OUT-OF-DISTRIBUTION DETECTION IN CLASS INCRE-MENTAL LEARNING

Anonymous authors

004

010 011

012

013

014

015

016

017

018

019

021

025

026

027

028

029

031 032

033 034

043

Paper under double-blind review

Abstract

Class incremental learning (CIL) aims to learn a model that can not only incrementally accommodate new classes, but also maintain the learned knowledge of old classes. Out-of-distribution (OOD) detection in CIL is to retain this incremental learning ability, while being able to reject unknown samples that are drawn from different distributions of the learned classes. This capability is crucial to the safety of deploying CIL models in open worlds. However, despite remarkable advancements in the respective CIL and OOD detection, there lacks a systematic and large-scale benchmark to assess the capability of advanced CIL models in detecting OOD samples. To fill this gap, in this study we design a comprehensive empirical study to establish such a benchmark, named **OpenCIL**, offering a unified protocol for enabling CIL models with different OOD detectors using two principled OOD detection frameworks. One key observation we find through our comprehensive evaluation is that the CIL models can be severely biased towards the OOD samples and newly added classes when they are exposed to open environments. Motivated by this, we further propose a novel approach for OOD detection in CIL, namely Bi-directional Energy Regularization (BER), which is specially designed to mitigate these two biases in different CIL models by having energy regularization on both old and new classes. Extensive experiments show that BER can substantially improve the OOD detection capability across a range of CIL models, achieving state-of-the-art performance on the OpenCIL benchmark.

1 INTRODUCTION

Training of deep neural networks (DNNs) heavily relies on large-scale data on a fixed set of classes (Russakovsky et al., 2015; Krizhevsky et al., 2017), but data in real-world applications is constantly changing, leading to continuous new classes in training data. Continual learning (CL) enables DNNs to continuously learn a sequence of tasks, with each task consisting of a set of unique classes. The samples for the learned/old tasks are assumed to be not accessible in such dynamic environments. Class incremental learning (CIL) is one type of CL where task identifiers are not known at testing time. Compared to another type of CL, task-incremental learning, CIL is often considered a more practical setting in real applications (Rebuffi et al., 2017; Li & Hoiem, 2017). Thus, we focus on CIL in this study.

Many CIL methods have been introduced over the years to overcome *Catastrophic Forgetting* (CF) 044 of the knowledge learned on old classes, *i.e.*, degraded classification accuracy on old classes due to model updating on new classes (Xiao et al., 2023; Wang et al., 2022a; Niu et al., 2024). They 046 have shown remarkable performance on the in-distribution (ID) classes in incremental tasks, but lack 047 the capability to recognize and reject out-of-distribution (OOD) samples that are drawn from non-048 training datasets during incremental learning (Huang & Li, 2021; Wang et al., 2020) (i.e., no class overlapping between ID and OOD samples). Such a capability is crucial to the safety of deploying CIL models in open environments in real-world application systems such as autonomous systems 051 (Kendall & Gal, 2017; Leibig et al., 2017). For example, in Unmanned Aerial Vehicles (UAVs), the DNNs are initially trained using a limited set of available trajectories and then continuously updated 052 based on newly available trajectories. Meanwhile, these DNNs need to be capable of recognizing OOD data to handle unexpected situations in every ongoing trajectory.



Figure 1: Results of the CIL model iCaRL (Rebuffi et al., 2017) on CIFAR100 (Krizhevsky et al., 2009). (a) Mean prediction confidence of iCaRL on test samples from all incremental classes. (b) Mean prediction confidence of iCaRL classifying six OOD datasets into one of the ID classes based on the final incremental task. (see Appendix D for the results of other CIL models)



066

067

068

072

073

074 Numerous OOD detection methods have been proposed (Sun et al., 2021; Wang et al., 2023b; 075 Li et al., 2024b; Miao et al., 2024), but these OOD methods are focused on static environments, 076 where the samples of all tasks are accessible during training, making them ineffective in dynamic 077 CIL environments. Therefore, it is non-trivial to combine the off-the-shelf CIL and OOD models. To justify this difficulty, a plausible evaluation protocol is needed to assess the OOD detection capability of different CIL models with the support of different types of OOD detectors. There are 079 such protocols on the respective CIL and OOD detection areas, e.g., OpenOOD (Yang et al., 2022; Zhang et al., 2023) for OOD detection and FACIL (Masana et al., 2022) for CIL, but no work has 081 been done on a systematic and large-scale benchmarking study to evaluate the synergy of existing 082 state-of-the-art (SOTA) CIL models and OOD detection methods. 083

To bridge this gap, we design a performance benchmark for OOD detection in CIL, called **Open**-084 CIL, offering a unified protocol for different CIL models with diverse OOD detectors. To achieve 085 this, OpenCIL introduces two principled frameworks for incorporating diverse OOD detection methods into CIL models and also introduces a new evaluation pipeline that enables fair comparison of 087 not only the OOD detection capability for different CIL models but also the ability of different OOD 880 detectors in the presence of CF. In particular, OpenCIL accommodates four representative CIL mod-089 els with 15 diverse OOD detection methods, resulting in 60 baseline models on two popular CIL 090 datasets and six commonly-used near/far OOD datasets. Based on the large-scale experiments on 091 OpenCIL, we provide a number of important observations, offering crucial insights into the design 092 of CIL models for open-world applications.

One key observation we find is that compared to OOD detection in static environments, the dynamic 094 environments in the CIL setting can lead to increasing biases towards OOD samples and newly 095 added classes with the growth of incremental learning steps. The underlying reasons are two-fold. 096 One main reason is that due to the CF problem, CIL models often have lower prediction confidence for samples of old ID classes (*i.e.*, classes seen in the old tasks), compared to new ID classes, as 098 illustrated in Fig. 1a. This leads to a difficulty in distinguishing between old ID class samples and 099 OOD samples. Furthermore, as illustrated in Fig. 1b, DNN-based CIL models typically exhibit an over-confident prediction on OOD samples, causing the misclassification of the OOD samples into 100 not only the old ID classes but also the new ID classes. Increasing the incremental learning steps 101 results in a larger classification semantic space and more severe CF, which continually amplifies the 102 biases towards the OOD samples and newly added classes. Motivated by these issues, we propose 103 a new approach for OOD detection in CIL, namely Bi-directional Energy Regularization (BER), 104 which jointly optimizes two energy regularization terms to modulate the energy prediction of OOD 105 samples w.r.t. samples of old and new class samples, respectively. This effectively reduces these 106 two biases, improving the OOD detection capability across a range of CIL models.

107

In summary, our main contributions are as follows:

- We investigate the synergy between CIL and OOD detection models, and establish the first benchmark, OpenCIL, for evaluating the OOD detection capability of CIL models and promoting the development of more advanced methods for this under-explored problem.
 - We further introduce BER, a novel approach that provides an effective framework for mitigating increasing biases of CIL models towards OOD samples and newly added classes with the growth of incremental steps. This helps largely improve the OOD detection capability of a wide range of CIL models.
 - Extensive experiments show that BER achieves state-of-the-art performance on the Open-CIL benchmark under varying incremental step sizes on popular CIL and OOD datasets.
- 118 119 2 RELATED WORK

Out-of-distribution (OOD) Detection. The objective of this task is to determine whether a 121 given input sample belongs to the learned classes (in-distribution) or unknown classes (out-of-122 distribution). In recent years, OOD detection has been extensively developed, including Post-hoc-123 based methods (Sun et al., 2021; Wang et al., 2023b; Zhang & Xiang, 2023) and fine-tuning-based 124 methods (Liu et al., 2020; Wei et al., 2022; Tian et al., 2022; Yu et al., 2023; Li et al., 2023; Liu 125 et al., 2023b; Miao et al., 2024; Li et al., 2024b;a). The post-hoc methods focus on devising new 126 OOD scoring functions in the inference stage. The fine-tuning-based methods focus on separating 127 OOD samples from ID samples by training a strong classifier as OOD detector. However, all these 128 methods are applied to non-CIL models. There lacks of exploration of their capability on CIL mod-129 els, resulting in poor performance when there is catastrophic forgetting. Our BER is a fine-tuning method that can be applied to different CIL models to improve their OOD detection performance. 130

131

108

110

111

112

113

114

115

116

117

120

Class Incremental Learning (CIL). CIL performs the learning procedure in an incremental man-132 ner with growing data samples. It focuses on alleviating the catastrophic forgetting problem, in 133 which the CIL models are required to remember the knowledge of the learned classes from old tasks 134 while learning the discriminative information for the newly coming classes. There are three main 135 lines of work in this area (Luo et al., 2023; Xiao et al., 2023). Regularization-based methods focus 136 on applying discrepancy (between old and new models) as penalization terms in their objective func-137 tions (Liu et al., 2021; Rebuffi et al., 2017; Xiao et al., 2023). Parameter-isolation-based methods 138 aim to increase the model parameters in each new incremental step to prevent knowledge forgetting 139 caused by parameter overwritten when learning new tasks (Xu & Zhu, 2018; Yan et al., 2021; Wang 140 et al., 2022a). Replay-based algorithms assume there is a memory budget allowing a handful of 141 old class examples in the memory. These memory examples can be used to re-train/fine-tune the 142 CIL model in each new incremental step (Rebuffi et al., 2017; Wu et al., 2019; Luo et al., 2023; Niu et al., 2024). However, all these methods focus on tackling the CIL problem in a closed world, 143 failing to take into account distinguishing ID data from unknown samples (e.g., OOD data), Our 144 BER baselines can be applied to different pre-trained CIL models, in which all new class data and 145 old class memory are used for fine-tuning a new OOD detector for the CIL models. It effectively 146 equips the CIL models with significantly improved capability for OOD detection. 147

148 149

150

153

3 EQUIPPING CIL MODELS WITH OOD DETECTION CAPABILITY

This section investigates the effectiveness of two principled OOD detection frameworks in enablingCIL models to reject unknown samples.

Problem Statement. Our goal is to equip CIL models with the capability of rejecting OOD samples. In this setting, a model is required to learn a sequence of tasks during training. At testing time, the model is used to classify samples into old/new classes while also rejecting unknown samples at each incremental step. The model is evaluated based on its effectiveness in preventing the CF problem and distinguishing samples of old/new classes from OOD data.

Formally, CIL models are learned from a sequence of c tasks ID data $T = \{T_1, T_2, ..., T_c\}$. For each t-th $(1 \le t \le c)$ task, we have $T_t = (X_t^{train}, X_t^{test}, Y_t)$, where X_t^{train} denotes the training ID data, X_t^{test} denotes the testing ID data, and Y_t denotes the classification semantic space consisting of a set of unique classes, *i.e.*, the label spaces between any two incremental tasks have no class

177

181



174 Figure 2: Left: Two principled frameworks are used in OpenCIL to incorporate OOD detection 175 methods into the different CIL models. Both frameworks are performed on pre-trained CIL models $\theta(\cdot)$, *i.e.*, $\theta(\cdot)$ is a composition of $\phi(\cdot)$ and $h(\cdot)$, which keep frozen throughout, ensuring that their 176 CIL classification performance is not affected. Post-hoc-based OOD methods are directly applied to pre-trained CIL models. Fine-tuning-based OOD methods train an additional classifier $f(\cdot)$, and 178 apply OOD detection based on this new classifier. Right: Our proposed BER aims to leverage 179 two sample synthesis methods to better train $f(\cdot)$, including interpolating samples of new classes to synthesize the pseudo OOD samples for enlarging their decision boundary margin to mitigate bias towards new classes, and interpolating samples of new and old classes to synthesize enhanced old 182 samples for expanding old class decision boundary to mitigate bias towards OOD samples. 183

overlapping: when $i \neq j, Y_i \cap Y_j = \emptyset$. Thus, the set of all seen classes at task t can be denoted 185 as $Q_t = \bigcup_{i=1}^t Y_i$. CIL assumes that data samples of the old classes are not accessible. Many 186 CIL methods, such as the replay-based methods, assume the availability of a memory buffer, in 187 which a memory block is assigned to each task to store a very small set of samples for the task, 188 *i.e.*, $B_i \subseteq (X_i^{train}, Y_i)$ is a small subset of training ID data sampled from task T_i . Thus, we 189 have the memory $M_t = \bigcup_{i=1}^{t-1} B_i$ that includes replay data from all t tasks. Let $\theta_t(\cdot)$ be the CIL 190 model at the incremental learning step t, then the memory and the training data of task t form the training ID data: $T_t^{train} = (X_t^{train}, Y_t) \cup M_t$ for training $\theta_t(\cdot)$ at the learning step t. For 191 replay-free CIL methods, $\theta_t(\cdot)$ is trained with (X_t^{train}, Y_t) only. During testing time at task t, let 192 $T_t^{test} = X_1^{test} \cup X_2^{test} \cup ... \cup X_t^{test}$ be the testing ID data from all t tasks, $\theta_t(\cdot)$ is used to classify 193 samples in T_t^{test} into one of the classes in Q_t . This is the setting for the conventional CIL problem. 194 195 For OOD detection in CIL, in addition to T_t^{test} , the CIL model $\theta_t(\cdot)$ is also presented with an 196 OOD dataset X_{t}^{ood} at task t, which is a set of samples of unknown classes drawn from a different

197 distribution as the ID data in T_t . Then given test data $x \in T_t^{test} \cup X_t^{ood}$, the goal of the CIL model $\theta_t(\cdot)$ is to either classify x into the correct ID class from old/new tasks, or detect it as OOD data. 199

200 201

3.1 OPENCIL BENCHMARK: ENABLING CIL MODELS WITH EXISTING OOD DETECTORS

It is challenging for CIL models to recognize OOD samples and to conduct a fair comparison of 202 OOD detection capabilities, given the diverse types of OOD detectors, the significant differences of 203 two settings, and the various incremental steps involved. Therefore, we introduce **OpenCIL**, the first 204 large-scale and systematic benchmark designed to enable CIL models with existing OOD detectors. 205 There are two types of OOD detection methods: post-hoc-based and fine-tuning-based methods. 206 Post-hoc methods calculate OOD scores based on features/logits derived from the pre-trained model, 207 which can be used for different pre-trained classification models. Fine-tuning methods require the 208 fine-tuning of part or all layers of the pre-trained models, and then calculating the OOD score based 209 on the fine-tuned models. Below we introduce two principled frameworks that OpenCIL uses to 210 incorporate these two types of OOD methods into CIL models, as also illustrated in Fig. 2 Left.

211

212 CIL models with post-hoc-based OOD detection methods. A SOTA CIL algorithm is first ap-213 plied to learn the *c*-tasks stream of ID data in an incremental manner, resulting in the standard CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot), \dots, \theta_c(\cdot)\}$. Subsequently, we directly perform the post-hoc OOD scoring 214 function on the features/logits extracted from the well-trained CIL model $\theta_t(\cdot), t \in \{1, 2, ..., c\}$ at 215 each incremental step to calculate the OOD score, without having any effect on the CIL models.

216 CIL models with fine-tuning-based OOD detection methods. Similar to the previous frame-217 work, fine-tuning-based OOD methods also apply a SOTA CIL algorithm to obtain the standard CIL 218 model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot), ..., \theta_c(\cdot)\}$ in the first place. As shown in Fig. 2, for each incremental 219 step, the well-trained CIL model $\theta_t(\cdot), t \in \{1, 2, ..., c\}$ contains a feature extractor $\phi_t(\cdot)$ and a clas-220 sifier $h_t(\cdot)$, *i.e.*, $\theta_t(\cdot)$ is a composition of $\phi_t(\cdot)$ and $h_t(\cdot)$. Then we freeze both $\phi_t(\cdot)$ and $h_t(\cdot)$, and fine-tune an additional classifier $f_t(\cdot)$ only on top of $\phi_t(\cdot)$ to avoid intensifying the catastrophic for-221 getting problem. This means that the fine-tuned CIL model contains the same feature extractor $\phi_t(\cdot)$ 222 and classifier $h_t(\cdot)$ as the standard pre-trained CIL model for incremental ID classification, but it also has the additional fine-tuned classifier $f_t(\cdot)$ for OOD detection at each incremental step t, with-224 out affecting any of the learning procedures of the CIL model $\theta_t(\cdot)$ (*i.e.*, $\phi_t(\cdot)$ and $h_t(\cdot)$). Notably, 225 the fine-tuning of $f_t(\cdot)$ is applied to each incremental step only with training ID data T_t^{train} . 226

With these two principled frameworks, different OOD detection methods can be easily incorporated
into three different types of CIL models without impairing the incremental learning accuracy at all.
The overall algorithm of these two principled frameworks is provided in Appendix F.

231 3.1.1 BENCHMARK SETUP

232 CIL Datasets. Following (Wang et al., 2022a; Wu et al., 2019; Wang et al., 2023a; Luo et al., 233 2023), we use two popular CIL datasets as the ID data in our benchmark: CIFAR100 (Krizhevsky 234 et al., 2009) and large-scale ImageNet1k (Russakovsky et al., 2015). Besides, these two datasets are 235 also widely used as ID datasets in the area of OOD detection. Following (Rebuffi et al., 2017; Wang 236 et al., 2022a; Wu et al., 2019), the splits of the two ID datasets are as follows. 1) For CIFAR100, we train the CIL model gradually with k classes per incremental step with a fixed memory size of 237 2,000 exemplars. We respectively evaluate the performance with the step size $k \in \{5, 10, 20\}$. 2) 238 For **ImageNet1K**, since it is a much larger dataset, we train the CIL model gradually with a larger 239 number of k classes per step that $k \in \{50, 100, 200\}$ with a fixed memory size of 20,000 exemplars. 240

241 **OOD Datasets.** Following the recent large-scale solely OOD detection benchmark OpenOOD 242 (Yang et al., 2022; Zhang et al., 2023), we select six datasets as OOD data for each ID dataset 243 respectively. 1) For the ID dataset CIFAR100, the OOD data includes two near OOD datasets 244 - CIFAR10 (Krizhevsky et al., 2009) and Tiny-ImageNet (TIN) (Le & Yang, 2015) - and four 245 far OOD datasets: MNIST (LeCun et al., 2010), Texture (Cimpoi et al., 2014), SVHN (Netzer 246 et al., 2011) and Places365 (Zhou et al., 2017). 2) For ImageNet1k, the OOD data includes four 247 near OOD datasets – ImageNet_O (Hendrycks et al., 2021), iNaturalist (Van Horn et al., 2018), 248 OpenImage_O (Wang et al., 2022b) and Species (Basart et al., 2022) – and two far OOD datasets: MNIST (LeCun et al., 2010) and Texture (Cimpoi et al., 2014). 249

It is notable that, although these datasets have been widely used in the CIL and OOD detection community respectively, there is still no readily accessible and unified protocol for combining them into one experimental setting. Our OpenCIL benchmark offers one way to unify them into a systematic evaluation setting, facilitating the application of diverse OOD detectors under different CIL models.

Evaluation Metrics. To fairly compare the OOD detection performance among different incremental steps, we keep the ratio of testing OOD data X_t^{ood} to testing ID data T_t^{test} fixed at each incremental step t. Specifically, we control the test OOD data fed into the CIL model in an incremental manner during the inference stage, increasing the number of testing OOD data X_t^{ood} with increasing number of tasks proportionally. Formally, the number of OOD samples for each OOD dataset at step t is defined as:

261 262

230

 $|X_t^{ood}| = |X^{ood}| \times \frac{t}{|T|},\tag{1}$

where $|X^{ood}|$ and $|X^{ood}_t|$ denote the number of OOD samples in the full OOD dataset and in the testing OOD data used at step t respectively, and T is the total number of CIL tasks/steps. In this way, the number of testing ID and OOD data grows at the same speed, keeping the ratio of them identical across all incremental steps, supporting the fair comparison of OOD detection performance among different incremental steps. Following *average incremental accuracy* (Rebuffi et al., 2017), a popular metric in the standard CIL evaluation that assesses the average classification accuracy across all incremental steps, we report the average OOD performance across all incremental steps.

More details about the datasets used and the performance metrics are provided in Appendix A.



Figure 3: (a) All four representative OOD detection methods experience a decreased AUC perfor-280 mance with increasing incremental steps, compared to themselves working on the full training data 281 of all steps. (b) ACC performance decreases quickly throughout all incremental steps, while the 282 AUC performance of all four OOD methods decreases slowly and then levels off. Both (a) and (b) 283 are average performance on six OOD datasets at each incremental step, where the CIL model iCaRL 284 (Rebuffi et al., 2017) is used. ACC is the accuracy of iCaRL on CIFAR100 at each step. The results 285 for the other three CIL models are provided in Appendix D. (c) Average performance of CIL models 286 with the OOD detector REGMIX (Pinto et al., 2023) on six OOD datasets at each incremental step 287 on CIFAR100. The results for the other OOD methods are provided in Appendix D. 288

209 290 291

3.1.2 BASELINE LIBRARY

292 15 OOD detection methods are used in OpenCIL, including nine post-hoc methods - MSP 293 (Hendrycks & Gimpel, 2017), ODIN (Liang et al., 2018), Energy (Liu et al., 2020), MaxLogit (Basart et al., 2022), GEN (Liu et al., 2023a), ReAct (Sun et al., 2021), KLM (Basart et al., 2022), 295 Relation (Kim et al., 2023), and NNGuide (Park et al., 2023) and six fine-tuning-based methods – 296 LogitNorm (Wei et al., 2022), T2FNorm (Regmi et al., 2023), AUGMIX (Hendrycks et al., 2020), 297 REGMIX (Pinto et al., 2023), VOS (Du et al., 2022) and our proposed BER. OpenCIL is based 298 on four CIL models, including two regularization-based methods - iCaRL (Rebuffi et al., 2017) 299 and WA (Zhao et al., 2020), one replay-based method BiC (Wu et al., 2019), and one parameterisolation-based method FOSTER (Wang et al., 2022a). More details are presented in Appendix 300 В. 301

To ensure a fair comparison across methods originating from different areas, we use unified settings
with common hyperparameters and architecture choices. Following the most commonly used architecture for the respective ID dataset in the CIL community (Wang et al., 2022a; Zhao et al., 2020),
the backbone ResNet32 is used when CIFAR100 is used as the ID dataset, and ResNet18 is used
whenever ImageNet1K is the ID dataset. All results are averaged over three independent runs using
different random seeds. More details are presented in Appendix C.

308

310 3.2 MAIN FINDINGS

We summarize our main findings and justification based on our OpenCIL benchmarking as follows.

OOD detectors suffer from catastrophic forgetting as well when applied to CIL models. This is
 illustrated in Fig. 3a, where the performance of different OOD detection methods decreases significantly with increasing incremental steps. This is because, with more incremental steps, more old
 class samples are wrongly detected as OOD samples, while at the same time, more OOD samples
 are misclassified into new classes, leading to fast downgraded OOD detection performance.

However, catastrophic forgetting is more persistent in CIL models than OOD detection methods. As
 shown in Fig. 3b, the AUC performance for OOD detection drops fast in early incremental steps, and
 then it slows down and levers off, but the ACC performance for CIL has a continuous, fast decrease
 throughout all incremental steps. This is because remembering the forgotten ID samples is more
 difficult than distinguishing them from OOD data. Also, after the CIL models reach a certain level
 of forgetting for some ID classes, the performance of distinguishing them from OOD data tends to be
 stable, resulting in relatively stable OOD detection performance at the end of incremental learning.

Consequently, CIL models are prone to misclassifying OOD samples into new classes. This is also
 shown in Fig. 1b, where the CIL models often yield significantly higher prediction confidence on
 misclassifying the OOD samples into the new classes than the old classes, due to the presence of
 large samples from these new classes at each incremental step. The reason is that due to more severe
 catastrophic forgetting in CIL, the CIL models predict the ID samples as the new classes with higher
 confidence than that for the old classes and the OOD samples.

In addition to catastrophic forgetting, CIL models need to handle new issues in the presence of OOD
 samples, since they exhibit stronger prediction confidence on OOD samples than Old class samples.
 As shown in Figs. 1a and 1b, the ID samples from old classes/tasks often have lower prediction
 confidence than different OOD samples, thereby being misclassified as the OOD samples. This
 phenomenon becomes more severe with an increasing number of incremental steps. This is because
 the CIL models tend to be less confident when predicting the samples of old classes due to the CF
 problem, making their OOD scores lower than samples from the OOD data.

337 On the other hand, CIL models with fine-tuning-based OOD methods show to be more advantageous 338 than those with post-hoc-based methods. This can be observed by looking at the average AUC 339 and FPR results of each CIL model for the nine post-hoc-based and the five previous fine-tuning 340 methods in Tables 1, 2, and 3. The observation holds for both ID datasets. This demonstrates the 341 advantage of tuning an additional classifier for OOD detection, but note that it is at the expense of some computational overhead. Besides, CIL models with higher CIL accuracy often gain better 342 OOD detection performance. This can be observed in Tables 1, 2, and 3, where, when averaged 343 over all the existing OOD methods used, the CIL models with higher CIL accuracy, e.g., WA and 344 FOSTER, often achieve better AUC and FPR performance than the other two CIL models, especially 345 on CIFAR100. This observation is consistent with the performance at each incremental step, as 346 shown in Fig. 3c. This is because better CIL algorithms can keep more essential information about 347 ID data to improve ID classification, which can also prevent OOD data from being misclassified into 348 ID classes.

349 350

351 352

362

4 OUR PROPOSED APPROACH BER

353 As summarized above, one key issue for the CIL models in the presence of OOD samples is the increasing biases of the CIL models towards OOD samples and newly added classes with the growth 354 of incremental steps due to more severe catastrophic forgetting. Further, fine-tuning-based OOD 355 methods are generally more effective than the post-hoc methods. Therefore, we introduce the novel 356 approach, Bi-directional Energy Regularization (BER), a fine-tuning-based OOD detection ap-357 proach, to tackle this bias issue, *i.e.*, avoiding the classification of the old and new class samples as 358 OOD samples. BER consists of two components, namely New Task Energy Regularization (NTER) 359 and Old Task Energy Regularization (OTER). NTER is designed to distinguish OOD data from sam-360 ples of new task classes, while OTER is designed to distinguish OOD data from samples of old task 361 classes. Below we introduce each component in detail.

363
 364
 4.1 New Task Energy Regularization (NTER)

365 Due to the overwhelming presence of new class samples, CIL models typically demonstrate a 366 strongly biased prediction on the OOD samples towards the new classes, *i.e.*, they have high predic-367 tion confidence on classifying OOD samples into the classes in the new task, as illustrated in Fig. 1b. To address this issue, NTER synthesizes the pseudo OOD samples that are distributed on the de-368 cision boundary of different new classes, and further utilizes them to enlarge the decision boundary 369 margin, as illustrated in Figure 2 Right. Specifically, NTER first randomly mixes up samples of dif-370 ferent new classes as the pseudo-OOD samples. Formally, let $\mathbf{x}_t = \{x_t^1, x_t^2, ..., x_t^b, \}, \mathbf{x}_t \in X_t^{train}$ 371 be one training batch of ID data from new classes at task t, where task t is the new task and b is 372 the batch size, $\mathbf{y}_t = \{y_t^1, y_t^2, ..., y_t^b,\}$ and $\mathbf{y}_t \in Y_t$ be the corresponding class label of \mathbf{x}_t , then a 373 pseudo-OOD sample \bar{x}_t is synthesized as follows:

- 374 375
- 376

$$\bar{x}_t = \beta x_t^i + (1 - \beta) x_t^j, y_t^i \neq y_t^j, \tag{2}$$

where $\beta \in [0, 1]$ is sampled from Beta distribution. Motivated by the success of energy-based methods (Liu et al., 2020), we further utilize these pseudo-OOD samples to regularize the classification of new class samples via the following energy loss function:

$$\mathcal{L}_{n} = \mathbb{E}_{x_{t} \sim X_{t}^{train}} [(max(0, p_{in} - E(x_{t})))^{2}] + \mathbb{E}_{(x_{t}^{i}, x_{t}^{j}) \sim X_{t}^{train}} [(max(0, E(\bar{x}_{t}) - p_{out}))^{2}], \quad (3)$$

where $E(x; f) = -\tau \cdot log(\sum_{j=1}^{Q_t} e^{f_j(x)/\tau})$, Q_t is the whole label space that the class set of all seen classes at task t, and τ is a temperature scaling hyperparameter. Note that we do not combine all possible pairs of (x_t^i, x_t^j) in the full training dataset X_t^{train} to form \bar{x}_t , which are produced within the mini-batches for efficiency consideration (Zhou et al., 2021). As a result, the time complexity is of the same magnitude as vanilla training, having minimal computational overhead.

386 387

388

401

405

380

4.2 OLD TASK ENERGY REGULARIZATION (OTER)

389 Due to the severe catastrophic forgetting problem, as illustrated in Fig. 1a, the CIL models exhibit 390 significantly lower prediction confidence on samples of old classes than new classes. As a result, the CIL models often misclassify the old class samples as the OOD samples. To address this issue, 391 OTER performs a different mixup operation from NTER to generate more samples of old classes. 392 In particular, it randomly mixes up new class samples with old class samples to synthesize old 393 class samples. This mixup operation not only increases the diversity of the small-sized old class 394 samples but also transfers information from new class samples to old class samples. Then using 395 these mixup samples in the fine-tuning stage largely enhances the prediction confidence of the old class samples and expands their decision boundary, as illustrated in Figure 2 Right. Formally, let 397 $\mathbf{x}_t = \{x_t^1, x_t^2, ..., x_t^b, \}, \mathbf{x}_t \in X_t^{train}$ be one training batch of ID data from new classes at task t, 398 where task t is the new task and b is the batch size, $M_t = \{m_t^1, m_t^2, ..., m_t^s\}$ be the samples of 399 old classes in the memory bank for task t, and s is the memory size, then the synthesize old class 400 samples \bar{m}_t are generated via:

$$\bar{m_t} = \lambda x_t + (1 - \lambda)m_t,\tag{4}$$

where λ is a mixup hyperparameter, similar to β in Eq. 2. We further leverage these augmented ID samples to boost the prediction confidence of old class samples via the following energy loss function:

$$\mathcal{L}_{o} = \mathbb{E}_{(x_{t}, m_{t}) \sim (X_{t}^{train}, M_{t})} [(max(0, E(\bar{m}_{t}) - p_{in}))^{2}],$$
(5)

406 where $E(x; f) = -\tau \cdot \log(\sum_{j=1}^{Q_t} e^{f_j(x)/\tau}), Q_t$ is the whole label space that the class set of all seen 407 classes at task t, and τ is the temperature scaling. Note that we only apply energy regularization 408 to the ID data in OTER, without the energy regularization term on the pseudo-OOD samples in 409 Eq. 3. The old class samples cannot be used for synthesizing the pseudo-OOD samples via, e.g., 410 mixup between old and new class samples, or mixup between samples of different old classes. This 411 is because, as the samples stored in the data replay memory, these old task samples are the most 412 representative samples of the old classes, which are typically located near the class center in their respective belonging classes in the feature space. Therefore, generating pseudo-OOD samples using 413 these old class samples will severely compress the decision boundary of old task classes, resulting 414 in a significant adverse impact on the OOD detection performance. 415

Lastly, we utilize the cross-entropy loss, together with the two energy regularization losses, to finetune the extra classifier $f_t(\cdot)$ for the CIL models following the fine-tuning OOD detection framework. Thus, the overall optimization objective of our BER approach at each task t is as follows:

419 420

$$\mathcal{L} = \mathbb{E}_{(x,y)\sim (T_t^{train}, Q_t)} [\ell(f(x), y] + \alpha(\mathcal{L}_n + \mathcal{L}_o), \tag{6}$$

where ℓ is a cross-entropy loss, α is the hyperparameter, \mathcal{L}_n is as defined in Eq. 3, and \mathcal{L}_o is as defined in Eq. 5. During inference, BER utilizes the Energy Score defined in (Liu et al., 2020) as the OOD score. The overall algorithm of our proposed BER is provided in Appendix F.

424 425 4.3 EMPIRICAL EVALUATION

We perform large-scale experiments that evaluate the composition of the proposed BER-based OOD detection method and four CIL models using three different incremental step sizes based on two ID datasets CIFAR100 and ImageNet1K. To ensure a fair comparison, BER is incorporated into the four CIL models in exactly the same way as the other fine-tuning-based methods. Due to the space limitation, we only report the average OOD detection metric values (AUC, FPR) on all six OOD datasets for each ID dataset, more fine-grained results (on near-OOD datasets and far-OOD datasets, respectively) and more OOD detection metric AP values are presented in Appendix E.

Table 1: Main results on OpenCIL benchmark for the step size of k = 5 for CIFAR 100 and the step size of k = 50 for ImageNet1K. The results are average over six OOD datasets and all incremental steps. Either the post-hoc-based or fine-tuning-based OOD methods do not affect the original CIL performance, so their average CIL accuracy on the ID data remains the same. The best and <u>second-best</u> performance per dataset in the fine-tuning-based methods are highlighted.

| | | | | ID | Dataset: | CIFAR1 | 00 | | | | | | | ID | Dataset: | ImageNe | t1K | | | |
|----------------|---------|-----------|--------------|--------------|--------------|--------|--------------|--------------|--------------|-------|--------------|-------|--------------|--------------|--------------|---------|--------------|--------------|--------------|-------|
| | iCa | RL | Bi | С | W | A | FOS | TER | Ave | rage | iCa | RL | B | iC | W | A | FOS | TER | Aver | rage |
| | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ |
| Average CIL ac | curacy | | | | | | | | | | | | | | | | | | | |
| | 58. | 20 | 55. | 87 | 61. | 44 | 63 | .51 | 59. | .76 | 40. | .86 | 42. | .26 | 45. | .99 | 45. | .96 | 43. | .77 |
| Post-hoc-based | OOD me | ethods | | | | | | | | | | | | | | | | | | |
| MSP | 66.89 | 87.86 | 68.66 | 87.22 | 69.19 | 86.11 | 70.01 | 85.68 | 68.69 | 86.72 | 61.87 | 91.03 | 63.06 | 92.46 | 61.69 | 92.85 | 65.45 | 90.89 | 63.02 | 91.81 |
| ODIN | 70.26 | 79.60 | 70.78 | 78.54 | 71.70 | 80.09 | 72.89 | 75.48 | 71.41 | 78.43 | 66.00 | 88.90 | 67.62 | 90.54 | 63.47 | 91.07 | 69.16 | 87.58 | 66.56 | 89.52 |
| Energy | 70.23 | 81.17 | 69.55 | 82.78 | 71.99 | 82.53 | 73.89 | 76.34 | 71.41 | 80.70 | 62.97 | 91.95 | 65.61 | 92.43 | 62.43 | 92.94 | 67.46 | 90.94 | 64.62 | 92.06 |
| MaxLogit | 70.16 | 81.87 | 69.87 | 83.31 | 71.89 | 82.96 | 73.82 | 77.34 | 71.44 | 81.37 | 63.94 | 91.58 | 64.67 | 93.32 | 63.51 | 91.44 | 67.31 | 91.30 | 64.86 | 91.91 |
| GEN | 70.39 | 82.07 | 70.89 | 79.54 | 72.22 | 82.14 | 74.21 | 76.95 | 71.93 | 80.17 | 55.67 | 93.51 | 60.65 | 92.82 | 61.27 | 93.38 | 59.94 | 94.26 | 59.38 | 93.49 |
| ReAct | 70.21 | 81.26 | 69.86 | 84.89 | 73.54 | 82.15 | 74.29 | 76.97 | 71.97 | 81.32 | 55.67 | 93.51 | 60.64 | 92.82 | 61.27 | 93.37 | 59.94 | 94.26 | 59.38 | 93.46 |
| KLM | 66.21 | 88.89 | 67.53 | 86.71 | 68.19 | 87.70 | 69.38 | 86.36 | 67.83 | 87.42 | 63.34 | 89.99 | 62.93 | 89.79 | 63.18 | 89.52 | 66.33 | 87.95 | 63.95 | 89.31 |
| Relation | 66.33 | 78.06 | 70.64 | 81.42 | 71.89 | 77.19 | 72.49 | 75.74 | 70.34 | 78.10 | 63.13 | 89.86 | 66.39 | 93.63 | 63.11 | 95.41 | 63.48 | 90.83 | 64.03 | 92.43 |
| NNGuide | 70.27 | 78.83 | 70.70 | 79.64 | 71.60 | 79.00 | 73.68 | 75.96 | 71.56 | 78.36 | 63.00 | 89.47 | 69.87 | 85.81 | 62.61 | 90.30 | 68.63 | 88.54 | 66.03 | 88.53 |
| Average | 68.99 | 82.18 | 69.83 | 82.67 | 71.36 | 82.21 | 72.74 | 78.54 | 70.73 | 81.40 | 62.58 | 90.70 | 64.86 | 91.18 | 62.72 | 91.83 | 66.01 | 90.03 | 64.04 | 90.94 |
| Fine-tuning-ba | sed OOD |) method. | 5 | | | | | | | | | | | | | | | | | |
| LogitNorm | 70.21 | 81.18 | 69.22 | 83.39 | 71.13 | 82.65 | 73.31 | <u>76.78</u> | 70.97 | 81.00 | 62.02 | 92.73 | <u>66.30</u> | 91.20 | 61.58 | 93.94 | 65.87 | 93.26 | 63.94 | 92.78 |
| T2FNorm | 70.45 | 81.50 | <u>69.59</u> | 83.29 | 70.90 | 83.26 | 73.26 | 77.35 | 71.05 | 81.35 | 62.82 | 92.12 | 65.80 | 92.30 | 62.40 | 92.84 | 66.55 | 92.01 | 64.39 | 92.32 |
| AUGMIX | 70.27 | 81.22 | 68.65 | 83.16 | 71.11 | 82.93 | 73.12 | 76.98 | 70.79 | 81.07 | 62.19 | 90.86 | 65.97 | <u>89.82</u> | 62.88 | 90.12 | 67.53 | <u>89.45</u> | 64.64 | 90.06 |
| REGMIX | 70.92 | 81.33 | 68.99 | 83.65 | <u>71.44</u> | 84.30 | <u>73.78</u> | 77.56 | <u>71.28</u> | 81.71 | 63.93 | 91.20 | 66.24 | 90.29 | <u>62.96</u> | 91.64 | <u>67.84</u> | 91.56 | <u>65.24</u> | 91.17 |
| VOS | 71.54 | 79.83 | 67.63 | 77.80 | 66.73 | 81.28 | 72.70 | 77.08 | 69.65 | 79.00 | 61.33 | 89.74 | 65.74 | 90.22 | 62.08 | 89.98 | 66.73 | 90.14 | 63.97 | 90.02 |
| BER (Ours) | 72.75 | 77.59 | 71.47 | <u>77.82</u> | 72.47 | 78.69 | 74.20 | 74.93 | 72.72 | 77.26 | <u>63.45</u> | 89.53 | 67.72 | 88.61 | 64.09 | 89.15 | 69.34 | 88.39 | 66.15 | 88.92 |
| Average | 70.68 | 81.01 | 68.82 | 82.26 | 70.26 | 82.88 | 73.23 | 77.15 | 70.75 | 80.83 | 62.46 | 91.33 | 66.01 | 90.77 | 62.38 | 91.70 | 66.90 | 91.28 | 64.44 | 91.27 |
| Average (All) | 69.59 | 81.76 | 69.47 | 82.52 | 70.97 | 82.45 | 72.91 | 78.04 | 70.74 | 81.20 | 62.54 | 90.92 | 65.27 | 91.03 | 62.60 | 91.78 | 66.33 | 90.48 | 64.18 | 91.06 |

Table 2: Main results on OpenCIL benchmark for the step size of k = 10 for CIFAR 100 and the step size of k = 100 for ImageNet1K. The results are averaged over six OOD datasets and all incremental steps. Either the post-hoc-based or fine-tuning-based OOD methods do not affect the original CIL performance, so their average CIL accuracy on the ID data remains the same. The best and second-best performance per dataset in the fine-tuning-based methods are highlighted.

| | | | | ID | Dataset: | CIFAR1 | 00 | | | | | | | ID | Dataset: 1 | ImageNe | t1K | | | |
|----------------|----------|----------|-------|-------|----------|--------|-------|-------|-------|-------|-------|-------|-------|-------|------------|---------|-------|-------|-------|-------|
| | iCa | RL | Bi | iC | W | Ά | FOS | TER | Ave | rage | iCa | RL | B | iC | W | Ά | FOS | TER | Ave | rage |
| | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ |
| Average CIL a | ccuracy | | | | | | | | | | | | | | | | | | | |
| | 60 | .08 | 61 | .68 | 65 | .88 | 66 | .01 | 63 | .41 | 44 | .44 | 49 | .63 | 52. | .22 | 52. | 29 | 49 | .65 |
| Post-hoc-based | l OOD m | ethods | | | | | | | | | | | | | | | | | | |
| MSP | 67.77 | 88.33 | 66.67 | 86.98 | 70.49 | 86.04 | 71.69 | 84.94 | 69.16 | 86.57 | 63.84 | 89.81 | 67.45 | 91.80 | 65.97 | 89.17 | 67.64 | 89.32 | 66.23 | 90.03 |
| ODIN | 70.05 | 81.93 | 69.76 | 80.52 | 71.86 | 80.18 | 74.70 | 75.68 | 71.59 | 79.58 | 68.41 | 87.91 | 71.09 | 87.45 | 67.86 | 89.08 | 70.24 | 88.15 | 69.40 | 88.15 |
| Energy | 70.35 | 83.74 | 69.76 | 82.01 | 73.03 | 81.22 | 75.76 | 77.24 | 72.23 | 81.05 | 65.64 | 91.88 | 70.16 | 89.57 | 66.57 | 92.22 | 69.39 | 90.49 | 67.94 | 91.04 |
| MaxLogit | 70.34 | 84.45 | 69.54 | 82.78 | 72.98 | 82.33 | 75.71 | 78.11 | 72.14 | 81.92 | 65.83 | 91.38 | 69.93 | 89.42 | 66.75 | 92.03 | 69.56 | 90.02 | 68.02 | 90.71 |
| GEN | 70.86 | 84.31 | 69.87 | 82.51 | 73.30 | 81.42 | 76.11 | 77.64 | 72.54 | 81.47 | 66.89 | 90.78 | 69.51 | 89.77 | 68.55 | 88.54 | 69.74 | 89.91 | 68.67 | 89.75 |
| ReAct | 68.79 | 84.43 | 70.62 | 82.60 | 74.87 | 81.82 | 76.15 | 78.03 | 72.61 | 81.72 | 55.66 | 94.47 | 66.57 | 89.86 | 63.85 | 90.94 | 65.15 | 91.16 | 62.81 | 91.61 |
| KLM | 67.51 | 88.00 | 66.27 | 87.66 | 69.89 | 87.65 | 71.34 | 85.90 | 68.75 | 87.30 | 66.48 | 88.28 | 67.32 | 87.17 | 65.24 | 89.59 | 70.23 | 88.45 | 67.32 | 88.37 |
| Relation | 64.40 | 89.57 | 65.48 | 80.85 | 76.12 | 88.37 | 72.17 | 75.29 | 69.54 | 83.52 | 66.38 | 89.92 | 69.10 | 89.63 | 67.51 | 94.59 | 67.93 | 88.99 | 67.73 | 90.78 |
| NNGuide | 72.14 | 79.65 | 68.74 | 78.79 | 75.60 | 76.00 | 75.90 | 74.70 | 73.09 | 77.28 | 66.53 | 89.88 | 70.64 | 87.33 | 68.69 | 88.82 | 67.63 | 88.32 | 68.37 | 88.59 |
| Average | 69.13 | 84.93 | 68.52 | 82.74 | 73.13 | 82.78 | 74.39 | 78.61 | 71.29 | 82.27 | 65.07 | 90.48 | 69.09 | 89.11 | 66.78 | 90.55 | 68.61 | 89.42 | 67.39 | 89.89 |
| Fine-tuning-bo | ised OOL |) method | s | | | | | | | | | | | | | | | | | |
| LogitNorm | 70.44 | 83.78 | 70.64 | 82.46 | 72.04 | 81.33 | 74.74 | 78.28 | 71.97 | 81.46 | 65.15 | 92.51 | 69.54 | 87.26 | 66.30 | 93.16 | 67.84 | 92.39 | 67.21 | 91.33 |
| T2FNorm | 70.65 | 84.14 | 71.75 | 82.72 | 71.98 | 82.33 | 74.65 | 78.83 | 72.26 | 82.00 | 66.09 | 91.58 | 69.53 | 89.09 | 67.39 | 91.83 | 68.19 | 90.95 | 67.80 | 90.86 |
| AUGMIX | 70.72 | 83.91 | 70.78 | 82.94 | 72.11 | 81.53 | 74.75 | 78.62 | 72.09 | 81.75 | 68.60 | 87.98 | 70.83 | 87.85 | 68.10 | 89.37 | 69.62 | 88.53 | 69.29 | 88.43 |
| REGMIX | 71.47 | 84.53 | 71.43 | 82.23 | 72.01 | 82.95 | 75.26 | 79.35 | 72.54 | 82.26 | 66.41 | 90.69 | 69.83 | 89.68 | 67.45 | 90.69 | 69.43 | 90.78 | 68.28 | 90.46 |
| VOS | 72.31 | 82.53 | 66.66 | 79.03 | 68.49 | 78.15 | 74.58 | 77.71 | 70.51 | 79.36 | 66.28 | 91.67 | 66.67 | 89.47 | 65.41 | 89.92 | 68.98 | 88.74 | 66.83 | 89.95 |
| BER (Ours) | 74.17 | 78.87 | 71.83 | 77.76 | 76.77 | 75.58 | 76.40 | 74.58 | 74.79 | 76.70 | 68.80 | 87.51 | 71.94 | 86.80 | 69.06 | 87.59 | 70.54 | 87.44 | 70.09 | 87.33 |
| Average | 71.12 | 83.78 | 70.25 | 81.88 | 71.33 | 81.26 | 74.80 | 78.56 | 71.87 | 81.37 | 66.51 | 90.89 | 69.28 | 88.67 | 66.93 | 90.99 | 68.81 | 90.28 | 67.88 | 90.21 |
| Average (All) | 69.84 | 84.52 | 69.14 | 82.43 | 72.49 | 82.24 | 74.54 | 78.59 | 71.50 | 81.95 | 65.58 | 90.63 | 69.16 | 88.95 | 66.83 | 90.71 | 68.68 | 89.73 | 67.56 | 90.00 |

Performance of BER. We compare the OOD detection capability of our proposed BER with the five fine-tuning-based OOD detectors among four CIL models based on two ID datasets CIFAR100 and ImageNet1K. Tables 1, 2, and 3 present the comparisons at the step size of $k = \{5, 10, 20\}$ for CIFAR100 on left side and at the step size of $k = \{50, 100, 200\}$ for ImageNet1K on right side, respectively. For the two metrics of four CIL models on CIFAR100 across three steps, there are 24 possible combinations for each OOD detector. In 23 out of 24 cases, BER achieves the best OOD detection performance across the combination of four representative CIL models and three types of step sizes in AUR and FPR. As for ImageNet1K, BER also achieves the best OOD detection performance in 21 out of 24 cases. Note that the other three best performers on ImageNet1K are achieved by three different previous OOD detectors combined with different CIL models. BER also achieves the second-best performance in the cases where it is not the best performer. This demon-strates the strong robustness of our proposed BER under diverse combined CIL and OOD dataset and model scenarios. Furthermore, BER also achieves the best performance in the average results across four CIL models for all three types of step sizes in AUC and FPR on both the CIFAR100 and ImageNet1K datasets. This consistent improvement indicates that our energy regularization on both the old and new classes helps effectively mitigate biases towards the OOD samples and new classes in the use of fine-tuning-based OOD detection methods in CIL models.

Table 3: Main results on OpenCIL benchmark for the step size of k = 20 for CIFAR 100 and the step size of k = 200 for ImageNet1K. The results are averaged over six OOD datasets and all incremental steps. Either the post-hoc-based or fine-tuning-based OOD methods do not affect the original CIL performance, so their average CIL accuracy on the ID data remains the same. The best and <u>second-best</u> performance per dataset in the fine-tuning-based methods are highlighted.

| | | | | ID | Dataset: | CIFAR1 | 00 | | | | | | | ID | Dataset: | ImageNe | t1K | | | |
|-----------------|---------|-----------|-------|-------|--------------|--------|-------|-------|--------------|-------|-------|-------|--------------|--------------|----------|---------|-------|-------|-------|-------|
| ſ | iCa | RL | Bi | iC | W | A | FOS | TER | Ave | rage | iCa | RL | Bi | C | W | A | FOS | TER | Aver | rage |
| | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ |
| Average CIL ac | curacy | | | | | | | | | | | | | | | | | | | |
| | 62. | 65 | 64 | .14 | 68. | 05 | 68 | .75 | 65. | .90 | 48. | .85 | 53. | 30 | 58. | .42 | 57. | .84 | 54. | .60 |
| Post-hoc-based | OOD me | ethods | | | | | | | | | | | | | | | | | | |
| MSP | 68.38 | 87.19 | 69.72 | 87.75 | 72.46 | 84.74 | 72.10 | 85.02 | 70.66 | 86.17 | 66.83 | 87.47 | 71.39 | 86.68 | 70.73 | 87.12 | 71.57 | 84.44 | 70.13 | 86.43 |
| ODIN | 71.34 | 81.20 | 71.46 | 80.62 | 74.27 | 78.08 | 74.45 | 76.73 | 72.88 | 79.16 | 71.36 | 83.61 | 75.27 | 82.55 | 71.69 | 84.60 | 73.56 | 82.93 | 72.97 | 83.42 |
| Energy | 72.38 | 81.99 | 71.48 | 82.45 | 75.22 | 79.00 | 76.14 | 78.54 | 73.81 | 80.50 | 69.52 | 88.30 | 74.36 | 84.92 | 70.56 | 87.73 | 72.67 | 87.09 | 71.78 | 87.01 |
| MaxLogit | 72.32 | 82.20 | 71.60 | 83.73 | 75.22 | 80.02 | 76.08 | 78.84 | 73.80 | 81.20 | 69.62 | 87.90 | 74.17 | 85.99 | 71.08 | 87.26 | 73.15 | 85.57 | 72.00 | 86.68 |
| GEN | 72.73 | 82.44 | 71.92 | 83.00 | 75.36 | 79.38 | 76.55 | 78.39 | 74.14 | 80.80 | 70.16 | 87.60 | 74.19 | 85.84 | 71.68 | 85.67 | 73.84 | 84.30 | 72.47 | 85.85 |
| ReAct | 71.59 | 82.99 | 73.69 | 82.87 | 76.82 | 78.49 | 77.33 | 78.85 | 74.86 | 80.80 | 59.48 | 92.57 | 69.67 | 85.93 | 63.12 | 88.28 | 64.82 | 89.13 | 64.27 | 88.98 |
| KLM | 69.10 | 85.53 | 70.19 | 86.76 | 72.47 | 86.06 | 72.47 | 84.39 | 71.06 | 85.69 | 69.51 | 84.99 | 70.55 | 85.72 | 69.61 | 87.49 | 73.55 | 82.10 | 70.81 | 85.07 |
| Relation | 65.37 | 80.08 | 69.10 | 79.46 | 76.23 | 77.46 | 71.27 | 76.53 | 70.49 | 78.38 | 71.24 | 87.09 | 75.12 | 83.34 | 71.45 | 92.75 | 73.27 | 83.31 | 72.77 | 86.62 |
| NNGuide | 74.14 | 77.05 | 71.24 | 78.89 | 76.28 | 73.78 | 76.20 | 76.13 | 74.47 | 76.46 | 71.41 | 88.03 | 77.95 | 86.53 | 71.90 | 86.90 | 73.27 | 82.74 | 73.63 | 86.05 |
| Average | 70.82 | 82.30 | 71.16 | 82.84 | 74.93 | 79.67 | 74.73 | 79.27 | 72.91 | 81.02 | 68.79 | 87.51 | 73.63 | 85.28 | 70.20 | 87.53 | 72.19 | 84.62 | 71.20 | 86.24 |
| Fine-tuning-bas | sed OOD |) method. | 5 | | | | | | | | | | | | | | | | | |
| LogitNorm | 72.41 | 82.32 | 71.27 | 83.80 | 74.29 | 79.49 | 75.38 | 79.28 | 73.34 | 81.22 | 69.40 | 89.17 | 73.18 | 84.05 | 69.14 | 89.57 | 71.62 | 89.30 | 70.84 | 88.02 |
| T2FNorm | 72.67 | 82.39 | 71.51 | 83.79 | 74.38 | 80.66 | 75.70 | 79.60 | 73.56 | 81.61 | 70.45 | 87.47 | 74.21 | 82.17 | 70.95 | 87.36 | 71.64 | 87.72 | 71.81 | 86.18 |
| AUGMIX | 72.50 | 82.62 | 70.70 | 84.33 | 74.35 | 79.97 | 75.54 | 79.61 | 73.27 | 81.63 | 71.32 | 87.91 | 76.72 | 84.63 | 70.84 | 87.20 | 72.81 | 83.66 | 72.92 | 85.85 |
| REGMIX | 72.74 | 82.81 | 71.62 | 84.65 | <u>74.78</u> | 80.45 | 75.69 | 80.09 | <u>73.71</u> | 82.00 | 69.56 | 89.61 | 72.78 | 87.57 | 70.68 | 88.13 | 71.84 | 88.85 | 71.22 | 88.54 |
| VOS | 72.93 | 81.39 | 68.77 | 78.28 | 72.21 | 76.57 | 75.07 | 78.52 | 72.24 | 78.69 | 62.94 | 89.63 | 64.19 | 88.20 | 62.86 | 89.02 | 65.61 | 88.09 | 63.90 | 88.73 |
| BER (Ours) | 74.55 | 76.75 | 71.90 | 77.97 | 76.64 | 71.72 | 76.52 | 75.04 | 74.90 | 75.37 | 71.88 | 87.00 | <u>76.46</u> | <u>83.01</u> | 71.98 | 86.24 | 73.95 | 82.66 | 73.57 | 84.73 |
| Average | 72.65 | 82.31 | 70.77 | 82.97 | 74.00 | 79.43 | 75.48 | 79.42 | 73.22 | 81.03 | 68.73 | 88.76 | 72.22 | 85.32 | 68.89 | 88.26 | 70.70 | 87.52 | 70.14 | 87.46 |
| Average (All) | 71.47 | 82.30 | 71.02 | 82.89 | 74.60 | 79.58 | 75.00 | 79.32 | 73.02 | 81.02 | 68.77 | 87.96 | 73.13 | 85.29 | 69.73 | 87.79 | 71.66 | 85.66 | 70.82 | 86.68 |

Table 4: Ablation study of our proposed BER based on the CIFAR100 ID dataset with the step size of k = 10. Energy (Liu et al., 2020) is used as the baseline that does not use both NTER and OTER.

| Loss Co | mnonent | N | lear OOI | D Dataset | s | | | | Far OOD | Datasets | | | |
|---------|---------|-------|----------|-----------|-------|-------|-------|-------|---------|----------|-------|-------|-------|
| LUSS CU | mponent | CIFA | R10 | TI | N | MN | IST | SV | HN | Tex | ture | Place | s365 |
| NTER | OTER | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ | AUC↑ | FPR↓ |
| X | X | 70.70 | 84.86 | 72.74 | 83.12 | 86.63 | 62.37 | 62.56 | 96.81 | 56.18 | 92.92 | 73.31 | 82.33 |
| 1 | X | 71.31 | 84.16 | 73.64 | 82.37 | 89.93 | 52.73 | 65.63 | 95.93 | 59.66 | 91.15 | 75.33 | 79.72 |
| × | 1 | 70.77 | 83.84 | 73.80 | 82.10 | 88.65 | 55.36 | 64.67 | 95.80 | 59.26 | 91.94 | 74.35 | 81.34 |
| _ ✓ | 1 | 71.60 | 83.17 | 75.84 | 81.06 | 91.56 | 46.02 | 67.52 | 93.08 | 62.59 | 90.52 | 75.90 | 79.35 |

Ablation Study. This section evaluates the importance of the two key components of BER, New Task Energy Regularization (NTER) and Old Task Energy Regularization (OTER). Table 4 presents the results of the ablation study conducted on these two components using six OOD datasets individ-ually based on the CIFAR 100 ID dataset with the step size of k = 10 using Energy (Liu et al., 2020) as the baseline. The results show that either NTER or OTER helps boost the AUC performance and reduce the FPR on both near and far OOD datasets, and they can achieve the best performance when the two components are combined. Since NTER is designed to alleviate the misclassification of OOD samples into new classes while OTER is designed to reduce the misclassification of old class samples as OOD samples, their combination results in a detection model that largely reduces both types of detection errors.

CONCLUSION

In this paper, we introduce OpenCIL, the first large-scale and systematic benchmark designed to enable CIL models with existing OOD detectors, regarding the CIL models in open world applica-tions. OpenCIL introduces two principled frameworks for incorporating diverse OOD detectors into CIL models and a new evaluation pipeline for fairly evaluating the capability of OOD detectors in incremental learning. In particular, OpenCIL accommodates four representative CIL models with 15 diverse OOD detection methods, resulting in 60 baseline models on two popular CIL datasets and six commonly-used near/far OOD datasets. Based on our large-scale experiments on OpenCIL, we offer several important insights into the design of CIL models for open-world applications. We fur-ther propose a novel approach, namely Bi-directional Energy Regularization (BER), which utilizes energy regularization based on two types of sample synthesis to effectively mitigate the increasing bias of CIL Models towards OOD samples and newly added classes with the growth of incremen-tal steps. Extensive experiments demonstrate that BER achieves state-of-the-art performance on the OpenCIL benchmark under varying incremental step sizes on popular CIL and OOD datasets, improving the OOD detection capability of a wide range of CIL models.

540 REFERENCES

570

571

572

579

586

- Steven Basart, Mazeika Mantas, Mostajabi Mohammadreza, Steinhardt Jacob, and Song Dawn.
 Scaling out-of-distribution detection for real-world settings. In *International Conference on Machine Learning*, 2022.
- Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3606–3613, 2014.
- 549 Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. Vos: Learning what you don't know by
 550 virtual outlier synthesis. In *International Conference on Learning Representations*, 2022.
- Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations*, 2017.
- Dan Hendrycks, Norman Mu, Ekin D. Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshmi narayanan. AugMix: A simple data processing method to improve robustness and uncertainty.
 Proceedings of the International Conference on Learning Representations (ICLR), 2020.
- Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. Natural adversarial
 examples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recogni tion*, pp. 15262–15271, 2021.
- Rui Huang and Yixuan Li. Mos: Towards scaling out-of-distribution detection for large semantic
 space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*,
 pp. 8710–8719, 2021.
- Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.
- Jang-Hyun Kim, Sangdoo Yun, and Hyun Oh Song. Neural relation graph: A unified framework
 for identifying label noise and outlier data. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
 - Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015.
- Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database. ATT Labs [Online].
 Available: http://yann.lecun.com/exdb/mnist, 2, 2010.
- Christian Leibig, Vaneeda Allken, Murat Seçkin Ayhan, Philipp Berens, and Siegfried Wahl. Leveraging uncertainty information from deep neural networks for disease detection. *Scientific reports*, 7(1):17816, 2017.
- Jingyao Li, Pengguang Chen, Zexin He, Shaozuo Yu, Shu Liu, and Jiaya Jia. Rethinking out-of distribution (ood) detection: Masked image modeling is all you need. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11578–11589, 2023.
- Tianqi Li, Guansong Pang, Xiao Bai, Wenjun Miao, and Jin Zheng. Learning transferable negative prompts for out-of-distribution detection. *arXiv preprint arXiv:2404.03248*, 2024a.
- Tianqi Li, Guansong Pang, Xiao Bai, Jin Zheng, Lei Zhou, and Xin Ning. Learning adversarial semantic embeddings for zero-shot recognition in open worlds. *Pattern Recognition*, 149:110258, 2024b.
- ⁵⁹³ Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017.

594 Shiyu Liang, Yixuan Li, and R Srikant. Enhancing the reliability of out-of-distribution image detec-595 tion in neural networks. In International Conference on Learning Representations, 2018. 596 Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detec-597 tion. Advances in neural information processing systems, 33:21464–21475, 2020. 598 Xixi Liu, Yaroslava Lochman, and Christopher Zach. Gen: Pushing the limits of softmax-based 600 out-of-distribution detection. In Proceedings of the IEEE/CVF Conference on Computer Vision 601 and Pattern Recognition, pp. 23946-23955, 2023a. 602 603 Yaoyao Liu, Bernt Schiele, and Qianru Sun. Rmm: Reinforced memory management for classincremental learning. Advances in Neural Information Processing Systems, 34:3478–3490, 2021. 604 605 Yuyuan Liu, Choubo Ding, Yu Tian, Guansong Pang, Vasileios Belagiannis, Ian Reid, and Gustavo 606 Carneiro. Residual pattern learning for pixel-wise out-of-distribution detection in semantic seg-607 mentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 608 1151–1161, 2023b. 609 610 Zilin Luo, Yaoyao Liu, Bernt Schiele, and Qianru Sun. Class-incremental exemplar compression 611 for class-incremental learning. In Proceedings of the IEEE/CVF Conference on Computer Vision 612 and Pattern Recognition, pp. 11371–11380, 2023. 613 Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D Bagdanov, and Joost 614 Van De Weijer. Class-incremental learning: survey and performance evaluation on image clas-615 sification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(5):5513–5533, 616 2022. 617 Wenjun Miao, Guansong Pang, Tianqi Li, Xiao Bai, and Jin Zheng. Out-of-distribution detection in 618 long-tailed recognition with calibrated outlier class learning. In *Proceedings of the 38th Annual* 619 AAAI Conference on Artificial Intelligence, 2024. 620 621 Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading 622 digits in natural images with unsupervised feature learning. 2011. 623 624 Chaoxi Niu, Guansong Pang, and Ling Chen. Graph continual learning with debiased lossless 625 memory replay. arXiv preprint arXiv:2404.10984, 2024. 626 Jaewoo Park, Yoon Gyo Jung, and Andrew Beng Jin Teoh. Nearest neighbor guidance for out-of-627 distribution detection. In Proceedings of the IEEE/CVF International Conference on Computer 628 Vision, pp. 1686–1695, 2023. 629 630 Francesco Pinto, Harry Yang, Ser-Nam Lim, Philip Torr, and Puneet K. Dokania. Regmixup: Mixup 631 as a regularizer can surprisingly improve accuracy and out distribution robustness. 2023. 632 Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: 633 Incremental classifier and representation learning. In Proceedings of the IEEE conference on 634 Computer Vision and Pattern Recognition, pp. 2001–2010, 2017. 635 636 Sudarshan Regmi, Bibek Panthi, Sakar Dotel, Prashnna K Gyawali, Danail Stoyanov, and Binod 637 Bhattarai. T2fnorm: Extremely simple scaled train-time feature normalization for ood detection. 638 arXiv preprint arXiv:2305.17797, 2023. 639 Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng 640 Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual 641 recognition challenge. International journal of computer vision, 115:211–252, 2015. 642 643 Yiyou Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activa-644 tions. Advances in Neural Information Processing Systems, 34:144–157, 2021. 645 Yu Tian, Yuyuan Liu, Guansong Pang, Fengbei Liu, Yuanhong Chen, and Gustavo Carneiro. Pixel-646 wise energy-biased abstention learning for anomaly segmentation on complex urban driving 647 scenes. In European Conference on Computer Vision, pp. 246-263. Springer, 2022.

685

686

687

| 040 | Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, |
|-----|---|
| 649 | Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In |
| 650 | Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8769–8778, |
| 651 | 2018. |

- Fu-Yun Wang, Da-Wei Zhou, Han-Jia Ye, and De-Chuan Zhan. Foster: Feature boosting and compression for class-incremental learning. In *European conference on computer vision*, pp. 398–414.
 Springer, 2022a.
- Fu-Yun Wang, Da-Wei Zhou, Liu Liu, Han-Jia Ye, Yatao Bian, De-Chuan Zhan, and Peilin Zhao.
 Beef: Bi-compatible class-incremental learning via energy-based expansion and fusion. In *The eleventh international conference on learning representations*, 2023a.
- 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*, pp. 4921–4930, 2022b.
- Kudong Wang, Long Lian, Zhongqi Miao, Ziwei Liu, and Stella X Yu. Long-tailed recognition by routing diverse distribution-aware experts. *arXiv preprint arXiv:2010.01809*, 2020.
- Zhenyu Wang, Yali Li, Xi Chen, Ser-Nam Lim, Antonio Torralba, Hengshuang Zhao, and Shengjin
 Wang. Detecting everything in the open world: Towards universal object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11433–11443, 2023b.
- 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*, pp. 23631–23644. PMLR, 2022.
- Yue Wu, Yinpeng Chen, Lijuan Wang, Yuancheng Ye, Zicheng Liu, Yandong Guo, and Yun Fu.
 Large scale incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision* and pattern recognition, pp. 374–382, 2019.
- Jia-Wen Xiao, Chang-Bin Zhang, Jiekang Feng, Xialei Liu, Joost van de Weijer, and Ming-Ming
 Cheng. Endpoints weight fusion for class incremental semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7204–7213, 2023.
- Ju Xu and Zhanxing Zhu. Reinforced continual learning. *Advances in neural information processing systems*, 31, 2018.
- Shipeng Yan, Jiangwei Xie, and Xuming He. Der: Dynamically expandable representation for class
 incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3014–3023, 2021.
 - Jingkang Yang, Haoqi Wang, Litong Feng, Xiaopeng Yan, Huabin Zheng, Wayne Zhang, and Ziwei Liu. Semantically coherent out-of-distribution detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8301–8309, 2021.
- Jingkang Yang, Pengyun Wang, Dejian Zou, Zitang Zhou, Kunyuan Ding, Wenxuan Peng, Haoqi
 Wang, Guangyao Chen, Bo Li, Yiyou Sun, et al. Openood: Benchmarking generalized out-ofdistribution detection. Advances in Neural Information Processing Systems, 35:32598–32611,
 2022.
- Yeonguk Yu, Sungho Shin, Seongju Lee, Changhyun Jun, and Kyoobin Lee. Block selection method
 for using feature norm in out-of-distribution detection. In *Proceedings of the IEEE/CVF Confer- ence on Computer Vision and Pattern Recognition*, pp. 15701–15711, 2023.
- Jingyang Zhang, Jingkang Yang, Pengyun Wang, Haoqi Wang, Yueqian Lin, Haoran Zhang, Yiyou
 Sun, Xuefeng Du, Kaiyang Zhou, Wayne Zhang, et al. Openood v1. 5: Enhanced benchmark for
 out-of-distribution detection. *arXiv preprint arXiv:2306.09301*, 2023.
- 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*, pp. 3388–3397, 2023.

| 702 703 704 | Bowen Zhao, Xi Xiao, Guojun Gan, Bin Zhang, and Shu-Tao Xia. Maintaining discrimination and fairness in class incremental learning. In <i>Proceedings of the IEEE/CVF conference on computer vision and pattern recognition</i> , pp. 13208–13217, 2020. |
|-------------------|--|
| 705 706 | Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. <i>IEEE transactions on pattern analysis and machine</i> |
| 707 708 | intelligence, 40(6):1452–1464, 2017. |
| 709 | Da-Wei Zhou, Han-Iia Ye, and De-Chuan Zhan. Learning placeholders for open-set recognition |
| 710 | In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. |
| 711 | 4401–4410, 2021. |
| 712 | |
| 713 | |
| 714 | |
| 715 | |
| 716 | |
| /1/ | |
| 710 | |
| 720 | |
| 721 | |
| 722 | |
| 723 | |
| 724 | |
| 725 | |
| 726 | |
| 727 | |
| 728 | |
| 729 | |
| 730 | |
| 732 | |
| 733 | |
| 734 | |
| 735 | |
| 736 | |
| 737 | |
| 738 | |
| 739 | |
| 740 | |
| 741 | |
| 742 | |
| 743 | |
| 745 | |
| 746 | |
| 747 | |
| 748 | |
| 749 | |
| 750 | |
| 751 | |
| 752 | |
| 753 | |
| 754 755 | |
| (33 | |

⁷⁵⁶ A More Evaluation Setting Details

758 A.1 DATASETS

The in-distribution (ID) datasets for class incremental learning (CIL) include two datasets. 1) CIFAR100 (Krizhevsky et al., 2009) contains 50,000 training images and 10,000 test images of size
32 × 32 with 100 classes, and 2) ImageNet (Russakovsky et al., 2015) is a large-scale classification
dataset, which contains 1.28M training images and 50K testing images with 1,000 classes sampled
from nature images.

765 We select six commonly used out-of-distribution (OOD) datasets for OOD detection on CIFRA100. 766 Following the recent large-scale OOD detection benchmark OpenOOD (Yang et al., 2022), our 767 OOD datasets include two near OOD datasets: CIFAR10 (Krizhevsky et al., 2009) and Tiny-ImageNet (TIN) (Le & Yang, 2015), and four far OOD datasets: MNIST (LeCun et al., 2010), 768 Texture (Cimpoi et al., 2014), Places365 (Zhou et al., 2017), and SVHN (Netzer et al., 2011). 769 CIFAR10 (Krizhevsky et al., 2009) contains 10,000 images with 10 classes. TIN (Le & Yang, 770 2015) contains 7,498 images, which removes the 2,502 images that have overlapping semantics 771 (Yang et al., 2021) with CIFAR100 classes. MNIST (LeCun et al., 2010) contains 10,000 images 772 with 10 classes. Texture (Cimpoi et al., 2014) contains 5,640 images with 47 classes. Places365 773 (Zhou et al., 2017) contains 33, 773 images, which removes the 2, 727 images that have overlapping 774 semantics (Yang et al., 2021) with CIFAR100 classes. SVHN (Netzer et al., 2011) contains 26, 032 775 images with 10 class. 776

We also select six commonly used OOD datasets for OOD detection on ImageNet1k. Following 777 OpenOOD (Yang et al., 2022), these includes four near OOD datasets: ImageNet_O (Hendrycks 778 et al., 2021), iNaturalist (Van Horn et al., 2018), OpenImage_O (Wang et al., 2022b) and Species 779 (Basart et al., 2022), and two far OOD datasets: MNIST (LeCun et al., 2010) and Texture (Cimpoi et al., 2014). ImageNet_O (Hendrycks et al., 2021) contains 2,000 images from categories not 781 found in the ImageNet1k dataset. We use a 10,000 image subset of iNaturalist (Van Horn et al., 782 2018), which is based on 110 manually selected plant classes not present in ImageNet1k. The OOD 783 samples are randomly sampled images from these 110 classes (Huang & Li, 2021). All images are 784 resized to have a max dimension of 800 pixels. OpenImage_O (Wang et al., 2022b) contains 15,869 images with the support of a manual filter. We use a 10,000 subset of 713K images Species (Basart 785 et al., 2022) with 10 classes. Two far OOD datasets are the same as CIFAR100. Near OOD datasets 786 mean that their OOD samples have small semantic shifts compared with the ID samples, while 787 far OOD datasets mean that their OOD samples are very different from the ID samples, typically 788 containing obvious covariate (domain) shift (Yang et al., 2022). A summary of the CIL ID and 789 OOD datasets is presented in Table 5. 790

Table 5: Key statistics of used CIL ID datasets and OOD datasets.

| Banchmark | CIFA | AR100 | | Ima | geNet | |
|--------------------|--------------|--------|-------|-------------|--------|-------|
| Deneminark | Dataset | Images | Class | Dataset | Images | Class |
| ID data (Training) | CIFAR100 | 50,000 | 100 | ImageNet | 1.28M | 1000 |
| ID data (Testing) | CIFAR100 | 10,000 | 100 | ImageNet | 50,000 | 1000 |
| | CIFAR10 | 10,000 | 10 | ImageNet_O | 2,000 | / |
| | TinyImageNet | 7,498 | / | iNaturalist | 10,000 | 110 |
| OOD data | MNIST | 10,000 | 10 | OpenImage_O | 15,869 | / |
| OOD data | Texture | 5,640 | 47 | Species | 10,000 | 10 |
| | Places365 | 33,773 | / | MNIST | 10,000 | 10 |
| | SVHN | 26,032 | 10 | Texture | 5,640 | 47 |

801 802

791

792 793 794

803 804

805

A.2 PERFORMANCE METRICS

This section provides more introduction to the performance measures used in our experiments. (1) FPR is the false positive rate of OOD examples when the true positive rate of ID examples is at 95%. It measures the portion of falsely recognized OOD when most of the ID samples are recalled. (2) AUC computes the area under the receiver operating characteristic curve of detecting OOD samples, evaluating the OOD detection performance. (3) AP measures the area under the precisionrecall curve, in which the OOD samples are treated as positive samples. (4) ACC calculates the classification accuracy of the ID data for the CIL models. Among all these metrics, only FPR95 is expected to have a lower value for a better model. Higher values indicate better performance for the other three metrics.

814 815

816

B BASELINE LIBRARY.

We include four different popular CIL models. iCaRL (Rebuffi et al., 2017) and WA (Zhao et al., 2020) are the regularization-based algorithms, BiC (Wu et al., 2019) is a replay-based algorithm, and FOSTER (Wang et al., 2022a) is a parameter-isolation-based method but it also uses data replay.

For OOD methods, there are two main categories, post-hoc-based and fine-tuning-based methods. 821 For post-hoc-based methods, we include nine OOD detection methods: MSP (Hendrycks & Gim-822 pel, 2017), ODIN (Liang et al., 2018), Energy (Liu et al., 2020), MaxLogit (Basart et al., 2022), 823 and GEN (Liu et al., 2023a) use the statistic based on prediction output of each test samples without 824 using any training ID information, while the other four methods: ReAct (Sun et al., 2021), KLM 825 (Basart et al., 2022), Relation (Kim et al., 2023), and NNGuide (Park et al., 2023) need the posterior 826 information of training ID samples to obtain the OOD score. Therefore, if they need a whole label 827 space of training ID data, we feed T_t^{train} to them, which is the combination of memory data and cur-828 rent task training data. For fine-tuning-based methods, we include five methods: LogitNorm (Wei 829 et al., 2022) and T2FNorm (Regmi et al., 2023) are the regularization-based methods that apply the normalization to calibrate the loss function, AUGMIX (Hendrycks et al., 2020) and REGMIX (Pinto 830 et al., 2023) are the augmentation-based methods that apply the data augmentation to ID samples 831 for enhancing the ID data training, while VOS (Du et al., 2022) are synthesis-based methods that 832 generate pseudo-OOD samples to assist the detector training. Notably, VOS focuses on outlier syn-833 thesis, but applying their original training method to fine-tune the extra classifier is difficult since it 834 was originally designed without the final linear classifier. Thus, to have a fair comparison, we apply 835 the same energy regularization as our BER for them, which replaces our synthesized pseudo-OOD 836 samples \bar{x}_t with their synthesized outlier samples.

837 838

839

C IMPLEMENTATION DETAILS.

840 For the CIL pre-training, we employ default hyperparameters of these CIL models as stated in their 841 original papers. For OOD fine-tuning, following (Yang et al., 2022; Zhang et al., 2023), we use 842 the common setting with SGD optimizer, using a learning rate of 0.1, a momentum of 0.9, a weight 843 decay of 0.0005, and adjusting the learning rate using a cosine annealing learning rate schedule. 844 The batch size is fixed at 128 for all experiments. We freeze the feature extractor and original 845 classifier, fine-tune the extra classifier for 10 epochs, and keep other hyperparameters the same as 846 the original paper. For our proposed BER baseline, following (Liu et al., 2020), we set $\tau = 1$, 847 $\alpha = 0.1, p_{in} = -5$ and $p_{out} = -27$ by default. $\lambda = 0.002$ is used throughout the experiments. All 848 results are averaged over three independent runs using different random seeds. All experiments are 849 performed using 8 NVIDIA RTX 3090.

850 851

852

D MORE RESULTS W.R.T. INCREASING INCREMENTAL STEPS

In this section, we provide more results for the different baselines w.r.t. increasing incremental steps
in our paper. Particularly, following the results of iCaRL (Rebuffi et al., 2017) with CIFAR100 (Krizhevsky et al., 2009) in our paper, we provide the results based on Bic (Wu et al., 2019) in Fig.
4, WA (Zhao et al., 2020) in Fig. 5, FOSTER (Wang et al., 2022a) in Fig. 6. Furthermore, we show
the relationship between ACC and OOD detection performance on these three CIL models in Fig.
7. We also show the average performance of all four CIL models with more OOD detectors (GEN (Liu et al., 2023a), KLM (Basart et al., 2022), and VOS (Du et al., 2022)) in Fig. 8.

- 860
- 861
- 862



Figure 4: Qualitative results of the CIL model BiC with CIFAR100. (a) All four representative OOD
detection methods experience a decreased AUC performance with increasing incremental steps,
compared to themselves working on the full training data of all steps. (b) Mean prediction confidence of BiC on test samples from all incremental classes. (c) Mean prediction confidence of BiC
classifying six OOD datasets into one of the ID classes based on the final incremental task.



Figure 5: Qualitative results of the CIL model WA with CIFAR100. (a) All four representative OOD detection methods experience a decreased AUC performance with increasing incremental steps, compared to themselves working on the full training data of all steps. (b) Mean prediction confidence of WA on test samples from all incremental classes. (c) Mean prediction confidence of WA classifying six OOD datasets into one of the ID classes based on the final incremental task.



Figure 6: Qualitative results of the CIL model FOSTER with CIFAR100. (a) All four representative OOD detection methods experience a decreased AUC performance with increasing incremental steps, compared to themselves working on the full training data of all steps. (b) Mean prediction confidence of FOSTER on test samples from all incremental classes. (c) Mean prediction confidence of FOSTER classifying six OOD datasets into one of the ID classes based on the final incremental task.



Figure 7: Average performance of four representative OOD methods on six OOD datasets at each incremental step, where the different CIL models are used. ACC is the accuracy of CIL models on CIFAR100 at each step.



Figure 8: Average performance of CIL models with different OOD detectors on six OOD datasets at each incremental step on CIFAR100.

FINE-GRAINED EXPERIMENTAL RESULTS ON NEAR- AND FAR-OOD Ε **DETECTION DATASETS**

Following the experiment results in our paper, we provide more fine-grained results (on near-OOD datasets and far-OOD datasets, respectively) at different step sizes in Tables 6, 7, and 8, in which we also add the AP results of the OOD detection performance.

Table 6: Detailed results for those in Table 1 on near- and far-OOD detection datasets at the step size of k = 5 for CIFAR 100 and at the step size of k = 50 for ImageNet1K. The best and second-best performance per dataset in the fine-tuning-based methods are highlighted. The upper, middle, lower parts of the table are for AUC, AP, and FPR performance, respectively.

| AUC↑ | | | | | - | | | | | _ | | | | | | | | | | |
|--|---|---|---|--|--|---|---|--|---|--|---|--|--|--|---|--|--|--|---|--|
| AUC↑ | | | | ID | Dataset: | CIFARI | 100 | | | | | | | IDI | Jataset: 1 | ImageNe | tlK | | | |
| | iCa | RL | Bi | iC | W | /A | FOS | TER | Aver | age | iCa | RL | Bi | C | W | A | FOS | TER | Ave | rage |
| | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far |
| Average CIL a | ccuracy | | | | | | | | | | | | | | | | | | | |
| | 58 | .20 | 55. | .87 | 61 | .44 | 63 | .51 | 59. | 76 | 40. | 86 | 42. | 26 | 45. | 99 | 45. | .96 | 43 | .77 |
| -Post-hoc-hase | d OOD 1 | nethods | | | | | | | | | | | | | | | | | | |
| MCD | 67.21 | 66 72 | 67.60 | 60.15 | 70.22 | 69 69 | 70.14 | 60.04 | 69.91 | 68 62 1 | 61.92 | 61.09 | 65.12 | 58.02.1 | 62.50 | 60.07 | 64.12 | 68.12 | 62 20 | 62.27 |
| NIST ODDA | 07.21 | 00.75 | 07.09 | 09.15 | 70.22 | 08.08 | 70.14 | 09.94 | 00.01 | 08.02 | 64.00 | 01.90 | 00.12 | 56.92 | 62.50 | 60.07 | 04.12 | 00.12 | 05.59 | 02.27 |
| ODIN | /0.97 | 69.91 | 69.22 | /1.56 | 72.98 | /1.06 | 72.25 | 73.21 | /1.36 | 71.44 | 64.80 | 68.40 | 68.11 | 66.65 | 63.29 | 63.83 | 66.19 | 75.10 | 65.60 | 68.50 |
| Energy | 71.23 | 69.73 | 68.56 | 70.05 | 73.19 | 71.38 | 72.98 | 74.34 | 71.49 | 71.38 | 62.62 | 63.07 | 67.58 | 63.37 | 62.00 | 63.50 | 64.59 | 73.29 | 64.20 | 65.81 |
| MaxLogit | 71.16 | 69.66 | 68.85 | 70.38 | 73.16 | 71.26 | 72.98 | 74.25 | 71.54 | 71.39 | 62.87 | 63.17 | 67.11 | 62.62 | 62.65 | 62.01 | 65.02 | 72.34 | 64.41 | 65.03 |
| GEN | 71.18 | 69.99 | 69.44 | 71.61 | 73.30 | 71.68 | 73.34 | 74.65 | 71.81 | 71.98 | 63.55 | 64.72 | 65.69 | 62.61 | 64.69 | 61.15 | 65.23 | 71.47 | 64.79 | 64.99 |
| ReAct | 68.67 | 70.98 | 66.20 | 71.65 | 72.30 | 74.16 | 70.81 | 76.03 | 69.52 | 73.20 | 57.02 | 51.17 | 63.42 | 55.00 | 62.11 | 50 50 | 50.26 | 61.31 | 60.68 | 56 79 |
| KIM | 66.70 | (5.00 | 66.27 | (0.00 | 12.50 | (0.01 | 10.01 | (0.05 | 67.01 | (7.04 | (2.20 | 0.40 | (2.(2 | (1.55 | 62.11 | (0.74 | 65.00 | (7.15 | 64.06 | (2.72 |
| KLM | 00.78 | 05.92 | 00.44 | 08.09 | 08.55 | 08.01 | 09.47 | 09.35 | 07.81 | 07.84 | 03.29 | 03.40 | 0.5.02 | 01.55 | 0.5.40 | 02.74 | 05.93 | 07.15 | 04.00 | 03.73 |
| Relation | 61.30 | 68.85 | 65.38 | 73.27 | 67.53 | 74.07 | 68.42 | 74.52 | 65.66 | 72.68 | 58.73 | 69.95 | 66.19 | 66.80 | 63.22 | 62.87 | 61.05 | 68.37 | 62.30 | 67.00 |
| NNGuide | 67.33 | 71.73 | 66.81 | 72.65 | 68.53 | 73.13 | 68.97 | 76.03 | 67.91 | 73.38 | 61.20 | 66.61 | 70.63 | 68.36 | 58.74 | 70.33 | 63.94 | 78.02 | 63.63 | 70.83 |
| Average | 68.43 | 69.28 | 67.63 | 70.93 | 71.08 | 71.49 | 71.04 | 73.59 | 69.55 | 71.32 | 61.87 | 63.61 | 66.39 | 62.89 | 62.51 | 62.90 | 63.93 | 70.57 | 63.67 | 64.99 |
| -Fine-tuning-h | ased OO | D metho | ds | | | | | | | | | | | | | | | | - | |
| La aitManna | 1 71 22 | 60 70 | 60.00 | 60 70 | 72.11 | 70.64 | 72.50 | 72 66 | 1 71 05 | 70.02 | 61.40 | 62.00 1 | 66 52 | 65.96 | 61.25 | 62.25 | 62 62 | 70.52 | 62.20 | 65 42 |
| LogiuNomi | 71.22 | 69.70 | 00.20 | 09.70 | 72.11 | 70.04 | 72.39 | 75.00 | 71.05 | 70.95 | 01.49 | 05.09 | 00.52 | 05.00 | 01.25 | 62.23 | 05.55 | 70.55 | 05.20 | 05.45 |
| T2FNorm | /1.5/ | 69.99 | 68.03 | /0.3/ | /1.45 | 70.62 | 72.50 | 73.64 | 70.84 | /1.16 | 62.65 | 63.16 | 66.47 | 64.45 | 62.00 | 63.21 | 64.16 | /1.34 | 63.82 | 65.54 |
| AUGMIX | 71.47 | 69.67 | 68.12 | 68.92 | 72.15 | 70.58 | 72.74 | 73.32 | 71.12 | 70.62 | 62.79 | 60.99 | 67.43 | 63.05 | 61.96 | 64.72 | 65.97 | 71.09 | 64.54 | 64.96 |
| REGMIX | 71.99 | 70.38 | 68.08 | 69.44 | 72.94 | 70.69 | 73.56 | 73.90 | 71.64 | 71.10 | 66.05 | 59.70 | 68.13 | 62.45 | 64.26 | 60.36 | 67.13 | 69.27 | 66.39 | 62.94 |
| VOS | 71.96 | 71.33 | 61.91 | 70.48 | 62.23 | 68.98 | 72.18 | 72.96 | 67.07 | 70.94 | 62.17 | 59.65 | 67.03 | 63.16 | 61.48 | 63.28 | 65.24 | 69.71 | 63.98 | 63.95 |
| BFR (Ours) | 72 54 | 72 85 | 69.87 | 72.28 | 68 21 | 74.60 | 73.60 | 74 40 | 71.06 | 73 55 | 64.04 | 62.20 | 68.96 | 65.25 | 64.00 | 64.30 | 67.16 | 73 70 | 66.04 | 66 38 |
| Automatics) | 71.60 | 70.21 | 66.00 | 60.78 | 70.12 | 79.00 | 73.00 | 72.50 | 70.24 | 70.05 | 62.02 | 61.22 | 67.12 | 62.70 | 62.10 | 62.76 | 65 22 | 70.20 | 64.20 | 64.50 |
| Average | /1.00 | /0.21 | 00.88 | 09.78 | /0.18 | /0.50 | 12.11 | /3.30 | /0.54 | /0.95 | 03.03 | 01.32 | 07.12 | 03.79 | 02.19 | 02.70 | 05.21 | /0.39 | 04.39 | 04.30 |
| Average (All) | 69.56 | 69.61 | 67.36 | 70.52 | 70.76 | 71.06 | 71.64 | 73.56 | 69.83 | 71.19 | 62.28 | 62.79 | 66.65 | 63.21 | 62.40 | 62.85 | 64.39 | 70.51 | 63.93 | 64.84 |
| | iCa | RL | Bi | iC | W | /A | FOS | TER | Aver | age | iCa | RL | Bi | С | W | Ά | FOS | TER | Ave | rage |
| AP↑ | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far |
| Average CH - | 1 | | | | | | 1.001 | | | | | | | | | | | | | |
| Average CIL a | ccuracy | 20 | | 07 | 61 | 44 | | £1 (| 1 60 | 76 | | 04 1 | 42 | ne 1 | 15 | 00 | | 06 | 1 42 | 77 |
| | 1 38 | .20 | | .0/ | 01. | .44 | 0.5 | .31 | 59. | /0 | 40. | 00 | 42. | 20 | 45. | 77 | 45. | .90 | 4.5 | .11 |
| -Post-hoc-base | d OOD 1 | nethods | | | | | | | | | | | | | | | | | | |
| MSP | 62.73 | 77.05 | 62.57 | 76.91 | 65.46 | 77.77 | 65.08 | 78.70 | 63.96 | 77.61 | 22.90 | 20.01 | 24.27 | 16.79 | 22.30 | 18.64 | 23.44 | 23.64 | 23.23 | 19.77 |
| ODIN | 66.72 | 76.69 | 64.51 | 77.01 | 68.34 | 77.22 | 67.73 | 78.80 | 66.83 | 77.43 | 24.11 | 26.47 | 26.44 | 20.77 | 22.84 | 21.73 | 24.24 | 31.43 | 24.41 | 25.10 |
| Energy | 66.01 | 76.88 | 63 27 | 76.01 | 68.42 | 77 48 | 68.44 | 70 37 | 66.76 | 77 44 | 22.40 | 21.21 | 25.73 | 17.03 | 21.63 | 19.61 | 22.01 | 27.15 | 23.17 | 21.48 |
| MaxL ogit | 66 70 | 76.04 | 62 72 | 76.49 | 68 24 | 77.60 | 69.29 | 70.46 | 66.91 | 77.64 | 22.40 | 21.21 | 25.75 | 17.72 | 22.10 | 10.25 | 22.91 | 26.19 | 22.24 | 21.40 |
| MaxLogit | 00.79 | 70.94 | 05.75 | 70.46 | 00.54 | 77.09 | 00.50 | 79.40 | 00.81 | 77.04 | 22.70 | 21.10 | 23.32 | 17.72 | 22.10 | 19.20 | 23.23 | 20.16 | 23.54 | 21.08 |
| GEN | 66.82 | 77.17 | 64.69 | 77.14 | 68.47 | //./6 | 68.63 | 79.70 | 67.15 | 77.94 | 23.10 | 21.71 | 23.11 | 18.57 | 24.54 | 18.68 | 23.54 | 24.96 | 23.74 | 20.98 |
| ReAct | 65.35 | 78.84 | 61.33 | 77.20 | 67.66 | 79.19 | 66.75 | 80.90 | 65.27 | 79.03 | 21.14 | 15.51 | 23.59 | 15.14 | 23.24 | 16.84 | 21.55 | 17.84 | 22.38 | 16.33 |
| KLM | 62.17 | 76.80 | 62.02 | 78.65 | 62.92 | 78.44 | 64.09 | 79.07 | 62.80 | 78.24 | 24.13 | 20.25 | 23.96 | 19.58 | 23.91 | 21.54 | 26.15 | 23.51 | 24.54 | 21.22 |
| Relation | 61.13 | 78.48 | 61.63 | 80.21 | 64.64 | 82.66 | 65.50 | 84.27 | 63.23 | 81.41 | 25.75 | 20.03 | 26.93 | 21.43 | 22.16 | 19.27 | 22.46 | 25.88 | 24.33 | 21.65 |
| NNGuide | 66.45 | 80.63 | 62.47 | 78 56 | 67.86 | 81.45 | 66.92 | 78.62 | 65.03 | 70.81 | 24.47 | 18 74 | 21.35 | 23.34 | 23.08 | 25.34 | 25.62 | 30.94 | 23.86 | 24.59 |
| Intoduce | 65.01 | 77.70 | (2.01 | 70.50 | 66.00 | 70.05 | 66.94 | 70.02 | 65.75 | 79.50 | 24.47 | 20.57 | 21.55 | 10.02 | 23.70 | 20.10 | 23.02 | 25.72 | 23.00 | 24.57 |
| Average | 05.01 | 11.12 | 62.91 | 11.57 | 00.90 | /8.85 | 00.84 | /9.88 | 05.41 | /8.50 | 25.41 | 20.57 | 24.00 | 19.03 | 22.97 | 20.10 | 23.08 | 25.75 | 23.00 | 21.30 |
| Fine-tuning-ba | ased OO | D method | ls | | | | | | | | | | | | | | | | | |
| LogitNorm | 66.92 | 76.86 | 63.39 | 76.09 | 67.54 | 77.16 | 68.07 | 79.06 | 66.48 | 77.29 | 21.68 | 20.97 | 25.67 | 19.59 | 21.05 | 19.16 | 22.21 | 23.25 | 22.65 | 20.74 |
| T2FNorm | 67.03 | 77.10 | 63.24 | 76.85 | 66.88 | 77.44 | 67.93 | 79.19 | 66.27 | 77.64 | 22.44 | 21.16 | 24.99 | 18.61 | 21.71 | 19.77 | 22.82 | 24.41 | 22.99 | 20.99 |
| AUGMIX | 67.12 | 76.84 | 63.17 | 75.63 | 67.52 | 77.13 | 68.22 | 78 88 | 66 51 | 77 12 | 24.63 | 19.31 | 23.96 | 20.14 | 23 70 | 18.54 | 24.83 | 22.17 | 24 30 | 20.04 |
| DECMIX | 67.65 | 77.00 | 62.17 | 75.00 | 60.34 | 77.00 | 60.22 | 70.00 | 67.00 | 77 42 | 25.00 | 17.51 | 23.90 | 17.62 | 24.26 | 17.61 | 25.02 | 20.54 | 25.00 | 10.24 |
| NOC | 07.05 | 77.50 | 50.00 | 13.92 | 00.31 | 77.10 | 09.44 | 79.52 | 07.09 | 11.43 | 23.96 | 1/.01 | 41.39 | 17.02 | 24.30 | 1/.01 | 23.92 | 20.54 | 23.90 | 10.54 |
| VOS | 67.50 | //.56 | 58.66 | 11.45 | 59.23 | //.19 | 07.59 | /8.69 | 03.24 | 11.12 | 23.13 | 19.60 | 26.89 | 20.23 | 23.48 | 19.63 | 24.86 | 24.53 | 24.74 | 21.00 |
| BER (Ours) | 67.81 | 78.63 | 64 43 | 79 57 | 69 56 | 81.00 | 68.89 | 79.69 | 67.67 | 79.72 | 26.86 | 33.65 | | 72.05 | | | | | | 22.74 |
| DER (Ours) | | | 01110 | 1,2101 | 07.50 | | | | | | 20.00 | 22.05 | 27.04 | 22.05 | 26.54 | 22.23 | 27.46 | 24.02 | 26.98 | 22.74 |
| Average | 67.24 | 77.12 | 62.33 | 76.39 | 65.90 | 77.23 | 68.21 | 79.03 | 65.92 | 77.44 | 23.69 | 19.73 | 27.04 25.82 | 19.24 | 26.54 22.88 | 22.23 18.94 | 27.46 24.13 | 24.02 22.98 | 26.98 24.13 | 20.22 |
| Average Average (All) | 67.24 65.81 | 77.12 | 62.33 62.70 | 76.39 | 65.90 66.54 | 77.23 78.27 | 68.21 67.33 | 79.03 79.58 | 65.92 65.59 | 77.44 78.12 | 23.69 23.51 | 19.73 20.27 | 27.04 25.82 25.04 | 19.24 19.11 | 26.54 22.88 22.94 | 22.23 18.94 19.69 | 27.46 24.13 23.84 | 24.02 22.98 24.75 | 26.98 24.13 23.83 | 20.22 |
| Average Average (All) | 67.24 65.81 | 77.12 77.51 | 62.33 62.70 | 76.39 77.15 | 65.90 66.54 | 77.23 78.27 | 68.21 67.33 | 79.03 79.58 | 65.92 65.59 | 77.44 78.12 | 23.69 23.51 | 22.65 19.73 20.27 | 27.04 25.82 25.04 | 19.24 19.11 | 26.54 22.88 22.94 | 22.23 18.94 19.69 | 27.46 24.13 23.84 | 24.02 22.98 24.75 | 26.98 24.13 23.83 | 20.22 20.95 |
| Average Average (All) | 67.24 65.81 iCa | 77.12 77.51 RL | 62.33 62.70 Bi | 76.39 77.15 | 65.90 66.54 W | 77.23 78.27 /A | 68.21 67.33 FOS | 79.03 79.58 TER | 65.92 65.59 Aver | 77.44 78.12 | 23.69 23.51 iCa | 22.65 19.73 20.27 RL | 27.04 25.82 25.04 Bi | 19.24 19.11 C | 26.54 22.88 22.94 W | 22.23 18.94 19.69 A | 27.46 24.13 23.84 FOS | 24.02 22.98 24.75 TER | 26.98 24.13 23.83 Ave | 20.22 20.95 rage |
| Average Average (All) | 67.24 65.81 iCa Near | 77.12 77.51 IRL Far | 62.33 62.70 Bi Near | 76.39 77.15 IC Far | 65.90 66.54 Wear | 77.23 78.27 /A Far | 68.21 67.33 FOS Near | 79.03 79.58 TER Far | 65.92 65.59 Aver Near | 77.44 78.12 Far | 23.69 23.51 iCa Near | 22.05 19.73 20.27 RL Far | 27.04 25.82 25.04 Bi Near | 19.24 19.11 C Far | 26.54 22.88 22.94 W Near | 22.23 18.94 19.69 A Far | 27.46 24.13 23.84 FOS Near | 24.02 22.98 24.75 TER Far | 26.98 24.13 23.83 Ave Near | 22.74 20.22 20.95 rage Far |
| Average (All) | 67.24 65.81 iCa Near <i>ccuracy</i> | 77.12 77.51 IRL Far | 62.33 62.70 Bi Near | 76.39 77.15 IC Far | 65.90 66.54 Wear | 77.23 78.27 /A Far | 68.21 67.33 FOS Near | 79.03 79.58 TER Far | 65.92 65.59 Aver Near | 77.44 78.12 rage Far | 23.69 23.51 iCa Near | 22.65 19.73 20.27 RL Far | 27.04 25.82 25.04 Bi Near | 19.24 19.11 C Far | 26.54 22.88 22.94 W. Near | 22.23 18.94 19.69 A Far | 27.46 24.13 23.84 FOS Near | 24.02 22.98 24.75 TER Far | 26.98 24.13 23.83 Ave Near | 20.22 20.95 rage Far |
| Average (All) FPR Average CIL ad | 67.24 65.81 iCa Near ccuracy 58 | 77.12 77.51 IRL Far .20 | 62.33 62.70 Bi Near 55. | 76.39 77.15 iC Far .87 | 65.90 66.54 W Near | 77.23 78.27 /A Far .44 | 68.21 67.33 FOS Near 63 | 79.03 79.58 TER Far .51 | 65.92 65.59 Aver Near 59. | 77.44 78.12 *age Far 76 | 23.69 23.51 iCa Near 40. | 22.65 19.73 20.27 RL Far 86 | 27.04 25.82 25.04 Bi Near 42. | 22.05 19.24 19.11 C Far 26 | 26.54 22.88 22.94 W Near 45. | 22.23 18.94 19.69 A Far 99 | 27.46 24.13 23.84 FOS Near 45. | 24.02 22.98 24.75 TER Far .96 | 26.98 24.13 23.83 Ave: Near 43 | 22.74 20.22 20.95 rage Far .77 |
| Average (All) FPR Average CIL ac | 67.24 65.81 Near ccuracy 58 | 77.12 77.51 IRL Far .20 methods | 62.33 62.70 Bi Near 55. | 76.39 77.15 IC Far 87 | 65.90 66.54 Wear 61. | 77.23 78.27 /A Far .44 | 68.21 67.33 FOS Near 63 | 79.03 79.58 TER Far .51 | 65.92 65.59 Aver Near 59. | 77.44 78.12 rage Far 76 | 23.69 23.51 iCa Near 40. | 22.65 19.73 20.27 RL Far 86 | 27.04 25.82 25.04 Bi Near 42. | 22.03 19.24 19.11 C Far 26 | 26.54 22.88 22.94 W. Near 45. | 22.23 18.94 19.69 A Far 99 | 27.46 24.13 23.84 FOS Near 45. | 24.02 22.98 24.75 TER Far .96 | 26.98 24.13 23.83 Ave: Near 43 | 22.74 20.22 20.95 rage Far .77 |
| Average (All) FPR↓ Average CIL ad -Post-hoc-base MSP | 67.24 65.81 iCa Near ccuracy 58 d OOD n | 77.12 77.51 RL Far 20 methods 88.12 | 62.33 62.70 Bi Near 55. | 76.39 77.15 iC Far 87 86.72 | 65.90 66.54 W Near 61. | 77.23 78.27 /A Far .44 | 68.21 67.33 FOS Near 63 | 79.03 79.58 TER Far .51 | 65.92 65.59 Aver Near 59. | 77.44 78.12 *age Far 76 | 23.69 23.51 iCa Near 40. | 22.65 19.73 20.27 RL Far 86 | 27.04 25.82 25.04 Bi Near 42. | 22.03 19.24 19.11 C Far 26 | 26.54 22.88 22.94 W Near 45. | 22.23 18.94 19.69 A Far 99 93.14 | 27.46 24.13 23.84 FOS Near 45. | 24.02 22.98 24.75 TER Far .96 | 26.98 24.13 23.83 Ave: Near 43 91.72 | 22.74 20.22 20.95 rage Far .77 |
| Average Average (All) FPR↓ Average CIL au -Post-hoc-base MSP | 67.24 65.81 iCa Near ccuracy 58 60 OOD n 87.34 | 77.12 77.51 RL Far 20 methods 88.12 77.57 | 62.33 62.70 Bi Near 55. 88.23 | 76.39 77.15 IC Far 87 86.72 | 65.90 66.54 W Near 61. 85.16 | 77.23 78.27 /A Far .44 86.59 78.60 | 68.21 67.33 FOS Near 63 86.10 | 79.03 79.58 TER Far .51 85.47 | 65.92 65.59 Aver Near 59. 86.71 | 77.44 78.12 *age Far 76 86.72 | 23.69 23.51 iCa Near 40. | 22.65 19.73 20.27 RL Far 86 90.83 84.01 | 27.04 25.82 25.04 Bi Near 42. 91.25 | 22.03 19.24 19.11 C Far 26 94.89 92.02 | 26.54 22.88 22.94 W. Near 45. 92.70 | 22.23 18.94 19.69 A Far 99 93.14 80.01 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.02 | 24.02 22.98 24.75 TER Far .96 89.02 | 26.98 24.13 23.83 Ave: Near 43 91.72 00.70 | 22.74 20.22 20.95 rage Far .77 91.97 97.16 |
| Average Average (All) FPR↓ Average CIL ad -Post-hoc-base MSP ODIN | 67.24 65.81 Near ccuracy 58 6 OOD 1 87.34 83.67 | 77.12 77.51 IRL Far .20 methods 88.12 77.57 | 62.33 62.70 Bi Near 55. 88.23 85.70 88.23 | 76.39 77.15 C Far 87 86.72 74.96 | 65.90 66.54 W Near 61. 85.16 82.91 | 77.23 78.27 /A Far .44 86.59 78.69 | 68.21 67.33 FOS Near 63 86.10 83.14 | 79.03 79.58 TER Far .51 85.47 71.65 | 65.92 65.59 Aver Near 59. 86.71 83.86 | 77.44 78.12 rage Far 76 86.72 75.72 | 23.69 23.51 iCa Near 40. 91.12 90.89 | 22.65 19.73 20.27 RL Far 86 90.83 84.91 90.55 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 | 22.03 19.24 19.11 C Far 26 94.89 92.93 95.55 | 26.54 22.88 22.94 W Near 45. 92.70 91.66 | 22.23 18.94 19.69 A Far 99 93.14 89.91 22.23 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 | 26.98 24.13 23.83 Ave: Near 43 91.72 90.70 92.75 | 22.74 20.22 20.95 rage Far .77 91.97 87.16 |
| Average Average (All) FPR↓ Average CIL ad -Post-hoc-base MSP ODIN Energy | 67.24 65.81 Near ccuracy 58 6 OOD n 87.34 83.67 83.78 | 77.12 77.51 IRL Far .20 methods 88.12 77.57 79.86 | 62.33 62.70 Bi Near 55. 88.23 85.70 87.19 | 76.39 77.15 C Far 87 86.72 74.96 80.57 | 65.90 66.54 W Near 61. 85.16 82.91 83.40 | 77.23 78.27 /A Far .44 86.59 78.69 82.10 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 | 65.92 65.59 Near 59. 86.71 83.86 84.27 | 77.44 78.12 rage Far 76 86.72 75.72 78.92 | 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 | 22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 | 22.03 19.24 19.11 C Far 26 94.89 92.93 95.36 | 26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 | 26.98 24.13 23.83 Near 43 91.72 90.70 92.17 | 22.74 20.22 20.95 rage Far .77 91.97 87.16 91.76 |
| Average Average (All) FPR↓ Average CIL au -Post-hoc-base MSP ODIN Energy MaxLogit | 67.24 65.81 Near ccuracy 58 6 OOD 1 87.34 83.67 83.78 83.91 | 77.12 77.51 IRL Far .20 methods 88.12 77.57 79.86 80.85 | 62.33 62.70 Bi Near 55. 88.23 85.70 87.19 86.80 | 76.39 77.15 IC Far 87 86.72 74.96 80.57 81.57 | 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 | 77.23 78.27 /A Far .44 86.59 78.69 82.10 82.90 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 74.54 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 | 77.44 78.12 *age Far 76 86.72 75.72 78.92 79.97 | 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 | 22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 | 22.03 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 | 26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 | 26.98 24.13 23.83 Near 43 91.72 90.70 92.17 92.20 | 22.74 20.22 20.95 rage Far .77 91.97 87.16 91.76 91.80 |
| Average Average (All) FPR↓ Average CIL ad -Post-hoc-base MSP ODIN Energy MaxLogit GEN | 67.24 iCa Near ccuracy 58 57.34 87.34 83.67 83.78 83.91 83.84 | 77.12 77.51 IRL Far .20 methods 88.12 77.57 79.86 80.85 81.19 | 62.33 62.70 Bi Near 55. 88.23 85.70 87.19 86.80 85.88 | 76.39 77.15 C Far 87 86.72 74.96 80.57 81.57 76.37 | 65.90 66.54 Wear 61. 85.16 82.91 83.40 83.07 82.77 | 77.23 78.27 74 Far .44 86.59 78.69 82.10 82.90 81.82 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 74.54 74.14 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 | 77.44 78.12 78.12 Far 76 86.72 75.72 78.92 79.97 78.38 | 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 | 22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.94 90.50 10.50 <th10.50< th=""> <th10.50< th=""> <th10.50< td="" th1<=""><td>27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30</td><td>22.03 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37</td><td>26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40</td><td>22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52</td><td>27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 92.00</td><td>24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 89.90</td><td>26.98 24.13 23.83 Near 43 91.72 90.70 92.17 92.20 91.71</td><td>22.74 20.22 20.95 rage Far .77 91.97 87.16 91.76 91.80 92.32</td></th10.50<></th10.50<></th10.50<> | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 | 22.03 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 | 26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 92.00 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 89.90 | 26.98 24.13 23.83 Near 43 91.72 90.70 92.17 92.20 91.71 | 22.74 20.22 20.95 rage Far .77 91.97 87.16 91.76 91.80 92.32 |
| Average Average (All) FPR↓ Average CIL ad -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct | 67.24 65.81 iCa Near ccuracy 58 d OOD n 87.34 83.67 83.78 83.91 83.84 83.84 83.84 | 77.12 77.51 RL Far 20 methods 88.12 77.57 79.86 80.85 81.19 79.97 | 62.33 62.70 Bi Near 555. 88.23 85.70 87.19 86.80 85.88 88.95 | 76.39 77.15 77.15 77.15 76.39 77.15 76.39 80.57 74.96 80.57 76.37 82.85 | 65.90 66.54 W Near 61. 85.16 85.16 83.00 83.07 82.77 83.64 | 77.23 78.27 78.27 74 Far 44 86.59 78.69 82.10 82.90 81.82 81.41 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 74.54 74.14 73.64 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 | 77.44 78.12 78.12 Far 76 86.72 75.72 78.92 79.97 78.38 79.47 | 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 | 22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.98 90.98 90.98 90.950 90.50 94.57 100.000 <td>27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42</td> <td>22.05 19.24 19.11 C Far 26 94.89 92.93 95.36 95.36 95.37 95.62</td> <td>26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 92.18</td> <td>22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77</td> <td>27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.54 92.01 92.00 93.90</td> <td>24.02 22.98 24.75 TER Far 96 89.02 80.90 87.43 88.00 89.90 94.98</td> <td>26.98 24.13 23.83 Near 43 91.72 90.70 92.17 92.20 91.71 92.62</td> <td>22.74 20.22 20.95 rage Far 77 91.97 87.16 91.80 92.32 95.23</td> | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 | 22.05 19.24 19.11 C Far 26 94.89 92.93 95.36 95.36 95.37 95.62 | 26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.54 92.01 92.00 93.90 | 24.02 22.98 24.75 TER Far 96 89.02 80.90 87.43 88.00 89.90 94.98 | 26.98 24.13 23.83 Near 43 91.72 90.70 92.17 92.20 91.71 92.62 | 22.74 20.22 20.95 rage Far 77 91.97 87.16 91.80 92.32 95.23 |
| Average (All) Average (All) FPR↓ Average CIL a -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct VIM | 67.24 65.81 iCa Near ccuracy 58 d OOD n 87.34 83.67 83.78 83.91 83.84 83.83 87.04 | 77.12 77.51 RL Far 20 methods 88.12 77.57 79.86 80.85 81.19 79.97 80.37 | 62.33 62.70 Bi Near 55. 88.23 85.70 87.19 86.80 85.88 88.95 89.12 | 76.39 77.15 77.15 77.15 Far 87 86.72 74.96 80.57 81.57 76.37 82.85 86.00 | 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 83.07 83.07 83.77 83.77 | 77.23 78.27 7A Far 44 86.59 78.69 82.10 82.90 81.82 81.82 81.41 87.66 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 74.54 74.14 73.64 74.14 73.69 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.72 | 77.44 78.12 rage Far 76 86.72 75.72 78.92 79.97 78.38 79.47 78.726 | 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 92.98 | 22.65 19.73 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.20 91.7 89.77 | 22.05 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.37 95.64 | 26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.58 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 92.00 93.90 89.30 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 89.90 94.98 87.25 | 26.98 24.13 23.83 Ave: Near 43 91.72 90.70 92.17 92.20 91.71 92.20 91.71 92.62 89.64 | 22.74 20.22 20.95 Far 77 91.97 87.16 91.76 91.80 92.32 95.23 88.67 |
| Average Average (All) FPR↓ Average CLL au -Post-hoc-base. MSP ODIN Energy MaxLogit GEN ReAct KLM | 67.24 65.81 iCa Near ccuracy 58 d OOD n 87.34 83.67 83.78 83.91 83.84 83.83 87.94 | 77.12 77.51 RL Far 20 nethods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 89.37 | 62.33 62.70 Bi Near 55. 88.23 85.70 87.19 86.80 85.88 88.95 88.13 89.66 | 76.39 77.15 77.15 77.15 77.15 77.15 76.37 80.57 80.57 81.57 76.37 82.85 86.00 | 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 | 77.23 78.27 /A Far .44 86.59 78.69 82.10 82.90 81.82 81.41 87.66 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 | 79.03 79.58 TER Far 5.1 85.47 71.65 73.14 74.54 74.54 74.14 73.64 86.00 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 | 77.44 78.12 rage Far 76 86.72 75.72 78.92 79.97 78.38 79.47 87.26 | 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 89.88 | 22.65 19.73 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 96.60 96.60 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 89.27 | 22.05 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.62 89.84 95.42 85.42 95.45 95.45 95.45 95.45 95.45 95.4 | 26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 87.48 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.54 92.41 92.00 93.90 88.30 | 24.02 22.98 24.75 TER Far 96 89.02 80.90 87.43 88.00 89.90 94.98 87.25 | 26.98 24.13 23.83 Near 43 91.72 90.70 92.17 92.20 91.71 92.62 89.64 89.64 | 22.74 20.22 20.95 rage Far 77 91.97 87.16 91.76 91.80 92.32 95.23 88.67 20.95 |
| Average Average (All) FPR↓ -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation | 67.24 65.81 iCa Near ccuracy 58 d OOD n 87.34 83.67 83.78 83.91 83.84 83.83 83.94 83.84 83.83 | 77.12 77.51 RL Far 20 methods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 | 62.33 62.70 Bi Near 55. 88.23 85.70 87.19 86.80 85.88 88.95 88.13 88.48 | 76.39 77.15 C Far 87 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.88 | 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 83.64 | 77.23 78.27 /A Far .44 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 | 68.21 67.33 FOS Near 63 86.10 83.14 82.94 82.96 83.63 87.06 84.10 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 74.54 74.54 73.64 86.00 71.57 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 | 77.44 78.12 'age Far 76 86.72 75.72 78.92 79.97 78.38 79.47 87.26 74.35 | 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 | 22.65 19.73 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.98 90.50 94.57 90.20 86.00 1000000000000000000000000000000000000 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 | 22.05 19.24 19.11 19.11 C Far 26 94.89 92.93 95.36 95.36 95.37 95.62 89.84 95.19 95.19 | 26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.54 92.41 92.00 93.90 88.30 92.56 | 24.02 22.98 24.75 TER 96 89.02 80.90 87.43 88.00 89.90 87.43 88.00 94.98 87.25 87.38 | 26.98 24.13 23.83 Near 43 91.72 90.70 92.17 92.20 91.71 92.20 91.71 92.62 89.64 93.18 | 22.74 20.22 20.95 rage Far 77 91.97 87.16 91.76 91.76 91.76 92.32 95.23 88.67 90.94 |
| Average Average (All) FPR↓ Average CIL au -Post-hoc-base. MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide | 67.24 65.81 iCa Near ccuracy 58 d OOD n 87.34 83.67 83.78 83.91 83.84 83.83 87.94 85.39 85.66 | 77.12 77.51 RL Far 20 nethods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 | 88.23 88.23 85.70 87.19 86.80 85.88 88.95 88.13 88.48 87.97 87.97 86.80 88.97 88.13 88.48 87.97 <th< td=""><td>76.39 77.15 IC Far 87 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.88 75.47</td><td>65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 84.44 86.10</td><td>77.23 78.27 /A Far .44 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 73.56</td><td>68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 84.10 87.47</td><td>79.03 79.58 TER Far 51 85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20</td><td>65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 86.80</td><td>77.44 78.12 Far 76 86.72 75.72 78.92 79.97 78.38 79.47 87.26 74.35 74.13</td><td>23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 90.00</td><td>22.65 19.73 20.27 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 86.00 88.40 100.00</td><td>27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.00 92.30 91.77 92.85 84.14</td><td>22.05 19.24 19.11 19.11 C Far 26 92.93 95.36 95.30 95.37 95.62 89.84 95.19 89.16 89.16</td><td>26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30</td><td>22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30</td><td>27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.92 92.54 92.00 93.90 88.30 92.56 89.94</td><td>24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 89.90 94.98 87.25 87.38 85.73</td><td>26.98 24.13 23.83 Near 43 91.72 90.70 92.17 92.20 91.71 92.20 91.71 92.62 89.64 93.18 88.59</td><td>22.74 20.22 20.95 rage Far 77 91.97 87.16 91.80 92.32 95.23 88.67 90.94 88.40</td></th<> | 76.39 77.15 IC Far 87 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.88 75.47 | 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 84.44 86.10 | 77.23 78.27 /A Far .44 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 73.56 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 84.10 87.47 | 79.03 79.58 TER Far 51 85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 86.80 | 77.44 78.12 Far 76 86.72 75.72 78.92 79.97 78.38 79.47 87.26 74.35 74.13 | 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 90.00 | 22.65 19.73 20.27 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 86.00 88.40 100.00 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.00 92.30 91.77 92.85 84.14 | 22.05 19.24 19.11 19.11 C Far 26 92.93 95.36 95.30 95.37 95.62 89.84 95.19 89.16 89.16 | 26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.92 92.54 92.00 93.90 88.30 92.56 89.94 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 89.90 94.98 87.25 87.38 85.73 | 26.98 24.13 23.83 Near 43 91.72 90.70 92.17 92.20 91.71 92.20 91.71 92.62 89.64 93.18 88.59 | 22.74 20.22 20.95 rage Far 77 91.97 87.16 91.80 92.32 95.23 88.67 90.94 88.40 |
| Average Average (All) FPR↓ Average CIL at -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide | 67.24 65.81 iCa Near ccuracy 58 d OOD n 87.34 83.67 83.78 83.91 83.84 83.83 83.94 83.84 83.83 87.94 85.39 85.66 85.04 | 77.12 77.51 RL Far 20 methods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 80.75 | 62.33 62.70 Bi Near 55. 88.23 85.70 87.19 86.80 85.88 88.95 88.13 88.48 87.97 87.48 | 76.39 77.15 IC Far 87 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.82 86.00 77.84 75.47 80.27 | 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 84.44 86.10 84.36 | 77.23 78.27 78.27 78 44 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 75.54 75.54 81.13 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 84.10 87.47 84.41 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20 70.20 75.59 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 85.60 85.32 | 77.44 78.12 Far 76 86.72 75.72 78.92 79.97 78.38 79.47 87.26 74.35 79.44 | 23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 90.78 90.78 91.54 | 22.65 19.73 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 86.00 88.40 88.40 89.70 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 91.42 89.77 92.85 84.14 90.29 | 22.05 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.62 89.84 95.19 89.16 93.74 | 26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 90.30 92.38 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 92.04 93.90 93.90 88.30 92.56 89.94 91.60 | 24.02 22.98 24.75 TER 96 89.02 80.90 87.43 88.00 89.90 94.98 87.25 87.38 85.73 | 26.98 24.13 23.83 Near 43 91.72 90.70 92.17 92.20 91.71 92.62 89.64 93.18 88.59 91.39 | 22.74 20.25 Fage Far 77 91.97 87.16 91.76 91.80 92.32 95.23 88.67 90.94 88.60 90.91 |
| Average Average (All) FPR↓ Average CILaa -Post-hoc-base. MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average | 67.24 65.81 iCa Near ccuracy 58 d OOD 87.34 83.67 83.78 83.91 83.84 83.83 87.94 85.39 85.66 85.06 | 77.12 77.51 RL Far 20 methods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 80.75 74.39 75.41 | 8 23 62.33 62.70 Bi 82.70 S55. 55. 88.23 85.70 87.19 86.80 85.88 88.95 88.13 88.48 87.97 87.48 | 76.39 76.39 77.15 IC Far 87 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.88 75.47 80.27 | 65.90 66.54 Wear 61. 85.16 82.91 83.40 83.07 83.07 83.64 83.07 83.64 84.44 86.10 84.36 | 77.23 78.27 78.27 78 44 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 75.44 81.13 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.94 83.63 87.06 84.10 87.47 84.41 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 74.54 74.14 74.54 74.14 86.00 71.57 70.20 75.59 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 86.80 85.32 | 77.44 78.12 age Far 76 86.72 75.72 78.92 79.97 78.98 79.47 87.26 74.35 74.13 79.44 | 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 90.00 91.54 | 22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 86.00 88.40 89.70 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 84.14 90.29 | 22.05 19.24 19.11 I C Far 26 94.89 92.93 95.36 95.30 95.37 95.62 89.84 95.19 89.16 93.74 10.11 | 26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.54 92.54 92.00 93.90 88.30 92.56 89.94 91.60 | 24.02 22.98 24.75 TER Far 96 89.02 80.90 87.43 88.00 89.90 87.43 88.00 89.90 87.43 88.725 87.38 87.25 | 26.98 24.13 23.83 Near 91.72 90.70 92.17 92.20 91.71 92.20 91.71 92.20 91.71 92.62 89.64 93.18 88.59 91.39 | 22.74 20.22 20.95 Far Far 91.97 87.16 91.76 91.80 92.32 95.23 88.67 90.94 88.40 90.91 |
| Average Average (All) FPR↓ Average CIL au -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average Fine-tuning-ba | 67.24 65.81 Near ccuracy 87.34 83.67 83.78 83.91 83.84 83.91 83.84 83.91 83.84 83.91 83.84 83.91 83.84 83.91 83.66 85.04 85.04 85.04 | 77.12 77.51 RL 20 methods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 80.75 07 methods 70.96 70.96 70.96 70.96 70.96 70.96 70.96 70.96 70.96 70.96 70.97 70 70.97 70 70.97 70 70 70 70 70 70 70 70 70 70 70 70 70 | 88.23 85.70 88.23 85.70 87.19 86.80 85.88 88.95 88.13 88.48 87.97 87.48 | 76.39 77.15 C Far 87 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.88 85.47 80.27 | 65.90 66.54 Wear 61. 85.16 82.91 83.40 83.07 83.64 83.07 83.64 87.78 84.44 86.10 84.36 | 77.23 78.27 7A Far .44 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 73.56 75.44 81.13 | 68.21 67.33 FOS Near 63 86.10 83.14 82.94 82.94 82.56 83.63 87.06 84.10 87.47 84.41 | 79.03 79.58 TER Far 51 85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20 75.59 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 85.32 | 77.44 78.12 age Far 76 86.72 75.72 79.97 78.38 79.47 87.26 74.35 74.13 79.44 | 23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 90.00 91.54 | 22.65 19.73 19.73 20.27 RL - Far - 86 - 90.83 84.91 90.98 90.94 90.90 90.50 94.57 90.20 86.00 88.40 89.70 - | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.75 84.14 90.29 | 22.05 19.24 19.21 19.21 19.24 91.01 C - Far - 26 - 94.89 92.93 95.36 95.36 95.37 95.62 89.84 95.19 95.16 93.74 | 26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 | 22.23 18.94 19.69 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 92.38 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 92.00 93.90 88.30 92.56 89.94 91.60 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 89.90 94.98 87.25 87.38 85.73 87.84 | 26.98 24.13 23.83 23.83 Ave: Near 43 91.72 90.70 92.17 92.070 92.17 92.62 89.64 93.18 88.59 91.39 | 22.74 20.95 Fage Far 77 91.97 87.16 91.97 87.16 91.80 92.32 95.23 88.67 90.94 88.40 90.91 |
| Average Average (All) FPR↓ -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average Fline-tuning-bot LogitNorm | 67.24 65.81 iCa Near ccuracy 58 d OOD r 87.34 83.67 83.78 83.91 83.84 83.83 83.91 83.84 83.83 87.94 85.39 85.66 85.04 85 | 77.12 77.51 RL Far 20 methods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 80.75 D method 79.88 | 8 23 62.33 62.70 Bi 86.70 85.23 85.70 87.19 86.80 85.88 88.95 88.13 88.48 87.97 87.48 86.75 86.75 | 76.39 77.15 77.15 77.15 77.15 77.15 87 80.72 74.96 80.57 81.57 81.57 82.85 86.00 77.88 75.47 80.27 81.70 | 85.16 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 84.44 86.10 84.36 83.71 | 77.23 78.27 78.27 74 Far 44 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 75.44 81.13 82.11 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.66 83.63 87.06 84.10 87.47 84.41 83.01 | 79.03 79.58 TER Far .51 .51 .51 .73.14 74.54 73.14 73.64 86.00 71.57 70.20 75.59 <u>73.66</u> | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 86.80 85.32 84.31 | 77.44 78.12 age Far 76 86.72 75.72 78.92 79.97 78.38 79.47 87.26 74.35 74.13 79.44 79.34 | 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.65 92.65 92.65 92.12 92.98 89.88 91.78 90.00 91.54 93.23 | 22.65 19.73 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 86.00 88.40 89.70 91.73 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 84.14 90.29 89.61 | 22.03 19.24 19.21 19.11 C - Far - 26 - 94.89 92.93 95.36 95.37 95.62 89.84 95.19 89.16 93.74 - 94.38 - | 26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 93.92 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 93.98 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.00 93.90 93.90 93.90 92.56 89.94 91.60 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 87.43 88.00 94.98 87.25 87.38 85.73 87.84 93.03 | 26.98 24.13 23.83 Near 91.72 90.70 92.17 92.20 91.71 92.20 91.71 92.20 91.71 92.20 91.71 92.20 91.71 92.54 | 22.74 20.22 20.95 Far Far 77 91.97 87.16 91.80 92.32 95.23 88.67 90.94 88.40 90.91 93.28 |
| Average (All) FPR↓ Average (ILa) FOR-t-hoc-base Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average Fine-tuning-ba LogitNorm T2FNorm | 67.24 65.81 iCa Near ccuracy 588 d OOD n 87.34 83.78 83.78 83.78 83.78 83.78 83.78 83.78 83.78 83.79 83.78 85.66 85.04 85.39 85.66 85.04 85.39 85.66 85.04 83.77 83.69 | 77.12 77.51 RL 20 nethods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 80.75 D method 79.88 80.39 | 8.2.33 62.33 62.70 Bi Bi 8.23 85.70 \$87.19 86.80 85.88 88.95 88.13 88.48 \$87.97 87.48 87.97 87.48 18 86.75 86.75 | 76.39 77.15 C 87 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.88 75.47 80.27 81.70 81.55 | 65.90 66.54 Wear 61 85.16 82.91 83.40 83.07 82.77 83.64 87.78 84.44 86.10 84.36 83.71 83.97 | 77.23 78.27 74 Far .44 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 73.56 73.56 73.54 81.13 82.11 82.91 | 68.21 67.33 FOS Near 63 86.10 83.14 82.94 82.94 82.96 83.63 87.06 84.10 87.47 84.41 83.01 82.95 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20 75.59 <u>73.66</u> 74.56 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 85.32 84.31 84.34 | 77.44 78.12 *age Far 76 86.72 75.72 78.92 78.92 79.97 78.38 79.47 87.26 74.35 74.13 79.44 79.34 79.34 | 23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 90.00 91.54 93.23 92.62 | 22.63 19.73 19.73 20.27 RL Far 86 90.83 84.91 90.94 90.50 94.57 90.50 86.00 88.40 89.70 91.73 91.12 | 27.04 25.82 25.04 Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.73 91.42 89.75 84.14 90.29 89.61 90.77 | 22.05 19.24 19.11 C Far 26 94.89 95.36 95.37 95.37 95.37 95.37 95.37 95.37 95.37 95.37 95.37 95.37 95.37 95.37 95.37 95.36 93.74 94.38 95.36 | 26.54 22.94 22.94 Wear 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 93.92 93.92 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 93.98 92.58 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 92.00 93.90 88.30 92.56 89.94 91.60 93.39 92.67 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 89.90 94.98 87.25 87.38 85.73 87.84 93.03 90.70 | 26.98 24.13 23.83 Aver Near 43 91.72 90.70 92.17 92.20 91.72 90.70 92.17 92.62 89.64 93.18 88.59 91.39 92.54 92.54 | 22.74 20.22 20.95 Far Far 77 91.97 87.16 91.76 91.76 91.76 91.80 92.32 95.23 88.67 90.94 88.40 90.91 93.28 92.44 |
| Average Average (All) FPR↓ -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average Fine-tuning-base LogitNorm T2FNorm | 67.24 65.81 iCa Near S8 d OOD r 87.34 83.67 83.78 83.91 83.84 83.83 87.94 83.84 83.83 87.94 85.66 85.04 ased OOI 83.77 83.69 83.69 | 77.12 77.51 RL Far 20 methods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 80.75 D method 79.88 80.39 79.98 | 88.23 62.33 62.70 Bi 88.23 85.70 87.19 \$55. 88.23 85.70 87.19 \$86.80 85.88 88.95 88.13 \$84.83 87.97 87.48 87.97 \$87.48 86.75 86.75 86.75 \$86.75 86.92 86.92 86.92 | 76.39 77.15 77.15 77.15 77.15 78.28 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.88 75.47 80.27 80.27 81.55 81.28 | 65.90 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 83.07 83.07 83.07 83.64 84.44 86.10 84.36 84.36 83.71 83.97 | 77.23 78.27 78.27 74 Far 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 75.44 81.13 81.11 82.51 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.63 87.06 84.10 87.47 84.41 82.95 82.55 | 79.03 79.58 TER Far .51 .85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20 75.59 73.66 74.20 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 85.32 84.31 84.34 84.21 | 77.44 78.12 *age Far 76 86.72 75.72 78.92 79.97 78.38 79.47 87.26 74.35 74.13 79.44 79.34 79.50 | 23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.92 92.98 92.65 92.46 92.12 92.98 91.78 90.00 91.54 93.23 92.62 90.16 | 22.63 19.73 20.27 RL Far 86 90.83 84.91 90.94 90.920 86.00 88.40 89.70 91.73 91.22 92.26 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 84.14 90.29 89.61 90.77 89.24 | 22.05 19.24 19.11 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.62 89.84 95.19 89.16 93.74 94.38 95.36 90.98 | 26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 93.92 92.97 90.73 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 93.98 92.58 88.90 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 92.41 92.41 92.41 92.56 89.94 91.60 93.90 93.90 93.90 93.39 92.67 90.02 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 89.90 87.43 88.00 94.98 87.25 87.38 85.73 87.84 93.03 90.70 88.31 | 26.98 24.13 23.83 Ave: Near 43 91.72 90.70 92.17 92.20 91.71 92.20 91.71 92.20 91.71 92.64 93.18 88.59 91.39 92.54 92.26 90.04 | 22.74 20.72 20.95 Far Far 77 91.97 87.16 91.97 87.16 91.80 92.32 95.23 88.67 90.94 88.40 90.91 90.91 93.28 92.44 90.11 |
| Average Average (All) FPR↓ -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average Fine-tuning-ba LogitNorm T2FNorm AUGMIX | 67.24 65.81 iCa Near ccuracy 58 d OOD r 87.34 83.91 83.84 83.83 83.94 83.84 83.83 85.04 85.04 ased OOI 83.77 83.69 83.69 83.69 83.69 | 77.12 77.51 RL Far 20 nethods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 80.75 D method 79.88 80.39 79.98 80.26 | 8.23 62.33 62.70 Bi Rear 85.70 85.82 \$55. \$8.23 \$85.70 \$87.19 \$86.80 \$85.88 \$8.95 \$81.13 \$88.48 \$8.95 \$81.13 \$84.48 \$87.97 \$87.48 \$85.92 \$87.06 | 76.39 77.15 C 87 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.88 75.47 80.27 81.28 81.70 81.55 81.28 81.94 | 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 83.07 82.77 83.64 84.34 84.34 84.36 83.71 83.97 83.71 83.97 83.71 83.31 | 77.23 78.27 78.27 78.27 78.69 78.69 82.10 82.90 81.82 81.41 87.66 73.56 73.56 73.56 73.56 73.54 81.13 82.11 82.91 82.91 82.55 84.80 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 84.10 87.47 84.41 83.01 82.95 82.93 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20 75.59 73.66 74.56 74.56 74.56 74.56 74.56 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 85.32 84.31 84.34 84.34 84.34 84.34 | 77.44 78.12 78.12 76 76 86.72 75.72 78.92 79.97 78.38 79.47 87.26 74.35 74.13 79.44 79.85 79.59 90.59 | 23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 89.88 91.74 90.00 91.54 93.23 92.62 90.16 89.56 | 22.65 19.73 19.73 90.83 86 90.83 84.91 90.94 90.950 94.57 90.20 86.00 88.40 89.70 91.73 91.12 92.26 94.48 | 27.04 25.02 25.02 15.04 10.05 10 | 22.05 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.36 95.37 95.62 89.84 89.84 89.84 95.19 89.16 93.74 94.38 95.36 90.98 94.54 | 26.54 22.88 22.94 W. Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 95.52 90.30 95.52 90.30 95.74 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 93.98 92.58 88.800 94.34 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 92.00 93.90 88.30 92.56 89.94 91.60 93.39 92.67 90.02 89.97 | 24.02 22.98 24.75 TER Far .96 89.02 80.90 87.43 88.00 94.98 87.25 87.38 85.73 87.84 93.03 90.70 88.31 94.73 | 26.98 24.13 23.83 Ave: Near 43 91.72 90.70 90.70 92.17 92.20 91.71 92.20 91.71 92.20 91.71 92.20 91.71 92.262 89.64 93.18 88.59 91.39 92.54 92.54 92.26 90.04 89.50 | 22.74 20.22 20.95 Far Far 77 91.97 87.16 91.76 91.76 91.76 91.76 91.76 91.76 91.76 92.32 95.23 88.67 90.94 88.40 90.91 93.28 92.44 90.21 94.52 |
| Average Average (All) FPR -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average Fine-tuning- TEPNorm AUGMIX REGMIX | 67.24 65.81 iCa Near ccuracy 58 d OOD 87.34 83.67 83.78 83.91 83.84 83.83 87.94 85.39 85.66 85.04 ascord OOI 83.77 83.69 83.69 83.69 83.69 83.69 83.67 83.69 83.69 83.69 83.67 83.69 8 | 77.12 77.51 RL Far 20 methods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 80.75 D methoo 79.88 80.39 79.98 80.26 D methoo | 88.23 62.33 62.70 Bi Near 55. 88.23 85.70 87.19 86.80 85.88 88.95 88.13 88.48 87.95 86.75 86.75 86.92 87.06 89.20 | 76.39 77.15 77.15 77.15 77.15 78.28 86.72 74.96 80.57 81.57 76.57 82.85 86.00 77.88 76.57 82.85 86.00 77.88 76.37 82.85 86.00 77.88 81.94 81.28 81.28 | 65.90 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 84.44 86.10 84.36 83.71 83.97 83.71 83.97 83.71 83.44 | 77.23 78.27 78.27 74 Far 4.4 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 75.44 81.13 82.11 82.91 82.55 84.80 77.70 77.70 77.70 77.70 77.70 78.70 77.70 77.70 78.70 78.70 78.70 77.70 77.70 78.70 77 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 84.10 87.41 83.01 82.55 82.53 81.93 83.55 | 79.03 79.58 TER Far .51 85.47 71.65 73.14 74.54 74.14 73.60 70.20 75.59 73.66 74.20 75.39 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 86.80 85.32 84.31 84.34 84.34 84.34 84.34 | 77.44 78.12 78.12 76 86.72 75.72 78.92 79.97 78.38 79.47 87.26 74.13 79.44 79.34 79.34 79.50 80.59 75.28 | 23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 90.00 91.54 93.23 92.62 90.16 89.42 | 22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.920 86.00 88.40 91.73 91.73 92.26 94.45 92.26 94.26 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 84.14 90.29 89.61 90.77 89.24 88.61 90.77 | 22.05 19.24 19.11 19.11 C Far 26 94.89 95.36 95.36 95.30 95.36 95.30 95.36 95.36 95.37 95.62 89.84 95.19 89.16 89.16 93.74 | 26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 93.92 92.97 90.73 90.73 90.74 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 93.98 92.58 88.90 94.34 88.46 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.54 92.54 92.54 92.54 92.56 89.94 91.60 91.60 93.39 92.67 90.02 89.97 89.67 | 24.02 22.98 24.75 7ER Far 96 89.02 80.90 87.43 88.00 89.90 94.98 87.25 87.38 87.34 87.84 87.33 87.84 87.84 87.84 87.83 87.84 93.03 90.70 88.31 94.73 91.10 | 26.98 24.13 23.83 Avei Near 43 91.72 90.70 92.17 92.20 91.71 92.20 91.71 92.62 89.64 93.18 88.59 91.39 91.38 92.26 90.04 89.254 92.54 92.26 90.04 89.59 | 22.74 20.22 20.95 Far Far 77 91.97 87.16 91.76 91.80 92.32 88.67 90.94 88.40 90.91 93.28 892.44 90.11 94.52 90.32 |
| Average Average (All) FPR4 Average CL au Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average Fine-tuning-ba LogitNorm AUGMIX NS BED (Journ) | 67.24 65.81 iCa Near ccuracy 58 d OOD r 87.34 83.67 83.78 83.91 83.84 83.83 87.94 85.39 85.66 85.06 85.06 83.77 83.69 83.69 83.67 83.67 83.67 83.69 83.67 83.67 83.67 83.69 83.67 83.67 83.69 83.67 83.67 83.69 83.67 83.67 83.69 83.67 83.67 83.69 83.67 83.67 83.67 83.77 83.69 83.67 83.67 83.77 83.69 83.67 83.67 83.77 83.69 83.67 83.67 83.77 83.69 83.67 83.67 83.67 83.77 83.69 83.67 83.67 83.77 83.69 83.67 83.67 83.67 83.77 83.69 83.67 83.67 83.67 83.67 83.77 83.69 83.67 83.67 83.77 83.69 83.67 83.67 83.67 83.77 83.69 83.67 83.67 83.77 83.69 83.67 83.67 83.69 83.67 83.67 83.69 83.67 83.67 83.69 83.67 83.67 83.69 83.67 83.67 83.67 83.69 83.67 83.67 83.67 83.69 83.67 83.67 83.67 83.69 83.67 83.67 83.67 83.69 83.67 83.67 83.67 83.69 83.67 83.67 83.67 83.69 83.67 83.78 83.77 83.78 83.78 83.78 83.78 83.78 83.7 | 77.12 77.51 RL Far 20 nethods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 80.75 74.39 75.41 80.75 00 method 79.88 80.39 79.98 80.26 77.91 75.00 | 81.00 81.00 <th< td=""><td>76.39 77.15 77.15 77.15 77.15 77.15 86.72 74.96 80.57 81.57 81.57 81.57 81.57 81.57 81.28 81.94 81.94 81.94 81.94 81.94 81.94 81.94 72.55</td><td>65.90 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 83.64 87.78 84.44 86.10 84.36 83.71 83.71 83.31 88.44 83.97</td><td>77.23 78.27 78.27 Far Far 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 73.56 73.56 75.44 81.13 82.11 82.91 82.91 82.91 82.59 84.80 77.72 76.27 76.27 76.27 77.27 77.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.29 78.69 82.10 82.90 82.90 82.90 82.90 82.90 82.90 82.90 82.90 82.90 82.90 82.90 82.91 78.56 73.56 73.56 73.56 73.56 73.56 73.56 75.44 82.91 82.91 82.91 76.27 77.77 77.777</td><td>68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 84.10 87.47 84.41 83.01 82.55 81.93 83.55 82.95</td><td>79.03 79.58 TER Far 51 85.47 71.65 73.14 74.54 74.54 73.14 74.54 73.14 73.64 86.00 71.57 70.20 75.59 73.66 74.56 74.56 74.56 74.56 74.58 73.84 73.84 73.84</td><td>65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 85.32 84.31 84.34 84.34 84.21 84.34 84.21 83.94 86.24 83.76</td><td>77.44 78.12 rage Far 76 86.72 75.72 79.97 79.97 78.38 79.47 87.26 74.35 74.13 79.44 79.34 79.34 79.50 80.59 75.59 74.09 75.92 79.50 80.59 74.09 75.92 75.72 78.92 79.97 78.38 79.41 79.44 79.44 79.44 79.44 79.44 79.55 79.50 80.59 79.50 80.59 74.09 74.09 75.72 75.72 75.72 78.92 79.97 78.38 79.44 79.44 79.44 79.44 79.44 79.44 79.44 79.55 79.50 79.50 79.55 79.50 79.57 79.57 70.55 70.50 70.50 70.59 70.50 70.50 70.50 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.50</td><td>23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.46 92.12 92.98 89.88 91.78 90.00 91.54 93.23 92.62 90.16 89.56 89.56 89.43</td><td>22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.98 90.98 90.98 90.98 90.98 90.98 90.98 90.98 90.98 90.90 94.57 90.20 86.00 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.76</td><td>27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 84.14 90.29 89.61 90.77 89.24 89.65 89.65</td><td>22.03 19.24 19.11 C Far 26 94.89 92.93 95.36 95.36 95.37 95.62 89.84 95.19 89.16 93.74 95.19 89.16 93.74 94.38 95.36 90.98 94.54 91.36 00 97</td><td>26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 93.92 92.97 90.73 90.73 90.30 90.74</td><td>22.23 18.94 19.69 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 95.98 88.90 94.34 88.46 90.99</td><td>27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.54 92.24 92.00 88.30 92.56 89.94 91.60 93.39 92.56 93.39 92.67 90.02 89.97 89.07 89.07</td><td>24.02 22.98 24.75 7ER Far .96 89.02 80.90 87.43 88.00 89.90 94.98 87.43 88.00 89.90 94.98 87.25 87.38 85.73 87.84 93.03 90.70 88.31 94.73 94.73 91.10 90.55</td><td>26,98 24,13 23,83 Ave: Near 91,72 90,70 92,17 92,20 91,71 92,20 91,71 92,62 89,64 93,18 88,59 91,39 92,26 90,004 89,87 92,54 92,26 90,04 89,87</td><td>22.74 20.22 20.95 Far Far 77 91.97 87.16 91.76 91.80 92.32 95.23 88.67 90.94 88.40 90.91 93.28 92.24 90.91 93.28 92.44 90.91 94.52 90.32</td></th<> | 76.39 77.15 77.15 77.15 77.15 77.15 86.72 74.96 80.57 81.57 81.57 81.57 81.57 81.57 81.28 81.94 81.94 81.94 81.94 81.94 81.94 81.94 72.55 | 65.90 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 83.64 87.78 84.44 86.10 84.36 83.71 83.71 83.31 88.44 83.97 | 77.23 78.27 78.27 Far Far 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 73.56 73.56 75.44 81.13 82.11 82.91 82.91 82.91 82.59 84.80 77.72 76.27 76.27 76.27 77.27 77.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.27 78.29 78.69 82.10 82.90 82.90 82.90 82.90 82.90 82.90 82.90 82.90 82.90 82.90 82.90 82.91 78.56 73.56 73.56 73.56 73.56 73.56 73.56 75.44 82.91 82.91 82.91 76.27 77.77 77.777 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 84.10 87.47 84.41 83.01 82.55 81.93 83.55 82.95 | 79.03 79.58 TER Far 51 85.47 71.65 73.14 74.54 74.54 73.14 74.54 73.14 73.64 86.00 71.57 70.20 75.59 73.66 74.56 74.56 74.56 74.56 74.58 73.84 73.84 73.84 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 85.32 84.31 84.34 84.34 84.21 84.34 84.21 83.94 86.24 83.76 | 77.44 78.12 rage Far 76 86.72 75.72 79.97 79.97 78.38 79.47 87.26 74.35 74.13 79.44 79.34 79.34 79.50 80.59 75.59 74.09 75.92 79.50 80.59 74.09 75.92 75.72 78.92 79.97 78.38 79.41 79.44 79.44 79.44 79.44 79.44 79.55 79.50 80.59 79.50 80.59 74.09 74.09 75.72 75.72 75.72 78.92 79.97 78.38 79.44 79.44 79.44 79.44 79.44 79.44 79.44 79.55 79.50 79.50 79.55 79.50 79.57 79.57 70.55 70.50 70.50 70.59 70.50 70.50 70.50 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.57 70.50 | 23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.46 92.12 92.98 89.88 91.78 90.00 91.54 93.23 92.62 90.16 89.56 89.56 89.43 | 22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.98 90.98 90.98 90.98 90.98 90.98 90.98 90.98 90.98 90.90 94.57 90.20 86.00 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.73 91.76 | 27.04 25.82 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 84.14 90.29 89.61 90.77 89.24 89.65 89.65 | 22.03 19.24 19.11 C Far 26 94.89 92.93 95.36 95.36 95.37 95.62 89.84 95.19 89.16 93.74 95.19 89.16 93.74 94.38 95.36 90.98 94.54 91.36 00 97 | 26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 93.92 92.97 90.73 90.73 90.30 90.74 | 22.23 18.94 19.69 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 95.98 88.90 94.34 88.46 90.99 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.54 92.24 92.00 88.30 92.56 89.94 91.60 93.39 92.56 93.39 92.67 90.02 89.97 89.07 89.07 | 24.02 22.98 24.75 7ER Far .96 89.02 80.90 87.43 88.00 89.90 94.98 87.43 88.00 89.90 94.98 87.25 87.38 85.73 87.84 93.03 90.70 88.31 94.73 94.73 91.10 90.55 | 26,98 24,13 23,83 Ave: Near 91,72 90,70 92,17 92,20 91,71 92,20 91,71 92,62 89,64 93,18 88,59 91,39 92,26 90,004 89,87 92,54 92,26 90,04 89,87 | 22.74 20.22 20.95 Far Far 77 91.97 87.16 91.76 91.80 92.32 95.23 88.67 90.94 88.40 90.91 93.28 92.24 90.91 93.28 92.44 90.91 94.52 90.32 |
| Average Average (All) FPR↓ Average CIL ad -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average Fine-tuning-base LogitNorm T2FNorm AUGMIX REGMIX VOS BER (Ours) | 67.24 65.81 iCa Near ccuracy 58 d OOD 87.34 83.67 83.78 83.91 83.84 83.83 87.94 85.66 85.04 85.04 85.04 85.04 85.04 85.04 85.69 83.78 83.69 83.69 83.78 83.69 83.78 83.69 83.78 | 77.12 77.51 RL Far 20 methods 88.12 77.57 79.86 80.85 81.19 79.97 80.37 74.39 75.41 80.75 79.88 80.39 79.98 80.36 79.98 80.39 79.98 | 8.23 62.33 62.33 62.70 Bi State State </td <td>76.39 77.15 77.15 77.15 77.15 77.15 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.88 87.47 82.85 86.00 77.88 87.47 80.27 81.70 81.55 81.28 81.94 72.05 73.51</td> <td>65.90 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 84.44 86.10 84.36 83.71 83.97 83.71 83.71 83.71 83.71 83.71 83.44 83.97 83.71 83.44 83.97 83.71 83.44 83.97</td> <td>77.23 78.27 78.27 78.27 78.4 86.59 78.69 82.90 81.82 81.41 87.66 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 75.20 76.20 77.20 7</td> <td>68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.66 83.63 87.06 84.10 83.01 82.95 82.53 81.93 83.55 82.81</td> <td>79.03 79.58 71ER Far .51 85.47 71.65 73.14 74.54 74.54 74.54 74.54 74.54 74.54 74.54 74.54 74.54 75.59 7.559 7.559 7.3.66 74.20 75.59 73.66 74.20 75.89 73.84 70.98</td> <td>65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.60 85.32 84.31 84.34 84.34 84.21 83.94 86.24 85.32</td> <td>77.44 78.12 *age Far 76 86.72 75.72 79.97 78.38 79.47 87.26 74.35 74.13 79.44 79.34 79.50 80.59 75.50 80.59 75.50 80.59 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50</td> <td>23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 90.00 91.54 91.54 91.54 91.54 91.54 91.68 91.68 91.68 91.68 91.68 91.68 91.69 91.23 91.69 91.23 91.69 91.23 91.69 91.23 91.69 91.23 91.69 91.68 91.69 91.68 91.69 91.68 92.65 91.78 90.00 91.52 91.52 92.65 91.52 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 92.65 92.65 91.54 91.54 91.54 92.53 92.62 90.16 89.55 89.88 92.62 90.16 89.55 89.88 92.62 90.16 89.55 89.55 92.62 90.16 89.55 89.55 89.55 92.62 90.16 91.54 91.54 92.55 92.55 92.55 91.55</td> <td>22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 88.40 89.70 91.73 91.12 92.26 94.48 90.36 91.73</td> <td>27.04 25.04 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 84.14 90.29 89.61 90.77 89.24 88.65 87.48 89.65 87.48 89.65</td> <td>22.05 19.24 19.11 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.62 89.84 95.37 95.62 89.84 95.16 93.74 94.38 95.36 90.98 94.38 95.36 90.98 94.38 95.36 90.98 94.38 95.36 90.98 94.38 95.36 95.36 95.37 95.36 95.36 95.37 95.36 95.37 95.36 95.37 95.36 95.37 95.36 95.37 95.36 95.37 95.36 95.36 95.37 95.36 95.37 95.36 95.36 95.37 95.36 95.36 95.36 95.37 95.36 95.37 95.3</td> <td>26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 93.92 92.97 90.73 90.73 90.74 88.28</td> <td>22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 93.98 92.58 88.90 94.34 88.46 90.85 94.55 94.55 92.58 88.90 94.58 88.46 90.85 94.55 94.55 95.58 95.5</td> <td>27.46 24.13 23.84 FOS Near 91.83 90.92 92.54 92.54 92.54 92.41 92.00 93.90 88.30 92.56 89.94 91.60 93.39 92.67 90.02 89.97 89.67 89.67 89.67 89.67</td> <td>24.02 22.98 24.75 7ER Far 996 89.02 80.90 87.43 88.00 87.43 88.00 87.43 88.00 87.43 88.00 87.43 88.00 87.43 88.00 89.90 87.43 87.84 93.03 90.70 88.31 94.73 91.10 90.58</td> <td>26.98 24.13 23.83 4we: Near 91.72 90.70 92.17 92.20 91.71 92.20 91.71 92.20 91.71 92.20 91.71 92.20 91.71 92.20 91.39 91.39 91.39</td> <td>22.74 20.22 20.95 Far Far 77 91.97 87.16 91.96 91.96 91.80 92.32 95.23 88.67 90.94 88.40 90.94 88.40 90.94 88.67 90.94 9</td> | 76.39 77.15 77.15 77.15 77.15 77.15 86.72 74.96 80.57 81.57 76.37 82.85 86.00 77.88 87.47 82.85 86.00 77.88 87.47 80.27 81.70 81.55 81.28 81.94 72.05 73.51 | 65.90 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 84.44 86.10 84.36 83.71 83.97 83.71 83.71 83.71 83.71 83.71 83.44 83.97 83.71 83.44 83.97 83.71 83.44 83.97 | 77.23 78.27 78.27 78.27 78.4 86.59 78.69 82.90 81.82 81.41 87.66 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 73.56 75.20 76.20 77.20 7 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.66 83.63 87.06 84.10 83.01 82.95 82.53 81.93 83.55 82.81 | 79.03 79.58 71ER Far .51 85.47 71.65 73.14 74.54 74.54 74.54 74.54 74.54 74.54 74.54 74.54 74.54 75.59 7.559 7.559 7.3.66 74.20 75.59 73.66 74.20 75.89 73.84 70.98 | 65.92 65.59 Near 59. 86.71 83.86 84.27 84.18 83.76 85.60 85.32 84.31 84.34 84.34 84.21 83.94 86.24 85.32 | 77.44 78.12 *age Far 76 86.72 75.72 79.97 78.38 79.47 87.26 74.35 74.13 79.44 79.34 79.50 80.59 75.50 80.59 75.50 80.59 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 80.50 75.50 | 23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.12 92.98 89.88 91.78 90.00 91.54 91.54 91.54 91.54 91.54 91.68 91.68 91.68 91.68 91.68 91.68 91.69 91.23 91.69 91.23 91.69 91.23 91.69 91.23 91.69 91.23 91.69 91.68 91.69 91.68 91.69 91.68 92.65 91.78 90.00 91.52 91.52 92.65 91.52 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 91.54 92.65 92.65 91.54 91.54 91.54 92.53 92.62 90.16 89.55 89.88 92.62 90.16 89.55 89.88 92.62 90.16 89.55 89.55 92.62 90.16 89.55 89.55 89.55 92.62 90.16 91.54 91.54 92.55 92.55 92.55 91.55 | 22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 88.40 89.70 91.73 91.12 92.26 94.48 90.36 91.73 | 27.04 25.04 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 84.14 90.29 89.61 90.77 89.24 88.65 87.48 89.65 87.48 89.65 | 22.05 19.24 19.11 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.62 89.84 95.37 95.62 89.84 95.16 93.74 94.38 95.36 90.98 94.38 95.36 90.98 94.38 95.36 90.98 94.38 95.36 90.98 94.38 95.36 95.36 95.37 95.36 95.36 95.37 95.36 95.37 95.36 95.37 95.36 95.37 95.36 95.37 95.36 95.37 95.36 95.36 95.37 95.36 95.37 95.36 95.36 95.37 95.36 95.36 95.36 95.37 95.36 95.37 95.3 | 26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 93.92 92.97 90.73 90.73 90.74 88.28 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 93.98 92.58 88.90 94.34 88.46 90.85 94.55 94.55 92.58 88.90 94.58 88.46 90.85 94.55 94.55 95.58 95.5 | 27.46 24.13 23.84 FOS Near 91.83 90.92 92.54 92.54 92.54 92.41 92.00 93.90 88.30 92.56 89.94 91.60 93.39 92.67 90.02 89.97 89.67 89.67 89.67 89.67 | 24.02 22.98 24.75 7ER Far 996 89.02 80.90 87.43 88.00 87.43 88.00 87.43 88.00 87.43 88.00 87.43 88.00 87.43 88.00 89.90 87.43 87.84 93.03 90.70 88.31 94.73 91.10 90.58 | 26.98 24.13 23.83 4we: Near 91.72 90.70 92.17 92.20 91.71 92.20 91.71 92.20 91.71 92.20 91.71 92.20 91.71 92.20 91.39 91.39 91.39 | 22.74 20.22 20.95 Far Far 77 91.97 87.16 91.96 91.96 91.80 92.32 95.23 88.67 90.94 88.40 90.94 88.40 90.94 88.67 90.94 9 |
| Average Average (All) FPR↓ Average CIL at Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average Fine-finame Average MaxLogit Average Average MaxLogit Average Average Average NNGuide Average NNGuide Average NNGuide Average NNGuide Average NNGuide Average NNGuide Average NNGuide Average NNGuide NNGuide Average NNGUID NNGuide NNGuide Average NNGUID NNGuide Average NNGUID NN | 67.24 65.81 iCa Near 58 d OOD n 87.34 83.67 83.78 83.84 83.83 87.94 85.36 85.04 ased OOD 83.69 83.67 83.69 83.69 83.67 83.69 83.67 83.69 83.67 83.67 83.67 83.67 | 77.12 77.51 RL Far 20 nethods 88.12 77.57 79.86 80.85 81.19 79.97 89.37 74.39 75.41 80.75 D method 79.88 80.39 79.98 80.26 77.91 75.00 79.68 | 81.00 81.00 <th< td=""><td>76.39 77.15 77.15 77.15 77.15 78.27 86.72 74.96 80.57 81.57 81.57 81.57 80.27 81.57 80.27 81.70 81.75 81.28 81.94 72.05 73.51 79.70</td><td>65.90 65.90 66.54 W Near 61 85.16 82.91 83.40 83.07 83.64 87.78 84.44 86.10 84.36 83.71 83.97 83.71 83.97 83.71 83.97 83.71 83.97 83.71 83.31 88.44 83.02 84.63</td><td>77.23 78.27 78.27 78.27 78.44 86.59 78.69 82.10 82.90 81.82 81.41 82.00 81.82 81.41 87.66 73.56 73.56 73.56 73.56 73.56 73.56 73.56 82.01 76.52 82.01</td><td>68.21 67.33 FOS Near 63 86.10 83.14 82.94 82.66 83.63 87.06 84.10 87.47 84.41 83.01 82.53 81.93 83.55 82.81 82.79</td><td>79.03 79.58 71ER Far 51 85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20 75.59 73.20 75.59 73.66 74.50 74.33</td><td>65.92 65.59 Aver Near 59. 86.71 83.86 84.27 84.18 85.01 87.73 85.60 86.80 85.32 84.31 84.31 84.31 84.31 84.21 83.94 86.24 83.76 84.61</td><td>77.44 78.12 rage Far 76 86.72 75.72 79.97 78.38 79.47 87.26 74.13 79.44 79.34 79.34 79.34 79.34 79.50 80.59 75.59 80.59 75.38 74.00 78.93</td><td>23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.46 92.42 92.48 92.46 92.12 92.98 89.88 89.88 91.78 90.00 91.54 93.23 92.62 90.16 89.56 89.56 89.56 89.56</td><td>22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.950 94.57 90.20 86.00 88.40 89.70 91.73 91.73 92.26 94.48 90.36 91.78 91.99</td><td>27.04 25.04 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.74 90.29 89.61 90.77 89.24 89.65 87.48 87.48 89.49</td><td>22.03 22.03 19.24 19.11 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.37 95.62 89.84 95.19 93.74 94.38 95.36 90.98 94.38 95.36 90.98 94.54 91.36 90.87 93.32 </td><td>26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 95.52 90.30 95.52 90.30 95.74 88.28 91.73</td><td>22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 73.52 95.77 87.38 95.19 90.30 92.38 93.98 95.28 95.88 <u>88.90</u> 94.34 88.46 90.65</td><td>27.46 24.13 23.84 FOS Near 45. 90.92 92.54 92.54 92.54 92.54 92.00 93.90 93.90 93.90 93.90 93.90 93.56 89.94 91.60 93.39 92.56 89.97 89.07 89.07 89.07 89.07 90.02 90.14</td><td>24.02 22.98 24.75 TER Far 96 89.02 80.90 87.43 88.00 87.43 88.00 87.43 88.00 87.43 88.00 89.90 94.98 87.25 87.38 85.73 87.84 93.03 90.70 88.31 94.73 91.10 90.58 91.57</td><td>26.98 24.13 23.83 Ave: 91.72 90.70 92.17 92.20 91.71 92.62 89.64 93.18 88.59 91.39 92.54 92.26 90.04 89.87 87.87 90.84</td><td>22.74 20.22 20.95 Far Far 91.97 87.16 91.97 87.16 91.80 92.32 95.23 88.67 90.94 88.40 90.91 93.28 90.91 93.28 90.91 93.28 90.91 93.28 90.11 94.52 91.03 92.13</td></th<> | 76.39 77.15 77.15 77.15 77.15 78.27 86.72 74.96 80.57 81.57 81.57 81.57 80.27 81.57 80.27 81.70 81.75 81.28 81.94 72.05 73.51 79.70 | 65.90 65.90 66.54 W Near 61 85.16 82.91 83.40 83.07 83.64 87.78 84.44 86.10 84.36 83.71 83.97 83.71 83.97 83.71 83.97 83.71 83.97 83.71 83.31 88.44 83.02 84.63 | 77.23 78.27 78.27 78.27 78.44 86.59 78.69 82.10 82.90 81.82 81.41 82.00 81.82 81.41 87.66 73.56 73.56 73.56 73.56 73.56 73.56 73.56 82.01 76.52 82.01 | 68.21 67.33 FOS Near 63 86.10 83.14 82.94 82.66 83.63 87.06 84.10 87.47 84.41 83.01 82.53 81.93 83.55 82.81 82.79 | 79.03 79.58 71ER Far 51 85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20 75.59 73.20 75.59 73.66 74.50 74.33 | 65.92 65.59 Aver Near 59. 86.71 83.86 84.27 84.18 85.01 87.73 85.60 86.80 85.32 84.31 84.31 84.31 84.31 84.21 83.94 86.24 83.76 84.61 | 77.44 78.12 rage Far 76 86.72 75.72 79.97 78.38 79.47 87.26 74.13 79.44 79.34 79.34 79.34 79.34 79.50 80.59 75.59 80.59 75.38 74.00 78.93 | 23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.46 92.42 92.48 92.46 92.12 92.98 89.88 89.88 91.78 90.00 91.54 93.23 92.62 90.16 89.56 89.56 89.56 89.56 | 22.65 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.950 94.57 90.20 86.00 88.40 89.70 91.73 91.73 92.26 94.48 90.36 91.78 91.99 | 27.04 25.04 25.04 Bi Near 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.74 90.29 89.61 90.77 89.24 89.65 87.48 87.48 89.49 | 22.03 22.03 19.24 19.11 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.37 95.62 89.84 95.19 93.74 94.38 95.36 90.98 94.38 95.36 90.98 94.54 91.36 90.87 93.32 | 26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 95.52 90.30 95.52 90.30 95.74 88.28 91.73 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 73.52 95.77 87.38 95.19 90.30 92.38 93.98 95.28 95.88 <u>88.90</u> 94.34 88.46 90.65 | 27.46 24.13 23.84 FOS Near 45. 90.92 92.54 92.54 92.54 92.54 92.00 93.90 93.90 93.90 93.90 93.90 93.56 89.94 91.60 93.39 92.56 89.97 89.07 89.07 89.07 89.07 90.02 90.14 | 24.02 22.98 24.75 TER Far 96 89.02 80.90 87.43 88.00 87.43 88.00 87.43 88.00 87.43 88.00 89.90 94.98 87.25 87.38 85.73 87.84 93.03 90.70 88.31 94.73 91.10 90.58 91.57 | 26.98 24.13 23.83 Ave: 91.72 90.70 92.17 92.20 91.71 92.62 89.64 93.18 88.59 91.39 92.54 92.26 90.04 89.87 87.87 90.84 | 22.74 20.22 20.95 Far Far 91.97 87.16 91.97 87.16 91.80 92.32 95.23 88.67 90.94 88.40 90.91 93.28 90.91 93.28 90.91 93.28 90.91 93.28 90.11 94.52 91.03 92.13 |
| Average Average (All) FPR; Average CIL at -Post-hoc-base MSP ODIN Energy MaxLogit GEN ReAct KLM Relation NNGuide Average Fine-tuning- Common AUGMIX REGMIX REGMIX VOS BER (Ours) Average (All) | 67.24 65.81 iCa Near ccuracy 58 d OOD r 87.34 83.67 83.78 83.91 83.84 83.83 87.94 85.66 85.04 85.04 85.04 85.04 83.67 83.78 83.69 83.67 83.69 83.67 83.67 83.78 83.69 83.67 83.69 83.67 83.69 83.67 83.69 83.67 83.69 83.67 83.69 83.67 83.69 83.67 83.69 83.69 83.67 83.69 83.69 83.67 83.69 83.69 83.67 83.69 83.69 83.67 83.69 83.69 83.69 83.67 83.69 83.69 83.69 83.69 83.69 83.69 83.69 83.69 83.69 83.67 83.69 83.6 | 77.12 77.51 RL Far 20 methods 88.12 77.57 79.86 80.85 81.19 79.86 80.85 81.19 79.87 89.37 74.39 75.41 80.75 D method 79.88 80.39 P (7.91) 75.00 79.00 77.01 70.00 70 70 70 70 70 70 70 70 70 70 70 70 7 | 8.23 62.33 62.33 62.70 Bi State State </td <td>76.39 76.39 77.15 T Far 87 86.72 74.96 80.57 76.37 82.85 86.00 77.88 75.47 80.27 81.55 81.28 81.94 72.05 73.51 79.70 80.07</td> <td>85.10 65.90 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 84.44 86.10 83.71 83.31 83.71 83.31 88.44 83.02 84.46 84.46</td> <td>77.23 78.27 78.27 78 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 73.56 73.56 73.56 81.41 82.91 82.91 82.91 82.55 84.80 77.70 76.52 82.01 81.44</td> <td>68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 84.41 83.01 82.95 82.53 81.93 83.55 82.81 82.78 83.83</td> <td>79.03 79.58 7128 Far .51 85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20 75.59 73.66 74.20 75.59 73.84 74.33 75.14</td> <td>65.92 65.59 Near 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 86.80 85.32 84.31 84.34 84.21 83.94 84.31 84.34 84.21 83.94 86.24 83.76</td> <td>77.44 78.12 *age Far 76 86.72 75.72 78.92 79.97 78.38 79.47 79.34 79.34 79.34 79.34 79.34 79.50 80.59 80.59 75.38 74.00 75.38 74.00 75.38 74.00 75.90</td> <td>23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.46 92.46 92.46 92.48 91.78 90.65 91.54 93.23 92.62 90.16 89.56 89.43 88.42 91.05</td> <td>22.65 19.73 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 86.00 88.40 91.73 91.78 91.78 90.52</td> <td>27.04 25.82 25.04 80 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 84.14 90.29 89.77 89.24 88.16 89.65 87.48 89.49 90.00</td> <td>22.03 22.03 19.24 19.11 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.62 93.74 94.38 95.36 93.74 94.38 91.36 90.87 93.32 93.32</td> <td>26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 90.59 92.14 90.59 92.14 93.92 92.97 90.30 92.14 93.92 92.97 90.73 90.30 90.74 88.28 91.73 91.93</td> <td>22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 92.58 88.46 90.88 91.65 92.12</td> <td>27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 92.00 93.90 88.30 92.56 89.94 91.60 93.39 92.56 89.94 91.60 88.30 92.67 90.02 89.96 87.29 91.14</td> <td>24.02 22.98 24.75 TER 96 89.02 80.90 87.43 89.00 87.43 89.00 89.90 94.98 87.25 87.38 87.84 87.84 87.84 87.84 87.84 93.03 90.70 88.31 94.73 91.10 90.58 91.57 89.17</td> <td>26.98 24.13 23.83 Ave: Near 91.72 90.70 92.17 92.17 92.17 92.17 92.17 92.26 90.70 91.71 92.62 89.64 93.18 88.59 91.39 91.39 91.39 92.54 92.26 90.04 89.87 87.87 90.04 89.87 87.87</td> <td>22.74 20.22 20.95 Far Far 91.97 87.16 91.97 87.16 91.76 91.80 92.32 95.23 88.67 90.94 88.60 90.91 93.28 92.44 90.32 91.03 91.35</td> | 76.39 76.39 77.15 T Far 87 86.72 74.96 80.57 76.37 82.85 86.00 77.88 75.47 80.27 81.55 81.28 81.94 72.05 73.51 79.70 80.07 | 85.10 65.90 65.90 66.54 W Near 61. 85.16 82.91 83.40 83.07 82.77 83.64 87.78 84.44 86.10 83.71 83.31 83.71 83.31 88.44 83.02 84.46 84.46 | 77.23 78.27 78.27 78 86.59 78.69 82.10 82.90 81.82 81.41 87.66 73.56 73.56 73.56 73.56 81.41 82.91 82.91 82.91 82.55 84.80 77.70 76.52 82.01 81.44 | 68.21 67.33 FOS Near 63 86.10 83.14 82.72 82.94 82.56 83.63 87.06 84.41 83.01 82.95 82.53 81.93 83.55 82.81 82.78 83.83 | 79.03 79.58 7128 Far .51 85.47 71.65 73.14 74.54 74.14 73.64 86.00 71.57 70.20 75.59 73.66 74.20 75.59 73.84 74.33 75.14 | 65.92 65.59 Near 86.71 83.86 84.27 84.18 83.76 85.01 87.73 85.60 86.80 85.32 84.31 84.34 84.21 83.94 84.31 84.34 84.21 83.94 86.24 83.76 | 77.44 78.12 *age Far 76 86.72 75.72 78.92 79.97 78.38 79.47 79.34 79.34 79.34 79.34 79.34 79.50 80.59 80.59 75.38 74.00 75.38 74.00 75.38 74.00 75.90 | 23.69 23.69 23.51 iCa Near 40. 91.12 90.89 92.65 92.46 92.46 92.46 92.46 92.48 91.78 90.65 91.54 93.23 92.62 90.16 89.56 89.43 88.42 91.05 | 22.65 19.73 19.73 20.27 RL Far 86 90.83 84.91 90.98 90.94 90.50 94.57 90.20 86.00 88.40 91.73 91.78 91.78 90.52 | 27.04 25.82 25.04 80 42. 91.25 89.34 90.53 91.00 92.30 91.42 89.77 92.85 84.14 90.29 89.77 89.24 88.16 89.65 87.48 89.49 90.00 | 22.03 22.03 19.24 19.11 19.24 19.11 C Far 26 94.89 92.93 95.36 95.30 95.37 95.62 93.74 94.38 95.36 93.74 94.38 91.36 90.87 93.32 93.32 | 26.54 22.88 22.94 W Near 45. 92.70 91.66 92.95 92.92 90.40 92.18 90.59 95.52 90.30 92.14 90.59 92.14 90.59 92.14 93.92 92.97 90.30 92.14 93.92 92.97 90.73 90.30 90.74 88.28 91.73 91.93 | 22.23 18.94 19.69 A Far 99 93.14 89.91 93.25 92.97 93.52 95.77 87.38 95.19 90.30 92.38 92.58 88.46 90.88 91.65 92.12 | 27.46 24.13 23.84 FOS Near 45. 91.83 90.92 92.54 92.41 92.00 93.90 88.30 92.56 89.94 91.60 93.39 92.56 89.94 91.60 88.30 92.67 90.02 89.96 87.29 91.14 | 24.02 22.98 24.75 TER 96 89.02 80.90 87.43 89.00 87.43 89.00 89.90 94.98 87.25 87.38 87.84 87.84 87.84 87.84 87.84 93.03 90.70 88.31 94.73 91.10 90.58 91.57 89.17 | 26.98 24.13 23.83 Ave: Near 91.72 90.70 92.17 92.17 92.17 92.17 92.17 92.26 90.70 91.71 92.62 89.64 93.18 88.59 91.39 91.39 91.39 92.54 92.26 90.04 89.87 87.87 90.04 89.87 87.87 | 22.74 20.22 20.95 Far Far 91.97 87.16 91.97 87.16 91.76 91.80 92.32 95.23 88.67 90.94 88.60 90.91 93.28 92.44 90.32 91.03 91.35 |

Table 7: Fine-grained results on near- and far-OOD detection datasets at the step size of k = 10 for CIFAR 100 and at the step size of k = 100 for ImageNet1K. The best and second-best performance per dataset in the fine-tuning-based methods are highlighted. The upper, middle, lower parts of the table are for AUC, AP, and FPR performance, respectively.

| 1038 | | | | | ID | Dataset | CIFAR | 100 | | | | | | | ID | Dataset: | ImageNe | t1K | | | |
|-------|----------------|---------------------------|-------------------|-------------|-----------------------|----------------|----------------|---------|----------------|-------|----------------|----------------|----------------|--------------|----------------|----------------|-----------------------|-----------------------|----------------|----------------|----------------|
| 1020 | AUC↑ | iCa | IRL | В | iC | W | /A | FOS | TER | Ave | rage | iCa | ıRL | В | iC | W | 'A | FOS | TER | Ave | rage |
| 1039 | Average CIL | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far |
| 1040 | -Average CIL t | 60 | .08 | 61 | .68 | 65 | .88 | 66. | .01 | 63 | 41 | 44 | .44 | 49 | .63 | 52 | .22 | 52. | .29 | 49 | .65 |
| 10/1 | -Post-hoc-base | d OOD 1 | nethods | 1 71 10 | (1.15 | 71.02 | 60.02 | 71.04 | 71.02 | 20.54 | (0.47 | 62.67 | (1.26 | | ((0) | | 0.00 | ((50 | (0.75 | ((1) | 66.40 |
| 1041 | ODIN | 71.10 | 69.53 | 73.95 | 67.67 | 74.18 | 70.69 | 73.37 | 75.37 | 73.15 | 70.81 | 67.06 | 71.13 | 70.54 | 72.19 | 67.62 | 68.34 | 67.61 | 75.49 | 68.21 | 71.79 |
| 1042 | Energy | 71.72 | 69.67 | 74.01 | 67.63 | 74.62 | 72.23 | 74.12 | 76.54 | 73.62 | 71.52 | 64.87 | 67.17 | 69.35 | 71.79 | 65.96 | 67.78 | 66.80 | 74.59 | 66.75 | 70.33 |
| 10.10 | MaxLogit | 71.65 | 69.69 | 74.04 | 67.29 | 74.62 | 72.16 | 74.23 | 76.45 | 73.64 | 71.40 | 65.13 | 67.22 | 69.50 | 70.81 | 66.80 | 66.64 | 67.36 | 73.94 | 67.20 | 69.65 |
| 1043 | ReAct | 67.63 | 69.38 | 73.24 | 69.31 | 73.54 | 75.54 | 75.16 | 76.64 | 72.39 | 72.72 | 58.88 | 49.21 | 66.53 | 66.63 | 65.11 | 61.31 | 65.10 | 65.25 | 63.90 | 60.60 |
| 10// | KLM | 67.82 | 67.36 | 69.91 | 64.45 | 70.46 | 69.61 | 70.98 | 71.52 | 69.79 | 68.23 | 65.77 | 67.91 | 68.62 | 64.72 | 65.99 | 63.72 | 69.18 | 72.31 | 67.39 | 67.16 |
| 1044 | Relation | 58.11 | 67.55 | 67.34 | 64.55 | 72.13 | 78.12 | 67.40 | 74.55 | 66.25 | 71.19 | 63.33 | 72.47 | 68.51 | 70.30 | 67.61 | 67.31 | 65.67 | 72.42 | 66.28 | 70.62 |
| 1045 | Average | 68.46 | 69.47 | 71.68 | 66.95 | 73.18 | 73.10 | 72.83 | 75.17 | 71.54 | 71.17 | 64.38 | 66.46 | 68.85 | 69.55 | 66.76 | 66.81 | 66.65 | 72.54 | 66.66 | 68.84 |
| 1040 | -Fine-tuning-b | ased OC | D metho | ods | | | | | | | | | | | | | | | | | |
| 1040 | LogitNorm | 71.88 | 69.72 | 73.52 | 69.20 | 73.56 | 71.28 | 73.86 | 75.18 | 73.20 | 71.34 | 64.30 | 66.85 | 68.25 | 72.13 | 65.31 | 68.28 | 66.36 | 70.78 | 66.06 | 69.51 |
| 1047 | AUGMIX | 72.50 | 69.82 | 73.19 | 69.58 | 73.63 | 71.35 | 74.33 | 74.96 | 73.41 | 71.43 | 64.94 | 67.91 75.91 | 67.75 | 76.97 | 65.71 | 72.87 | 67.23 | 72.30 69.06 | 66.41 | 70.62 73.70 |
| 1017 | REGMIX | 73.12 | 70.64 | 72.97 | 69.91 | 73.81 | 71.11 | 74.74 | 75.51 | 73.66 | 71.79 | 67.86 | 63.51 | <u>69.41</u> | 70.68 | 67.70 | 66.97 | 68.96 | 70.36 | 68.48 | 67.88 |
| 1048 | VOS | 73.41 | 71.76 | 67.55 | 66.23 | 65.31 | 70.08 | 73.80 | 74.97 | 70.02 | 70.76 | 63.69 | 71.47 | 64.94 | 70.13 | 64.14 | 67.94 | 67.19 | 72.57 | 64.99 | 70.53 |
| 1040 | Average | 72.56 | 70.39 | 75.08 | <u>70.20</u> 69.24 | 75.43 | 71.44 | 75.20 | 75.14 | 72.62 | 74.76 | 65.20 | 69.13 | 67.71 | 72.43 | 65.79 | <u>70.16</u> 69.21 | 67.17 | 75.41 | 66.74 66.47 | 70.45 |
| 1049 | Average (All) | 69.92 | 69.80 | 71.78 | 67.77 | 72.71 | 72.37 | 73.28 | 75.16 | 71.93 | 71.27 | 64.67 | 67.41 | 68.44 | 70.58 | 66.41 | 67.67 | 66.84 | 72.00 | 66.59 | 69.42 |
| 1050 | AP↑ | iCa | ıRL | В | iC | W | /A | FOS | TER | Ave | rage | iCa | ıRL | В | iC | W | Ά | FOS | TER | Ave | rage |
| 1051 | Average CIL | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far |
| 1051 | -Average CIL i | 60 | .08 | 61 | .68 | 65 | .88 | 66. | .01 | 63 | 41 | 44 | .44 | 49 | .63 | 52 | .22 | 52. | .29 | 49 | .65 |
| 1052 | -Post-hoc-base | d OOD I | nethods | 1 66 42 | 75.72 | | 70.47 | 1 ((22 | 70.07 | 6.0 | 77.01 | 1 24 27 | 21.20 | 1 26 97 | 22.72 | 05.14 | 10.00 | 25.20 | 25.00 | 25.41 | 22.24 |
| 1050 | ODIN | 66.69 | 76.57 | 69.50 | 75.62 | 69.80 | 77.20 | 68.69 | 80.05 | 68.67 | 77.36 | 24.27 | 21.29 | 28.84 | 26.58 | 25.14 | 23.63 | 25.38 | 25.00 | 25.41 | 22.24 |
| 1053 | Energy | 67.16 | 76.79 | 69.45 | 75.66 | 70.12 | 78.12 | 69.40 | 80.78 | 69.03 | 77.84 | 23.79 | 22.11 | 26.69 | 25.28 | 24.12 | 21.18 | 24.37 | 27.24 | 24.74 | 23.95 |
| 1054 | MaxLogit | 67.02 | 76.91 | 69.47 | 75.73 | 70.03 | 78.37 | 69.41 | 80.90 | 68.98 | 77.98 | 24.18 | 22.41 | 27.17 | 24.60 | 24.90 | 20.69 | 25.00 | 26.82 | 25.31 | 23.63 |
| 1001 | ReAct | 64.32 | 78.22 | 68.84 | 77.00 | 69.24 | 80.52 | 67.72 | 82.49 | 67.53 | 79.56 | 24.77 | 13.04 | 25.82 | 20.62 | 25.94 | 18.00 | 25.85 | 20.09 | 20.39 | 17.94 |
| 1055 | KLM | 62.71 | 78.06 | 64.73 | 77.12 | 64.53 | 79.28 | 65.79 | 80.35 | 64.44 | 78.70 | 26.60 | 24.14 | 28.25 | 21.19 | 25.73 | 20.66 | 30.06 | 27.00 | 27.66 | 23.25 |
| 1056 | Relation | 58.70 | 79.28 | 65.41 | 78.51 | 68.32 | 80.80 | 65.20 | 80.88 | 64.41 | 79.87 | 27.17 | 30.24 | 28.16 | 25.52 | 28.12 | 24.41 | 24.89 | 27.99 | 27.09 | 27.04 |
| 1000 | Average | 64.78 | 77.67 | 68.08 | 76.60 | 68.81 | 79.07 | 67.82 | 80.14 | 67.37 | 79.52 | 27.25 | 28.60 | 28.61 | 24.23 | 28.10 | 24.66 | 26.13 | 26.78 | 26.27 | 27.88 |
| 1057 | Fine-tuning-ba | ised OO | D metho | ds | | | | | | | | | | | | | | | | | |
| | LogitNorm | 67.34 | 76.81 | 69.06 | 76.32 | 69.11 | 77.66 | 69.05 | 80.11 | 68.64 | 77.72 | 23.42 | 21.69 | 26.54 | 29.55 | 23.43 | 21.51 | 24.01 | 23.52 | 24.35 | 24.07 |
| 1058 | AUGMIX | 67.89 | 76.84 | 68.75 | 76.51 | 69.13 | 77.69 | 69.52 | 80.24 79.90 | 68.82 | 77.74 | 24.12 | 33.75 | 25.87 | 27.30 34.07 | 24.15 | 26.99 | 24.25 | 24.90 | 24.60 | 24.41 30.51 |
| 1050 | REGMIX | 68.56 | 77.26 | 68.57 | 76.82 | 69.36 | 77.65 | 69.12 | 80.13 | 68.90 | 77.97 | 27.42 | 18.63 | 27.65 | 22.81 | 26.48 | 19.82 | 27.03 | 21.45 | 27.14 | 20.68 |
| 1055 | VOS | 67.64 | 77.73 | 67.63 | 75.52 | 61.74 | 77.84 | 69.02 | 80.09 | 66.51 | 77.80 | 21.51 | 22.55 | 25.32 | 27.85 | 24.27 | 23.24 | 24.12 | 26.19 | 23.80 | 24.96 |
| 1060 | Average | 67.74 | 77.14 | 68.44 | 76.61 | 67.53 | 80.55 | 69.04 | 81.39 | 68.21 | 77.91 | 24.69 | 23.84 | 26.62 | 27.81 | 28.38 | 22.87 | 28.45 | 24.66 | 25.21 | 28.04 |
| 1001 | Average (All) | 65.84 | 77.48 | 68.21 | 76.60 | 68.35 | 78.61 | 68.29 | 80.51 | 67.67 | 78.30 | 24.93 | 23.68 | 27.14 | 25.69 | 25.77 | 22.08 | 25.72 | 26.02 | 25.89 | 24.37 |
| 1061 | FDD | iCa | ıRL | B | iC | W | /A | FOS | TER | Ave | rage | iCa | ıRL | В | iC | W | A | FOS | TER | Ave | rage |
| 1062 | -Average CIL | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far |
| 1062 | -Average CIL1 | 60 | .08 | 61 | .68 | 65 | .88 | 66. | .01 | 63 | 41 | 44 | .44 | 49 | .63 | 52 | .22 | 52. | .29 | 49 | .65 |
| 1003 | -Post-hoc-base | d OOD 1 | nethods | 85.24 | 97.90 | 95.29 | 86.27 | 85.07 | 84.42 | 86.25 | 86.72 | 1 00 01 | 80.42 | 1 80.26 | 88.00 | 01.20 | 03.00 | 00.37 | 87.20 | 00.21 | 80.65 |
| 1064 | ODIN | 84.48 | 88.54 80.66 | 85.34 | 87.80 | 85.58 | 80.37 | 82.86 | 84.45 72.09 | 80.25 | 86.73 | 90.01 88.87 | 89.42 | 89.20 | 86.88 | 91.20 | 93.00 88.91 | 90.37 | 87.20 | 90.21 88.83 | 89.65 |
| | Energy | 83.99 | 83.61 | 82.41 | 81.82 | 81.75 | 80.95 | 82.58 | 74.57 | 82.68 | 80.24 | 92.13 | 91.38 | 90.36 | 87.98 | 91.81 | 93.03 | 91.78 | 87.91 | 91.52 | 90.07 |
| 1065 | MaxLogit | 84.57 | 84.39 | 82.26 | 83.04 | 82.12 | 82.44 | 82.72 | 75.80 | 82.92 | 81.42 | 91.72 | 90.72 | 89.80 | 88.68 | 91.53 | 93.03 | 91.26 | 87.53 | 91.08 | 89.99 |
| 1066 | ReAct | 85.04 | 84.12 | 82.62 | 82.59 | 82.22 | 81.62 | 83.73 | 75.17 | 83.40 | 80.81 | 93.09 | 97.23 | 89.67 | 90.88 | 90.02 | 92.80 | 90.04 | 92.61 | 90.81 | 93.22 |
| | KLM | 88.24 | 87.89 | 86.78 | 88.09 | 88.03 | 87.46 | 86.18 | 85.76 | 87.31 | 87.30 | 88.79 | 87.27 | 86.30 | 88.93 | 89.25 | 90.26 | 88.33 | 88.69 | 88.17 | 88.79 |
| 1067 | NNGuide | 84.23 | 77.32 | 82.26 | 79.24 | 90.80 82.06 | 72.97 | 83.57 | 70.82 | 83.03 | 74.41 | 91.08 | 86.84 | 87.44 | 87.11 | 94.14 89.02 | 88.42 | 89.19 | 86.63 | 89.26 | 89.01 |
| 1068 | Average | 86.00 | 84.40 | 83.37 | 82.43 | 84.00 | 82.17 | 83.82 | 76.01 | 84.30 | 81.25 | 90.99 | 89.27 | 88.85 | 88.76 | 90.35 | 91.84 | 90.15 | 87.64 | 90.08 | 89.38 |
| | Fine-tuning-ba | <i>ised OO</i> 83.94 | D methoo 83.70 | ds 82.26 | 82.56 | 82.37 | 80.81 | 82.88 | 75.98 | 82.86 | 80.76 | 92.56 | 92.39 | 89.75 | 82.28 | 92.84 | 93.78 | 92.58 | 92.02 | 91.93 | 90.12 |
| 1069 | T2FNorm | 84.16 | 84.13 | 83.15 | 82.50 | 83.16 | 81.91 | 82.75 | 76.87 | 83.31 | 81.35 | 91.88 | 91.00 | 90.87 | 85.53 | 91.79 | 91.92 | 91.57 | 89.71 | 91.53 | 89.54 |
| 1070 | AUGMIX | 83.61 | 84.07 | 82.48 | 83.17 | 82.33 | 81.14 | 82.47 | 76.70 | 82.72 | 81.27 | 89.26 | 85.42 | 89.94 | 83.66 | 89.29 | 89.55 | 89.42 | 86.75 | 89.48 | 86.34 |
| | VOS | 83.34 | 82.12 | 85.81 | 85.49 75.64 | 84.71 | 85.52 74.86 | 82.75 | 75.18 | 84.15 | 02.42 76.95 | 92.36 | 90.29 | 91.35 | 85.71 | 90.14 | 95.57 89.48 | <u>89.55</u> 90.24 | 95.02 85.72 | 91.02 | 95.20 87.80 |
| 1071 | BER (Ours) | 83.61 | 76.49 | 85.30 | 73.99 | 80.94 | 72.89 | 82.17 | 70.78 | 83.00 | 73.54 | 88.31 | 85.91 | 85.32 | 89.75 | 87.18 | 88.40 | 88.45 | 85.39 | 87.31 | 87.36 |
| 1072 | Average (All) | 83.73 | 83.80 | 83.28 | 81.47 | 82.95 | 80.41 | 82.62 | 76.52 | 83.15 | 80.55 | 90.99 | 90.68 | 90.15 | 85.70 | 90.66 | 91.66 | 90.63 | 89.56 | 90.61 | 89.40 |
| 1012 | | 05.19 | 34.17 | 00.04 | 02.09 | 00.02 | 51.54 | 35.59 | 70.19 | 35.09 | 51.00 | ,0.79 | 37.11 | 57.51 | 57.07 | 70.40 | 71.70 | 70.52 | 30.55 | 70.27 | 37.37 |

1088Table 8: Detailed results for those in Table 3 on near- and far-OOD detection datasets at the step1089size of k = 20 for CIFAR 100 and at the step size of k = 200 for ImageNet1K. The best and1090second-best performance per dataset in the fine-tuning-based methods are highlighted. The upper,1091middle, lower parts of the table are for AUC, AP, and FPR performance, respectively.

| 1092 | | | | | ID | Dataset | CIEARI | 00 | | | | | | | ID | Dataset: | ImageNe | t1K | | | |
|-----------|-------------------|--|------------------|----------------|----------------|----------------|----------------|----------------|----------------|---------|----------------|----------------|----------------|----------------|-----------------------|-----------------------|----------------|----------------|----------------|----------------|----------------|
| 1000 | AUC↑ | iCa | IRL | В | iC | Wataset. | /A | FOS | TER | Ave | rage | iCa | ıRL | Bi | iC | W | 'A | FOS | TER | Ave | rage |
| 1093 | | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far |
| 1094 | Average CIL a | <i>ccuracy</i> 62 | .65 | 64 | .14 | 68 | .05 | 68. | .75 | 65 | .90 | 48 | .85 | 53 | .30 | 58 | .42 | 57 | .84 | 54 | .60 |
| 1095 | MSP | 69.42 | 67.86 | 71.44 | 68.87 | 73.14 | 72.11 | 72.33 | 71.99 | 71.58 | 70.21 | 66.14 | 68.23 | 70.66 | 72.84 | 69.22 | 73.77 | 69.35 | 76.01 | 68.84 | 72.71 |
| | ODIN | 72.67 | 70.68 | 74.14 | 70.11 | 75.06 | 73.87 | 74.02 | 74.67 | 73.97 | 72.33 | 69.56 | 74.95 | 72.70 | 80.40 | 69.88 | 75.29 | 70.83 | 79.00 | 70.74 | 77.41 |
| 1096 | Energy | 73.62 | 71.75 | 74.19 | 70.13 | 75.32 | 75.17 | 75.64 | 76.39 | 74.69 | 73.36 | 67.58 | 73.41 | 71.44 | 80.20 | 68.36 | 74.95 | 69.92 | 78.17 | 69.33 | 76.68 |
| 1007 | GEN | 73.58 | 71.69 | 74.31 | 70.25 | 75.45 | 75.10 | 75.05 | 76.30 | 74.75 | 73.33 | 67.88 | 73.09 | 71.87 | 78.76 | 69.29 | 73.00 | 70.50 | 78.46 | 69.89 | 76.24 |
| 1097 | ReAct | 70.33 | 72.22 | 73.64 | 73.72 | 74.16 | 78.15 | 74.81 | 78.60 | 73.23 | 75.67 | 60.87 | 56.69 | 68.14 | 72.73 | 63.74 | 61.87 | 66.82 | 60.83 | 64.89 | 63.03 |
| 1098 | KLM Relation | 69.38 60.26 | 68.96 67.92 | 71.08 67.36 | 69.75 69.97 | 72.76 72.90 | 72.33 77.89 | 72.62 66.61 | 72.40 73.61 | 71.46 | 70.86 72.35 | 68.67 69.03 | 71.19 75.66 | 71.37 72.58 | 68.92 80.21 | 69.94 69.52 | 68.94 75.30 | 72.37 69.76 | 75.89 80.29 | 70.59 70.22 | 71.23 77.87 |
| 1000 | NNGuide | 71.30 | 75.56 | 71.33 | 71.20 | 73.40 | 77.72 | 72.02 | 78.30 | 72.01 | 75.69 | 68.64 | 76.97 | 75.00 | 83.84 | 68.53 | 78.64 | 70.03 | 79.75 | 70.55 | 79.80 |
| 1035 | Average | 70.48 | 70.98 | 72.45 | 70.51 | 74.20 | 75.29 | 73.29 | 75.46 | 72.61 | 73.06 | 67.46 | 71.46 | 71.78 | 77.33 | 68.83 | 72.95 | 70.09 | 76.38 | 69.54 | 74.53 |
| 1100 | -rine-iuning-b | <i>asea 00</i> 74.05 | 71 59 | 1 73 34 | 70.23 | 74.81 | 74.03 | 75 63 | 75 25 | 74 46 | 72 78 | 68 55 | 71.11 | 70.12 | 79 30 | 67 75 | 71.92 | 70.00 | 74 85 | 69.11 | 74 29 |
| | T2FNorm | 74.00 | 72.01 | 72.31 | 71.12 | 74.69 | 74.23 | 75.80 | 75.65 | 74.20 | 73.25 | 68.68 | 73.98 | 70.42 | 81.78 | 69.08 | 74.69 | 69.34 | 76.24 | 69.38 | 76.67 |
| 1101 | AUGMIX | 74.65 | 71.43 | <u>73.06</u> | 69.52 | 75.30 | 73.87 | 76.22 | 75.20 | 74.81 | 72.50 | 67.42 | 79.12 | <u>71.80</u> | 86.56 | 67.22 | 78.09 | 68.86 | 80.72 | 68.83 | 81.12 |
| 1100 | REGMIX | 74.54 | 71.84 | 72.97 | 70.94 | 75.19 | 74.57 | 76.00 | 75.53 | 74.67 | 73.22 | 69.64 | 69.41 | 70.05 | 78.26 | 69.44 | 73.16 | 70.85 | 73.83 | 70.00 | 73.67 |
| 1102 | VOS REP (Ourc) | 74.16 | 74.81 | 70.53 | 67.89 72.40 | 09.55 | 73.55 | 75.17 | 75.01 | 72.55 | 72.19 | 60.12 | 08.59 | 59.60 | 73.30 | 60.57 70.61 | 67.45 | 61.50 70.80 | 73.81 | 60.45 71.01 | 70.80 |
| 1102 | Average | 74.03 | 71.84 | 72.44 | 69.94 | 73.91 | 74.05 | 75.76 | 75.33 | 74.10 | 72.79 | 66.88 | 72.44 | 68.40 | 79.85 | 66.81 | 73.06 | 68.11 | 75.89 | 67.55 | 75.31 |
| 1105 | Average (All) | 71.84 | 71.29 | 72.45 | 70.31 | 74.10 | 74.85 | 74.17 | 75.41 | 73.14 | 72.96 | 67.25 | 71.81 | 70.57 | 78.23 | 68.11 | 72.99 | 69.38 | 76.20 | 68.83 | 74.81 |
| 1104 | AP↑ | iCa | ıRL | В | iC | W | /A | FOS | TER | Ave | rage | iCa | ıRL | Bi | iC | W | Ά | FOS | TER | Ave | rage |
| 1105 | Average CIL a | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far |
| 1100 | Post hos has | $\begin{vmatrix} 62 \\ 400 \\ 1 \end{vmatrix}$ | .65 nathods | 64 | .14 | 68 | .05 | 68. | .75 | 65 | .90 | 48 | .85 | 53 | .30 | 58 | .42 | 57 | .84 | 54 | .60 |
| 1106 | MSP | 64.69 | 77.88 | 66.09 | 77.36 | 68.06 | 79.42 | 67.19 | 79.81 | 66.51 | 78.62 | 26.39 | 25.10 | 29.95 | 27.19 | 27.46 | 28.99 | 28.24 | 33.60 | 28.01 | 28.72 |
| 1107 | ODIN | 68.22 | 76.99 | 68.99 | 75.95 | 70.69 | 78.59 | 69.52 | 79.47 | 69.35 | 77.75 | 29.08 | 32.24 | 31.33 | 37.17 | 29.02 | 31.30 | 29.01 | 36.89 | 29.61 | 34.40 |
| 1107 | Energy | 69.22 | 77.93 | 68.94 | 76.22 | 70.91 | 79.51 | 71.12 | 80.44 | 70.05 | 78.53 | 25.81 | 28.64 | 28.82 | 35.39 | 26.01 | 29.60 | 27.01 | 32.20 | 26.91 | 31.46 |
| 1108 | MaxLogit | 69.11 | 78.06 | 68.98 | 76.48 | 70.94 | 79.75 | 71.03 | 80.58 | 70.02 | 78.72 | 26.38 | 28.39 | 29.68 | 32.27 | 27.10 | 29.57 | 27.87 | 33.98 | 27.76 | 31.05 |
| 1100 | GEN | 66.60 | 79.64 | 68.56 | 78.62 | 70.19 | /9.00 | 70.37 | 80.91 | 68.95 | 78.80 | 27.02 | 27.81 | 30.20 | 26.53 | 25.82 | 27.10 | 28.07 | 30.42 18.63 | 29.10 | 20.75 |
| 1109 | KLM | 65.01 | 79.43 | 65.97 | 79.45 | 67.14 | 80.87 | 67.58 | 81.20 | 66.42 | 80.24 | 29.76 | 26.16 | 30.73 | 24.36 | 29.31 | 23.31 | 34.15 | 28.59 | 30.99 | 25.61 |
| | Relation | 58.89 | 81.39 | 65.14 | 80.71 | 69.00 | 85.23 | 65.80 | 82.40 | 64.71 | 82.43 | 28.20 | 32.80 | 31.11 | 30.76 | 33.80 | 27.37 | 28.26 | 28.12 | 30.34 | 29.76 |
| 1110 | NNGuide | 69.13 | 82.22 | 69.16 | 79.13 | 71.16 | 83.33 | 71.80 | 82.63 | 70.31 | 81.83 | 31.55 | 28.89 | 25.52 | 29.69 | 32.09 | 29.09 | 33.44 | 29.75 | 30.65 | 29.36 |
| 4444 | Average | 00.09 | 79.10 D metho | 67.90 ds | 77.83 | 69.90 | 80.92 | 69.52 | 81.10 | 68.50 | 79.74 | 27.39 | 27.46 | 29.57 | 30.56 | 29.01 | 27.44 | 29.42 | 30.91 | 28.85 | 29.09 |
| 1111 | LogitNorm | 69.65 | 77.80 | 68.22 | 76.29 | 70.29 | 78.78 | 71.03 | 79.71 | 69.80 | 78.14 | 26.76 | 26.25 | 27.62 | 37.77 | 25.39 | 27.70 | 27.39 | 27.84 | 26.79 | 29.89 |
| 1112 | T2FNorm | 69.53 | 78.22 | 67.29 | 77.07 | 70.04 | 79.23 | 71.12 | 80.11 | 69.50 | 78.66 | 26.88 | 29.40 | 27.96 | 41.73 | 26.59 | 30.40 | 27.04 | 29.86 | 27.12 | 32.85 |
| 1112 | AUGMIX | 70.16 | 77.73 | 67.84 | 76.12 | 70.68 | 78.69 | 71.57 | 79.62 | 70.06 | 78.04 | <u>29.41</u> | 31.62 | 28.77 | 28.31 | 27.78 | 27.66 | <u>29.20</u> | 29.28 | <u>28.79</u> | 29.22 |
| 1113 | REGMIX | 70.27 | 78.14 | 68.12 | 76.98 | 71.06 | 79.33 | 71.58 | 79.97 | 70.26 | 78.60 | 28.46 | 22.69 | 28.06 | 29.67 | 27.67 | 25.48 | 29.05 | 24.29 | 28.31 | 25.53 |
| | BER (Ours) | 70.63 | 81.41 | 68.86 | 79.98 | 71.41 | 83.73 | 72.86 | 82.60 | 70.94 | 81.93 | 30.26 | 31.87 | 30.71 | 36.50 | 29.82 | 33.39 | 30.66 | 32.12 | 30.36 | 33.47 |
| 1114 | Average | 69.85 | 78.00 | 67.50 | 76.44 | 69.41 | 79.07 | 71.15 | 79.80 | 69.48 | 78.32 | 26.87 | 27.20 | 26.80 | 33.71 | 25.92 | 27.53 | 27.18 | 28.15 | 26.69 | 29.15 |
| 4445 | Average (All) | 67.82 | 78.71 | 67.76 | 77.33 | 69.72 | 80.26 | 70.10 | 80.64 | 68.85 | 79.23 | 27.20 | 27.37 | 28.58 | 31.68 | 27.91 | 27.47 | 28.62 | 29.92 | 28.08 | 29.11 |
| CIII | FPR | iCa | ıRL | B | iC | W | /A | FOS | TER | Ave | rage | iCa | ıRL | Bi | iC | W | 'A | FOS | TER | Ave | rage |
| 1116 | Anorra CII a | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far | Near | Far |
| | Average CIL a | 62 | .65 | 64 | .14 | 68 | .05 | 68. | .75 | 65 | .90 | 48 | .85 | 53 | .30 | 58 | .42 | 57 | .84 | 54 | .60 |
| 1117 | -Post-hoc-base | d OOD 1 | nethods | | | | | | | | | | | | | | | | | | |
| 1118 | MSP | 86.59 | 87.49 | 86.66 | 88.30 | 84.62 | 84.80 | 85.18 | 84.94 | 85.76 | 86.38 | 88.25 | 85.89 | 86.92 | 86.19 | 89.12 | 83.14 | 87.95 | 77.41 | 88.06 | 83.16 |
| 1110 | Energy | 82 74 | 79.84 81.62 | 84.56 | 81.39 | 82.02 | 77 47 | 81.72 | 76.95 | 82 77 | 79.36 | 90.62 | 78.80 83.67 | 88 72 | 77 34 | 90.36 | 82.46 | 90.03 | 74.00 | 89.93 | 81 17 |
| 1119 | MaxLogit | 82.62 | 81.99 | 84.72 | 83.23 | 81.85 | 79.10 | 81.44 | 77.54 | 82.66 | 80.47 | 89.86 | 84.00 | 87.78 | 82.40 | 89.54 | 82.69 | 89.14 | 78.44 | 89.08 | 81.88 |
| | GEN | 82.86 | 82.23 | 84.43 | 82.28 | 81.67 | 78.23 | 81.24 | 76.96 | 82.55 | 79.92 | 89.24 | 84.32 | 87.31 | 82.91 | 85.46 | 86.07 | 88.44 | 76.02 | 87.61 | 82.33 |
| 1120 | ReAct | 84.18 | 82.40 | 84.02 | 82.30 | 81.51 | 76.98 | 81.86 | 77.35 | 82.89 | 79.76 | 92.37 | 92.98 | 86.31 | 85.15 | 88.68 | 87.47 | 87.89 | 91.61 | 88.81 | 89.30 |
| 1101 | Relation | 87.14 | 76.55 | 84.52 | 76.94 | 83.35 | 74.51 | 83.91 | 72.84 | 84.73 | 75.21 | 89.50 | 82.30 | 86.36 | 77.31 | 93.46 | 91.33 | 88.11 | 73.72 | 89.36 | 81.16 |
| 1121 | NNGuide | 82.70 | 74.23 | 84.31 | 76.17 | 81.69 | 69.83 | 85.19 | 71.60 | 83.47 | 72.96 | 91.91 | 80.27 | 91.24 | 77.11 | 91.99 | 76.72 | 87.52 | 73.19 | 90.66 | 76.82 |
| 1122 | Average | 84.35 | 81.27 | 84.91 | 81.80 | 82.78 | 78.11 | 83.10 | 77.35 | 83.78 | 79.63 | 89.16 | 84.21 | 87.30 | 81.23 | 89.09 | 84.43 | 87.45 | 78.98 | 88.25 | 82.21 |
| 1 1 An An | Fine-tuning-bo | <i>sed OO</i> | D method | ds 84 77 | 92.21 | 1 82 22 | 78.12 | 9197 | 77 08 | 1 82 87 | 80.40 | 00.18 | 87.16 | 80.50 | 72.00 | 01.40 | 85 72 | 00.21 | 97 79 | 00.20 | 82.20 |
| 1123 | T2FNorm | 82.63 | 62.17 82.28 | 85.28 | 83.05 | 82.63 | 79.67 | 81.59 | 77.98 78.60 | 83.03 | 80.40 80.90 | 90.18 89.78 | 82.85 | 89.59 | <u>12.99</u> 68.00 | 90.11 | 6J./J 81.85 | 90.31 89.61 | 07.28 83.92 | 90.39 89.69 | 65.29 79.16 |
| | AUGMIX | 82.34 | 82.77 | 84.96 | 84.02 | 81.91 | 78.99 | 81.34 | 78.74 | 82.64 | 81.13 | 89.08 | 85.59 | 88.27 | 77.35 | 92.17 | 77.26 | 88.27 | 74.44 | 89.45 | 78.66 |
| 1124 | REGMIX | 81.88 | 83.27 | 85.08 | 84.43 | <u>81.84</u> | 79.75 | 81.07 | 79.61 | 82.47 | 81.77 | 88.40 | 92.04 | 88.75 | 85.22 | 88.44 | 87.50 | 87.88 | 90.78 | 88.37 | 88.88 |
| 1105 | VOS | 82.83 | 80.67 | 85.24 | 74.80 | 86.31 | 71.70 | 82.50 | 76.54 | 84.22 | 75.93 | 91.28 | 86.32 | 88.06 | 88.46 | 91.31 | 84.44 | 90.04 | 84.19 | 90.17 | 85.85 |
| 1120 | Average | 82.10 | 82.23 | 85.96 | 81.97 | 82.98 | 77.65 | 81.67 | 78.29 | 82.85 | 80.03 | 89 74 | 86 79 | 88 78 | 78.40 | 85.40 90.70 | 83.36 | 89.22 | 84.12 | 89.61 | 83.17 |
| 1126 | Average (All) | 83.67 | 81.61 | 84.97 | 81.84 | 82.85 | 77.95 | 82.59 | 77.69 | 83.52 | 79.77 | 89.37 | 85.13 | 87.83 | 80.22 | 89.66 | 84.05 | 88.08 | 80.82 | 88.74 | 82.55 |
| | | | | | | | | - | | | | | | | | | | | | - | |

¹¹³⁴ F THE OPENCIL BENCHMARK FRAMEWORK AND THE BER ALGORITHM

The full framework of the OpenCIL Benchmark for post-hoc-based OOD methods are given in Algorithm 1, and that for the fine-tuning-based OOD methods is given in Algorithm 2 below. Furthermore, the algorithm of our proposed method BER is given in Algorithm 3.

```
1140
           Algorithm 1 : Post-hoc-based OOD Detection Framework in the OpenCIL Benchmark
1141
           Class Incremental Training
1142
           Data: A data memory M; A sequence of c tasks ID data T = {T_1, T_2, ..., T_c}, T_t =
           (X_t^{train}, X_t^{test}, Y_t)
1143
1144
            1: for each task do
1145
            2:
                   Obtain t-th (1 \le t \le c) task training ID data T_t^{train} = (X_t^{train}, Y_t) \cup M
1146
            3:
                   for each iteration do
            4:
                       Sample a mini-batch of ID training data from T_t^{train}
1147
            5:
                       Perform different CIL algorithms
1148
            6:
                   end for
1149
                   Update data memory M //*for replay-based CIL models only*
            7:
1150
                   Save current CIL model \theta_t(\cdot)
            8:
1151
            9: end for
1152
           Output: A well-trained CIL model \theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot), ..., \theta_c(\cdot)\}
1153
1154
           Inference
1155
          Input: A well-trained CIL model \theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot), ..., \theta_c(\cdot)\}
Data: A sequence of c tasks OOD data X^{ood} = \{X_1^{ood}, X_2^{ood}, ..., X_c^{ood}\}; A sequence of c tasks ID data T = \{T_1, T_2, ..., T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)
1156
1157
1158
            1: for each task do
                   Obtain t-th (1 \le t \le c) task testing ID data T_t^{test} = X_1^{test} \cup X_2^{test} \cup ... \cup X_t^{test}
Obtain t-th (1 \le t \le c) task testing OOD data X_t^{ood}
1159
            2:
1160
            3:
                   Perform ID classification on CIL model \theta_t(\cdot) based on T_t^{test}
1161
            4:
                   Perform different post-hoc OOD detection methods on CIL model \theta_t(\cdot) based on T_t^{test} \cup X_t^{ood}
1162
            5:
            6: end for
1163
           Output: Average incremental accuracy; Average OOD detection performance
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
```

| Algorithm 2 : Fine-tuning-based OOD Detection Framework in the OpenCIL Benchmark Class Incremental Training Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}$, T_t ($X_t^{train}, X_t^{test}, Y_t$) 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: for each iteration do 4: Sample a mini-batch of ID training data from T_t^{train} 5: Perform different CIL algorithms 6: end for 7: Update data memory M //*for replay-based CIL models only* 8: Save current CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ OUtput: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}$, T_t ($X_t^{train}, X_t^{test}, Y_t$) If or each task do Cobtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ Sace care task do Cobtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ Sace care task do Cobtain t -th $(1 \le t \le c)$ task training ID data T_t^{train} | | |
|---|----------------|---|
| Algorithm 2 : Fine-tuning-based OOD Detection Framework in the OpenCIL BenchmarkClass Incremental TrainingData: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: for each iteration do4: Sample a mini-batch of ID training data from T_t^{train} 5: Perform different CIL algorithms6: end for7: Update data memory M //*for replay-based CIL models only*8: Save current CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Output: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 4: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do6: for each iteration do7: for each iteration do6: for each iteration do7: for each iteration do7: for each iteration d | | |
| Algorithm 2 : Fine-tuning-based OOD Detection Framework in the OpenCIL BenchmarkClass Incremental Training Data: A data memory M; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do2: Obtain t-th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: for each iteration do4: Sample a mini-batch of ID training data from T_t^{train} 5: Perform different CIL algorithms6: end for7: Update data memory M //*for replay-based CIL models only*8: Save current CIL model $\theta_t(\cdot)$ 9: end forOutput: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do2: Obtain t-th CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do2: Obtain t-th CIL model $\theta(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(0)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $h_t(\cdot)$ 5: for each iteration do6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 7: Gupta data memory M //*for replay-based CIL models | | |
| Algorithm 2 : Fine-tuning-based OOD Detection Framework in the OpenCIL BenchmarkClass Incremental TrainingData: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: for each iteration do4: Sample a mini-batch of ID training data from T_t^{train} 5: Perform different CIL algorithms6: end for7: Update data memory M //*for replay-based CIL models only*8: Save current CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Output: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do2: Obtain t -th CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do2: Obtain t -th CIL model $\theta(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ ad a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for10: Update data memory M //*for replay-based CIL models only*11: Save current CIL model $\theta_t(\cdot$ | | |
| Agointim 2.1 intertaining Dataci CoD Detection Frankwork in the Optic Detection Function X Class Incremental Training Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: for each iteration do 4: Sample a mini-batch of ID training data from T_t^{train} 5: Perform different CIL algorithms 6: end for 7: Update data memory M //*for replay-based CIL models only* 8: Save current CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Output: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier ($\theta(\cdot), f(\cdot)$) | λΙα | arithm 2 : Fine tuning based OOD Detection Framework in the OpenCII Benchmark |
| Class inderential training Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: for each iteration do 4: Sample a mini-batch of ID training data from T_t^{train} 5: Perform different CIL algorithms 6: end for 7: Update data memory M //*for replay-based CIL models only* 8: Save current CIL model $\theta_t(\cdot)$ 9: end for Output: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Detat: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training D data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 0: Update data memory M //*for replay-based CIL | Cloc | Incremental Training |
| $\begin{aligned} & \text{Statis} \ R \text{ dual } \mathbf{T} = (T_1, T_2,, T_{c_1}, T_t) \\ & \mathbf{T}_t^{train}, \mathbf{X}_t^{test}, \mathbf{Y}_t) \\ & \mathbf{T}_t^{train}, \mathbf{X}_t^{test}, \mathbf{Y}_t) \\ & \text{Statis} \ \mathbf{T}_t^{train} = (\mathbf{X}_t^{train}, \mathbf{Y}_t) \cup M \\ & \text{Stample a mini-batch of ID training data from } T_t^{train} \\ & \text{Sample a mini-batch of ID training data from } T_t^{train} \\ & \text{Sample a mini-batch of ID training data from } T_t^{train} \\ & \text{Stample a mini-batch of ID training data from } T_t^{train} \\ & \text{Sample a mini-batch of ID training data from } T_t^{train} \\ & \text{Sample a mini-batch of ID training data from } T_t^{train} \\ & \text{Sample a mini-batch of ID training data from } T_t^{train} \\ & \text{Save current CIL model } \theta_t(\cdot) \\ & \text{9: end for} \\ & \text{Output: A well-trained CIL model } \theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\} \\ & \text{Dotat A data memory } M; \ A \text{ sequence of } c \text{ tasks ID data } T = \{T_1, T_2,, T_c\}, \ T_t \\ & (X_t^{train}, X_t^{test}, Y_t) \\ & \text{1: for each task do} \\ & \text{2: Obtain } t + \text{th } (1 \leq t \leq c) \text{ task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ & \text{3: Obtain } t + \text{th } (1 \leq t \leq c) \text{ task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ & \text{3: Obtain } t + \text{th } (I \leq t \leq c) \text{ task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ & \text{3: Obtain } t + \text{th } (I \leq t \leq c) \text{ task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ & \text{3: Obtain } t + \text{th } (I \leq t \leq c) \text{ task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ & \text{3: Obtain } t + \text{th } (I \leq t \leq c) \text{ task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ & \text{3: Obtain } t + \text{th } (I \in t \leq c) \text{ task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ & \text{3: Obtain } t + \text{th } (I \in t \leq c) \text{ task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ & \text{3: Obtain } t + \text{th } (I \in t \leq c) \text{ task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ & \text{3: Obtain } t + \text{th } (I \in t \in c) \text{task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ & \text{3: Obtain } t + $ | Cias Date | as A data memory M: A sequence of c tasks ID data $T = \{T_1, T_2, \dots, T_n\}$ |
| 1: for each task do 2: Obtain t-th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: for each iteration do 4: Sample a mini-batch of ID training data from T_t^{train} 5: Perform different CIL algorithms 6: end for 7: Update data memory M //*for replay-based CIL models only* 8: Save current CIL model $\theta_t(\cdot)$ 9: end for Dutput: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ DOD Method - Fine-tuning Input: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t-th ($1 \le t \le c$) task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t-th CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 2: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 2: end for 2: Dutput: A well-trained CIL model (the same as Input) with a finetuned classifier ($\theta(\cdot), f(\cdot)$) | Xti | T_{rain} X test Y_{1} |
| 1. In the dath as do 2. Obtain t-th ($1 \le t \le c$) task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3. for each iteration do 4. Sample a mini-batch of ID training data from T_t^{train} 5. Perform different CIL algorithms 6. end for 7. Update data memory M //*for replay-based CIL models only* 8. Save current CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Output: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Dutput: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ ($X_t^{train}, X_t^{test}, Y_t$) 1: for each task do 2: Obtain t-th ($1 \le t \le c$) task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t-th ($1 \le t \le c$) task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t-th ($1 \le t \le c$) task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t-th ($1 \le t \le c$) task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t-th ($1 \le t \le c$) task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t-th ($1 \le t \le c$) task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t-th ($1 \le t \le c$) task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t-th ($1 \le t \le c$) task training ID data from T_t^{train} 3: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier ($\theta(\cdot), f(\cdot)$) | $\frac{1}{1}$ | for each task do |
| a) for each iteration do a) for each iteration do b) the term of term of the term of term of term of the term of ter | 2: | Obtain t-th $(1 \le t \le c)$ task training ID data $T_{train}^{train} = (X_{train}^{train}, Y_t) \cup M$ |
| 4: Sample a mini-batch of ID training data from T_t^{train} 5: Perform different CIL algorithms 6: end for 7: Update data memory M //*for replay-based CIL models only* 8: Save current CIL model $\theta_t(\cdot)$ 9: end for Dutput: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ DOD Method - Fine-tuning Input: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Dota: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Dutput: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | 3: | for each iteration do |
| 5: Perform different CIL algorithms 6: end for 7: Update data memory M //*for replay-based CIL models only* 8: Save current CIL model $\theta_t(\cdot)$ 9: end for Output: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ OOD Method - Fine-tuning Input: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ ($X_t^{train}, X_t^{test}, Y_t$) 1: for each task do 2: Obtain t-th (I $\leq t \leq c$) task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t-th CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier ($\theta(\cdot), f(\cdot)$) | 4: | Sample a mini-batch of ID training data from T_t^{train} |
| 6: end for 7: Update data memory M //*for replay-based CIL models only* 8: Save current CIL model θ_t(·) 9: end for Output: A well-trained CIL model θ(·) = {θ₁(·), θ₂(·),, θ_c(·)} OOD Method - Fine-tuning Input: A well-trained CIL model θ(·) = {θ₁(·), θ₂(·),, θ_c(·)} Data: A data memory M; A sequence of c tasks ID data T = {T₁, T₂,, T_c}, T_t (X^{train}_t, X^{test}_t, Y_t) 1: for each task do 2: Obtain t-th (1 ≤ t ≤ c) task training ID data T^{train}_t = (X^{train}_t, Y_t) ∪ M 3: Obtain t-th CIL model θ_t(·) = {φ_t(·), h_t(·)} which is composed of a feature extractor φ_t(and a original classifier h_t(·) 4: Initialize an extra classifier f_t(·) 5: for each iteration do 6: freeze the φ_t(·) and h_t(·) 7: Sample a mini-batch of ID training data from T^{train}_t 8: Perform different training-time OOD detection method to finetune this extra classifier f_t(·) only on top of φ_t(·). 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model θ_t(·) with extra classifier f_t(·) 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier (θ(·), f(·)) | 5: | Perform different CIL algorithms |
| 7: Update data memory $M = //*for replay-based CIL models only* 8: Save current CIL model \theta_t(\cdot)9: end forOutput: A well-trained CIL model \theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}OOD Method - Fine-tuningInput: A well-trained CIL model \theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}Data: A data memory M; A sequence of c tasks ID data T = \{T_1, T_2,, T_c\}, T_t(X_t^{train}, X_t^{test}, Y_t)1: for each task do2: Obtain t-th (1 \le t \le c) task training ID data T_t^{train} = (X_t^{train}, Y_t) \cup M3: Obtain t-th CIL model \theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\} which is composed of a feature extractor \phi_t(\cdot)and a original classifier h_t(\cdot)4: Initialize an extra classifier f_t(\cdot)5: for each iteration do6: freeze the \phi_t(\cdot) and h_t(\cdot)7: Sample a mini-batch of ID training data from T_t^{train}8: Perform different training-time OOD detection method to finetune this extra classifier f_t(\cdot)0: update data memory M = //*for replay-based CIL models only*11: Save current CIL model \theta_t(\cdot) with extra classifier f_t(\cdot)12: end forOutput: A well-trained CIL model (the same as Input) with a finetuned classifier (\theta(\cdot), f(\cdot))$ | 6: | end for |
| 8: Save current CIL model $\theta_t(\cdot)$ 9: end for Output: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ DOD Method - Fine-tuning Input: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(IL model \theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | 7: | Update data memory M //*for replay-based CIL models only* |
| 9: end for Output: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ DOD Method - Fine-tuning Input: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th $(IL model \theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | 8: | Save current CIL model $\theta_t(\cdot)$ |
| Output: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Input: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ ($X_t^{train}, X_t^{test}, Y_t$) 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot) = t(\cdot)$ | 9: (| |
| OOD Method - Fine-tuning Input: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | Out | put: A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ |
| Input : A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot),, \theta_c(\cdot)\}$ Data : A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ ($X_t^{train}, X_t^{test}, Y_t$) 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for 10: Update data memory M <i>#for replay-based CIL models only</i> * 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | 001 | D Method - Fine-tuning |
| Data: A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t$ $(X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ 3: Obtain t -th CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ 4: Initialize an extra classifier $f_t(\cdot)$ 5: for each iteration do 6: freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier $f_t(\cdot)$ 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | Inni | if A well-trained CIL model $\theta(\cdot) = \{\theta_1(\cdot), \theta_2(\cdot), \theta_3(\cdot)\}$ |
| $\begin{aligned} &(X_t^{train}, X_t^{test}, Y_t) \\ &1: \text{ for each task do} \\ &2: \text{Obtain } t\text{-th } (1 \leq t \leq c) \text{ task training ID data } T_t^{train} = (X_t^{train}, Y_t) \cup M \\ &3: \text{Obtain } t\text{-th CIL model } \theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\} \text{ which is composed of a feature extractor } \phi_t(\cdot) \\ ∧ a \text{ original classifier } h_t(\cdot) \\ &4: \text{Initialize an extra classifier } f_t(\cdot) \\ &5: \text{for each iteration do} \\ &6: \text{freeze the } \phi_t(\cdot) \text{ and } h_t(\cdot) \\ &7: \text{Sample a mini-batch of ID training data from } T_t^{train} \\ &8: \text{Perform different training-time OOD detection method to finetune this extra classifier } f_t(\cdot) \\ &9: \text{end for} \\ &10: \text{Update data memory } M //*for replay-based CIL models only* \\ &11: \text{Save current CIL model } \theta_t(\cdot) \text{ with extra classifier } f_t(\cdot) \\ &12: \text{ end for} \\ & \text{Output: A well-trained CIL model (the same as Input) with a finetuned classifier } (\theta(\cdot), f(\cdot)) \\ &12: \left(\theta_t(t), \theta_t(t)\right) = \left(\theta_t(t), \theta_t(t)\right) \\ &13: \theta_t(t) = f_t(t) \\ &13: \theta_t(t) \\ &13: \theta_t(t) = f_t(t) \\ &13: \theta_t(t) \\ &13: \theta_t(t)$ | Data | a : A data memory M: A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t =$ |
| for each task do Obtain t-th (1 ≤ t ≤ c) task training ID data T_t^{train} = (X_t^{train}, Y_t) ∪ M Obtain t-th CIL model θ_t(·) = {φ_t(·), h_t(·)} which is composed of a feature extractor φ_t() and a original classifier h_t(·) Initialize an extra classifier f_t(·) Initialize an extra classifier f_t(·) for each iteration do freeze the φ_t(·) and h_t(·) Sample a mini-batch of ID training data from T_t^{train} Perform different training-time OOD detection method to finetune this extra classifier f_t() only on top of φ_t(·). end for Update data memory M //*for replay-based CIL models only* Save current CIL model θ_t(·) with extra classifier f_t(·) and for Output: A well-trained CIL model (the same as Input) with a finetuned classifier (θ(·), f(·)) | $(X_{t}^{tr}$ | $(r_1, r_2, \dots, r_c), r_t$ |
| Obtain t-th (1 ≤ t ≤ c) task training ID data T^{train}_t = (X^{train}_t, Y_t) ∪ M Obtain t-th CIL model θ_t(·) = {φ_t(·), h_t(·)} which is composed of a feature extractor φ_t(and a original classifier h_t(·) Initialize an extra classifier f_t(·) for each iteration do freeze the φ_t(·) and h_t(·) Sample a mini-batch of ID training data from T^{train}_t Perform different training-time OOD detection method to finetune this extra classifier f_t(only on top of φ_t(·). end for Update data memory M //*for replay-based CIL models only* Save current CIL model θ_t(·) with extra classifier f_t(·) end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier (θ(·), f(·)) | 1: 1 | for each task do |
| 3: Obtain t-th CIL model θ_t(·) = {φ_t(·), h_t(·)} which is composed of a feature extractor φ_t(and a original classifier h_t(·) 4: Initialize an extra classifier f_t(·) 5: for each iteration do 6: freeze the φ_t(·) and h_t(·) 7: Sample a mini-batch of ID training data from T^{train}_t 8: Perform different training-time OOD detection method to finetune this extra classifier f_t(only on top of φ_t(·). 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model θ_t(·) with extra classifier f_t(·) 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier (θ(·), f(·)) | 2: | Obtain t-th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ |
| and a original classifier h_t(·) 4: Initialize an extra classifier f_t(·) 5: for each iteration do 6: freeze the φ_t(·) and h_t(·) 7: Sample a mini-batch of ID training data from T^{train}_t 8: Perform different training-time OOD detection method to finetune this extra classifier f_t(only on top of φ_t(·). 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model θ_t(·) with extra classifier f_t(·) 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier (θ(·), f(·)) | 3: | Obtain t-th CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ |
| 4: Initialize an extra classifier f_t(·) 5: for each iteration do 6: freeze the φ_t(·) and h_t(·) 7: Sample a mini-batch of ID training data from T_t^{train} 8: Perform different training-time OOD detection method to finetune this extra classifier f_t(only on top of φ_t(·). 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model θ_t(·) with extra classifier f_t(·) 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier (θ(·), f(·)) | | and a original classifier $h_t(\cdot)$ |
| 5: for each iteration do 6: freeze the φ_t(·) and h_t(·) 7: Sample a mini-batch of ID training data from T^{train}_t 8: Perform different training-time OOD detection method to finetune this extra classifier f_t(only on top of φ_t(·). 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model θ_t(·) with extra classifier f_t(·) 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier (θ(·), f(·)) | 4: | Initialize an extra classifier $f_t(\cdot)$ |
| 6: freeze the φ_t(·) and h_t(·) 7: Sample a mini-batch of ID training data from T^{train}_t 8: Perform different training-time OOD detection method to finetune this extra classifier f_t() only on top of φ_t(·). 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model θ_t(·) with extra classifier f_t(·) 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier (θ(·), f(·)) | 5: | for each iteration do |
| 8: Perform different training-time OOD detection method to finetune this extra classifier f_t (only on top of φ_t(·). 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model θ_t(·) with extra classifier f_t(·) 12: end for Dutput: A well-trained CIL model (the same as Input) with a finetuned classifier (θ(·), f(·)) | 6: | freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ |
| 8. Perform different training-time OOD detection method to method | /: o. | Sample a mini-batch of 1D training data from T_t^{t} and T_t^{t} |
| 9: end for 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | 0. | only on top of $\phi_t(\cdot)$ |
| 10: Update data memory M //*for replay-based CIL models only* 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | 9٠ | end for |
| 11: Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | 10: | Update data memory M //*for replay-based CIL models only* |
| 12: end for Output: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | 11: | Save current CIL model $\theta_t(\cdot)$ with extra classifier $f_t(\cdot)$ |
| Output: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot))$ | 12: | end for |
| $\left(\begin{array}{c} \left(\begin{array}{c} 0 \end{array}\right) \\ \left(\begin{array}{c} 0 \end{array}\right)$ | Out | put : A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot)) =$ |
| $\{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)), \dots, (\theta_c(\cdot), f_c(\cdot))\}$ | $\{(\theta_1)$ | $(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)), \dots, (\theta_c(\cdot), f_c(\cdot)))$ |
| | Inne | rence ut: A finetuned CII model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)), (\theta_1(\cdot), f_1(\cdot))\}$ |
| Interence Input: A finetuned CII model $(\theta(x), f(y)) = \{(\theta_1(y), f_1(y)), (\theta_2(y), f_2(y))\}$ $(\theta_1(y), f_2(y))\}$ | Data | a : A sequence of c tasks OOD data $X^{ood} = \{X^{ood}, X^{ood}, \dots, X^{ood}\}$: A sequence of c tasks I |
| Interence Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of c tasks OOD data $X^{ood} = \{X^{ood}, X^{ood}, X^{ood}\}$: A sequence of c tasks l | data | $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ |
| Interence Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of c tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of c tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ | 1. | for each task do |
| Interence Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of c tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of c tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do | 2: | Obtain t-th $(1 \le t \le c)$ task testing ID data $T_{t}^{test} = X_{1}^{test} \cup X_{2}^{test} \cup \cup X_{t}^{test}$ |
| Interence Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of c tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of c tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t-th $(1 \le t \le c)$ task testing ID data $T_t^{test} = X_t^{test} \cup X_t^{test} \cup \cup X_t^{test}$ | 3: | Obtain t-th $(1 \le t \le c)$ task testing OOD data X_t^{ood} |
| Interence Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of c tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of c tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t-th $(1 \le t \le c)$ task testing ID data $T_t^{test} = X_1^{test} \cup X_2^{test} \cup \cup X_t^{test}$ 3: Obtain t-th $(1 \le t \le c)$ task testing OOD data X_t^{ood} | 4: | Perform ID classification on original CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ based on T_t^{test} |
| Inference Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of c tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of c tasks b data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t-th $(1 \le t \le c)$ task testing ID data $T_t^{test} = X_1^{test} \cup X_2^{test} \cup \cup X_t^{test}$ 3: Obtain t-th $(1 \le t \le c)$ task testing OOD data X_t^{ood} 4: Perform ID classification on original CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ based on T_t^{test} | 5: | Perform different post-hoc OOD scoring function on finetuned classifier $\{\phi_t(\cdot), f_t(\cdot)\}$ base |
| Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of c tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of c tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t -th $(1 \le t \le c)$ task testing ID data $T_t^{test} = X_1^{test} \cup X_2^{test} \cup \cup X_t^{test}$ 3: Obtain t -th $(1 \le t \le c)$ task testing OOD data X_t^{ood} 4: Perform ID classification on original CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ based on T_t^{test} 5: Perform different post-hoc OOD scoring function on finetuned classifier $\{\phi_t(\cdot), f_t(\cdot)\}$ based | | on $T_t^{test} \cup X_t^{ood}$ |
| Interence Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of <i>c</i> tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of <i>c</i> tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain <i>t</i> -th $(1 \le t \le c)$ task testing ID data $T_t^{test} = X_1^{test} \cup X_2^{test} \cup \cup X_t^{test}$ 3: Obtain <i>t</i> -th $(1 \le t \le c)$ task testing OOD data X_t^{ood} 4: Perform ID classification on original CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ based on T_t^{test} 5: Perform different post-hoc OOD scoring function on finetuned classifier $\{\phi_t(\cdot), f_t(\cdot)\}$ based on $T_t^{test} \cup X_t^{ood}$ | 6: | end for |
| Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of c tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of c tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t-th $(1 \le t \le c)$ task testing ID data $T_t^{test} = X_1^{test} \cup X_2^{test} \cup \cup X_t^{test}$ 3: Obtain t-th $(1 \le t \le c)$ task testing OOD data X_t^{ood} 4: Perform ID classification on original CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ based on T_t^{test} 5: Perform different post-hoc OOD scoring function on finetuned classifier $\{\phi_t(\cdot), f_t(\cdot)\}$ based of $T_t^{test} \cup X_t^{ood}$ 6: end for | Out | put: Average incremental accuracy; Average OOD detection performance |
| Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of c tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of c tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t-th $(1 \le t \le c)$ task testing ID data $T_t^{test} = X_1^{test} \cup X_2^{test} \cup \cup X_t^{test}$ 3: Obtain t-th $(1 \le t \le c)$ task testing OOD data X_t^{ood} 4: Perform ID classification on original CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ based on T_t^{test} 5: Perform different post-hoc OOD scoring function on finetuned classifier $\{\phi_t(\cdot), f_t(\cdot)\}$ base 6: end for Output: Average incremental accuracy; Average OOD detection performance | | · - * |
| Interence Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of c tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of c tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain t-th $(1 \le t \le c)$ task testing ID data $T_t^{test} = X_1^{test} \cup X_2^{test} \cup \cup X_t^{test}$ 3: Obtain t-th $(1 \le t \le c)$ task testing OOD data X_t^{ood} 4: Perform ID classification on original CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ based on T_t^{test} 5: Perform different post-hoc OOD scoring function on finetuned classifier $\{\phi_t(\cdot), f_t(\cdot)\}$ based of $T_t^{test} \cup X_t^{ood}$ 6: end for Output: Average incremental accuracy; Average OOD detection performance | | |
| Interence Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of <i>c</i> tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of <i>c</i> tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain <i>t</i> -th $(1 \le t \le c)$ task testing ID data $T_t^{test} = X_1^{test} \cup X_2^{test} \cup \cup X_t^{test}$ 3: Obtain <i>t</i> -th $(1 \le t \le c)$ task testing OOD data X_t^{ood} 4: Perform ID classification on original CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ based on T_t^{test} 5: Perform different post-hoc OOD scoring function on finetuned classifier $\{\phi_t(\cdot), f_t(\cdot)\}$ based 6: end for Output: Average incremental accuracy; Average OOD detection performance | | |
| Interence Input: A finetuned CIL model $(\theta(\cdot), f(\cdot)) = \{(\theta_1(\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot))\}$ Data: A sequence of <i>c</i> tasks OOD data $X^{ood} = \{X_1^{ood}, X_2^{ood},, X_c^{ood}\}$; A sequence of <i>c</i> tasks I data $T = \{T_1, T_2,, T_c\}, T_t = (X_t^{train}, X_t^{test}, Y_t)$ 1: for each task do 2: Obtain <i>t</i> -th $(1 \le t \le c)$ task testing ID data $T_t^{test} = X_1^{test} \cup X_2^{test} \cup \cup X_t^{test}$ 3: Obtain <i>t</i> -th $(1 \le t \le c)$ task testing OOD data X_t^{ood} 4: Perform ID classification on original CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ based on T_t^{test} 5: Perform different post-hoc OOD scoring function on finetuned classifier $\{\phi_t(\cdot), f_t(\cdot)\}$ based 6: end for Output: Average incremental accuracy; Average OOD detection performance | | |

| lgo | rithm 3 : Bi-directional Energy Regularization (BER) |
|-------------------|--|
| ata X_t^{tr} | The A data memory M ; A sequence of c tasks ID data $T = \{T_1, T_2,, T_c\}, T_t = ain, X_t^{test}, Y_t\}$ for each task do Obtain t -th $(1 \le t \le c)$ task training ID data $T_t^{train} = (X_t^{train}, Y_t) \cup M$ Obtain t -th CIL model $\theta_t(\cdot) = \{\phi_t(\cdot), h_t(\cdot)\}$ which is composed of a feature extractor $\phi_t(\cdot)$ and a original classifier $h_t(\cdot)$ |
| 4: | Initialize an extra classifier $f_t(\cdot)$ |
| 5: | for each iteration do |
| 6: | Freeze the $\phi_t(\cdot)$ and $h_t(\cdot)$ |
| | Sample a mini-batch of current task ID training data $\{(x_i^i, y_i^i)\}^B$ from (X_i^{train}, Y_i) |
| : | Separate the current task batch into two identical parts with equal size |
|): | Conduct mixup on the second part based on Eq.1 |
|): | Apply energy regularization on these two parts based on Eq.2 with extra classifier $f_t(\cdot)$ |
| 1: | Sample a mini-batch of old task ID training data $\{(x_i^i, y_i^i)\}_{i=1}^B$, $(1 \le o \le t)$ from M |
| 2: | Conduct mixup between the old task and current task ID training data based on Eq.3 |
| 3: | Apply energy regularization on this mixed data based on Eq.4 with extra classifier $f_t(\cdot)$ |
| 4: | end for |
| 5: | Update data memory M |
| 6: | Save current CIL model $	heta_t(\cdot)$ with extra classifier $f_t(\cdot)$ |
| 7: (| end for |
| Duty $[(\theta_1$ | put: A well-trained CIL model (the same as Input) with a finetuned classifier $(\theta(\cdot), f(\cdot)) = (\cdot), f_1(\cdot)), (\theta_2(\cdot), f_2(\cdot)),, (\theta_c(\cdot), f_c(\cdot)) \}$ |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |