We deeply appreciate the reviewers' careful comments. We hope all concerns can be resolved through our clarifications.

<Reviewer 1>

Q: I'd recommend picking an OOD detection threshould at 95% TPR for a more even comparison.

A: Thank you for your great suggestion. Previously we set the threshold at 0.5 as a default value for binary classification. Following your suggestion, we re-performed the experiments by setting the threshold at 95% TPR. More specifically, we first meta-trained OOD-MAML and chose 1000 different OOD-detection tasks from $D_{meta-train}$, for each of which we adapted our base classifier and then calculated in-dist probability for each of positive instances (i.e., in-dist samples) in the test data. Finally, based on all calculated in-dist probabilities, we picked 95% TPR threshold. In this way, we obtained the threshold as larger than 0.5, which allowed a tighter decision boundary for in-dist samples (0.9892 (Omniglot), 0.6183 (CIFAR-FS), 0.5255 (*mini*Imagenet)). With the new threshold, we could ensure more even comparison, and even could improve the performance. The following table compares the new (first row) and previous (second row) results. We will incorporate these new results in the final version.

	Omniglot		CIFAR-FS		*mini*ImageNet	
	detect.acc	TNR	detect.acc	TNR	detect.acc	TNR
OOD-MAML (M=3, threshold at 95% TPR)	0.9712 (0.0297)	0.9924 (0.0224)	0.6752 (0.0738)	0.5491 (0.1250)	0.6207 (0.0753)	0.6770 (0.1182)
OOD-MAML (M=3, threshold=0.5)	0.9683 (0.0339)	0.9380 (0.0674)	0.6637 (0.0737)	0.4558 (0.1295)	0.6218 (0.1099)	0.6386 (0.1204)

Q: The table in Appendix D shows the opposite trend (a typo perhaps?) of the text.

A: We are sorry that it is a typo: the name of the second and third rows should be switched, i.e., the second row is for "random-(*ini*)-OOD-MAML" and the third row is for "random-OOD-MAML." We will fix this typo in the final version.

Q: The reason for the decreased performance upon increasing the number of negative samples (M=3 to 5).

A: It is possible that overfitting led to decreased performance in testing phase, as more parameters are used when M=5. By the way, we found that there is a typo for *mini*ImageNet results in Table 1: TNR of OOD-MAML with M=5 should be 0.6372 (0.1196). We will fix the type in the final version.

Q: How were the hyper-parameters chosen?

A: For OOD-MAML, we heuristically chose the hyper-parameters by evaluating meta-training loss. For other methods, we chose them according to the settings reported in the MAML paper. We will clarify this in the final version.

Q: It should be more cautious to claim L45, L61 without some evidence.

A: We understand your concern. In the final version, we will modify L45 as "the algorithms in previous studies are not designed for few-shot settings." For L61, we actually checked this behavior (i.e. MAML generates trivial classifiers for OOD detection) by running experiments. We will add these experimental results in Appendix in the final version.

<Reviewer 2>

Q: How to avoid the issue of the vanishing gradient and mode collapsing of the proposed method is less presented.

A: We used sign-gradient of adversarial loss, which provides the direction for adversarial learning, and meta-SGD, which provides the amount of perturbation (L185-189). By using them in combination for adapting θ_{fake}, we found that both vanishing gradient and mode collapsing issues (see Figure 2(c): different adaptation results) could be avoided. We will discuss this in more detail in the method section in the final version.

Q: Related work of the open-set problem should be reported and compared with.

A: Surely, we will discuss the related work of open-set studies (e.g., [1,2]) in the final version.

[1] Boult, Terrance E., et al. (2019) "Learning and the unknown: Surveying steps toward open world recognition."

[2] Sehwag, Vikash, et al. (2019) "Analyzing the robustness of open-world machine learning."

<Reviewer 3>

Q: MAML, not designed for OOD detection, is not fair enough for comparison. Do you have any other baselines, e.g., many-shot OOD detection algorithms?

A: We agree that direct comparison with MAML is not fair enough because it is not designed for OOD detection; our comparison with MAML is rather to show that we effectively extend MAML for ODD detection problems. In Table 1, we have ODIN and MAH as other baselines of OOD detection methods. These methods require a pre-trained classifier from many-shot training data in general and can be considered as many-shot OOD detection algorithms. For fair comparison, we used an adapted classifier of MAML as the pre-trained classifier for ODIN and MAH.

<Reviewer 4>

Q: Need more clear discussion and comparison to Meta-GAN (lack of a generator for adv examples).

A: Fundamentally, OOD-MAML and MetaGAN have different objectives in meta-training phase. In MetaGAN, GAN is (meta-) trained to generate adversarial samples across all tasks, while meta parameters for initial base model are (meta-) trained to classify the instances in test data set after adaptation. Thus, the parameters for GAN and base model are trained with different objectives. In contrast, in OOD-MAML, all meta parameters share the same objective (Eq.(4)), and thus θ and θ_{fake} are interactively trained to minimize the same loss across tasks and collaboratively updated in each adaptation phase. Thus, OOD-MAML generates adversarial samples that are helpful for OOD-detection. We will clarify this in the final version. We will also compare MetaGAN and OOD-MAML via experiments.