CTR: Contrastive Training Recognition Classifier for Few-Shot Open-World Recognition

Citation for published version:
Dionelis, N, Tsaftaris, SA & Yaghoobi Vaighan, M 2022, CTR: Contrastive Training Recognition Classifier for Few-Shot Open-World Recognition. in *2022 26th International Conference on Pattern Recognition (ICPR).* Institute of Electrical and Electronics Engineers, pp. 1792-1799, 26TH International Conference on Pattern Recognition, Montréal Québec, Canada, 21/08/22.
https://doi.org/10.1109/ICPR56361.2022.9956206

Digital Object Identifier (DOI):
10.1109/ICPR56361.2022.9956206

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
2022 26th International Conference on Pattern Recognition (ICPR)

CTR: Contrastive Training Recognition Classifier for Few-Shot Open-World Recognition

Nikolaos Dionelis
School of Engineering
University of Edinburgh, UK
Contact email: nikolaos.dionelis@ed.ac.uk

Sotirios A. Tsaftaris
School of Engineering
Institute for Digital Communications
University of Edinburgh, UK

Mehrdad Yaghoobi
School of Engineering
Institute for Digital Communications
University of Edinburgh, UK

***Abstract*—AI-enabled systems in security, autonomous systems, safety, and healthcare do not only need to effectively detect Out-of-Distribution (OoD) samples, but also to recognize Objects of Concern (OoC), e.g. multiple thorax diseases, efficiently with few-shots. Detecting OoD samples is crucial, because reporting an out-of-domain input as abnormal is better than falsely classifying it. Data samples, during inference, are not confined to a finite labelled set, and thus closed-set approaches are limiting, as they misclassify OoD inputs, and they may assign them high prediction confidence. Furthermore, although anomaly detection is possible, recognizing new OoC fast using only few-shot samples remains challenging. There is a lack of methodologies for joint anomaly detection and few-shot OoC classification. Our contribution is the development of a framework for joint few-shot OoC detection and classification and anomaly detection in the unknown previously-unseen, in the wild, environment, which is known as Open-World Recognition (OWR). We propose a novel methodology, the data distribution boundary Contrastive Training Recognition (CTR) classifier for few-shot OWR. CTR takes advantage of labels and classes to learn the normal (and few-shot abnormal) data better, to more accurately detect OoD. The proposed model: (i) reduces failures to detect anomalies in health- and safety-critical applications for avoiding unfavourable consequences, (ii) decreases false alarms, and (iii) improves performance overall. Our framework differs from existing approaches because: (a) anomaly and OoC detection are combined, which has several benefits, including improved OoD performance, (b) the performance, accuracy, and robustness of OoD and few-shot OoC detection are improved by strengthening the estimation of the normal class distribution at the boundary of its support, self-generating samples and setting them as abnormal, and (c) the knowledge base of models is also augmented by learning class-incrementally, alleviating forgetting. CTR outperforms baselines in several settings, including on the SVHN, CIFAR-FS, and BSCD-FS ChestX and ISIC datasets.**

I. Introduction

Joint anomaly detection and classification. The simultaneous detection of Out-of-Distribution (OoD) samples and multi-class classification is crucial for systems to be deployed in real-world settings, because declaring an out-of-domain input as an OoD sample is better than misclassifying and possibly assigning a high prediction confidence to it. Joint OoD detection and classification, which refers to the classification of K existing classes and OoD detection with respect to these K classes, to effectively discern between $K+1$ classes, i.e. Open-Set Recognition (OSR), is important because in the real world, classifiers should know *when* they do not know, and have the capability to reject inputs. If an input belongs to an existing class, classification is performed. Otherwise, the input data sample is flagged as OoD. This is inspired by how humans, e.g. children [18], are dealing with seen (normal and abnormal) data and unseen unknown data, which also approximates even how expert security screening operators at airport checkpoints [10] act. Correspondingly, our brain can learn from limited seen (normal and abnormal) data, can discern between classes, and discern between the OoD class and the learned classes.

Moreover, knowing the normal and the (few-shot) abnormal data better, i.e. by taking advantage of labels, which provide additional information that we can take advantage of, helps us to more accurately detect OoD [33] and perform few-shot learning. For applications that need zero failures to detect OoD data, e.g. false negatives in medical imaging, labelled data and classes can be used to improve accuracy and robustness.

Joint few-shot classification and OoD detection. In several real-world settings, e.g. security screening at critical national infrastructure, anomaly detection on its own is not robust to adversarial and near-OoD samples [27] due to occlusion, clutter, concealment, disassembled objects, small items, and the problem's adversarial nature, i.e. humans might want to make threat objects look normal. We hence propose to combine few-shot learning and anomaly detection. Robust class-incremental few-shot learning is performed because OoD objects are rare, not confined to a finite labelled set, and learning such OoD items in an object- and class-incremental manner fast with few-shots is an ideal way to widen the knowledge base of models. We are interested in both low- and few-shot learning and in learning quickly with few-shots. Both these cases comprise classifying new classes and modifying the knowledge base of models. Augmenting the knowledge base of models, within a robust class-incremental framework that tries to alleviate forgetting, is also of our great interest. Enlarging the knowledge base of models efficiently with few-shots, while maintaining their desirable OoD detection capability, is known as few-shot Open-World classification. While class-incremental learning is useful in this setting, the problem remains challenging. Prior art in this context is limited, [1]–[3]. Generalised few-shot learning, to prevent *forgetting*, and cross-domain learning have also recently been attracting considerable interest, [46].

Few-shot OWR. Open-World Recognition (OWR) is defined by class-incremental OSR and refers to Open-Set adaptation and classification, [4]– [9]. Few-shot OWR has applica-

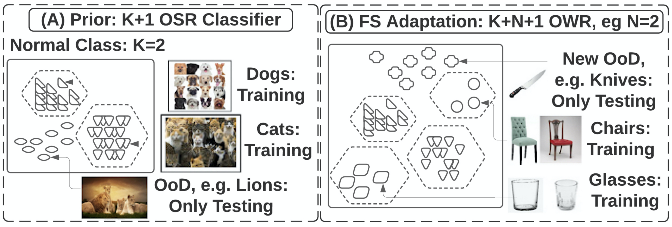

Fig. 1. $K+1$ OSR prior and new few-shot $K+N+1$ OWR adaptation.

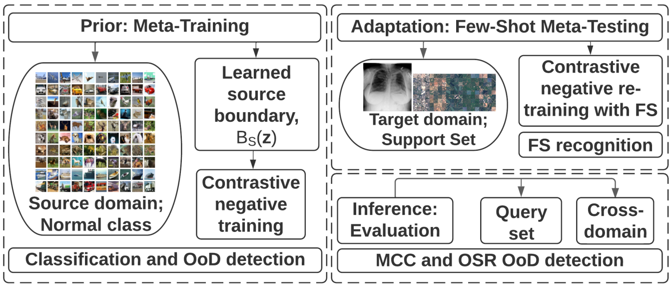

Fig. 2. Prior and few-shot OWR adaptation. *MCC is multi-class classification.*

tions in safety, surveillance and reconnaissance, and medical imaging. For unmanned vehicles, learning novel classes and recognizing OoD samples is crucial because the consequences of not accurately learning the new multi-class few-shots might be severe, leading to car accidents. Open-World classification is important for bacteria class recognition, for discovering new bacteria classes [81], where unsupervised training and masking are utilized for near-OoD detection of genomic sequences.

Proposed methodology. We devise the Contrastive Training Recognition (CTR) classifier for few-shot class-incremental joint Objects of Concern (OoC) classification and OoD detection in the unknown, in the wild, environment, which is known as Open-World classification. CTR is a methodology to first perform Open-Set classification and then few-shot OWR, when novel few-shot classes are introduced. Our contribution is the development of the CTR framework. To illustrate our setting, we present an example in Fig. 1: (A) Prior for OSR. For two base classes (normal class), (a) Cat with 5000 samples, and (b) Dog with 5000 samples, we classify the *three* classes of OoD, Cat, and Dog. This prior performs joint OoD detection and classification. (B) Adaptation. We learn new object classes quickly with low- and few-shots. For two novel classes, (i) Chair (5 samples), and (ii) Glass (5 samples), we classify the five classes of new OoD, Chair, Glass, Cat, and Dog. For few-shot OWR, the new OoD is with respect to the new and base classes. We discern between the novel classes, and learn class-incrementally, learning both the new classes and the OoD.

Our contribution. To address the important few-shot OWR setting and use few-shot learning models in the real world, we perform joint OoD detection and few-shot classification. To the best of our knowledge, we are the first to address this Open-World classification setting within a novel robust class-incremental learning framework, detecting unknown, in the wild, new anomalies, [1]–[3]. From a methodological point of view, we use negative training via our distribution boundary contrastive loss, and our Prediction Confidence f-Divergence criterion with the prediction confidence, for few-shot OWR. No other methodology performs few-shot contrastive training enhanced by our self-supervised learning with normal class data distribution boundary generation algorithm, followed by our class-assignment criterion. CTR outperforms baselines, and we achieve improved performance, accuracy, and robustness. CTR differs from existing methods because we: (a) combine anomaly detection and OoC detection, which has several benefits, including improved operational capability because of declaring out-of-domain inputs as abnormal, to avoid being falsely classified, (b) improve the performance, accuracy, and robustness for OoD and few-shot OoC detection, strengthening the estimation of the normal class distribution at its support boundary, self-generating samples and setting them as OoD, and (c) augment the knowledge base of models learning class-incrementally, alleviating forgetting: classes initially learned are effectively detected after learning new few-shot classes.

II. Related Work

Open-set classification and adaptation. Most supervised-learning methods classify base classes in a closed-set; this is limiting due to no anomaly detection capability and new class inclusion functionality, [3]. Such models are not applicable for the real world, as they classify unseen classes as base classes. In contrast, CTR for cross-domain few-shot OWR addresses such limitations, performs robust joint classification and OoD detection, and learns new classes efficiently with few-shots.

Few-shot learning. Including new few-shot classes is hard for supervised learning that needs many samples. It is difficult to generalize from few-shots, due to overfitting. It is expensive to label data for novel classes, and train from scratch. Gradient Magnitude [19] learns new classes, without training again from the start. CTR also effectively addresses such challenges.

Contrastive training. A generalized version of contrastive learning with data augmentation in [28] performs representation learning by computing the cosine similarity in the feature space to group same-class samples and repel different classes. In this context, several approaches propose contrastive training procedures, including [30], [1], [34]–[36], and [61], [62]. CTR performs contrastive training with the proposed algorithm based on the self-supervised data distribution boundary and the new multi-class few-shots to avoid them. To improve few-shot generalization and the robustness of our model for Open-World simultaneous classification and anomaly detection, CTR uses our proposed distractor/ negative training loss term to move away from our self-supervised data distribution boundary and the new multi-class few-shots. The proposed data distribution boundary contrastive algorithm sets the samples from the base classes as positive samples, and the self-generated boundary samples and the new few-shot classes as *negative* samples.

Models operating in the feature space may show limitations, e.g. dimensional [77] and complete collapse, which resembles mode collapse, [60]. Another challenge is representation collapse [78], which is like forgetting and it is the performance degradation of generalizable representations of the prior during *adaptation*. For classes with same characteristics, logits may be on top of others, [26]. Such models [24] do not perform few-shot OWR and OoD detection. [25] uses 75 queries in the evaluation, which leads to better results [59]. Instead, we compute the prediction confidence and distribution metrics for *few-shot OWR*, and enforce our model to avoid the learned source data distribution boundary. For adaptation, CTR classifies multi-class few-shots jointly with OoD detection using our distribution boundary contrastive training and our inference.

OSR. For K base classes, OSR discerns between $K+1$ classes: OoD + base, [31]–[42]. Classification and OoD detection has been tackled in [2], [31]–[38], few-shot OSR in [1], [43], [44], and cross-domain class recognition in [46]–[51]. To the best of our knowledge, the few-shot class-incremental OWR setting has not been solved. Existing models are lacking for few-shot OWR and may suffer from forgetting [1], [43]–[45]. oPen sEt mEta LEaRning (PEELER) [1] takes advantage of additional abnormal data, provided by the user, and it does not perform few-shot Open-World classification, because the class-incremental learning capability is missing. PEELER may generalize only to the provided OoD, and potentially fails to generalize to OoD data that are disjoint from the ad hoc selected OoD, as the provided anomalies might be far away from the distribution boundary, *i.e.* they are neither *task-specific* [18], nor well scattered. In contrast, CTR does not randomly choose OoD data, as we use our learned distribution boundary. CTR discerns between OoD, the novel few-shot OoC and the existing classes, and performs Open-World classification.

In the few-shot setting, Open woRld object dEtector (ORE) [3] may overfit the new few-shot classes, and may encounter representation collapse. It is not tested in the few-shot setting. Instead, it uses approximately 800 data samples to learn new classes. It is primarily designed for many-sample Open-World classification. In contrast, CTR performs few-shot OWR.

Cross-domain. Broader Study of Cross-Domain Few-Shots (BSCD-FS) is a satellite, medical, and plant dataset, [52], [48]. Many prior classes can lead to high accuracy for new classes, [53], [46]. Domain adaptation in [49] uses meta-learning [54], [45], while [51] stores features. In contrast, CTR for cross-domain OWR obviates data diversity memorization issues.

III. CTR for Few-Shot OWR

CTR for few-shot class-incremental Open-World classification and learning is presented in Fig. 2. CTR performs N-way R-shots classification [55], even for cross-domains. We first learn a prior on the base classes, and perform OoD detection. When novel few-shot classes are provided, we refine the prior to perform joint few-shot classification and OoD detection.

A. Prior for the base. Our prior learns the source dataset, where the base classes are the normal class, e.g. from CIFAR-10 (C10). Joint classification and OoD detection is performed using our self-supervised distribution normal class boundary. We denote this learned boundary by $B_S(\mathbf{z})$, where $\mathbf{z}$ are samples from a standard Gaussian distribution, i.e. $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$. We use $B_S(\mathbf{z})$ for contrastive training, for OoD detection, by setting them as self-generated negative data samples. Our prior performs OoD detection, using the prediction confidence.

B. Adaptation. We learn the target classes efficiently with few-shots, e.g. SVHN or BSCD-FS datasets for cross-domain and CIFAR-FS for intra-domain. Class-incremental learning and inclusion of the new classes are performed. OoD detection, with respect to the new classes and the base, is achieved. CTR discerns between OoD, the base, and the new few-shot classes. Few-shot contrastive retraining using the target-domain classes is employed to make the decision boundary tighter, because the new few-shot classes are OoD with respect to the base.

Contrastive training of the OSR prior with our learned distribution boundary. For N-way R-shot Open-World classification, CTR first trains an Open-Set classifier for joint classification and OoD detection with K base classes. Thus, $K+1$ classification is performed. We first minimize cross-entropy, jointly with negative training. In contrastive training, we set our generated distribution boundary, $B_S(\mathbf{z})$, as OoD data [27], [33], and we use the following negative-data optimization,

$$\arg\min_f \; -\frac{1}{Q}\sum_{j=1}^{Q}\log\frac{\exp(f_{y_j}(\mathbf{x}_j))}{\sum_{k=1}^{K}\exp(f_k(\mathbf{x}_j))} - \kappa\frac{1}{R}\times \tag{1}$$
$$\sum_{m=1}^{R}\log\left(1-\max_{r=1,2,\dots,K}\frac{\exp(f_r(B_S(\mathbf{z}_m)))}{\sum_{k=1}^{K}\exp(f_k(B_S(\mathbf{z}_m)))}\right),$$

$$\text{where } \arg\min_{B_S} \; \frac{1}{R-1}\sum_{m=1,\,\mathbf{z}_m\neq\mathbf{z}}^{R}\frac{||\mathbf{z}-\mathbf{z}_m||_2}{||B_S(\mathbf{z})-B_S(\mathbf{z}_m)||_2}$$
$$+\;\mu \max_{l=1,2,\dots,K}\frac{\exp(f_l(B_S(\mathbf{z}))-f_l(\mathbf{x}))}{\sum_{k=1}^{K}\exp(f_k(B_S(\mathbf{z}))-f_k(\mathbf{x}))} \tag{2}$$
$$+\;\nu\min_{j=1,2,\dots,Q}||B_S(\mathbf{z})-\mathbf{x}_j||_2,$$

where the normal class is $\mathbf{x}$, and $\mathbf{x}_j$ are the labeled data with labels y_j, [34], [35]. For the within-distribution data in the first term in (1), Q, j and k are respectively the batch size, sample and class indices. Here, the index j takes values from 1 to Q. We denote our data by $(\mathbf{x}_j, y_j)_{j=1}^{Q}$, e.g. $\mathbf{x}_j$ is a vector of length 3072 for C10. For the OoD data samples in the second term, R is the batch size, m a sample, and κ a hyper-parameter which is found by using a validation dataset, [33]–[35].

For the nested optimization, the inner optimization is our learned distribution boundary, B_S, in (2). The outer is cross-entropy with contrastive training in (1). In (2), we generate the boundary of the normal base classes [27], [60]. The first term is for scattering the $B_S(\mathbf{z})$ samples. This measure preserves distance proportionality in the $\mathbf{z}$ and $\mathbf{x}$ spaces, alleviating mode collapse. The second term guides to find the boundary, B_S, by penalizing the prediction confidence, and by pushing the generated samples OoD. The third term penalizes deviations from normality, using the distance from a point to a set.

Multi-class few-shot contrastive loss. The focal point of this research is multi-class few-shot Open-World classification. CTR contrastively trains our network to enforce the normal class to avoid the multi-class few-shots, FS, where FSi are the labeled few-shot data with labels i, where i is between 1 and N. Our objective is based on our self-supervised distribution boundary and the new multi-class few-shots, to avoid them and repel them from the base classes. We, thus, minimize:

$$\text{argmin}_f \; -\frac{1}{Q}\sum_{j=1}^{Q}\log\frac{\exp(f_{y_j}(\mathbf{x}_j))}{\sum_{k=1}^{K}\exp(f_k(\mathbf{x}_j))} - \lambda\frac{1}{NR}\times \tag{3}$$
$$\sum_{i=1}^{N}\sum_{m=1}^{R}\log\left(1-\max_{r=1,2,...,K}\frac{\exp(f_r(\text{FS}i_m))}{\sum_{k=1}^{K}\exp(f_k(\text{FS}i_m))}\right),$$

where (3) has two terms that operate on different samples for positive and negative training, respectively. The first term is the cross-entropy between y_j and the predictions to penalize deviation from existing classes and *reward* correct within-distribution classification of the labelled data. We use $f_{y_j}(\mathbf{x}_j)$, where j is the base-class sample index (1 to Q). The second term repels the provided new multi-class few-shots from the base classes. We penalize the prediction confidence on the multi-class few-shots, for FSi detection. CTR uses $f_r(\text{FS}i_m)$ and enforces our model to move away from the N new classes. A level of robustness is enforced with the second term, separating the new classes from the base. An improved decision boundary is learned by setting the few-shots as OoD. We learn new items efficiently with few-shots, using contrastive training (refinement loss) and the separateness objective, alerting for OoD. In (3), $f(\cdot)$ is a Convolutional Neural Network (CNN) with the final layer being linear, followed by softmax. We denote the sample index of the new class by m (1 to R). r and k are the class indexes of the base classes (1 to K).

Prediction Confidence f-Divergence. We propose the Prediction Confidence based f-Divergence (PCD) criterion to classify novel classes for OWR efficiently with few-shots. CTR computes the metric, m_0, between the queried test sample, $\tilde{\mathbf{x}}$, and the base classes. The metrics, m_1 and m_2, based on confidence and Kullback-Leibler (KL) divergence, between the first new few-shot class, FS1, and $\tilde{\mathbf{x}}$ and between the second new class, FS2, and $\tilde{\mathbf{x}}$, respectively, are then found. We repeat this for any new class i: we find the metrics, m_i, between $\tilde{\mathbf{x}}$ and the new classes. For the final class, the metric, m_N, is between FSN and $\tilde{\mathbf{x}}$. We denote the assigned class by c. The metric on which we set a threshold on for OoD detection, is m_c. If $m_c > \tau$ where τ is a class-independent threshold, then $\tilde{\mathbf{x}}$ is OoD. Our inference criterion [27], [85] is given by

$$m_c = \max(m_0, m_1, ..., m_N), \tag{4}$$
$$c = \arg\max(m_0, m_1, ..., m_N) \text{ if } m_c < \tau \tag{5}$$
$$\text{and } c = \text{ OoD if } m_c \geq \tau,$$
$$\text{where } m_i = \frac{M(\mathbf{x}, \text{FS}i)}{\exp(L(\mathbf{x}, \text{FS}i, \tilde{\mathbf{x}}))}, \tag{6}$$
$$\text{and where } L(\mathbf{x}, \text{FS}i, \tilde{\mathbf{x}}) = \frac{\text{KL}(\text{FS}i, \tilde{\mathbf{x}})}{M(\mathbf{x}, \text{FS}i)}, \tag{7}$$
$$\text{where } M(\mathbf{x}, \tilde{\mathbf{x}}) = \max_{r=1,2,...,K}\frac{\exp(f_r(\tilde{\mathbf{x}}))}{\sum_{k=1}^{K}\exp(f_k(\tilde{\mathbf{x}}))}, \tag{8}$$

where KL is f-divergence and M is prediction confidence. In (6)-(8), we use $\mathbf{x}$ to denote the labelled data from the base classes, and $\tilde{\mathbf{x}}$ is the queried sample. Here, we have N new and K base classes. In (5), the threshold τ is set at 95% True Positive Rate (TPR), [18]. For recognizing the multi-class few-shots, OWR class assignment on OoD, base and new classes, we utilize our proposed PCD criterion. The reference point in the metric computations is the base classes, $\mathbf{x}$. CTR uses (4)-(8) for few-shot classification using distribution metrics, i.e. KL divergence, and the prediction confidence with the normal class as *our reference*. We devise meaningful divergence metrics between the new classes using the prediction confidences of new-class samples and distribution metrics. We denote our metric by m_i in (6), where the numerator is confidence. We discern between OoD and few-shot and base classes using m_i, by analogy to prediction confidence with *(arg)max*, [33]–[35], [32]. Here, we denote our confidence-based f-divergence by $L(\cdot)$ in (7), where the numerator is KL and the denominator is confidence based on the definition of the KL divergence.

We adapt to the *ever-evolving OoD*, and we add new classes efficiently with few-shots. To the best of our knowledge, there is no other few-shot OWR method that uses contrastive training followed by our PCD criterion, [6]–[9]. This differs from Minimum Euclidean Distance in the feature space [57], which is suboptimal for OSR, [1]. In [26], nearest is maximum cosine similarity. We use (4)-(8) for OWR instead of implicit distributions, [27]. In (6)-(8), when $\tilde{\mathbf{x}}$ is near FSi, then m_i is close to $M(\mathbf{x}, \text{FS}i)$, while if $\tilde{\mathbf{x}}$ is far from FSi, then m_i is much less than $M(\mathbf{x}, \text{FS}i)$. In (4) and (5), $m_0 = (\exp(\text{KL}(\mathbf{x}, \tilde{\mathbf{x}})))^{-1}$. Here, when $\tilde{\mathbf{x}}$ is near the initial base classes, then m_0 is close to 1. If $\tilde{\mathbf{x}}$ is far from the base classes, then m_0 is much less than 1. The KL f-divergence metric is computed with a discrete distribution and Dirac functions, [86], [87]. The images of the novel class i, FSi, (discrete distribution with Dirac functions) and the unknown test image, $\tilde{\mathbf{x}}$, (Dirac function) are used.

IV. Evaluation of CTR and Results

CTR is evaluated and compared to benchmarks. Here, we test CTR's performance with respect to classification and OoD detection. The source dataset is C10 or Mini-ImageNet (MIN). The target set is SVHN, CIFAR-FS, BSCD-FS, or MIN.

Architecture. We implement CTR using CNNs for our classifier and for our data distribution boundary generator.

CTR results in the cross-domain setting. The top part of Table I presents the evaluation of CTR on SVHN, denoting N-way R-shots by Nw Rs. In these results, the source dataset is C10, while the target domain dataset (i.e. both the support and query sets) is SVHN. We compare CTR to second-order MAML and to Siamese using 5-way 5-shots [54], [58]. The percentage accuracy improvement of CTR here is at least 54%, consistently outperforming the examined baseline models.

Cross-domain robustness. In the bottom part of Table I, we show CTR's robustness to the number of few-shots. For $\geq$ 10-shots, CTR achieves its best performance: accuracy 0.98. CTR is robust to few-shots, achieving accuracy 0.91 with 5-shots. Regarding source generalization on the C10 dataset, the accuracy of CTR on C10 before and after SVHN adaptation is similar. This shows reduced forgetting despite the class-incremental setting, thanks to contrastive training which uses the learned distribution boundary, $B_S(\mathbf{z})$, [33], [45], [88].

Intra-domain results on CIFAR-FS. We evaluate CTR in Table II with 10- and 5-way R-shots, for R between 1 and

TABLE I
CTR CROSS-DOMAIN ROBUSTNESS TO THE NUMBER OF MULTI-CLASS FEW-SHOTS FROM SVHN (SOURCE DOMAIN C10) IN FEW-SHOT CLASSIFICATION METRICS, AND COMPARISON TO BASELINES, [58].

MODEL	ACC.	PREC.	REC.	F1
CTR 5W 5S	**0.83**	**0.75**	**0.83**	**0.78**
MAML 5W 5S	0.54	0.45	0.54	0.47
SIAMESE 5W 5S	0.42	0.34	0.42	0.37
CTR 10W 1S	0.55	0.38	0.55	0.42
CTR 10W 5S	0.91	0.86	0.91	0.88
CTR 10W 10S (TO 200S)	0.98	0.94	0.96	0.95

TABLE II
CTR ROBUSTNESS TO THE NUMBER OF MULTI-CLASS FEW-SHOTS FROM CIFAR-FS (SOURCE DOMAIN C10). COMPARISON TO BASELINES, [1].

CTR | PEELER	ACCURACY	PRECISION	RECALL
10W 1S	**0.69** | 0.42	**0.60** | 0.35	**0.68** | 0.42
5W 1S	**0.83** | 0.58	**0.75** | 0.50	**0.83** | 0.57
5W 5S	**0.83** | 0.75	**0.75** | 0.67	**0.83** | 0.74
5W 10S TO 20S	**0.95** | 0.86	**0.84** | 0.78	**0.92** | 0.83

TABLE III
CTR ROBUSTNESS TO THE NUMBER OF MULTI-CLASS FEW-SHOTS FROM CIFAR-FS (SOURCE DOMAIN C10). COMPARISON TO BASELINES, [59].

MODEL	ACC.	PREC.	REC.	F1
CTR 5W 1S	**0.83**	**0.75**	**0.83**	**0.78**
PROTONET1 5W 1S	0.72	0.61	0.68	0.65
METAOP1 5W 1S	0.71	0.64	0.70	0.68
PROTONET5 5W 1S *	0.83	0.73	0.81	0.77
METAOP5 5W 1S *	**0.83**	0.74	0.82	**0.79**
CTR 5W 5S	**0.83**	**0.75**	**0.83**	**0.78**
PROTONET1 5W 5S	0.71	0.61	0.68	0.65
METAOP1 5W 5S	0.72	0.64	0.70	0.68
PROTONET5 5W 5S *	0.84	0.73	0.81	0.77
METAOP5 5W 5S *	**0.85**	0.74	0.82	**0.79**
CTR 5W 10S (TO 20S)	**0.95**	**0.84**	**0.92**	**0.89**
PROTONET1 5W 10S (TO 20S)	0.71	0.62	0.67	0.64
METAOP1 5W 10S (TO 20S)	0.72	0.64	0.71	0.69
PROTONET5 5W 10S (T 20S)*	0.84	0.74	0.83	0.79
METAOP5 5W 10S (TO 20S)*	0.84	0.75	0.84	0.81

$*$ METAOP5 AND PROTONET5 USE 5 QUERY SAMPLES IN THE EVALUATION.

TABLE IV
CTR CROSS-DOMAIN ROBUSTNESS TO THE NUMBER OF MULTI-CLASS FEW-SHOTS ON THE BSCD-FS TARGET DATASET (SOURCE DOMAIN MIN) IN ACCURACY. COMPARISON TO MAML AND PROTONET, [46], [47].

MODEL	ESAT	ISIC	CHEX	CDIS.
CTR 5W 5S	**0.85**	**0.52**	**0.48**	**0.86**
MAML 5W 5S	0.72	0.40	0.23	0.78
PROTONET 5W 5S	0.73	0.40	0.24	0.80
CTR 5W 20S	**0.94**	**0.55**	**0.58**	**0.92**
MAML 5W 20S	0.82	0.52	0.23	0.90
PROTONET 5W 20S	0.82	0.50	0.28	0.88
CTR 5W 50S	**0.97**	**0.81**	**0.85**	**0.92**
MAML 5W 50S	0.85	0.63	0.49	0.92
PROTONET 5W 50S	0.80	0.41	0.29	0.91

20. We train the prior on C10. CTR in Table II outperforms PEELER in [1] by a large margin. In accuracy, for the 10-way 1-shot setting, the percentage improvement is 64% for CTR, compared to PEELER. The *raw improvement* is 0.27 points. PEELER begins from a stronger position, since it uses more information during training, due to the provided OoD samples, e.g. 2 C10 classes, used for Outlier Exposure. For intra-domain evaluation, CTR achieves improved accuracy: for 1-shot, 0.83. For 10-shots, CTR achieves high accuracy, i.e. 0.95. PEELER, which may suffer from forgetting [45], also employs a very different base model, i.e. ProtoNet [1], while our base model is [33]. Now, the relative percentage improvement, over the base model, for PEELER is 7% [1], while CTR's relative percentage improvement over our base model is approximately 11%.

Intra-domain results. CTR, in Table III, outperforms the baselines that start from the *same* position as CTR: ProtoNet1 and MetaOp1, [59]. CTR achieves comparable performance to ProtoNet5 and MetaOp5, which take advantage of more evaluation queries (five), while we use 1. ProtoNet5 and MetaOp5 cannot detect OoD, do not perform class-incremental learning, may suffer from forgetting, and do not train contrastively. In contrast, CTR performs OWR, learns class-incrementally, and performs contrastive negative training. CTR *outperforms* MetaOp5 for 5-way and for 10-shots. CTR also outperforms ProtoNet1 as we use distribution metrics and confidence.

CTR outperformance on BSCD-FS. CTR significantly outperforms MAML and ProtoNet for cross-domain few-shots in Table IV, in Acc. The prior is on MIN. MAML lacks OoD detection and may encounter catastrophic forgetting, [45].

CTR compared to benchmarks on BSCD-FS. In Table V, for 5-way 1-shot, CTR outperforms the benchmarks on ISIC and ChestX (CheX). For 5-shots, CTR outperforms the benchmarks on CheX. Here, the percentage improvement of CTR over STARTUP [46], which uses expensive student-teacher retraining, is 78%. STARTUP starts from a more advantageous position, using additional unlabeled target data. The baseline models also lack OoD detection. For 5-way 20-shots, CTR outperforms the baselines on CheX, by approximately 61%, and yields comparable results on EuroSAT (ESAT). For 5-way 50-shots, CTR *outperforms* the baselines on CheX data and on ISIC by approximately 73% and 8%, respectively.

CTR results on same source-target sets. CTR in Table VI performs classification on the same source and target datasets,

TABLE V

CTR CROSS-DOMAIN FEW-SHOT ROBUSTNESS ON BSCD-FS IN ACC. SOURCE DOMAIN MIN. COMPARISON TO BENCHMARKS [46], [47], [49].

MODEL	ESAT	ISIC	CHEX	CDIS.
CTR 5w 1s	**0.70**	**0.52**	**0.38**	**0.72**
STARTUP 5w 1s * [46]	0.64	0.33	0.27	0.76
SSLFS 5w 1s * [47]	0.61	0.39	0.28	0.78
CDFSL 5w 1s * [49]	0.60	0.37	0.25	0.74
CDFSMFT 5w 1s * [49]	0.69	0.48	0.26	0.79
CDFSREC 5w 1s * [50]	0.69	0.35	0.23	**0.83**
HVMFSLCD 5w 1s* [51]	0.54	0.31	0.20	0.72
CTR 5w 5s	**0.85**	**0.52**	**0.48**	**0.86**
STARTUP 5w 5s *	0.82	0.47	0.27	0.93
SSLFS 5w 5s *	0.83	0.50	0.29	0.92
CDFSL 5w 5s *	0.80	0.47	0.26	0.89
CDFSMFT 5w 5s *	**0.90**	**0.62**	0.28	**0.96**
CDFSREC 5w 5s *	0.88	0.48	0.28	**0.96**
HVMFSLCD 5w 5s *	0.75	0.42	0.27	0.88
CTR 5w 20s	**0.94**	**0.55**	**0.58**	**0.92**
STARTUP 5w 20s *	0.88	0.51	0.33	0.95
SSLFS 5w 20s *	0.90	0.61	0.34	0.97
CDFSL 5w 20s *	0.88	0.60	0.32	0.96
CDFSMFT 5w 20s *	**0.94**	**0.65**	0.36	**0.99**
CDFSREC 5w 20s *	0.91	0.57	0.32	0.98
HVMFSLCD 5w 20s *	0.85	0.55	0.31	0.95
CTR 5w 50s	**0.97**	**0.81**	**0.85**	**0.92**
STARTUP 5w 50s *	0.92	0.63	0.37	0.98
SSLFS 5w 50s *	0.95	0.67	0.39	**0.99**
CDFSL 5w 50s *	0.91	0.65	0.37	0.98
CDFSMFT 5w 50s *	0.96	0.75	0.45	**0.99**
CDFSREC 5w 50s *	0.93	0.68	0.38	**0.99**
HVMFSLCD 5w 50s *	0.87	0.62	0.33	0.98

$*$ STARTUP [46] USES AN ADDITIONAL UNLABELED TARGET DOMAIN SET.

C10 and MIN, and detects OoD. We test CTR on C10 using leave-4-out evaluation with 6 classes as the closed-set, and the rest 4 as open-set. Now, for anomaly detection, we compute the Area Under the Receiver Operating Characteristics Curve (AUROC), [60]. CTR on C10, compared to MIN, achieves an improvement of about 18% in accuracy and 3.5% in AUROC. On C10, we compare CTR to PEELER and to SoftMax [1]. We consistently outperform the benchmarks in accuracy and AUROC with an improvement of more than about 13%. Likewise, on MIN, CTR *outperforms* the benchmarks in accuracy and AUROC, with an improvement of more than about 4%. In contrast to PEELER, which avoids the provided anomalies using its negative entropy loss, CTR moves away from the new classes and computes our PCD criterion in (4)-(8). PEELER, which does not perform class-incremental learning and OWR, may fail to generalize to OoD data that are *disjoint* from the given OoD, due to ad hoc selection of OoD during training [1], [61], [63]. CTR offers improved convenience for the user using our learned distribution boundary for contrastive training.

TABLE VI

CTR PERFORMANCE TO RECOGNIZE NEW MANY-SHOT CLASSES AND PERFORM OWR. COMPARISON TO BASELINES, [1]. *C10 is CIFAR-10 with 6 closed-set classes for classification and 4 open-set classes, and MIN is Mini-Imagenet with 64 classes for classification and 20 open-set classes.*

SOURCE	MODEL	ACCURACY	AUROC
C10	**CTR**	**0.93**	**0.88**
C10	SOFTMAX	0.80	0.70
C10	BASICPEELER	0.82	0.75
C10	PEELER	0.82	0.77
MIN	**CTR**	**0.79**	**0.85**
MIN	SOFTMAX	0.76	0.78
MIN	BASICPEELER	0.76	0.81
MIN	PEELER	0.76	0.82

CTR contributions. According to all the above evaluations, CTR using (i) our proposed self-learned distribution boundary contrastive training, and (ii) recognition of the new few-shot classes with our inference criterion, outperforms baselines.

V. CONCLUSION

Few-shot Open-World classification is crucial in the real world. The CTR classifier was presented in this paper, for multi-class few-shot class-incremental learning, cross-domain adaptation, and few-shot OWR. CTR performs contrastive training using the learned data distribution boundary, as well as negative training, followed by the Prediction Confidence f-Divergence criterion. Avoiding and moving away from the new classes and the self-generated distribution boundary, proved to be beneficial. The evaluation of CTR on several benchmark datasets shows its superiority in various settings, including on the SVHN, CIFAR-FS, BSCD-FS, EuroSAT, ChestX, ISIC, C10, and MIN datasets. CTR, when trained on C10, outperforms benchmarks in accuracy on SVHN by at least 54% (Table I), and on CIFAR-FS by approximately 11% for 5-ways and for more than 10-shots (Table II). CTR, when trained on MIN, *outperforms* baselines in terms of accuracy on BSCD-FS ChestX by approximately 36%, 66%, 61%, and 73% for 1, 5, 20, and 50 few-shots, respectively (Tables IV and V). On the BSCD-FS ISIC dataset, for 1, and 50 few-shots, CTR outperforms benchmarks by approximately 8%. For 5-way 50-shots, CTR achieves high accuracy on EuroSAT, ISIC, ChestX and CropDisease data, 0.97, 0.81, 0.85 and 0.92, respectively. In AUROC, using leave-4-out evaluation on C10 and MIN, CTR outperforms recent benchmarks in Table VI. We think CTR opens the road to *few-shot Open-World classification* and will inspire others to adopt and use this real-world setting.

ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council of the UK (EPSRC) Grant number EP/S000631/1 and the UK MOD University Defence Research Collaboration (UDRC) in Signal Processing.

REFERENCES

[1] B. Liu, et al. Few-Shot Open-Set Recognition using Meta-Learning. In Proc. IEEE/CVF Computer Vision Pattern Recognition (CVPR). 2020.

[2] L. Neal, et al. Open set learning with counterfactual images. In Proc. European Conference on Computer Vision (ECCV), p. 613-628. 2018.

[3] K. Joseph, et al. Towards Open World Object Detection. In Proc. IEEE/CVF Computer Vision Pattern Recognition (CVPR). 2021.

[4] T. Boult, et al. Learning and the unknown: Surveying steps toward open world recognition. In Proc. AAAI, p. 9801-9807. 2019.

[5] A. Bendale and T. Boult. Towards Open Set Deep Networks. In Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2016-Decem:1563-1572. 2016.

[6] A. Bendale and T. Boult. Towards Open World Recognition. In Proc. Computer Vision Pattern Recognition (CVPR), p. 1893-1902. 2015.

[7] C. Geng, et al. Recent Advances in Open Set Recognition: A Survey. IEEE Transactions Pattern Analysis TPAMI. 2020.

[8] W. Wang, V. W. Zheng, H. Yu, and C. Miao. A Survey of Zero-Shot Learning: Settings, Methods, and Applications. TIST. 2019.

[9] Z. Fang, J. Lu, A. Liu, F. Liu, and G. Zhang. Learning Bounds for Open-Set Learning. In Proc. ICML. 2021.

[10] S. Akcay. Towards Automatic Threat Detection: A Survey of Advances of Deep Learning within X-ray Security Imaging. arXiv:2001.01293.

[11] J.-M. Perez-Rua, X. Zhu, T. Hospedales, and T. Xiang. Incremental few-shot object detection. In Proc. CVPR. 2020.

[12] C. Zhang, et al. Few-shot incremental learning with continually evolved classifiers. In Proc. IEEE CVF Conference Computer Vision Pattern Recognition (CVPR), p. 12455-12464. 2021.

[13] A. Akyürek, et al. Subspace regularizers for few-shot class incremental learning. arXiv:2110.07059. 2021.

[14] A. Dhamija, et al. Self-Supervised Features Improve Open-World Learning. arXiv:2102.07848. 2021.

[15] Ge. Generative OpenMax for Multi-Class Open Set Classification. 2019.

[16] M. Jafarzadeh. Automatic Open-World Reliability Assessment. In Proc. Winter Conference on Applications of Computer Vision (WACV), 2021.

[17] X.-Y. Zhang, C.-L. Liu, and C. Suen. Towards robust pattern recognition: A review. In Proc. IEEE, v. 108, n. 6, p. 894-922. 2020.

[18] T. Jeong and H. Kim. OoD-MAML: Meta-Learning for Few-Shot Out-of-Distribution Detection and Classification. In Proc. NeurIPS. 2020.

[19] B. Hariharan. Low-Shot Visual Recognition by Shrinking and Hallucinating Features. In Proc. International Conference ICCC. 2017.

[20] A. Kumagai and T. Iwata. Unsupervised Domain Adaptation by Matching Distributions Based on the Maximum Mean Discrepancy via Unilateral Transformations. In Proc. AAAI. 2019.

[21] C. Miao. SIXray: Large-scale Security Inspection X-ray Benchmark for Prohibited Item Discovery in Overlapping Images. In Proc. CVPR. 2019.

[22] Y. Wei, et al. Occluded Prohibited Items Detection: An X-ray Security Inspection Benchmark and De-occlusion Attention Module. In Proc. ACM International Conference on Multimedia, p. 138-146. 2020.

[23] D. Mery, et al. GDXray: The database of X-ray images for nondestructive testing. Journal of Nondestructive Evaluation. 2015.

[24] E. Triantafillou. Learning a Universal Template for Few-shot Dataset Generalization. In Proc. ICML. 2021.

[25] I. Ziko. Laplacian Regularized Few-Shot Learning. In Proc. ICML. 2020.

[26] J. Tack, et al. CSI: Novelty detection via contrastive learning on distributionally shifted instances. In Proc. NeurIPS. 2020.

[27] N. Dionelis, M. Yaghoobi, and S. Tsaftaris. Few-Shot Adaptive Detection of Objects of Concern Using Generative Models with Negative Retraining. In Proc. IEEE International Conference on Tools with Artificial Intelligence (ICTAI). 2021.

[28] T. Chen, et al. Intriguing Properties of Contrastive Losses. In Proc. Advances in Neural Information Processing Systems (NeurIPS), 2021.

[29] T. Chen, et al. A Simple Framework for Contrastive Learning of Visual Representations. In Proc. International Conference ICML. 2020.

[30] X. Luo. Boosting Few-Shot Classification with View-Learnable Contrastive Learning. In Proc. International Conference Multimedia. 2021.

[31] D. Zhou, et al. Learning Placeholders for Open-Set Recognition. In Proc. IEEE/CVF Computer Vision Pattern Recognition (CVPR). 2021.

[32] S. Vaze, et al. Open-Set Recognition: A Good Closed-Set Classifier is all you Need. arXiv:2110.06207. 2021.

[33] N. Dionelis, M. Yaghoobi, and S. Tsaftaris. FROB: Few-shot ROBust Model for Classification with Out-of-Distribution Detection. arXiv:2111.15487. 2021.

[34] M. Hein, et al. Why ReLU networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In Proc. CVPR. 2019.

[35] J. Bitterwolf, et al. Certifiably adversarially robust detection of out-of-distribution data. In Proc. NeurIPS. 2020.

[36] D. Hendrycks, et al. Deep anomaly detection with outlier exposure. In Proc. ICLR. 2019.

[37] J. Barria and S. Thajchayapong. Detection and Classification of Traffic Anomalies Using Microscopic Traffic Variables. IEEE T-ITS. 2011.

[38] S. Thajchayapong, et al. Distributed Classification of Traffic Anomalies Using Microscopic Traffic Variables. IEEE Transactions T-ITS. 2013.

[39] A. Fisch, et al. Few-Shot Conformal Prediction with Auxiliary Tasks. In Proc. ICML. 2021.

[40] M. Messoudi. Deep Conformal Prediction for Robust Models. In Proc. Information Processing and Management of Uncertainty. 2020.

[41] A. Angelopoulos. A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification. arXiv:2107.07511. 2021.

[42] D. Stutz. Learning Optimal Conformal Classifiers. arXiv:2110.09192.

[43] S. Huang, et al. Task-Adaptive Negative Class Envision for Few-Shot Open-Set Recognition. arXiv:2012.13073. 2020.

[44] M. Jeong, et al. Few-shot Open-set Recognition by Transformation Consistency. In Proc. CVPR. 2021.

[45] P. Yap, et al. Addressing Catastrophic Forgetting in Few-Shot Problems. In Proc. ICML. 2021

[46] C. Phoo and B. Hariharan. Self-training for few-shot transfer across extreme task differences. In Proc. ICLR. 2021.

[47] D. Chen, et al. Self-supervised learning for few-shot image classification. arXiv:1911.06045. 2019.

[48] Y. Guo, et al. A new benchmark for evaluation of cross-domain few-shot learning. In Proc. ECCV, 2020.

[49] J. Cai and S. Shen. Cross-domain few-shot learning with meta fine-tuning. arXiv:2005.10544. 2020.

[50] A. Islam, et al. A Simple Framework for Cross-Domain Few-Shot Recognition with Unlabeled Data. 2021.

[51] Author. Hierarchical Variational Memory for Few-shot Learning Across Domains. Submitted, ICLR, (also arXiv:2112.08181). 2022.

[52] Y. Guo, et al. A Broader Study of Cross-Domain Few-Shot Learning. In Proc. ECCV. 2020.

[53] G. Dhillon. A baseline for few-shot image classification. In ICLR. 2020.

[54] C. Finn. Model-Agnostic Meta-Learning for Fast Adaptation of DNNs. In Proc. International Conference Machine Learning (ICML). 2017.

[55] W. Chen, et al. A Closer Look at Few-Shot Classificatio. In Proc. International Conference Learning Representations (ICLR). 2019.

[56] K. Tang, et al. CODEs: Chamfer Out-of-Distribution Examples against Overconfidence Issue. In Proc. International Conference ICCV. 2021.

[57] P. Goyal, et al. Self-supervised Pretraining of Visual Features in the Wild. arXiv:2103.01988. 2021.

[58] R. Yang, et al. Few Shot Learning Based on the Street View House Numbers (SVHN) Dataset. In Proc. Edge Computing and IoT: Systems, Management and Security, ICECI 2020, v. 368. Springer. 2021.

[59] K. Lee, et al. Meta-learning with differentiable convex optimization. In Proc. CVPR. 2019.

[60] N. Dionelis, M. Yaghoobi, and S. Tsaftaris. Boundary of Distribution Support Generator (BDSG): Sample Generation on the Boundary. In Proc. IEEE International Conference on Image Processing (ICIP), p. 803-807. 2020.

[61] K. Wang, et al. Few-shot Out-of-Distribution Detection. ICML Workshop on Uncertainty and Robustness in Deep Learning. 2020.

[62] K. Wang, et al. Out-of-Distribution Detection in Few-Shot Classification. OpenReview. 2019.

[63] R. Das, Y. Wang, and J. Moura. On the importance of distractors for few-shot classification. In Proc. ICCV, p. 9030-9040. 2021.

[64] K. Tang, et al. Long-tailed classification by keeping the good and removing the bad momentum causal effect. In Proc. NeurIPS. 2020.

[65] S. Bhatia, et al. ExGAN: Adversarial Generation of Extreme Samples. In Proc. Advancement of Artificial Intelligence (AAAI). 2021.

[66] N. Dionelis, M. Yaghoobi, and S. Tsaftaris. OMASGAN: Out-of-Distribution Minimum Anomaly Score GAN for Sample Generation on the Boundary. arXiv:2110.15273. 2021.
[67] N. Dionelis, M. Yaghoobi, and S. Tsaftaris. Tail of Distribution GAN (TailGAN): Generative-Adversarial-Network-Based Boundary Formation. In Proc. Sensor Signal Processing for Defence (SSPD) IEEE Conference, p. 41-45. 2020.
[68] H. Zhang, et al. Hybrid Models for Open Set Recognition. In Proc. European Conference on Computer Vision (ECCV), Springer. 2020.
[69] M. Sensoy, et al. Uncertainty-Aware Deep Classifiers Using Generative Models. In Proc. AAAI Conference on Artificial Intelligence. 2020.
[70] E. Nalisnick, et al. Do Deep Generative Models Know What They Don't Know?. In Proc. ICLR. 2019.
[71] P. Kirichenko, et al. Why Normalizing Flows Fail to Detect Out-of-Distribution Data. In Proc. NeurIPS. 2020.
[72] Y. Ovadia, et al. Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. In Proc. NeurIPS. 2019.
[73] A. D'Amour, et al. Underspecification presents challenges for credibility in modern machine learning. arXiv:2011.03395, 2020.
[74] X. Chen, et al. A Boundary Based Out-Of-Distribution Classifier for Generalized Zero-Shot Learning. In Proc. Conference ECCV. 2020.
[75] Y. Lu, et al. Few-shot scene-adaptive anomaly detection. In Proc. European Conference on Computer Vision (ECCV). 2020.
[76] A. Kukleva, et al. Generalized and incremental few-shot learning by explicit learning and calibration without forgetting. In Proc. ICCV. 2021.
[77] L. Jing, P. Vincent, Y. LeCun, et al. Understanding dimensional collapse in contrastive self-supervised learning. arXiv:2110.09348. 2021.
[78] A. Aghajanyan, A. Shrivastava, A. Gupta, et al. Better fine-tuning by reducing representational collapse. arXiv:2008.03156. 2020.
[79] D. Le, K. Nguyen, et al. POODLE: Improving Few-shot Learning via Penalizing Out-of-Distribution Samples. In Proc. Neural Information Processing Systems (NeurIPS). 2021.
[80] C. Liu, et al. An Embarrassingly Simple Baseline to One-Shot Learning. Workshop Computer Vision and Pattern Recognition (CVPR). 2020.
[81] S. Fort, et al. Exploring the limits of out-of-distribution detection. In Proc. NeurIPS, 2021.
[82] Sinha, A., et al., "Negative Data Augmentation," in Proc. ICLR, 2021.
[83] M. Pourreza, "G2D: Generate to Detect Anomaly," Proc. WACV, 2021.
[84] Zaheer, M., et al., "Old is Gold: Redefining the Adversarially Learned One-Class Classifier Training Paradigm," in Proceedings CVPR, 2020.
[85] T. Dietterich. Anomaly Detection for OoD and Novel Category Detection. Amazon, Invited Talk. Annual machine learning conference featured presentations from thought leaders within academia. 2021.
[86] K. Lee, et al, "Training confidence-calibrated classifiers for detecting out-of-distribution samples," In Proc. ICLR, 2018a.
[87] S. Nowozin, et al, "f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization," in Proc. NeurIPS, 2016.
[88] A. Ayub and A. Wagner. Cognitively-inspired model for incremental learning using a few examples. In Proc. CVPR Workshop, 2020.