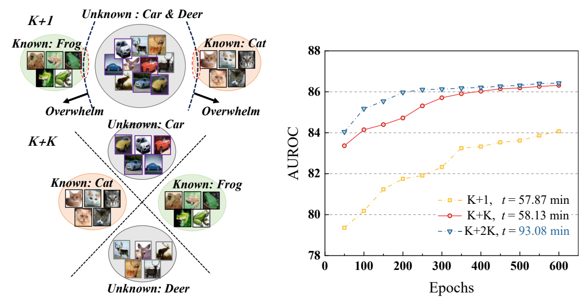
All Beings Are Equal in Open Set Recognition

Chaohua Li[1,2*], Enhao Zhang[1,2*], Chuanxing Geng[1,2], Songcan Chen[1,2†]

[1]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics
[2]MIIT Key Laboratory of Pattern Analysis and Machine Intelligence
{chaohuali, zhangeh, gengchuanxing, s.chen}@nuaa.edu.cn

Abstract

In open-set recognition (OSR), a promising strategy is exploiting pseudo-unknown data outside given K known classes as an additional K+1-th class to explicitly model potential open space. However, treating unknown classes without distinction is *unequal* for them relative to known classes due to the category-agnostic and scale-agnostic of the unknowns. This inevitably not only disrupts the inherent distributions of unknown classes but also incurs both class-wise and instance-wise imbalances between known and unknown classes. Ideally, the OSR problem should model the whole class space as K+∞, but enumerating all unknowns is impractical. Since the core of OSR is to effectively model the boundaries of known classes, this means just focusing on the unknowns nearing the boundaries of targeted known classes seems sufficient. Thus, as a compromise, we convert the open classes from infinite to K, with a novel concept *Target-Aware Universum* (TAU) and propose a simple yet effective framework **D**ual **C**ontrastive Learning with **T**arget-**A**ware **U**niversum (DCTAU). In details, guided by the targeted known classes, TAU automatically expands the unknown classes from the previous 1 to K, effectively alleviating the distribution disruption and the imbalance issues mentioned above. Then, a novel *Dual Contrastive* (DC) loss is designed, where all instances irrespective of known or TAU are considered as positives to contrast with their respective negatives. Experimental results indicate DCTAU sets a new state-of-the-art.



(a) (b)

Figure 1: (a) The K+1 stereotype (*top*) inevitably disrupts the inherent distributions of unknowns and incurs a *bigger* unknown class overwhelming other known classes. The K+K strategy introduced (*below*) can alleviate the issues existing in K+1; (b) An illustrated experiment on partial data of CIFAR10 indicates K+K can be as a compromise. K+K (*red*) outperforms K+1 (*yellow*) by a wide margin and shows comparable performance to K+2K (*blue*), while requiring less time cost.

Introduction

In real-world classification tasks, the deployed model may encounter data from unknown classes as the incomplete knowledge obtained during training(Geng, Huang, and Chen 2020). Thus, a more realistic recognition scenario has been introduced, namely open-set recognition (OSR), aiming to classify known data and reject unknown data simultaneously (Scheirer et al. 2012).

With the rapid emergence of algorithms crafted for OSR, the strategy of utilizing pseudo-unknown data participated in training has garnered increasing attention, which involves obtaining the pseudo-unknown data through GANs or augmentation methods (Ge et al. 2017; Neal et al. 2018; Kong and Ramanan 2021; Du et al. 2022; Zhou, Ye, and Zhan 2021; Dhamija, Günther, and Boult 2018; Gunther et al. 2017; Perera and Patel 2019; Chen et al. 2021; Xu, Shen, and Zhao 2023). This strategy treats the pseudo-unknown classes as an additional K+1-th class, excluding K known classes, and improves performance to varying degrees. However, consider a realistic scenario, even at the first sight of animals and vehicles, we would never categorize them as one class without distinction at all. This reveals that it is counter-intuitive to treat the category-agnostic and scale-agnostic unknowns (*i.e.,* the scale involves the number of instances in each unknown class and the number of classes of unknowns) as the K+1-th class. Such an extreme K+1 strategy will cause two issues. First, the inherent class distributions of unknowns will be disrupted since the initially specific categories of unknowns are fused into a single mixed class. Second, imbalance occurs on both class- and instance-wise: there are K classes for knowns but only one class for unknowns; on the latter, copious unknown instances forced

[*]These authors contributed equally.
[†]Corresponding author.

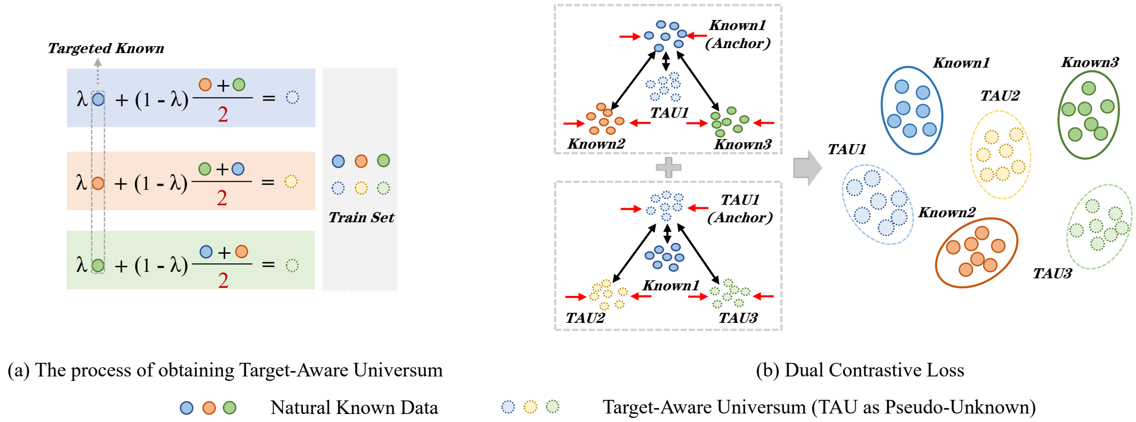


Figure 2: Two components of the proposed framework **D**ual **C**ontrastive Learning with **T**arget-**A**ware **U**niversum (DCTAU). (a) An illustration about how *Target-Aware Universum*(TAU) is generated; (b) The *Dual Contrastive* Loss is defined between the i-th targeted known class and other known classes & the i-th TAU class (*top*) and between the i-th TAU class and other TAU classes & the i-th targeted known class (*below*).

into a single mixed class will overwhelm the known classes, as illustrated in Fig. 1(a) (*top*). The ideal (yet extreme) strategy should adopt the K+∞, where modeling the whole class space involves K classes of known classes and an infinite number of unknown classes. However, in reality, it is impractical to enumerate all unknown classes. To further clarify such two extremes, we design an illustrated experiment on partial data of CIFAR10. As shown in Fig.1 (b), with K+1 as a baseline (*yellow line*), performance improves significantly as the number of pseudo-unknown classes grows to K (*red line*). Afterwards, with the further increase of it, performance plateaus and comes with substantial time consumption. For instance, when reached to $2K$, time consumption surprisingly rises by 60% (*blue line*). Therefore, we believe that using K pseudo-unknown classes to approximate the ∞ open classes is a roughly reasonable option.

Given this, the challenge converts to how to generate K pseudo-unknown classes. Since the core of OSR is recognized to effectively model the boundaries of known classes (Geng, Huang, and Chen 2020; Vaze et al. 2021), here comes an intuition: as seen in Fig. 1(a) (*below*), we should focus on generating K pseudo-unknown classes for open classes nearing the boundaries of K targeted known classes, which naturally converts to a K+K strategy.

Upon the above, we first introduce a novel concept, *Target-Aware Universum*, to serve as these K pseudo-unknown classes. Then we design *Dual Contrastive* loss to learn more discriminative representations around the boundaries between K known classes and K pseudo-unknown classes. We name this framework as Dual Contrastive Learning with Target-Aware Universum (DCTAU). Two main components of the framework are illustrated in Fig. 2. **(1) Target-Aware Universum** (TAU). The concept of TAU originates from the *Universum Learning*, which introduces an external dataset that does not belong to any classes in the task (Weston et al. 2006; Chapelle et al. 2007). To align perfectly with OSR task, TAU is the result of modifying the universum through *Targeted Mixup*, which interpolates a targeted known and the average of the remaining knowns, as seen in Fig. 2(a). By this, each TAU nears to its targeted known since the targeted known contributes more information whereas the remaining known classes are averaged out. Meanwhile, TAU automatically expands the pseudo-unknowns from the previous 1 to K classes which effectively alleviates the distribution disruption and the imbalance issues mentioned above. **(2) Dual Contrastive loss** (DC). To endow unknowns with the equal opportunity to be used as the positives, we design a novel DC loss, where all instances irrespective of known or TAU, aiming to simultaneously optimize greater inter-class margins and intra-class compactness for K TAUs in a similar manner as K known classes, as in Fig. 2(b). In DC loss, each anchor known class as the positive contrasts both other known classes and its TAU class firstly, meanwhile each TAU class contrasts both other TAU classes and its corresponding targeted known class. Further, theoretical analyses provide that DC loss can adaptively adjust the weights between knowns and TAUs in gradient update process to achieve fairly contrast and inherit the inner ability of hard negative mining. Ultimately, these two components enable DCTAU framework to achieve all beings (*i.e.,* known and unknown classes) are equal in OSR. The contributions of this paper are summarized as follows:

- We emphasize the comprehension of unknowns and break the K+1 stereotype to a K+K strategy, achieving all beings are equal in OSR.
- We propose a novel framework Dual Contrastive Learning with Target-Aware Universum (DCTAU) involving Target-Aware Universum and Dual Contrastive loss.
- We theoretically analyzed the effectiveness of DCTAU and extensive experiments show it surpasses the state-of-the-art performance.

Related Work

Open Set Recognition

(Scheirer et al. 2012) pioneered the formalization of open set recognition. (Bendale and Boult 2016) first integrates deep learning into the OSR task and proposed OpenMax. Subsequently, deep learning has gained significant attention in the OSR task. One of those focuses on transforming closed training into open. Apart from directly using natural images, it is broadly classified into two categories:

GANs Based Methods. Seminal work in (Ge et al. 2017) extended Openmax, using a conditional GAN to synthesize unknown data to train the DNNs. Following this, (Neal et al. 2018; Perera et al. 2020; Kong and Ramanan 2021) enhanced the performance by different strategies based on GANs. Recently, (Chen et al. 2021) is an extension of (Chen et al. 2020a) and adds confusing training samples from a generator. However, the training process of GANs is complex and unstable as it needs to train an additional network.

Augmentation Based Methods. (Zhou, Ye, and Zhan 2021) introduced Data Placeholder, which anticipates novel class patterns by Manifold Mixup (Verma et al. 2019). (Cho and Choo 2022) used background-class as *Known Unknown Classes* during training. (Xu, Shen, and Zhao 2023) synthesized vague virtual instances and augmented negatives to enhance representation learning. (Zhu et al. 2023) proposed OpenMix, which learned to reject pseudo data mixed through outlier samples. However, these methods not only consider the scale and quality of augmented instances but also treat the pseudo-unknown as the K+1-th class.

Contrastive Learning

Contrastive learning has become a dominant component in Self-Supervised Learning (SSL) (Duan et al. 2018; Hendrycks et al. 2019; Jaiswal et al. 2020; Misra and Maaten 2020). The standard SSL methods based on InfoNCE loss (Oord, Li, and Vinyals 2018), such as SimCLR(Chen et al. 2020b), MOCO(He et al. 2020) and Self-Con(Bae et al. 2023), has already demonstrated outstanding performance. Further, SupCon(Khosla et al. 2020) extends contrastive learning to the fully supervised setting.

(Vaze et al. 2021) heuristically discovered that a good closed-set accuracy always benefits for open set recognition. Leveraging this viewpoint to the representation learning, (Xu, Shen, and Zhao 2023) made an initial venture to apply Supervised Contrastive Learning to OSR task, and developed ConOSR which utilizes data augmentation and soft label technologies to representation learning. However, progress in this paradigm is bottlenecked by the absence of a contrastive loss specially designed for pseudo-unknowns.

Universum Learning

The initial concept of universum was introduced by (Weston et al. 2006) as a collection cannot be assigned to any target classes in task. In recent years, (Zhang and LeCun 2017) extended this to deep learning. (Nguyen, Morell, and De Baets 2017) introduced a distance metric learning approach that leverages universum data. (Xiao, Feng, and Liu 2021) tackled the challenge of Transductive learning with universum.

The most recent advancements in this field explored the acquisition of universum through the application of Mixup (Zhang et al. 2017). (Han, Geng, and Chen 2023) produced hard negatives in contrastive learning by Mixup-induced universum. (Zhang, Geng, and Chen 2022) used *High-order Mixup* for universum to re-balance the classes in long-tailed recognition. In spite of its efficiency, a lack of designing a novel universum suitable for OSR still persists.

Dual Contrastive Learning with Target-Aware Universum

Preliminaries and Problem Statement

We denote $\mathcal{D}_{tr}$ as a training set consisting n labeled instances $\{(\boldsymbol{x_i}, \boldsymbol{y_i})\}_{i=1}^{n}$, where $\boldsymbol{y_i} \in \{1, ..., K\}$ is the corresponding class label, and N_i denotes the number of instances in each class. Additionally, $\mathcal{D}_{tau}$ represents the set of TAU data derived from $\mathcal{D}_{tr}$, which contains an equal number of n instances$\{\boldsymbol{x_u^i}, \boldsymbol{y_u^i}\}_{i=1}^{n}$, where the labels of pseudo-unknowns $\boldsymbol{y_u^i} \in \{K+1, ..., K+K\}$.

We employ a two-step training strategy. In the contrastive learning step, an encoder network $E(\cdot)$ and a projection network $\psi(\cdot)$ are optimized by the contrastive loss based on the features of $\psi(E(\boldsymbol{x_i}))$ and $\psi(E(\boldsymbol{x_u^i}))$. In the classifier training step, only $\mathcal{D}_{tr}$ is involved in training, and a classifier $f(\cdot)$ is optimized by the cross entropy loss. $\mathcal{D}_{te} = \{\boldsymbol{t_1}, ..., \boldsymbol{t_u}\}_{i=1}^{n}$ denotes a test set which contains the instances drawn from unknown classes.

Target-Aware Universum

Different from existing methods for OSR, we explore the potential to generate universum as the pseudo-unknown data through instances of K known classes straight off the shelf.

Mixup. Traditional Mixup generates data by linear interpolating pairs of training instances. Given $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$, a pair of training instances randomly sampled from $\mathcal{D}_{tr}$, the synthesized instance is defined as (Zhang et al. 2017):

$$\tilde{\boldsymbol{x}} = \lambda \cdot \boldsymbol{x_i} + (1-\lambda)\,\boldsymbol{x_j}, \tag{1}$$

where $\lambda \in [0, 1]$ is sampled from the Beta distribution. However, it only combines the instance-pairs information of two images which simply captures the *local information* result in generating ambiguous samples. Thus, we introduce a variant of Mixup, namely *Targeted Mixup*, to acquire *global information* among all classes. The outcome is referred to as *Target-Aware Universum* (TAU), which eliminates ambiguous samples to a large extent. As shown in Fig. 3.

Definition 1. ***Target-Aware Universum.*** *Given a subset $\mathcal{D}_{sub} = \{(\boldsymbol{x_k}, \boldsymbol{y_k})\}_{k=1}^{K_B}$ where each instance is randomly sampled from each of the K_B known classes in a batch. For $\forall \boldsymbol{x_i} \in \mathcal{D}_{sub}$, its Target-Aware Universum is defined as:*

$$\boldsymbol{x_u^i} = \lambda \cdot \boldsymbol{x_i} + (1-\lambda) \cdot \frac{1}{K_B - 1} \sum\nolimits_{y_j \neq y_i}^{K_B - 1} \boldsymbol{x_j}, \tag{2}$$

where K_B is the number of known classes in a batch.

Compared with the traditional Mixup, besides its ability to eliminate ambiguous samples, the greatest characteristic of Targeted Mixup is targeted known $\boldsymbol{x_i}$ contributes significantly more information to TAU, whereas that of instances in the remaining K-1 known classes is averaged out. Therefore, TAU could be regarded as the high-quality hard negative (Kalantidis et al. 2020), as there is a great overlap of

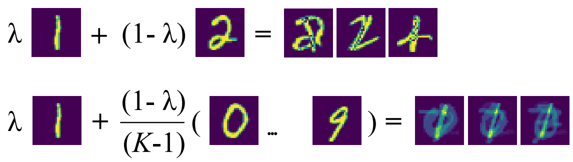


Figure 3: The generated images from mixing digital "1" and "2" may belong to digital "2", "4" and "8" images of known classes(*top*). TAU can highlight the targeted known digital "1" and avoid the *ambiguous samples* (*below*).

semantic information between it and its targeted known. We will discuss why TAU is suitable for K+K in OSR in the latter experimental results.

Further, a more insightful understanding of the essence of *Target-Aware* is that guided by targeted known classes, TAU is endowed with the concept of *independent classes* possessing distinct K classes and N_i samples in per class. Consequently, TAU automatically expands the unknown classes from the previous 1 to K and alleviates the distribution disruption and the imbalance issues on both class-wise and instance-wise.

Dual Contrastive Learning Loss

The contrastive learning loss was introduced to pull an anchor and its positives closer while the negatives are pushed apart(Duan et al. 2018). Particularly, SupCon loss introduces label information into learning. In this framework, given a training instance $\boldsymbol{x_i}$, the network maps it to a representation vector, $\psi(E(\boldsymbol{x_i})) = \boldsymbol{z}_i \in \mathbb{R}^{D_P}$, then the contrastive loss is defined as(Khosla et al. 2020):

$$\mathcal{L}^{sup} = \sum_{i} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_p/\tau)}{\sum_{k \neq i} \exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_k/\tau)}, \quad (3)$$

where $P(i)$ is the set of all positive data in a batch from the class i, and $|P(i)|$ is its cardinality. τ is the scalar temperature hyper-parameter.

With the inclusion of label information, SupCon has resulted in a remarkable improvement in the performance of discrimination. However, given the absence of a dedicated contrastive learning loss designed for unlabeled data, it hinders the comprehensive of unknowns' representation, which unfairly contrasts to unknowns. Thus, we propose an optimized version, namely *Dual Contrastive* (DC) loss, which inherently adapts TAU classes to the same contrastive form of known classes.

The DC loss is composed of two components. In the first part, known classes act as anchors:

$$\mathcal{L}^{k} = \sum_{i \in \mathcal{D}_{tr}} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_p/\tau)}{\sum_{k \neq i} \exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_k/\tau) + U}, \quad (4)$$

where $U = \sum_{\boldsymbol{x_u^i} \in U(i)} \exp(\boldsymbol{z_i} \cdot \boldsymbol{u_i}/\tau)$, $U(i) \in \mathcal{D}_{tau}$ represents the set of TAU relative to i-th targeted known class, and $\psi(E(\boldsymbol{x_u^i})) = \boldsymbol{u}_i \in \mathbb{R}^{D_P}$.

The objective of $\mathcal{L}^k$ is to improve the discrimination between the targeted known class and other known classes as well as its TAU class. Each targeted known class acts as an anchor, minimizing the intra-class distance between positively labeled classes while simultaneously maximizing the inter-class margins from negatively labeled classes and its TAU class.

From the insights provided before, we also hope to treat TAU as fairly as known data, allowing them to enhance discriminability through the contrastive learning framework. Then the second part for TAU is defined as:

$$\mathcal{L}^{u} = \sum_{i \in \mathcal{D}_{tau}} \frac{-1}{|U(i)|} \sum_{p \in U(i)} \log \frac{\exp(\boldsymbol{u}_i \cdot \boldsymbol{u}_p/\tau)}{\sum_{k \neq i} \exp(\boldsymbol{u}_i \cdot \boldsymbol{u}_k/\tau) + K}, \quad (5)$$

where $K = \sum_{\boldsymbol{x_i} \in P(i)} \exp(\boldsymbol{u_i} \cdot \boldsymbol{z_i}/\tau)$, $P(i) \in \mathcal{D}_{tr}$ is the set of data of the targeted known class relative to $\boldsymbol{x_u^i}$ and $U(i)$ is the set of all positive data belongs to $\boldsymbol{x_u^i}$.

$\mathcal{L}^u$, as the dual form of $\mathcal{L}^k$, differs in the following aspects: (1) anchors are replaced by TAU classes; (2)positively labeled classes are replaced by TAU classes, which share the same label with the anchor x_u^i, boosting a compressed intra-class space among them; (3) negatively labeled classes are replaced by TAU classes with different labels, increasing the inter-class margins; (4) meanwhile, the inter-class distance between $\boldsymbol{x_u^i}$ and its targeted known class $\boldsymbol{x_i}$ is enlarged.

Finally, the loss in the contrastive learning step is a combination:

$$\mathcal{L} = \mathcal{L}^k + \gamma \mathcal{L}^u, \quad (6)$$

where γ is a balancing coefficient. $\mathcal{L}$ can be regarded as the comprehensive results of fairly contrasting all instances irrespective of known or TAU with their negatives.

Theoretical Analysis

In this subsection, we conduct theoretical analyses from *Fairly Contrast* and *Hard Negative Mining* perspectives. Since $\mathcal{L}^k$ and $\mathcal{L}^u$ are dual forms of each other, these analyses are similar for both, for simplicity, we consider the $\mathcal{L}^k$ as an example. The details can be found in Appendix A[1].

The gradient for $\mathcal{L}^k$ with respect to $\boldsymbol{z}_i$ can be depicted as:

$$\begin{aligned} \frac{\partial \mathcal{L}^k}{\partial \boldsymbol{z}_i} &= -\frac{1}{\tau}\left(\boldsymbol{z}_p + \frac{1}{S}\sum_{k \neq i} C_k \cdot \boldsymbol{z}_k + \frac{1}{S}\sum_{\boldsymbol{x_u^i} \in U(i)} O_i \cdot \boldsymbol{u}_i\right) \\ &= -\frac{1}{\tau}\left(\boldsymbol{z}_p + G_{NK} + G_{TAU}\right), \end{aligned} \quad (7)$$

where $C_k = \exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_k/\tau)$, $O_i = \exp(\boldsymbol{z}_i \cdot \boldsymbol{u}_i/\tau)$ and $S = \sum_{k \neq i} C_k + \sum_{\boldsymbol{x_u^i} \in U(i)} O_i$. It can be divided into three parts involved in the gradient update process: the representation of positively labeled data, the gradient of negatively labeled known data and the gradient of TAU data.

If TAU is omitted, $\mathcal{L}^k$ reverts to $\mathcal{L}^{sup}$, and the gradient for $\mathcal{L}^{sup}$ with respect to $\boldsymbol{z}_i$ will be written as:

$$\frac{\partial \mathcal{L}^{sup}}{\partial \boldsymbol{z}_i} = -\frac{1}{\tau}\left(\boldsymbol{z}_p + \frac{\sum_{k \neq i} C_k \cdot \boldsymbol{z}_k}{\sum_{k \neq i} C_k}\right), \quad (8)$$

Dataset	MNIST	SVHN	CIFAR10	CIFAR+10	CIFAR+50	TinyImageNet
Softmax	97.8	88.6	67.7	81.6	80.5	57.7
OpenMax	98.1	89.4	69.5	81.7	79.6	57.6
G-OpenMax	98.4	89.6	67.5	82.7	81.9	58.0
OSRCI	98.9	91.0	69.9	83.8	82.7	58.6
CPN	99.0	92.6	82.8	88.1	87.9	63.9
RPL++	99.3	95.1	86.1	85.6	85.0	70.2
GFROSR	-	93.5	80.7	92.8	92.6	60.8
PROSER	-	94.3	89.1	96.0	95.3	69.3
OpenHybrid	99.5	94.7	95.0	96.2	95.5	79.3
ARPL	**99.7**	96.7	91.0	97.1	95.1	78.2
Class-inclusion	-	95.6	94.8	96.1	95.7	78.5
PMAL	99.5	96.3	94.6	96.0	94.3	81.8
All-U-Need(MLS)	99.3	97.1	93.6	97.9	96.5	83.0
Vanilla SupCon	**99.7**	98.8	93.7	97.9	97.0	79.6
ConOSR	**99.7**	99.1	94.2	98.1	97.3	80.9
DCTAU (w/o DC)	95.6	94.9	92.3	95.1	93.7	71.0
DCTAU	**99.7**	**99.2**	**95.6**	**98.5**	**98.1**	**83.6**

Table 1: Open Set recognition results in terms of the AUROC(%). DCTAU (w/o DC) means $\mathcal{L}$ only contains $\mathcal{L}^k$. ”-” means the original paper does not report the corresponding result. Results are averaged among five randomized trials.

Fairly Contrast. Compared with $\partial\mathcal{L}^{sup}/\partial\boldsymbol{z}_i$, the differences in $\partial\mathcal{L}^k/\partial\boldsymbol{z}_i$ primarily reflected in: (1) the denominator of G_{NK} contains $\sum_{\boldsymbol{x}_u^i\in U(i)}O_i$, causing the weight of negatively labeled known data decreases; (2) G_{TAU} added increases the weight of TAU explicitly. In a nutshell, DC loss can inherently adjust the weights between known data and TAU in the gradient update process to pay more attention to unknowns, achieving fairly contrast in OSR.

Hard Negative Mining. Within Eq. (7), our DC loss can adaptively adjust the weights of G_{NK} and G_{TAU} in which all negatives exist. This adjustment is based on whether the hard negatives attributed to G_{NK} or G_{TAU}, which reflects the inner ability of hard negative mining. We further analyze this ability across three situations, and more details will be presented in Appendix A.

Rejecting Unknowns

In this step, a light-weight classifier $f(\cdot)$ is trained by minimizing the cross entropy loss.

Given a training instance $(\boldsymbol{x_i}, \boldsymbol{y_i})$, a conventional Softmax classifier outputs the posterior probability of $\boldsymbol{x_i}$ belonging to the k-th known class by:

$$\hat{y_i} = P(y_i = k|\boldsymbol{x_i}) = \frac{\exp(f_k(E(\boldsymbol{x_i})))}{\sum_{c=1}^{K}\exp(f_c(E(\boldsymbol{x_i})))}, \tag{9}$$

The $f(\cdot)$ is optimized by minimizing the cross entropy loss, $\mathcal{L}^{ce} = -\sum_i y_i \log \hat{y}_i$. Followed the (Xu, Shen, and Zhao 2023), the rejection thresholds percentile ε_i are estimated for detecting unknown data.

During the test, the max posterior probability as confidence score, $\max_{k\in\{1,\cdots,K\}} P(y = k|\boldsymbol{t_i})$, and then a test instance t_i can be estimated as one of the known classes or recognized as the unknown data by:

$$\widehat{y} = \begin{cases} \arg\max_{k\in\{1,\cdots,K\}} P(y = k|\boldsymbol{t_i}), & \text{if } conf \geq \varepsilon, \\ \text{unknown class}, & \text{otherwise.} \end{cases} \tag{10}$$

[1]All technical appendices: https://github.com/SuperL7/DCTAU

Experiments

Experiments for Open Set Recognition

Datasets. Following the protocol defined in (Neal et al. 2018) and the dataset splits with (Chen et al. 2021; Xu, Shen, and Zhao 2023), a summary of 6 benchmark datasets is provided:

- **MNIST,SVHN,CIFAR10.** MNIST(Lake, Salakhutdinov, and Tenenbaum 2015), SVHN(Netzer et al. 2011) and CIFAR10(Krizhevsky, Hinton et al. 2009) all consist 10 of classes, of which 6 classes are randomly selected as known classes and the other 4 classes as unknown.
- **CIFAR+10,CIFAR+50.** In this group of experiments, 4 classes are selected from CIFAR10 as known classes for training, and 10\50 classes sampled from CIFAR100(Krizhevsky, Hinton et al. 2009) as unknown.
- **TinyImageNet.** TinyImageNet is a subset derived from ImageNet(Russakovsky et al. 2015) consisting of 200 classes. 20 known classes and the left 180 unknown classes are randomly sampled for evaluation.

Implementations. In the contrastive learning step, the feature encoder backbone is the same with (Neal et al. 2018), and an MLP with two fully connected layers is employed as the projection network. In the classifier training step, the network is also an MLP with a 128-node fully connected layer. Followed with (Xu, Shen, and Zhao 2023), the training epochs of these two steps are 600 and 20 respectively. More details will be found in Appendix B.

Dataset	MNIST	SVHN	CIFAR10	CIFAR+10	CIFAR+50	TinyImageNet
Softmax	99.2	92.8	83.8	90.9	88.5	60.8
GCPL	99.1	93.4	84.3	91.0	88.3	59.3
RPL	99.4	93.6	85.2	91.8	89.6	53.2
ARPL	99.4	94.0	86.6	93.5	91.6	62.3
ARPL+CS	99.5	94.3	87.9	94.7	92.9	65.9
Class-inclusion	-	85.4	87.0	88.1	86.5	49.3
DCTAU	**99.6**	**96.2**	**93.9**	**97.2**	**97.1**	**77.6**

Table 2: The open set classification rate OSCR(%) curve results of open set recognition. ”-” means the original paper does not report the corresponding result. Results are averaged among five randomized trials.

Evaluation Metrics. Area Under the Receiver Operating Characteristic (AUROC) curve(Fawcett 2006) is used for detecting unknown data from test set. Open Set Classification Rate (OSCR) (Dhamija, Günther, and Boult 2018; Wang et al. 2022) is employed for evaluating correct classifications of known classes. Details will be shown in Appendix B.

Results Comparison. The baselines compared with DCTAU includes Softmax Thresholding(Hendrycks and Gimpel 2016), OpenMax(Bendale and Boult 2016), GOpenMax(Ge et al. 2017), OSRCI(Neal et al. 2018), CPN(Yang et al. 2020), C2AE(Oza and Patel 2019), RPL++(Chen et al. 2020a), GFROSR(Perera et al. 2020), PROSER(Zhou, Ye, and Zhan 2021), OpenHybrid(Zhang et al. 2020), ARPL(Chen et al. 2021), Class-inclusion(Cho and Choo 2022), PMAL(Lu et al. 2022), All-U-Need(MLS)(Vaze et al. 2021), and ConOSR(Xu, Shen, and Zhao 2023).

We provide the AUROC performance of different methods in Table 1. DCTAU shows a significant superiority in performance over almost all the methods. Compared to the method employing the contrastive framework, our DCTAU outperforms the recently proposed ConOSR across all datasets, in particular, on the most challenging dataset TinyImageNet by a margin of 2.7%.

The results of OSCR are shown in Table 2. DCTAU shows a remarkable performance improvement on all datasets. Surprisingly, compared to the second best, the OSCR of DCTAU increased by 11.7% on TinyImageNet.

Experiments for Intrinsic Mechanism of DCTAU

In this group of experiments, we follow the protocol in (Yoshihashi et al. 2019; Zhou, Ye, and Zhan 2021) to explore the intrinsic mechanism for the efficacy of DCTAU through Out-of-distribution detection(OOD). Here, all training classes of the original dataset are used as In-distribution(ID) data. While instances from another dataset are added to the test set as the OOD data.

Datasets.

- **ID:MNIST/OOD:Omniglot,Noise,MNIST-Noise.** Omniglot(Lake, Salakhutdinov, and Tenenbaum 2015) is a dataset of hand-written alphabet characters. Noise is a set of images synthesized by sampling each pixel value from a uniform distribution on [0, 1]. MNIST-Noise is also a synthesized set by superimposing MNIST's test images on Noise. The number of OOD data is 10,000, equal to the test data of MNIST.
- **ID:CIFAR10/OOD:ImageNet,LSUN.** The OOD data is sampled from ImageNet and LSUN(Yu et al. 2015). Since the image size of ImageNet and LSUN is different from CIFAR10, we use two different ways to process them. CIFAR10, ImageNet and LSUN all have 10,000 images of their test set. More details are in Appendix C.

Dataset	Omniglot	MNIST-Noise	Noise
Softmax	59.5	64.1	82.9
OpenMax	68.0	72.0	82.6
CROSR	79.3	82.7	82.6
PRESER	86.2	87.4	88.2
ConSOR	95.4	98.7	98.8
DCTAU	**96.5**	**99.3**	**99.3**

Table 3: Out-of-Distribution Detection on MNIST with various datasets added to the test set as unknowns. We report macro F1-scores.

Dataset	ImageNet (Crop)	ImageNet (Resize)	LSUN (Crop)	LSUN (Resize)
Softmax	63.9	65.3	64.2	64.7
OpenMax	66.0	68.4	65.7	66.8
OSRCI	63.6	63.5	65.0	64.8
CROSR	72.1	73.5	72.0	74.9
GFROSR	75.7	79.2	75.1	80.5
PRESER	84.9	82.4	86.7	85.6
ConOSR	89.1	84.3	91.2	88.1
DCTAU	**94.7**	**93.2**	**95.1**	**94.5**

Table 4: Out-of-Distribution Detection on CIFAR10 with various datasets added to the test set as unknowns. We report macro F1-scores.

Evaluation Metrics. The macro-averaged F1-scores over all ID and OOD class is used to measure the performance.

Results Comparison. The results of the ID:MNIST setting are reported in Table 3. DCTAU significantly outperforms other methods. It achieves 99.3% for detecting noisy OOD images, while 96.5% in Omniglot dataset for detecting semantic OOD images. The results of the ID:CIFAR10 setting

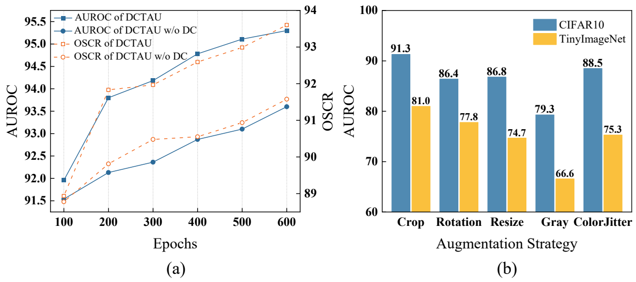


Figure 4: (a) AUROC and OSCR of DCTAU and DCTAU(w/o DC) with varying epochs. The experiments are conducted on CIFAR10; (b) AUROC of DCTAU with varying augmentation techniques. The experiments are conducted on CIFAR10 and TinyImageNet.

are reported in Table 4. We can see DCTAU also handles these scenarios with the best performance. Especially, for the ImageNet-Resize and LSUN-Resize datasets, DCTAU excels more than 8.9% and 6.4%. Further elaborations about the intrinsic mechanism for the efficacy of DCTAU from two aspects based on *Familiarity Hypothesis* (Dietterich and Guyer 2022) can be found in Appendix C.

In summary, the contrastive learning framework reinforces the grasp of learning the most distinctive information of features among known classes. And our DCTAU enhances this ability, since TAU inherently introduces the distinctiveness to pseudo-unknowns (*i.e.,* regards them as K independent classes) and DC loss specifically enhances the contrastive learning for more distinctive information of features of pseudo-unknowns.

Detailed Analysis

Why Target-Aware Universum is Effective. In this subsection, we perform experiments to shed light on the reasons behind the performance boost attained by employing TAU data as the pseudo-unknowns in the K+K strategy.

We conduct a group of experiments on known classes from CIFAR10, to compare the AUROC and OSCR of TAU data with other images including Noise-Gaussian, Noise-Uniform, Natural Images and images generated by traditional Mixup serving as the pseudo-unknown data. The Noise-Gaussian and Noise-Uniform images are the pure noisy images synthesized from a Gaussian distribution and a uniform distribution, respectively. The SVHN dataset is used as the Natural Images and traditional Mixup refers to the images generated based on Eq. (1). The results are shown in Table 5.

We interpret the results from the perspective of *Hard negative* in contrastive learning. The hard negatives are crucial in learning highly transferable visual representations(Kalantidis et al. 2020). The main reason of TAU can be regarded as high-quality hard negatives is that while maintaining better quality of semantic information about their targeted known class, TAU data also possess a certain degree of visual ambiguity. In contrast, Noise-Gaussian and Noise-Uniform guarantee visual ambiguity but are severely limited in the quality of semantic information. Despite the Natural Images can ensure quality, its substantial semantic shift makes them unsuitable as high-quality hard negatives. The decrease in performance of Plain Mixup images is primarily attributed to it may generate ambiguous samples.

Method	AUROC	OSCR
Noise-Gaussian	83.2	77.1
Noise-Uniform	85.1	81.4
Natural Images	91.1	88.0
Mixup Images	93.8	92.2
TAU	**95.6**	**93.9**

Table 5: Different data as the Pseudo-unknown employed in the K+K strategy.

λ	0.1	0.3	0.5	0.7	0.9
AUROC	93.5	94.2	**95.6**	95.1	95.0

Table 6: AUROC under different λ for TAU.

Effect of Dual Contrastive loss. Table 1 presents the results of the performance of DCTAU without $\mathcal{L}^u$ which solely emphasizes the discriminative nature of known classes. There is a significant decline of DCTAU (w/o DC) compared to the complete DCTAU, and it even lags behind other baselines. The main reason is the TAU nears to the targeted known class in the feature space. Without a specifically designed contrastive loss to constrain them, they may be easily confused with targeted known classes. This validates the importance of DC loss for our proposed DCTAU framework, and highlights our main viewpoint, that is to handle both known and unknown classes equally in the feature space.

Ablation of Hyper-parameters. (1)About Contrastive Learning. We modify epochs and the augmentation techniques, both of which are sensitive to contrastive learning. Fig. 4(a) showcases the influence of epochs on CIFAR10. The augmentation technique used in ConOSR is RandAugment(Cubuk et al. 2020). To investigate the impact of different augmentation techniques, we conduct a series of experiments on CIFAR10 and TinyImageNet datasets. The results are shown in Fig. 4(b). **(2) λ for TAU.** For Targeted Mixup weight λ, we vary it from 0.1 to 0.9 and 0.5 achieves the best performance. **(3)Threshold ϵ and Visualization.** Details about both can be found in Appendix D.

Conclusion

This paper analyzes the drawbacks of the K+1 stereotype, and first introduces a K+K strategy that emphasizes the comprehension of unknowns during training. Based on this guidance, we propose a novel framework involving two components to achieve all beings are equal in OSR. And extensive experiments on various benchmarks show ours outperforms the state-of-the-art approaches. In future work, we will further explore approaches following the K+K strategy and establish the relevant theoretical foundation.

Acknowledgments

This work was Supported by the National Natural Science Foundation of China (Grant No. 62076124, 62106102), the Natural Science Foundation of Jiangsu Province (Grant No. BK20210292) and the Scientific and technological achievements transferring project of Jiangsu Province (Grant No. BA2021005).

References

Bae, S.; Kim, S.; Ko, J.; Lee, G.; Noh, S.; and Yun, S.-Y. 2023. Self-Contrastive Learning: Single-Viewed Supervised Contrastive Framework Using Sub-network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 197–205.

Bendale, A.; and Boult, T. E. 2016. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1563–1572.

Chapelle, O.; Agarwal, A.; Sinz, F.; and Schölkopf, B. 2007. An analysis of inference with the universum. *Advances in neural information processing systems*, 20.

Chen, G.; Peng, P.; Wang, X.; and Tian, Y. 2021. Adversarial reciprocal points learning for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 8065–8081.

Chen, G.; Qiao, L.; Shi, Y.; Peng, P.; Li, J.; Huang, T.; Pu, S.; and Tian, Y. 2020a. Learning open set network with discriminative reciprocal points. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, 507–522. Springer.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020b. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607. PMLR.

Cho, W.; and Choo, J. 2022. Towards accurate open-set recognition via background-class regularization. In *European Conference on Computer Vision*, 658–674. Springer.

Cubuk, E. D.; Zoph, B.; Shlens, J.; and Le, Q. V. 2020. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 702–703.

Dhamija, A. R.; Günther, M.; and Boult, T. 2018. Reducing network agnostophobia. *Advances in Neural Information Processing Systems*, 31.

Dietterich, T. G.; and Guyer, A. 2022. The familiarity hypothesis: Explaining the behavior of deep open set methods. *Pattern Recognition*, 132: 108931.

Du, X.; Wang, Z.; Cai, M.; and Li, Y. 2022. Vos: Learning what you don't know by virtual outlier synthesis. *arXiv preprint arXiv:2202.01197*.

Duan, Y.; Zheng, W.; Lin, X.; Lu, J.; and Zhou, J. 2018. Deep adversarial metric learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2780–2789.

Fawcett, T. 2006. An introduction to ROC analysis. *Pattern recognition letters*, 27(8): 861–874.

Ge, Z.; Demyanov, S.; Chen, Z.; and Garnavi, R. 2017. Generative openmax for multi-class open set classification. *arXiv preprint arXiv:1707.07418*.

Geng, C.; Huang, S.-j.; and Chen, S. 2020. Recent advances in open set recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10): 3614–3631.

Gunther, M.; Cruz, S.; Rudd, E. M.; and Boult, T. E. 2017. Toward open-set face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 71–80.

Han, A.; Geng, C.; and Chen, S. 2023. Universum-inspired Supervised Contrastive Learning. *IEEE Transactions on Image Processing*.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9729–9738.

Hendrycks, D.; and Gimpel, K. 2016. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*.

Hendrycks, D.; Mazeika, M.; Kadavath, S.; and Song, D. 2019. Using self-supervised learning can improve model robustness and uncertainty. *Advances in neural information processing systems*, 32.

Jaiswal, A.; Babu, A. R.; Zadeh, M. Z.; Banerjee, D.; and Makedon, F. 2020. A survey on contrastive self-supervised learning. *Technologies*, 9(1): 2.

Kalantidis, Y.; Sariyildiz, M. B.; Pion, N.; Weinzaepfel, P.; and Larlus, D. 2020. Hard negative mixing for contrastive learning. *Advances in Neural Information Processing Systems*, 33: 21798–21809.

Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33: 18661–18673.

Kong, S.; and Ramanan, D. 2021. Opengan: Open-set recognition via open data generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 813–822.

Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.

Lake, B. M.; Salakhutdinov, R.; and Tenenbaum, J. B. 2015. Human-level concept learning through probabilistic program induction. *Science*, 350(6266): 1332–1338.

Lu, J.; Xu, Y.; Li, H.; Cheng, Z.; and Niu, Y. 2022. Pmal: Open set recognition via robust prototype mining. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 1872–1880.

Misra, I.; and Maaten, L. v. d. 2020. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6707–6717.

Neal, L.; Olson, M.; Fern, X.; Wong, W.-K.; and Li, F. 2018. Open set learning with counterfactual images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 613–628.

Netzer, Y.; Wang, T.; Coates, A.; Bissacco, A.; Wu, B.; and Ng, A. Y. 2011. Reading digits in natural images with unsupervised feature learning.

Nguyen, B.; Morell, C.; and De Baets, B. 2017. Distance metric learning with the Universum. *Pattern Recognition Letters*, 100: 37–43.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Oza, P.; and Patel, V. M. 2019. C2ae: Class conditioned auto-encoder for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2307–2316.

Perera, P.; Morariu, V. I.; Jain, R.; Manjunatha, V.; Wigington, C.; Ordonez, V.; and Patel, V. M. 2020. Generative-discriminative feature representations for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11814–11823.

Perera, P.; and Patel, V. M. 2019. Deep transfer learning for multiple class novelty detection. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, 11544–11552.

Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115: 211–252.

Scheirer, W. J.; de Rezende Rocha, A.; Sapkota, A.; and Boult, T. E. 2012. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7): 1757–1772.

Vaze, S.; Han, K.; Vedaldi, A.; and Zisserman, A. 2021. Open-set recognition: A good closed-set classifier is all you need? *arXiv preprint arXiv:2110.06207*.

Verma, V.; Lamb, A.; Beckham, C.; Najafi, A.; Mitliagkas, I.; Lopez-Paz, D.; and Bengio, Y. 2019. Manifold mixup: Better representations by interpolating hidden states. In *International conference on machine learning*, 6438–6447. PMLR.

Wang, Z.; Xu, Q.; Yang, Z.; He, Y.; Cao, X.; and Huang, Q. 2022. OpenAUC: Towards AUC-Oriented Open-Set Recognition. *Advances in Neural Information Processing Systems*, 35: 25033–25045.

Weston, J.; Collobert, R.; Sinz, F.; Bottou, L.; and Vapnik, V. 2006. Inference with the universum. In *Proceedings of the 23rd international conference on Machine learning*, 1009–1016.

Xiao, Y.; Feng, J.; and Liu, B. 2021. A new transductive learning method with universum data. *Applied Intelligence*, 1–13.

Xu, B.; Shen, F.; and Zhao, J. 2023. Contrastive Open Set Recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 10546–10556.

Yang, H.-M.; Zhang, X.-Y.; Yin, F.; Yang, Q.; and Liu, C.-L. 2020. Convolutional prototype network for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5): 2358–2370.

Yoshihashi, R.; Shao, W.; Kawakami, R.; You, S.; Iida, M.; and Naemura, T. 2019. Classification-reconstruction learning for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4016–4025.

Yu, F.; Seff, A.; Zhang, Y.; Song, S.; Funkhouser, T.; and Xiao, J. 2015. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*.

Zhang, E.; Geng, C.; and Chen, S. 2022. Class-Aware Universum Inspired Re-Balance Learning for Long-Tailed Recognition. *arXiv preprint arXiv:2207.12808*.

Zhang, H.; Cisse, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2017. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*.

Zhang, H.; Li, A.; Guo, J.; and Guo, Y. 2020. Hybrid models for open set recognition. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, 102–117. Springer.

Zhang, X.; and LeCun, Y. 2017. Universum prescription: Regularization using unlabeled data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Zhou, D.-W.; Ye, H.-J.; and Zhan, D.-C. 2021. Learning placeholders for open-set recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4401–4410.

Zhu, F.; Cheng, Z.; Zhang, X.-Y.; and Liu, C.-L. 2023. OpenMix: Exploring Outlier Samples for Misclassification Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12074–12083.