Joint Class-level and Instance-level Relationship Modeling for Novel Class Discovery

Jiaying Zhou[1, 2], **Qingchao Chen**[1, 2, 3] *

[1]National Institute of Health Data Science, Peking University, Beijing, China
[2]Institute of Medical Technology, Peking University Health Science Center, Beijing, China
[3] State Key Laboratory of General Artificial Intelligence, Peking University, Beijing, China
{zhoujiaying, qingchao.chen}@pku.edu.cn

Abstract

Novel class discovery(NCD) aims to cluster the unlabeled data with the help of a labeled set containing different but related classes. The key to solving NCD is the knowledge transfer between labeled and unlabeled sets. Since NCD requires that known classes and unknown classes are related, it is significant to explore class-level relationships between known and unknown for more effective knowledge transfer. However, most existing methods either facilitate knowledge transfer by learning a shared representation space or by modeling coarse-grained or asymmetric relationships between known and unknown, neglecting class-level relationships. To tackle these challenges, we propose a symmetric class-to-class relationship modeling and knowledge transfer method, achieving bidirectional knowledge transfer at class-level. Considering that class-level modeling often overlooks the subtle distinctions between samples, we propose pairwise similarity-based relationship modeling and consistency constraint for instance-level knowledge transfer. Extensive experiments on CIFAR100 and three fine-grained datasets demonstrate that our method achieves significant performance improvements compared to state-of-the-art methods.

Introduction

The remarkable success of deep learning in image classification relies heavily on abundant labeled data. However, data-labeling is cost-prohibitive at scale. In addition, labeled data cannot encompass all potential categories. This means that the trained model can only recognize a limited set of predefined categories, lacking the ability to discover and identify novel classes. Inspired by this, *Novel Class Discovery(NCD)* has been proposed and garnered significant attention(Troisemaine et al. 2023; Zhu et al. 2024).

NCD refers to the process of identifying previously unknown classes within an unlabeled set with the help of knowledge from a labeled set. Unlike semi-supervised learning, NCD typically assumes that the samples in the unlabeled set are all from unknown classes and the number of unknown classes is known. Existing NCD methods can be divided into two categories: two-stage methods and one-stage methods. Two-stage methods typically involve training a network on labeled data and then applying the learned network to unlabeled data(Hsu, Lv, and Kira 2017; Hsu et al. 2019). In two-stage methods, labeled data and unlabeled data streams remain disjoint. On the contrary, one-stage methods focus on the simultaneous utilization of labeled and unlabeled data(Fini et al. 2021; Li et al. 2023a).

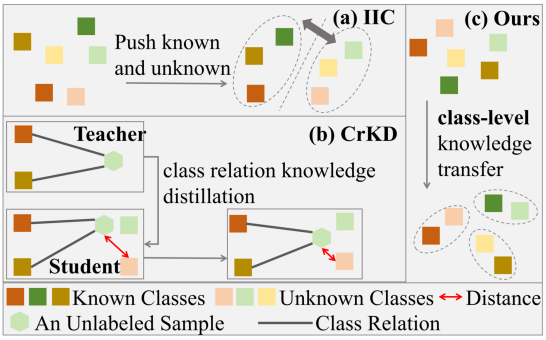


Figure 1: Differences between IIC, CrKD and our method. In figure, different colors represent different classes, classes with the same color family(such as dark green and light green) have higher semantic similarity. Unlabeled sample in light green belongs to the class in same color.

The basic hypothesis of NCD is that unknown classes are related to known classes(Zhao and Han 2021). Therefore, a widely accepted key to solving NCD is the knowledge transfer from known classes to unknown classes(Li et al. 2023b, 2022). Unfortunately, most existing work achieves knowledge transfer by learning a shared representation space, without considering how to leverage the the relationships between known and unknown classes to enhance performance. Recently, IIC(Li et al. 2023a) and CrKD(Gu et al. 2023) proposes to achieve knowledge transfer through relationship modeling. In detail, IIC emphasizes the "disjoint" characteristic of known and unknown classes, and forces known classes to be distant from unknown ones. As shown in Figure.1(a), this "forced separation" may cause classes with high semantic similarity to be pushed apart.

CrKD, on the other hand, takes the unlabeled samples' predicted logits output by classification head trained on known classes as class relation representation. It claims that such class relation should be maintained during discovery stage and proposes a knowledge distillation framework for

*Corresponding author.

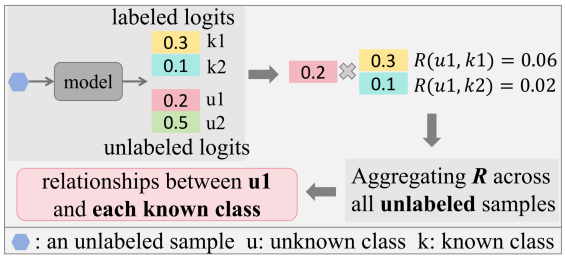


Figure 2: Calculation of relationships between an unknown class and each known class. "**R**" means "Relationship".

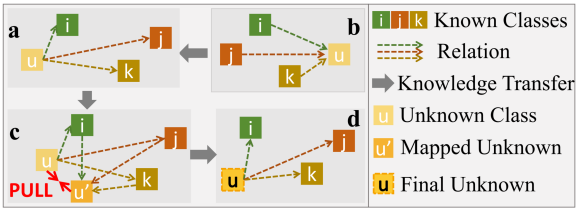


Figure 3: Knowledge transfer at class-level. (a) Transition matrix from unknown to known; (b) Transition matrix from known to unknown. Through "cyclic transition" by using known classes as transfer medium, we can obtain relationships between unknown classes in (c) and then we maximize the similarity between the unknown class **u** and itself(**u'**), which corresponds to "**PULL**" operation. In (d), unknown class **u** is adjusted to an appropriate location.

this purpose. However, class relation encodes relationships between unlabeled samples and known classes. This asymmetric relationship modeling means that the known(labeled) end is adjusted at class-level, while the unknown(unlabeled) end is adjusted at sample-level during knowledge distillation. As shown in Figure.1(b), to maintain the class relation, some unlabeled samples may approach the prototypes of other unknown classes, causing incorrect classification results, which exacerbates bias towards known classes. To sum up, although both IIC and CrKD attempts to model the relationships between known and unknown, ***the relationships they model are either coarse-grained(set of known classes – set of unknown classes) or asymmetric(unlabeled samples – known classes), rather than class-level relationships emphasized in the basic hypothesis of NCD, which may lead to sub-optimal results***.

To address the above issues, we propose a method to ***explicitly model the class-level relationships in a symmetric and fine-grained manner***, enabling more effective knowledge transfer. Considering that NCD does not require correlations between known classes or correlations between unknown classes, we propose to only model the symmetric class-to-class relationships between known and unknown classes. Furthermore, while class-level knowledge transfer is meaningful, it relies on the average of all samples, overlooking the fine-grained distinctions between samples, which limits its effectiveness for individual cases. Thus we propose instance-level relationship modeling and knowledge transfer to reveal subtle characteristics that might be overlooked by averaged class semantic relationships.

To achieve these goals, we need to clarify how to measure the class-level and instance-level relationships. Following CrKD, we employ predicted logits output by classifier to measure relationships. For convenience, we will refer to the logits on known classes as labeled logits and the logits on unknown classes as unlabeled logits. And readers can equate known with labeled and unknown with unlabeled. For class-level relationships, we have following assumptions: (1) For a sample, excluding its ground-truth class, the larger the logit is on a class, the higher the similarity is between the sample and this class, which indicates a stronger correlation. (2) The relationship between a class and others can be measured by the average similarity between all samples belonging to this class and other classes.

Based on the above two assumptions, a known class's relationship with each unknown class can be measured by averaging the unlabeled logits of all samples belonging to this class. However, this straight-forward idea can not be applied to an unknown class because we are unable to identify which samples belonging to it. Therefore, as shown in Figure.2, to measure the relationships between an unknown class and each known class, we propose to use all unlabeled samples' logits on this unknown class as weights to compute the weighted sum of labeled logits. Then by weighting unlabeled samples' labeled logits with unlabeled logits, we can obtain relationships between unknown classes and each known class, i.e., transition matrix from unknown to known. To ensure fairness, we can also obtain transition matrix from known to unknown by weighting labeled samples' unlabeled logits with labeled logits.

Given the estimated two transition matrices, we need to consider how to leverage these relationships to achieve class-level knowledge transfer. We believe that transition probability from a class to itself should be the largest among other transition directions. Inspired by this, we perform "cyclic transition" by multiplying the two transition matrices. In this way, as shown in Figure.3, using known classes as transfer medium, we can obtain relationships between unknown classes, i.e., transition matrix from unknown to unknown. We propose to maximize the transfer probability from a class to itself, which corresponds to the "**PULL**" operation in Figure.3(c). In this way, the prototype of the unknown class is further adjusted. Similarly, using unknown classes as transfer medium, the prototypes of known classes can be adjusted continuously. Thus we achieve ***bidirectional class-level knowledge transfer***.

For instance-level relationships, we use widely-applied pairwise similarity as the measurement and transfer knowledge within and between sets(labeled set and unlabeled set). The information encoded in labeled logits and unlabeled logits differs for labeled and unlabeled samples. For example, labeled samples' labeled logits encode the predictions of the ground truth class, while unlabeled samples' labeled logits encode the relationships between unlabeled samples and known classes. Taking this into consideration, we decouple the logits-based pairwise similarity into labeled-logits-based

pairwise similarity and unlabeled-logits-based pairwise similarity. We believe they should be consistent and we use KL-divergence to constrain this consistency.

To validate the effectiveness of our method, we conduct extensive experiments on four benchmarks, including CIFAR100, Stanford Cars, CUB, and FGVC-Aircraft. The results show that our method can significantly boost the performance compared to existing methods. In summary, our contributions are as follows:

- We propose a novel method that models symmetric class-to-class relationships between known and unknown, achieving bidirectional class-level knowledge transfer.
- We propose to constrain the consistency of pairwise similarity among samples, achieving intra-set and cross-set instance-level knowledge transfer.
- We evaluate our method on four datasets and achieve notable improvements compared to state-of-the-art methods, highlighting the effectiveness of our design.

Related Work

Novel Class Discovery. Novel class discovey(NCD) aims to discover undefined classes in an unlabeled set with the help of labeled set. The early work of NCD mostly employs two-stage approach, including KCL(Hsu, Lv, and Kira 2017), MCL(Hsu et al. 2019) and DTC(Han, Vedaldi, and Zisserman 2019) that apply networks trained on labeled set to unlabeled set. Specifically, KCL and MCL learn similarity prediction networks and cluster based on pairwise similary. DTC learns a feature extraction network and then uses DEC(Xie, Girshick, and Farhadi 2016) to cluster.

Subsequent work mostly adopts a single-stage approach. RS(Han et al. 2020) uses rank statistics to transfer the knowledge to the unlabeled set. DualRank(Zhao and Han 2021) expands RS to a two-branch framework focusing on both local and global features. Afterward, NCL(Zhong et al. 2021a) uses contrastive loss to learn discriminatory features. OpenMix(Zhong et al. 2021b) mixes the unlabeled and labeled samples to build relationship between labeled and unlabeled set. UNO(Fini et al. 2021) introduces a unified objective function to collaborate supervised and unsupervised learning. SNCD(Wang et al. 2024) proposes a semantic-guided method that introduces unknown classes' category name. There are also some work that focus on debias(Feng et al. 2024) and forgetting(Joseph et al. 2022).

Some methods propose to model the relationships to boost performance. IIC(Li et al. 2023a) focuses on the "disjoint" nature of labeled set and unlabeled set. CrKD(Gu et al. 2023) only cares about relationships between unlabeled samples and known classes. Although both consider relationships, neither of them establishes relationships between known and unknown at class-level. Different from them, we propose to model class-level and instance-level relationships to achieve knowledge transfer for novel class discovery.

Method

Overall

Problem Formulation. In NCD, the dataset is divided into two subsets: labeled set $D^l = \{(\mathbf{x}_1^l, y_1^l), ..., (\mathbf{x}_N^l, y_N^l)\}$ and unlabeled set $D^u = \{\mathbf{x}_1^u, ..., \mathbf{x}_M^u\}$, where each $\mathbf{x}_i^l$ or $\mathbf{x}_j^u$ is an image, y_i^l is corresponding label for $\mathbf{x}_i^l$, and N and M represent the numbers of labeled and unlabeled samples respectively. Known classes in the labeled set and unknown classes in the unlabeled set are disjoint, thus known class labels $\mathcal{Y}^l$ and unknown class labels $\mathcal{Y}^u$ can be respectively represented as: $\mathcal{Y}^l = \{1, ..., C^l\}$ and $\mathcal{Y}^u = \{C^l+1, ..., C^l+C^u\}$, where C^l is the number of known classes and C^u is the number of unknown classes.

Framework. As shown in Figure.4(a), our model consists of an encoder E and two classification heads: labeled head h for known classes and unlabeled head g for unknown classes. The encoder is a standard CNN network or a ViT that converts input images into features. h is a linear classifier with C^l output neurons, and g consists of a MLP and a linear classifier with C^u output neurons. Our model training is divided into two stages: in the pretraining stage, we train the encoder and labeled head using labeled data in a supervised manner; in the discovery stage, we train the entire model with both labeled and unlabeled data.

In the discovery stage, given an image $\mathbf{x}_i$, we first project it to the feature $\mathbf{z}_i$. Then, we feed the feature into the two classification heads, regardless of whether $\mathbf{x}_i$ is labeled or unlabeled. The logits $\boldsymbol{l}_h \in \mathcal{R}^{C^l}$ generated by h and $\boldsymbol{l}_g \in \mathcal{R}^{C^u}$ generated by g are concatenated: $\boldsymbol{l} = [\boldsymbol{l}_h, \boldsymbol{l}_g] \in \mathcal{R}^{C^l+C^u}$. Next, concatenated logits $\boldsymbol{l}$ is fed into softmax layer σ which produces the class probability distribution: $\mathbf{p} = \sigma(\boldsymbol{l}/\tau)$, where τ is the temperature coefficient.

We use cross-entropy loss to simultaneously classify labeled samples and cluster unlabeled samples. The assignment of pseudo-labels can be solved following conventional procedures in UNO(Fini et al. 2021). Once we get pseudo-labels, we can calculate cross-entropy loss as follows:

$$\mathcal{L}_{ce} = -\frac{1}{B}\sum_{b=1}^{B}\sum_{c=1}^{C}\boldsymbol{y}^b(c)log(\boldsymbol{p}^b(c)), \tag{1}$$

where $C = C^u + C^l$, $\boldsymbol{y}^b$ is the b-th image's zero-padded ground-truth label or pseudo-label in a batch, $\boldsymbol{p}^b$ is the b-th image's prediction probability, and (c) represents the c-th element of the vector.

According to the basic hypothesis and prior research(Li et al. 2023b), the correlation between known and unknown classes is the premise, and knowledge transfer is the key to solving NCD. Therefore, analyzing the relationship between known and unknown classes is of great importance. However, existing work either focus on constructing a shared representation space while *neglecting* how known and unknown classes are related and how to utilize class-level relationships for more effective knowledge transfer; or attempt to construct *coarse-grained* and *asymmetric* relationships without adequately addressing the class-to-class relationships between known and unknown. In this paper, we propose a class-level relationship modeling method through predicted logits. The relationships we model are "**directional**" and "**symmetric**", specifically from known to unknown and from unknown to known. By performing "cyclic transition" on relationships, we optimize the learning of unknown classes using known classes as medium

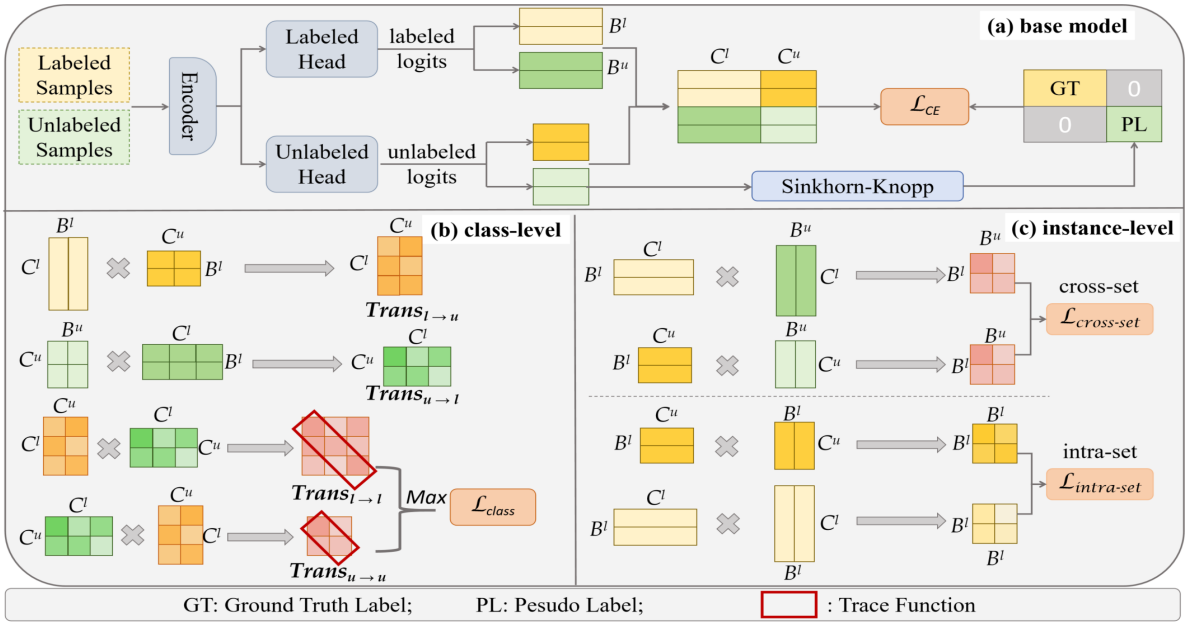


Figure 4: Overview of our proposed method. In (a), we show the architecture of our model. The cross-entropy serves as the main loss for classification and clustering. In the upper of (b), we demonstrate class-level relationship modeling; in the lower part, we achieve class-level knowledge transfer by modeled relationships. In (c), we show the instance-level pairwise similarity computation and consistency constraint, and we omit the relationships within unlabeled set due to space limits.

and optimize the learning of known classes using unknown classes as medium simultaneously. Thus, we achieve ***symmetric bidirectional class-level knowledge transfer***.

Class-level relationship modeling may obscure distinctions between samples within closely related classes by emphasizing general similarities over specific differences. Therefore, instance-level relationship modeling is essential for capturing finer-grained distinctions and refining decision boundaries. In this paper, we use pairwise similarity to model the relationships between samples within and between labeled set and unlabeled set. Considering the information encoded in labeled logits and unlabeled logits differs for labeled and unlabeled sample, we decouple logits-based pairwise similarity into labeled-logits-based pairwise similarity and unlabeled-logits-based pairwise similarity. We believe that decoupled pairwise similarity should be consistent, and we use KL divergence to enforce this consistency.

Class-level Relationship Modeling and Knowledge Transfer

NCD requires that known classes and unknown classes are related, so analyzing the relationships between known and unknown classes is crucial for achieving knowledge transfer and solving the NCD problem. However, previous work largely focus on learning a shared representation space, while neglecting how known and unknown classes are related. CrKD, for the first time, proposed that the logits output by the classification head encode the similarity structure between samples and classes, but it is limited to asymmetric relationships between unlabeled samples and known classes rather than relationships between known and unknown classes, which may exacerbate bias and lead to sub-optimal results as shown in Figure.1(b).

In this paper, we take it a step further by proposing a class-level relationship modeling method based on logits. Based on the assumptions regarding class-level relationship measurement presented in *Introduction*, the relationships we model are **directional and symmetric**. We represent relationships between known classes and each unknown class as transition matrix from known to unknown, and relationships between unknown classes and each known class as transition matrix from unknown to known. Transition matrix from unknown to known can be roughly measured by weighting unlabeled samples' labeled logits with unlabeled logits, as shown in Figure.2; while transition matrix from known to unknown can be measured by weighting labeled samples' unlabeled logits with labeled logits. Once we get these two transition matrices, we perform "cyclic transition" by multiplying them. Using unknown classes as medium, we can get transition matrix from known to known; using known classes as medium, we can get transition matrix from unknown to unknown. We believe that transition probability from a class to itself should be the largest among other transition directions, and we designate maximizing this probability as our optimization objective. In this way, the learning of known and unknown classes is optimized continuously. Thus we achieve ***bidirectional class-level knowledge transfer*** between known and unknown as shown in Figure.3.

Specifically, as shown in Figure.4 (a), given a batch of data, let $\boldsymbol{L}_h^l = [\boldsymbol{l}_h^1, ..., \boldsymbol{l}_h^{B^l}] \in \mathcal{R}^{C^l \times B^l}$ be the labeled logits of B^l labeled samples, $\boldsymbol{L}_g^l \in \mathcal{R}^{C^u \times B^l}$ be the unlabeled logits

of B^l labeled samples, $\boldsymbol{L}_h^u \in \mathcal{R}^{C^l \times B^u}$ and $\boldsymbol{L}_g^u \in \mathcal{R}^{C^u \times B^u}$ be the labeled and unlabeled logits of B^u unlabeled samples respectively. Then we can obtain transition matrix from known to unknown classes by using $\boldsymbol{L}_h^l$ as weights to compute the weighted sum of $\boldsymbol{L}_g^l$:

$$\boldsymbol{Trans}_{l\to u} = \boldsymbol{L}_h^l \cdot {\boldsymbol{L}_g^l}^\top \in \mathcal{R}^{C^l \times C^u}, \tag{2}$$

where the rows of $\boldsymbol{Trans}_{l\to u}$ indicate the relationships between a known class and each unknown class. Subsequently, we can calculate transition matrix from unknown to known classes in the same manner:

$$\boldsymbol{Trans}_{u\to l} = \boldsymbol{L}_g^u \cdot {\boldsymbol{L}_h^u}^\top \in \mathcal{R}^{C^u \times C^l}, \tag{3}$$

Then, we perform "cyclic transition" by multiplying these two matrices. Taking unknown classes as medium, we can obtain transition matrix from known to known by multiplying the transition matrix from known to unknown with transition matrix from unknown to known:

$$\boldsymbol{Trans}_{l\to l} = Norm(\boldsymbol{Trans}_{l\to u} \cdot \boldsymbol{Trans}_{u\to l}) \tag{4}$$

where $Norm(\cdot)$ denotes the normalization performed on the rows. Taking known classes as medium, we can obtain transition matrix from unknown to unknown as:

$$\boldsymbol{Trans}_{u\to u} = Norm(\boldsymbol{Trans}_{u\to l} \cdot \boldsymbol{Trans}_{l\to u}), \tag{5}$$

Transition probability from a class to itself should be the largest among other transition directions. That is, the element on the diagonal of the corresponding transition matrix should be the largest. Inspired by this, we maximize the sum of the elements on the diagonal of the transition matrices after cyclic transition:

$$\mathcal{L}_{class} = max[Tr(\boldsymbol{Trans}_{l\to l}) + Tr(\boldsymbol{Trans}_{u\to u})] \tag{6}$$

Instance-level Relationship Modeling and Knowledge Transfer

Although class relationship modeling is reasonable and sensible, it relies on the average of all samples, making it difficult to abstract more valid information at the instance-level. Instance-level relationship modeling can help the model to understand the structures within the data, allowing for recognition at a finer-grained granularity.

We use logits-based pairwise similarity to measure the relationships between samples. Based on samples' origin, instance-level relationships can be classified into cross-set modeling that concentrate on the relationships between labeled and unlabeled samples, and within-set modeling that focus on the relationships among labeled samples or among unlabeled samples. Additionally, the information encoded in labeled logits and unlabeled logits differs for samples from different origins. For example, labeled samples' labeled logits primarily encode the prediction of the ground-truth classes, while unlabeled samples' labeled logits encode the relationships with each known class. Therefore, we decouple the logits-based pairwise similarity into unlabeled-logits-based pairwise similarity and labeled-logits-based pairwise similarity, which is also referred to as the decoupled pairwise similarity. We believe that decoupled pairwise similarity should be consistent. By using KL divergence to enforce consistency, the predictions for the samples are further refined. Moreover, the relationships encoded in certain logits plays a significant role in pairwise similarity calculations, which also facilitates class-level relationship modeling.

As mentioned above, let $\boldsymbol{L}_h^l$ and $\boldsymbol{L}_h^u$ be the labeled logits of labeled samples and unlabeled samples, $\boldsymbol{L}_g^l$ and $\boldsymbol{L}_g^u$ be the unlabeled logits of labeled and unlabeled samples, respectively. Then for labeled samples, decoupled intra-set pairwise similarities can be computed as:

$$\boldsymbol{S}_h^l = {\boldsymbol{L}_h^l}^\top \cdot \boldsymbol{L}_h^l, \quad \boldsymbol{S}_g^l = {\boldsymbol{L}_g^l}^\top \cdot \boldsymbol{L}_g^l, \tag{7}$$

where the rows of $\boldsymbol{S}_h^l$ and $\boldsymbol{S}_g^l$ denote the similarity between labeled samples. And $\boldsymbol{S}_h^l$ is more influenced by the logits on ground-truth classes, while $\boldsymbol{S}_g^l$ depends more on relationships between labeled samples and each unknown class. We use KL-Divergence to constrain their consistency:

$$\mathcal{L}_l = KL(\boldsymbol{S}_h^l || \boldsymbol{S}_g^l) + KL(\boldsymbol{S}_g^l || \boldsymbol{S}_h^l), \tag{8}$$

Similarly, for unlabeled samples, decoupled intra-set pairwise similarities and consistency constraint are as follows:

$$\boldsymbol{S}_h^u = {\boldsymbol{L}_h^u}^\top \cdot \boldsymbol{L}_h^u, \quad \boldsymbol{S}_g^u = {\boldsymbol{L}_g^u}^\top \cdot \boldsymbol{L}_g^u, \tag{9}$$

$$\mathcal{L}_u = KL(\boldsymbol{S}_h^u || \boldsymbol{S}_g^u) + KL(\boldsymbol{S}_g^u || \boldsymbol{S}_h^u), \tag{10}$$

Therefore, the intra-set pairwise similarity consistency constraint can be defined as:

$$\mathcal{L}_{intra-set} = \frac{1}{2}(\mathcal{L}_l + \mathcal{L}_u). \tag{11}$$

For cross-set instance relationships, we similarly compute decoupled pairwise similarity, and use KL divergence to enforce their consistency:

$$\boldsymbol{S}_h^c = {\boldsymbol{L}_h^l}^\top \cdot \boldsymbol{L}_h^u, \quad \boldsymbol{S}_g^c = {\boldsymbol{L}_g^l}^\top \cdot \boldsymbol{L}_g^u, \tag{12}$$

$$\mathcal{L}_{cross-set} = KL(\boldsymbol{S}_h^c || \boldsymbol{S}_g^c) + KL(\boldsymbol{S}_g^c || \boldsymbol{S}_h^c). \tag{13}$$

In conclusion, instance-level relationship consistency can be expressed as:

$$\mathcal{L}_{instance} = \mathcal{L}_{cross-set} + \mathcal{L}_{intra-set} \tag{14}$$

Overall Objective

In summary, to maximize the sum of the elements on the diagonal of class relationship transition matrices and the instance-level pairwise similarity consistency, while minimizing cross-entropy, the overall objective function is:

$$\mathcal{L} = \mathcal{L}_{ce} - \alpha\mathcal{L}_{class} + \beta\mathcal{L}_{instance}, \tag{15}$$

where α, β are two hyperparameters.

Experiments

Experimental Setup

Datasets. We conduct experiments on *CIFAR100* dataset and three fine-grained datasets: *Stanford Cars*, *CUB*, and *FGVC-AirCraft*. Following CrKD(Gu et al. 2023), we do not conduct experiments on the widely-used *CIFAR10* and *ImageNet-882* benchmarks in previous work, as the results

Method	CIFAR100-50	CIFAR100-80	Stanford Cars	CUB	Aircraft
Kmeans	28.3±0.7	56.3±1.7	13.1±1.0	42.2±0.5	18.5±0.3
DTC(Han, Vedaldi, and Zisserman 2019)	35.9±1.0	67.3±1.2	-	-	-
RS+(Han et al. 2020)	44.1±3.7	75.2±4.2	36.5±0.6	55.3±0.8	38.4±0.6
NCL(Zhong et al. 2021a)	52.7±1.2	86.6±0.4	43.5±1.2	48.1±0.9	43.0±0.5
UNO(Fini et al. 2021)	62.3±1.4	90.5±0.7	49.8±1.4	59.2±0.4	52.1±0.7
ComEx(Yang et al. 2022)	53.4±0.7	85.7±1.3	-	-	-
IIC(Li et al. 2023a)	65.8±0.9	92.4±0.2	**53.8±0.9***	68.4±1.4*	55.2±2.8*
CrKD(Gu et al. 2023)	65.3±0.6	91.2±0.1	53.5±0.8	65.7±0.6	55.8±0.9
SNCD(Wang et al. 2024)	62.2±0.0†	**93.7±0.0**†	-	-	-
Ours	**68.2±1.8**	92.4±1.0	53.5±2.0	**69.1±2.7**	**56.1±2.1**

Table 1: Comparison with the SOTA methods on the unlabeled training set under task-aware protocol. Average clustering accuracy(%) is reported as *mean* ± *standard deviation* (averaged over 5 runs). Results marked with '*' are the ones we reproduced. In the results marked with †, the standard deviation is set to 0 as the authors did not report it in paper.

Method	CIFAR100-50			Stanford Cars			CUB			FGVC-Aircraft		
	Lab	Unlab	All	Lab	Unlab	All	Lab	Unlab	All	Lab	Unlab	All
RS+(Han et al. 2020)	69.7	40.9	55.3	81.8	31.7	56.8	80.7	51.8	66.3	66.4	36.5	51.5
NCL(Zhong et al. 2021a)	72.4	25.7	49.1	83.5	24.4	54.0	79.8	13.1	46.5	62.8	26.5	44.7
UNO(Fini et al. 2021)	75.0	57.6	66.3	81.7	46.7	64.2	78.7	62.1	70.4	71.2	52.4	61.8
IIC(Li et al. 2023a)	75.1	61.0	68.0	80.1*	**51.5***	65.8	75.5*	67.7*	71.6	71.3*	53.6*	62.5
CrKD(Gu et al. 2023)	**78.6**	59.4	69.0	83.9	51.3	67.6	**81.1**	67.5	**74.3**	**72.2**	55.2	**63.7**
SNCD(Wang et al. 2024)	77.2	60.5	68.9	-	-	-	-	-	-	-	-	-
Ours	76.3	**64.2**	**70.3**	**84.2**	51.4	**67.8**	80.1	**68.5**	**74.3**	71.3	**56.1**	**63.7**

Table 2: Comparison with the SOTA methods under task-agnostic protocol. Classification accuracy(%) and clustering accuracy(%) are reported on the labeled test set and unlabeled test set, respectively. Results marked with '*' are our reproductions.

Dataset	Labeled		Unlabeled	
	Images	Classes	Images	Classes
CIFAR100-20	$40.0k$	80	$10.0k$	20
CIFAR100-50	$25.0k$	50	$25.0k$	50
Stanford Cars	$\approx 4.0k$	98	$\approx 4.1k$	98
CUB	$\approx 3.0k$	100	$\approx 3.0k$	100
FGVC-Aircraft	$\approx 3.3k$	50	$\approx 3.3k$	50

Table 3: Details of dataset splits.

on these two datasets are nearly saturated. We divide each dataset into two parts: a labeled set and an unlabeled set. The number of classes and samples in each split are shown in Tab.3. Note that we assume that the number of unknown classes is known during the experiments.

Evaluation Metrics. We evaluate our method using both *task-aware* and *task-agnostic* protocols. In the task-aware protocol, we can identify if a sample is labeled or unlabeled, allowing labeled samples to be classified as known classes and unlabeled samples as unknown classes. In contrast, there is no prior about which subset a sample comes from in task-agnostic protocol. For labeled set, we use classification accuracy as metric. And for unlabeled set, we use average clustering accuracy, which is defined as:

$$ClusterAcc = \max_{perm \in P} \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{y_i = perm(\hat{y}_i)\} \quad (16)$$

where y_i and $\hat{y}_i$ represent ground-truth label and predicted label, respectively. P is the set of all permutations and the best one can be found by Hungarian algorithm(Kuhn 1955).

Implementation Details. For a fair comparison with the existing methods, we use the same experimental setup as previous methods. Due to space limitations, more details will be included in the supplementary materials.

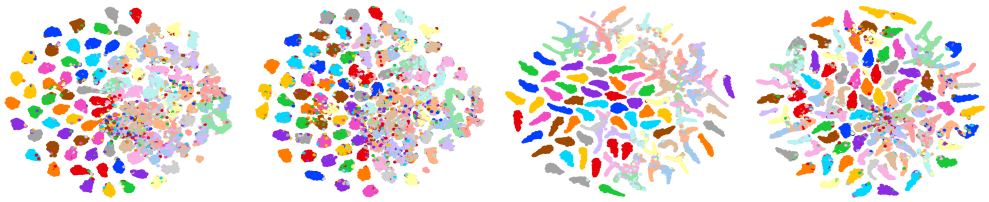
Experimental Results

Comparison with State of the Arts. We compare our method with current state-of-the-art methods, including DTC(Han, Vedaldi, and Zisserman 2019), RS+(Han et al. 2020), NCL(Zhong et al. 2021a), UNO(Fini et al. 2021), ComEx(Yang et al. 2022), IIC(Li et al. 2023a), CrKD(Gu et al. 2023) and SNCD(Wang et al. 2024). We report results on two wide-used benchmarks and three fine-grained datasets in Tab.1 and Tab.2.

In Tab.1, we report the average clustering accuracy on the unlabeled training set under task-aware protocol. It can be observed that our method performs better than other methods. In particular, our method outperforms UNO by 5.9%, IIC by 2.4%, CrKD by 2.9% on CIFAR100-50 benchmark.

$\mathcal{L}_{class}$	$\mathcal{L}_{instance}$	CIFAR100-50			Stanford Cars			CUB			FGVC-Aircraft		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
✗	✗	62.3	-	-	49.8	-	-	59.2	-	-	52.1	-	-
✓	✗	67.0	**0.71**	**0.53**	51.3	0.70	0.37	67.0	**0.81**	0.55	53.0	0.70	0.41
✗	✓	63.1	0.68	0.50	52.0	0.70	0.38	62.0	0.77	0.47	53.6	0.70	0.42
✓	✓	**68.2**	**0.71**	**0.53**	**53.5**	**0.71**	**0.39**	**69.1**	**0.81**	**0.57**	**56.1**	**0.72**	**0.44**

Table 4: Ablation study on CIFAR100-50 and three fine-grained datasets. Results are reported on the unlabeled training set.



(a) UNO (b) IIC (c) CrKD (d) Ours

Figure 5: T-SNE visualization on CIFAR100-50 benchmark.

And on the three fine-grained datasets, our method outperforms UNO by 3.7%, 9.9% and 4.0%, respectively. On the Stanford Cars and Aircraft datasets, our approach does not differ much from CrKD and IIC, which may be due to the generally high semantic similarity of the classes.

In Tab.2, we report results on the labeled unlabeled test sets under task-agnostic protocol. Our method shows a significant performance improvement over UNO, averaging 1.68% on the labeled split and 4.73% on the unlabeled split. Notably, it outperforms IIC by an average of 2.83% on the labeled split and 1.03% on the unlabeled split. In addition, our method outperforms CrKD on the unlabeled split by 1.2%, but it is 0.63% lower on the labeled split, which may be due to CrKD's preservation of supervised information from the teacher model.

Ablation Study. To evaluate the effectiveness of class-level and instance-level relationship modeling, we conduct ablation study on CIFAR100-50 benchmark and three fine-grained datasets. Following IIC, we introduce normalized mutual information(NMI) and adjusted rand index(ARI) as metrics, which measure the similarity between clustering results and the ground-truth distributions. The results are shown in Tab.4. Both class-level and instance-level relationship modeling improve clustering performance, with the best results achieved by combining them.

Results on ImageNet and Herbarium-19. To evaluate our method on more challenging datasets, we conduct experiments on a larger dataset, ImageNet, and a more unbalanced dataset, Herbarium-19k. As shown in Tab.5, our methods outperforms the SOTA methods.

T-SNE Visualization. To illustrate the differences between our method and others, we visualize the predicted logits using t-SNE on the CIFAR100-50 benchmark in Figure.5, comparing our method with UNO, IIC and CrKD. Known classes are shown in "bright" style and unknown classes in "pastel" style, appearing lighter in color. Compared to

Method	ImageNet-1k		ImageNet-100		Herbarium-19	
	Lab	Unlab	Lab	Unlab	Lab	Unlab
UNO	12.8	14.9	65.0	45.1	30.9	33.1
IIC	13.1	15.5	67.2	45.6	30.0	32.7
CrKD	**15.5**	15.0	68.4	46.3	31.0	32.1
Ours	13.5	**15.8**	**69.5**	**47.7**	**32.1**	**33.8**

Table 5: Results on more challenging datasets. The number of labeled and unlabeled classes are: ImageNet-1k(500/500), ImageNet-100(50/50), Herbarium-19(341/342).

UNO, IIC separate the labeled set from the unlabeled set, which makes a clear distinction between known and unknown classes. CrKD maintains the distribution of known classes but shows significant confusion among unknown classes, indicating bias towards known classes. In contrast, our method achieves a more balanced class distribution, with known and unknown classes intermingling, which demonstrates the superiority of class-level modeling.

Conclusion

In this paper, we propose a novel class-level relationship modeling and knowledge transfer method. Based on logits, we model symmetric and directional relationships for known and unknown classes, represented as transition matrix from known to unknown and transition matrix from unknown to known. Then, by "cyclic transition" conducted on transition matrices, we achieve bidirectional knowledge transfer at class-level. Considering the subtle distinctions between samples, we also model the relationships between samples through pairwise similarity. Thus, we achieve knowledge at both class-level and instance-level, which is more balanced and fine-grained. Extensive experiments demonstrate the effectiveness of our method.

Acknowledgments

This work was supported by the grants from the National Natural Science Foundation of China (62201014), Clinical Medicine Plus X - Young Scholars Project of Peking University PKU2024LCXQ028 , the Fundamental Research Funds for the Central Universities.

References

Asano, Y. M.; Rupprecht, C.; and Vedaldi, A. 2019. Self-labelling via simultaneous clustering and representation learning. *arXiv preprint arXiv:1911.05371*.

Caron, M.; Misra, I.; Mairal, J.; Goyal, P.; Bojanowski, P.; and Joulin, A. 2020. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33: 9912–9924.

Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; and Joulin, A. 2021. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9650–9660.

Cuturi, M. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Feng, J.; Yang, Y.; Xie, Y.; Li, Y.; Guo, Y.; Guo, Y.; He, Y.; Xiang, L.; and Ding, G. 2024. Debiased Novel Category Discovering and Localization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 1753–1760.

Fini, E.; Sangineto, E.; Lathuiliere, S.; Zhong, Z.; Nabi, M.; and Ricci, E. 2021. A unified objective for novel class discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9284–9292.

Gu, P.; Zhang, C.; Xu, R.; and He, X. 2023. Class-relation knowledge distillation for novel class discovery. *lamp*, 12(15.0): 17–5.

Han, K.; Rebuffi, S.-A.; Ehrhardt, S.; Vedaldi, A.; and Zisserman, A. 2020. Automatically discovering and learning new visual categories with ranking statistics. *arXiv preprint arXiv:2002.05714*.

Han, K.; Vedaldi, A.; and Zisserman, A. 2019. Learning to discover novel visual categories via deep transfer clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8401–8409.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hsu, Y.-C.; Lv, Z.; and Kira, Z. 2017. Learning to cluster in order to transfer across domains and tasks. *arXiv preprint arXiv:1711.10125*.

Hsu, Y.-C.; Lv, Z.; Schlosser, J.; Odom, P.; and Kira, Z. 2019. Multi-class classification without multi-class labels. *arXiv preprint arXiv:1901.00544*.

Huang, Z.; Yang, J.; and Gong, C. 2022. They are not completely useless: Towards recycling transferable unlabeled data for class-mismatched semi-supervised learning. *IEEE Transactions on Multimedia*, 25: 1844–1857.

Joseph, K.; Paul, S.; Aggarwal, G.; Biswas, S.; Rai, P.; Han, K.; and Balasubramanian, V. N. 2022. Novel class discovery without forgetting. In *European Conference on Computer Vision*, 570–586. Springer.

Khetan, A.; Lipton, Z. C.; and Anandkumar, A. 2017. Learning from noisy singly-labeled data. *arXiv preprint arXiv:1712.04577*.

Krause, J.; Stark, M.; Deng, J.; and Fei-Fei, L. 2013. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision workshops*, 554–561.

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

Kuhn, H. W. 1955. The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2): 83–97.

Li, M.; Zhou, T.; Huang, Z.; Yang, J.; Yang, J.; and Gong, C. 2024. Dynamic Weighted Adversarial Learning for Semi-Supervised Classification under Intersectional Class Mismatch. *ACM Transactions on Multimedia Computing, Communications and Applications*, 20(4): 1–24.

Li, W.; Fan, Z.; Huo, J.; and Gao, Y. 2023a. Modeling Inter-Class and Intra-Class Constraints in Novel Class Discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3449–3458.

Li, Z.; Otholt, J.; Dai, B.; Hu, D.; Meinel, C.; and Yang, H. 2023b. Supervised Knowledge May Hurt Novel Class Discovery Performance. *arXiv preprint arXiv:2306.03648*.

Li, Z.; Otholt, J.; Dai, B.; Meinel, C.; Yang, H.; et al. 2022. A closer look at novel class discovery from the labeled set. *arXiv preprint arXiv:2209.09120*.

Liu, M.; Roy, S.; Li, W.; Zhong, Z.; Sebe, N.; and Ricci, E. 2024. Democratizing fine-grained visual recognition with large language models. *arXiv preprint arXiv:2401.13837*.

Maji, S.; Rahtu, E.; Kannala, J.; Blaschko, M.; and Vedaldi, A. 2013. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*.

Roy, S.; Liu, M.; Zhong, Z.; Sebe, N.; and Ricci, E. 2022. Class-incremental novel class discovery. In *European Conference on Computer Vision*, 317–333. Springer.

Saito, K.; Kim, D.; and Saenko, K. 2021. Openmatch: Open-set semi-supervised learning with open-set consistency regularization. *Advances in Neural Information Processing Systems*, 34: 25956–25967.

Sun, Y.; Shi, Z.; Liang, Y.; and Li, Y. 2023. When and how does known class help discover unknown ones? provable understanding through spectral analysis. *arXiv preprint arXiv:2308.05017*.

Troisemaine, C.; Lemaire, V.; Gosselin, S.; Reiffers-Masson, A.; Flocon-Cholet, J.; and Vaton, S. 2023. Novel class discovery: an introduction and key concepts. *arXiv preprint arXiv:2302.12028*.

Wah, C.; Branson, S.; Welinder, P.; Perona, P.; and Belongie, S. 2011. The caltech-ucsd birds-200-2011 dataset. *California Institute of Technology*.

Wang, W.; Lei, T.; Chen, Q.; and Liu, Y. 2024. Semantic-Guided Novel Category Discovery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 5607–5614.

Weng, J.; Luo, Z.; Zhong, Z.; Lin, D.; and Li, S. 2023. Exploring non-target knowledge for improving ensemble universal adversarial attacks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 2768–2775.

Xie, J.; Girshick, R.; and Farhadi, A. 2016. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, 478–487. PMLR.

Yang, M.; Zhu, Y.; Yu, J.; Wu, A.; and Deng, C. 2022. Divide and conquer: Compositional experts for generalized novel class discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14268–14277.

Yao, Y.; Liu, T.; Han, B.; Gong, M.; Deng, J.; Niu, G.; and Sugiyama, M. 2020. Dual t: Reducing estimation error for transition matrix in label-noise learning. *Advances in neural information processing systems*, 33: 7260–7271.

Yu, X.; Liu, T.; Gong, M.; and Tao, D. 2018. Learning with biased complementary labels. In *Proceedings of the European conference on computer vision (ECCV)*, 68–83.

Zang, Z.; Shang, L.; Yang, S.; Wang, F.; Sun, B.; Xie, X.; and Li, S. Z. 2023. Boosting novel category discovery over domains with soft contrastive learning and all in one classifier. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 11858–11867.

Zhao, B.; Cui, Q.; Song, R.; Qiu, Y.; and Liang, J. 2022. Decoupled knowledge distillation. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 11953–11962.

Zhao, B.; and Han, K. 2021. Novel visual category discovery with dual ranking statistics and mutual knowledge distillation. *Advances in Neural Information Processing Systems*, 34: 22982–22994.

Zhong, Z.; Fini, E.; Roy, S.; Luo, Z.; Ricci, E.; and Sebe, N. 2021a. Neighborhood contrastive learning for novel class discovery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10867–10875.

Zhong, Z.; Zhu, L.; Luo, Z.; Li, S.; Yang, Y.; and Sebe, N. 2021b. Openmix: Reviving known knowledge for discovering novel visual categories in an open world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9462–9470.

Zhu, F.; Ma, S.; Cheng, Z.; Zhang, X.-Y.; Zhang, Z.; and Liu, C.-L. 2024. Open-world machine learning: A review and new outlooks. *arXiv preprint arXiv:2403.01759*.