEI (Ours)	67.02	77.96	**48.23**	**84.23**	57.63	81.10	57.99	**83.91**	**42.88**	83.09	**53.74**	82.98	57.76	80.55	54.29	81.24

Table 15. Comparison of FIDEI with baseline approaches on CIFAR100 benchmark for ViT-Base.

	CIFAR10		TIN		Near OOD		MNIST		SVHN		Textures		Places365		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
MSP	44.21	88.21	35.96	89.43	40.08	88.82	81.06	61.89	33.84	89.40	33.94	90.40	44.26	87.06	48.28	82.19
Gradnorm	81.80	73.75	85.87	70.54	83.83	72.14	74.74	67.90	81.67	72.08	90.86	71.54	95.50	57.00	85.69	67.13
React	54.67	81.02	46.88	84.21	50.78	82.61	52.14	80.44	48.90	83.12	44.37	85.66	49.02	82.93	48.61	83.04
ViM	49.38	83.76	41.55	86.92	45.47	85.34	47.02	83.21	43.18	86.04	38.91	88.47	44.56	85.27	43.42	85.75
KNN	46.91	84.15	39.64	87.28	43.28	85.71	45.83	83.59	41.12	86.73	36.94	89.11	42.39	86.02	41.57	86.36
ASH	62.77	78.44	58.39	80.91	60.58	79.68	65.11	77.52	61.44	79.83	57.02	82.06	63.18	78.94	61.69	79.59
GEN	53.09	82.03	45.92	85.14	49.51	83.59	51.33	81.02	46.70	84.61	42.88	86.93	48.39	84.12	47.33	84.17
FDBD	42.98	91.10	28.50	93.16	35.74	92.13	72.90	69.69	23.29	94.07	24.31	94.56	36.60	91.13	39.28	87.36
SCALE	42.50	91.82	35.00	92.49	38.75	92.15	83.68	64.21	23.56	94.30	24.42	94.64	47.56	89.79	44.80	85.74
RMDS	**33.78**	92.70	**28.03**	92.34	**30.91**	**92.52**	72.40	67.42	24.33	93.81	**23.16**	93.27	**36.99**	89.92	**39.22**	86.11
FIDEI (Ours)	44.21	**93.40**	36.97	**93.50**	40.59	92.40	**43.08**	**90.15**	**22.80**	**94.80**	35.61	**95.10**	41.02	**92.00**	39.75	**93.20**

Table 16. Comparison of FIDEI with baseline approaches on ImageNet-1K benchmark for ConvNeXt-Small.

	SSB-Hard		NINCO		Near OOD		iNaturalist		Textures		OpenImage-O		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
ViM	81.67	73.21	58.59	84.11	70.13	78.66	17.29	93.95	34.65	91.07	23.35	92.90	**25.10**	92.64
GEN	**69.81**	**78.50**	50.76	86.67	60.28	82.58	17.78	95.50	81.16	84.06	38.90	90.67	45.95	90.08
ASH	92.81	66.03	97.14	59.93	94.98	62.98	90.62	71.95	95.18	69.54	97.89	56.79	94.56	66.09
KNN	76.48	72.67	54.78	83.96	65.63	78.31	21.04	94.17	**32.93**	**91.23**	26.44	92.92	26.81	92.77
FDBD	77.76	73.49	54.90	84.18	66.33	78.84	22.03	93.47	35.84	90.41	27.08	92.42	28.32	92.10
SCALE	87.05	65.19	88.79	60.74	87.92	62.96	79.08	63.59	95.29	55.41	92.91	54.90	89.09	57.97
FIDEI (Ours)	70.24	77.67	45.28	**87.82**	**57.76**	**82.75**	**14.56**	**96.87**	38.46	89.97	**22.93**	**94.17**	25.31	**93.67**

Table 17. Comparison of FIDEI with baseline approaches on ImageNet-1K benchmark for ViT-B/16.

	SSB-Hard		NINCO		Near OOD		iNaturalist		Textures		OpenImage-O		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
ViM	62.12	82.66	37.34	91.83	49.73	87.24	**0.42**	**99.81**	42.56	86.00	23.87	95.33	22.28	93.71
GEN	**57.54**	**84.56**	**35.49**	**91.63**	**46.51**	**88.10**	1.82	99.46	**34.98**	**93.04**	**15.22**	**96.80**	**17.34**	**96.43**
KNN	78.52	68.71	52.32	84.23	65.42	76.47	2.42	99.27	52.97	87.03	30.61	93.49	28.67	93.26
ASH	76.22	72.92	88.51	68.40	82.37	70.66	48.06	86.30	82.56	71.21	84.59	70.21	71.74	75.91
FDBD	70.26	79.68	37.30	91.04	53.78	85.36	3.08	99.22	38.49	90.35	18.75	96.15	20.11	95.24
SCALE	57.79	84.57	35.34	91.70	46.56	88.14	1.91	99.45	34.79	93.06	15.42	96.80	17.37	96.44
FIDEI (Ours)	70.75	77.28	41.90	89.43	56.33	83.36	1.14	99.57	39.39	90.58	18.95	96.16	19.83	95.43

Table 18. Comparison of FIDEI with baseline approaches on ImageNet-200 benchmark for ConvNeXt-Small.

	SSB-Hard		NINCO		Near OOD		iNaturalist		Textures		OpenImage-O		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
MSP	84.76	74.70	74.80	82.21	79.78	78.45	39.72	91.46	96.13	76.28	79.71	83.82	71.86	83.86
EBO	95.36	56.74	95.38	57.29	95.37	57.02	93.76	56.83	99.23	51.60	97.97	51.45	96.99	53.30
React	85.24	68.33	72.06	76.56	78.65	72.45	59.71	77.63	88.03	76.81	72.82	77.72	73.52	77.39
ViM	51.31	86.62	28.30	91.05	39.81	88.83	11.44	95.54	10.79	96.09	11.80	96.12	11.34	95.92
KNN	51.47	89.98	29.50	93.08	40.48	91.53	7.46	97.37	**7.74**	**97.28**	9.06	97.31	8.09	97.32
ASH	94.37	61.68	97.52	56.95	95.94	59.31	91.71	68.61	93.86	71.56	98.38	54.19	94.65	64.79
GEN	72.52	79.91	51.08	88.09	61.80	84.00	14.70	96.31	84.50	85.02	37.57	91.83	45.59	91.06
FDBD	71.91	78.21	47.29	87.70	59.60	82.95	16.24	95.27	27.40	92.82	20.03	94.49	21.23	94.19
SCALE	89.07	58.07	91.21	50.62	90.14	54.34	78.78	54.44	71.77	68.68	84.21	54.33	78.25	59.15
RMDS	46.91	90.21	24.83	93.90	35.87	92.05	6.68	97.76	9.32	96.48	8.51	97.27	8.17	97.17
FIDEI (Ours)	**45.53**	**91.66**	**24.08**	**94.46**	**34.81**	**93.06**	**4.97**	**98.55**	9.70	96.43	**8.23**	**97.57**	**7.63**	**97.52**

Table 19. Comparison of FIDEI with baseline approaches on ImageNet-200 benchmark for ViT-Base/16.

	SSB-Hard		NINCO		Near OOD		iNaturalist		Textures		OpenImage-O		Far OOD	
	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC	FPR	AUROC
MSP	70.14	79.60	44.61	88.79	57.38	84.20	12.52	97.70	46.87	89.44	29.69	93.61	29.69	93.59
EBO	62.82	85.21	35.87	92.31	49.34	88.76	1.28	99.54	34.26	93.62	13.12	97.20	16.22	96.79
React	64.67	84.59	37.04	92.23	50.86	88.41	1.09	99.63	34.60	93.26	13.81	97.25	16.50	96.71
ViM	**43.28**	**90.18**	**22.28**	**95.35**	**32.78**	**92.77**	**0.29**	**99.91**	32.33	91.23	16.37	96.94	16.33	96.03
KNN	54.21	87.73	27.70	93.49	40.96	90.61	1.08	99.74	16.58	96.47	10.87	97.93	9.51	98.04
ASH	64.01	84.73	35.63	92.51	49.82	88.62	1.63	99.53	32.87	93.64	12.71	97.26	15.74	96.81
GEN	63.91	84.82	35.22	92.41	49.57	88.61	1.28	99.54	32.73	93.71	12.73	97.26	15.58	96.84
FDBD	65.82	83.99	27.56	93.81	46.69	88.90	1.70	99.49	29.87	93.00	11.96	97.48	14.51	96.66
SCALE	63.93	84.72	35.56	92.51	49.74	88.61	1.63	99.53	33.00	93.63	12.73	97.25	15.79	96.80
RMDS	54.51	88.73	25.20	94.06	39.86	91.39	2.53	98.89	17.72	92.56	11.71	96.17	10.66	95.87
FIDEI (Ours)	55.14	88.85	24.24	94.75	39.69	91.80	0.82	99.69	**13.36**	**96.17**	**8.83**	**97.93**	**7.67**	**97.93**

Listing 1. PyTorch implementation of the FIDEI postprocessor (part 1).

```
from typing import Any, Dict, List, Optional
import torch
import torch.nn as nn
import torch.nn.functional as F
from tqdm import tqdm
from .base_postprocessor import BasePostprocessor
from .info import num_classes_dict
class FIDEIPostprocessor(BasePostprocessor):
    def __init__(self, config):
        super().__init__(config)
        self.config = config
        self.num_classes = num_classes_dict[self.config.dataset.name]
        self.max_per_class = 5000
        self.eps = 1e-6
        self.device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
        self.setup_flag = False
        self.u: Optional[torch.Tensor] = None
        self.class_mean: Optional[torch.Tensor] = None
        self.whole_mean: Optional[torch.Tensor] = None
        self.alpha: float = 1.0
```

Listing 2. PyTorch implementation of the FIDEI postprocessor (part 2).

```
    def setup(self, net: nn.Module, id_loader_dict: Dict[str, Any], ood_loader_dict=None):
        if self.setup_flag:
            return
        net.eval()
        train_loader = id_loader_dict["train"]
        w_np, b_np = net.get_fc()
        w = torch.from_numpy(w_np).to(self.device)
        b = torch.from_numpy(b_np).to(self.device)
        self.u = -(torch.linalg.pinv(w) @ b)
        feats_by_class = [[] for _ in range(self.num_classes)]
        sum_maxlogit, count = 0.0, 0
        tmp_feats, tmp_labels = [], []
        with torch.no_grad():
            for batch in tqdm(train_loader):
                data = batch["data"].to(self.device).float()
                labels = batch["label"]
                logits, feats = net(data, return_feature=True)
                feats, logits = feats.cpu(), logits.cpu()
                for i, y in enumerate(labels):
                    c = int(y)
                    if len(feats_by_class[c]) >= self.max_per_class:
                        continue
                    feats_by_class[c].append(feats[i])
                    tmp_feats.append(feats[i])
                    tmp_labels.append(c)
                    sum_maxlogit += logits[i].max().item()
                    count += 1
                if all(len(v) >= self.max_per_class for v in feats_by_class):
                    break
        all_feats = torch.stack(tmp_feats)
        all_labels = torch.tensor(tmp_labels)
        z = all_feats.to(self.device) - self.u
        f = F.normalize(z, p=2, dim=-1).cpu()
        self.class_mean = torch.stack(
            [f[all_labels == c].mean(0) for c in range(self.num_classes)]
        )
        self.whole_mean = f.mean(0)
        bg = ((f - self.whole_mean)**2).sum(1)
        cd = ((f[:, None, :] - self.class_mean[None, :, :])**2).sum(2)
        conf_g = (bg[:, None] - cd).max(1).values
        self.alpha = (sum_maxlogit / count) / (conf_g.mean().item() + self.eps)
        self.setup_flag = True
```

Listing 3. PyTorch implementation of the FIDEI postprocessor (part 3).

```
    @torch.no_grad()
    def postprocess(self, net: nn.Module, data: Any):
        logits, features = net(data, return_feature=True)
        pred = logits.argmax(1)
        z = features - self.u.to(features.device)
        f = F.normalize(z, p=2, dim=-1)
        bg = ((f - self.whole_mean.to(f.device))**2).sum(1)
        cd = ((f[:, None, :] - self.class_mean.to(f.device)[None, :, :])**2).sum(2)
        conf_g = (bg[:, None] - cd).max(1).values
        energy = torch.logsumexp(logits, dim=1)
        conf = energy + self.alpha * conf_g
        return pred, conf
```