

## 见多识广：为新类别发现提供更丰富的先验知识

彭志茂<sup>1</sup>, 王恩光<sup>1</sup>, 刘夏雷<sup>1</sup>(✉), 程明明<sup>1</sup>

© The Author(s)

**Abstract** 新类别发现旨在使用一个在带标签的数据集上训练的模型, 在无标签的数据集中发现新的类别, 该模型的标签类别与新类别不同, 但是在语义上是相关的。这一任务的挑战在于, 模型需要从已知类别中学习区分性的表示, 从而准确地对未见过的类别进行分组。现有的方法通常在仅包含有限已知类别上预训练模型, 导致所学习到的表示对未来可能遇到的各种未见过的类别的区分性较差。本文提出了一种新的更丰富先验知识 (Richer Prior Knowledge (PRK)) 模块, 通过让模型接触大量合成的视觉类别, 以此为未来的新类别学习到多样且区分性的表示。我们认为模型在预训练阶段接触到的类别越多, 那么所学习到的表示空间对基类别的偏向就会更小。为了证明我们的方法的有效性, 我们在各种数据集和设置上进行了广泛的实验, 验证了我们提出的方法的有效性。此外, 我们的方法可以轻松地集成到其他方法中, 并实现优越的性能。

**Keywords** 新类别发现, 开放世界, 深度迁移聚类, 图像分类。

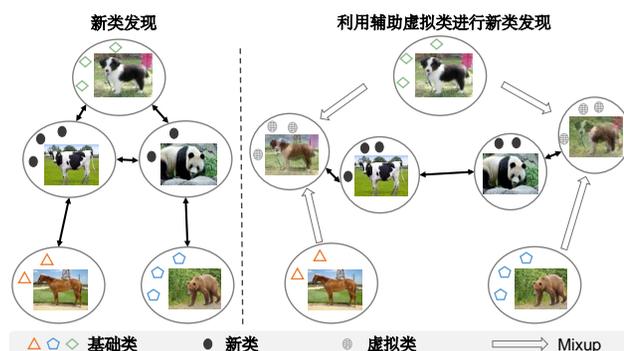
### 1 引言

深度学习在许多视觉识别任务中占据了主导地位 [1–3]。然而, 这些任务大多数遵循封闭集假设: 模型在带有指定类别的标签数据上进行训练, 测试数据仅来自这些已知类别。在这种设置下, 模型可以准确识别各种视觉类别, 但代价是需要大量的标签样本。相比之下, 在现实世界场景中, 模型将不可避免地遇到许多从未见过的新类别。为这些新类别进行标注不仅费力且昂贵, 而且提前标注世界上所有类别也是不切实际的。

为了使模型能够识别未见过的类别, 且无需在这些类别的标签数据上进行训练, 新类别发现 (NCD) [4, 5] 在最近被提出, 并引发了广泛的关注。NCD 的目标是借助

1 TKLNDST, College of Computer Science, Nankai University, Tianjin, China. E-mail: Z. Peng, zhi-mao796@mail.nankai.edu.cn; E. Wang, enguang-wang@mail.nankai.edu.cn; X. Liu, xialei@nankai.edu.cn. M.-M. Cheng, cmm@nankai.edu.cn

Manuscript received: 2023-06-26; accepted: 2024-06-10



**Fig. 1** 对比仅使用基础类别进行新类别发现 (NCD) 训练 (左图) 与使用基础类别和辅助虚拟类别进行 NCD 训练 (右图)。在左图中, 由于基础类别数量有限, 所学习的表示可能会强烈偏向这些类别 (例如, 黑白颜色的“边境牧羊犬”), (例如, 黑白颜色的“边境牧羊犬”), 使得在视觉上相关的新类别 (例如, “奶牛”和“熊猫”) 在特征表示空间中过于接近, 难以区分。我们的想法体现在右图, 通过使用 Mixup 生成的额外合成虚拟类别进行训练, 模型可以学习到更加多样化和可迁移的表示 (例如, “形状”), 这使得视觉上相关的类别在特征表示空间中保持更大的距离, 从而提高 NCD 聚类的准确性。

从已知类别中学习到的视觉先验知识来识别新类别。尽管已知类别和新类别的标签空间是不相交的, 但它们在某种程度上具有语义相似性。因此, NCD 任务假设在已知类别上学习到的特征表示也可以在新类别上具有区分能力。

为了将从已知类别学习到的视觉先验知识迁移到新类别用于 NCD, 当前大多数方法遵循两阶段的训练 [4–9]: 在第一阶段, 对带标签的样本进行监督训练, 以学习视觉特征表示; 在第二阶段, 通过自训练生成的伪标签, 对无标签样本进行无监督聚类。然而, 由于在监督训练中仅使用了类别标签作为监督信息, 所学到的视觉先验知识将偏向于这些已知类别。尽管在监督训练前引入额外的自监督预训练 [5] 可以缓解这一问题, 但模型从有限的视觉类别中学到的先验知识仍然难以应对未来可能遇到的大量视觉概念。

如图 1 (左) 所示, 当对基础类别进行监督训练时, 模型学习到的特征表示对这些类别具有区分性, 使得它们在特征空间中彼此远离。然而, 所学到的表示可能不适

用于描述新类别,使得它们在特征空间中的表示非常接近,然后在无监督聚类过程中生成过多的噪声伪标签。为了使所学的表示对从未见过的新类别具有区分性,本文提出了一种新颖的丰富先验知识 (Richer Prior Knowledge(RPK)) 模块,通过让模型接触更多的视觉概念,使所学的表示更加多样化和可迁移。如图 1 (右) 所示,我们合成了辅助虚拟类别,以模拟未来可能遇到的新类别的分布。通过这些虚拟类别的训练,所学的表示将对已知类别的偏向减小,并对新类别更加适应。如此一来,具有不同语义的新类别在特征空间中的表示可以更好地分开。

为了快速高效地合成大量具有新模式的虚拟类别样本,我们使用带有约束权重参数的 Mixup[10] 线性插值来自不同已知类别的两个样本。为了解决虚拟类别数量随着已知类别数量呈二次方增加的问题,我们提出了一种基于聚类的虚拟类别合成策略,以大幅降低训练成本,并使其在类别数量较多的数据集上可行。该策略可以生成语义超类别,通过聚类算法合并语义上相似的类别。因此,虚拟类别的合成可以通过在超类之间进行采样,从而避免语义相似的类别被过度合成,以此提高 NCD 性能。

本文的贡献总结如下:

- 我们提出了一种更丰富的先验知识 (Richer Prior Knowledge(RPK)) 模块,生成辅助虚拟类别以学习更丰富的先验知识,以便在 NCD 任务中,模型可以在无监督聚类中生成更少的噪声伪标签。
- 我们引入了一种聚类引导的虚拟类别合成策略,该策略可以显著减少合成的虚拟类别数量,并进一步提高 NCD 性能。
- 所提出的 RPK 模块是即插即用的,可以应用于主流的两阶段 NCD 方法。通过在四个公共数据集上进行广泛实验,我们验证了所提出的方法可以显著超越现有方法。

## 2 相关工作

### 2.1 新类别发现

新类别发现 (NCD) 的研究可以追溯到 [11, 12], 与传统的迁移学习任务 [13, 14] 不同,其尝试解决一种跨任务的迁移学习任务,即从已知类别的监督分类到未知类别的无监督聚类。通过引入预测的成对相似性,利用已知类别的先验知识来帮助对未知类别进行聚类。NCD 任务的正式定义由 Han 等人 [4] 提出,此外,他们还提出了一种两阶段的 NCD 方法:首先在有标签的数据集上进行监督预训练,提出改进的深度嵌入聚类方法,然后在无标签数据集上同时执行聚类和表示学习。随后, Han 等人 [5] 提出通过使用特征向量的稳健排名统计作为相似性度量来决定两个无标签样本是否属于同一类

别。值得注意的是,他们在预训练阶段引入了额外的自监督损失 [15],以减轻所学表示对已知类别的偏向。

[6] 提出了 OpenMix 方法,该方法在无监督聚类阶段将有标签样本与无标签样本混合,使用混合样本进行训练,减少模型对带有错误伪标签的无标签样本的过拟合。[7] 提出了一种邻域对比学习方法,通过引入局部邻域查询机制来选择伪正样本,提高对比学习在 NCD 任务中的效果。[16] 提出了一种端到端的对比学习框架,用于学习 NCD 的区分性表示空间。在 [9] 中, Zhao 等人改进了 [5] 的工作,提出了一个双分支框架,专注于全局和局部信息。[8] 提出了 UNO 框架,通过利用单一的交叉熵损失,在有标签和无标签样本上训练模型,使得模型能够识别已知和未知类别。在 [17] 中, Yang 等人提出利用组合专家的方法,以互补的方式处理有标签和无标签数据集,改进 NCD 的广义设定。在 [18] 中, Vaze 等人提出了一种新的 NCD 设定,称为“广义类别发现”,并使用预训练的视觉 Transformer[19] 进行对比表示学习来解决这一问题。随后, Guo 等人 [20] 提出了类似的设定。最近,类增量 [21, 22] 和分布无关 [23] 的 NCD 设定也被提出,以模拟现实世界部署中的挑战。与这些方法不同的是,我们专注于在监督预训练阶段让模型接触更多的视觉类别,以学习多样化和可迁移的特征,使模型在未来可能遇到的新类别上具有更强的区分能力。此外, NCD 任务也在领域自适应 [24] 和细粒度文本分类 [25] 领域进行了探索。

### 2.2 半监督学习

与 NCD 任务类似,半监督学习 (SSL) [26] 也使用有标签和无标签本来训练模型,但两者的目的不同。在 NCD 中,有标签数据集和无标签数据集的标签空间是不相交的。有标签数据集用于学习视觉先验知识,以帮助对无标签数据集进行聚类。在 SSL 中,无标签数据集的标签空间与有标签数据集相同,SSL 的目标是将无标签数据集作为辅助数据来提高模型的性能。在深度学习的背景下,有两种广受关注的半监督学习范式:一致性正则化 [27–31] 和伪标签化 [32–39]。一致性正则化假设模型的预测输出对于扰动输入不应发生变化。受深度去噪自编码器的启发, Rasmus 等人提出了 Ladder Network,通过计算编码器输出与去噪函数重建输出之间的均方误差 (MSE) 来构建一致性正则化损失 [27]。 $\Pi$  模型通过数据增强和 dropout 引入扰动,其训练目标是鼓励对同一输入样本的不同增强版本产生一致的预测输出 [28]。作为  $\Pi$  模型的改进版本, Temporal Ensembling [29] 通过指数移动平均 (EMA) 计算前几轮训练的预测输出,将其作为当前轮次的一致性损失目标向量,从而加快训练速度并减少噪声。与 Temporal Ensembling 在每一轮训练中计算预测输出的 EMA 不同, Mean Teacher [30] 在每个训练步骤中计算模型权重的 EMA,以产生更准确的预测输

出。UDA[31] 采用在监督学习中使用的更强大的数据增强方法 [40, 41] 来改进半监督学习性能，这些数据增强方法用于生成一致性损失的目标向量。伪标签化通常使用自学习方法从现有的有标签数据中生成无标签数据的伪标签，以获取更多的训练数据。EntMin[32] 和伪标签 [33] 使用分类置信度最高的类别作为无标签数据伪标签的真实类别。Noisy student[35] 首先在有标签数据上训练教师模型，然后为无标签数据生成伪标签，最后用有标签数据和伪标签数据训练学生模型。在迭代训练过程中，学生模型将成为新的教师模型。EnAET [36] 引入一些空间和非空间变换，然后使用原始样本和变换样本训练编码器，再使用解码器估计原始样本和变换样本的变换参数作为 AET 损失，最后将 AET 损失作为 SSL 的附加正则化项。SimCLRv2[37] 在无监督预训练、监督微调和无标签自训练三个步骤中表现出强大的半监督学习性能。除了上述方法，还有结合一致性正则化和伪标签化思想的混合方法，如 mixmatch[42] 和 fixmatch [43]。虽然伪标签化和自训练的想法很巧妙，但由于模型的校准较差，生成的伪标签噪声较大。为了解决这个问题，一些近期的研究 [44–46] 探索了置信度阈值方法，以选择可靠的伪标签样本用于半监督学习。

相比之下，NCD 通过聚类分配（通常是成对相似性 [5, 7] 或在线聚类 [8]）。为无标签数据生成伪标签。然而，由于已知类别标签的强监督，新类别的无标签数据特征不可避免地会偏向已知类别，导致生成的伪标签比半监督学习 (SSL) 中的伪标签更不可靠（在 SSL 中，有标签数据和无标签数据属于相同的类别空间）。通过合成虚拟类，我们让模型接触更多的视觉概念，使模型学习到的表示对新类别更具区分性，减少生成的噪声伪标签，提高 NCD 的性能。

### 2.3 Mixup 方法

Mixup [10] 引领了一系列空间数据增强方法，并在计算机视觉领域取得了令人满意的结果。在原始的 Mixup [10] 中，可以通过对一个小批量中的两张图像进行线性插值得到增强后的图像。Mixup 有各种变体，可以分为全局混合 [10, 47] 和局部混合 [48, 49]。最近，Mixup 已被用于解决一些开放世界视觉问题。在 [50] 中，Mixup 被用于减少长尾识别中的过度自信预测。在 [51] 中，Zhu 等人提出了一种类增产技术，通过使用带有约束权重因子的 Mixup 来混合两张图像。通过训练混合样本，模型能够有效缓解类别增量学习任务中旧类别的灾难性遗忘。在 [6] 中，OpenMix 混合了有标签和无标签的图像，以提高新类别发现的性能。在 [7] 中，manifold Mixup 被用于通过在特征空间中混合有标签和无标签的样本来合成对比学习中的困难负样本。尽管我们的方法与这些方法密切相关，但我们的目标是学习更多样化且具有区

分性的表示，以便在 NCD 的无监督聚类阶段生成更少噪声的伪标签。

## 3 更丰富的先验知识 (RPK) 用于 NCD

在本节中，我们首先介绍新类别发现 (NCD) 的定义，并回顾 NCD 中最常用的两阶段框架，指出其局限性 (3.1节)。接下来，我们详细说明如何通过合成额外的虚拟类别来构建更丰富的先验知识 (RPK) 模块，以学习多样化且可迁移的表示用于 NCD (3.2节)。最后，我们展示如何压缩虚拟类别的数量，使 RPK 模块在训练中更高效，并提高评估性能 (3.3节)。

### 3.1 预备知识

#### 3.1.1 问题定义

给定用于新类发现 (NCD) 的训练集，其包含两个部分：一个有标签数据集  $D^l$  和一个无标签数据集  $D^u$ 。我们将  $D^l$  中的所有类别定义为基类  $Y^l$ ，将  $D^u$  中的所有类别定义为  $Y^u$ 。具体来说， $D^l$  和  $D^u$  在类别空间是不相交的，即  $Y^l \cap Y^u = \emptyset$ 。 $Y^l$  和  $Y^u$  中类的数量定义分别为  $C^l$  和  $C^u$ 。NCD 的目的是通过在  $D^l$  上预训练模型，将  $D^u$  归类  $C^u$  个簇。最近，提出了一个更具挑战性且更实用的广义 NCD 问题 [8]，在该问题中，训练好的 NCD 模型需要在额外的测试集  $D^t$  上进行评估，该测试集中的样本来自基类  $Y^l$  和新类  $Y^u$ 。

#### 3.1.2 用于 NCD 的两阶段框架

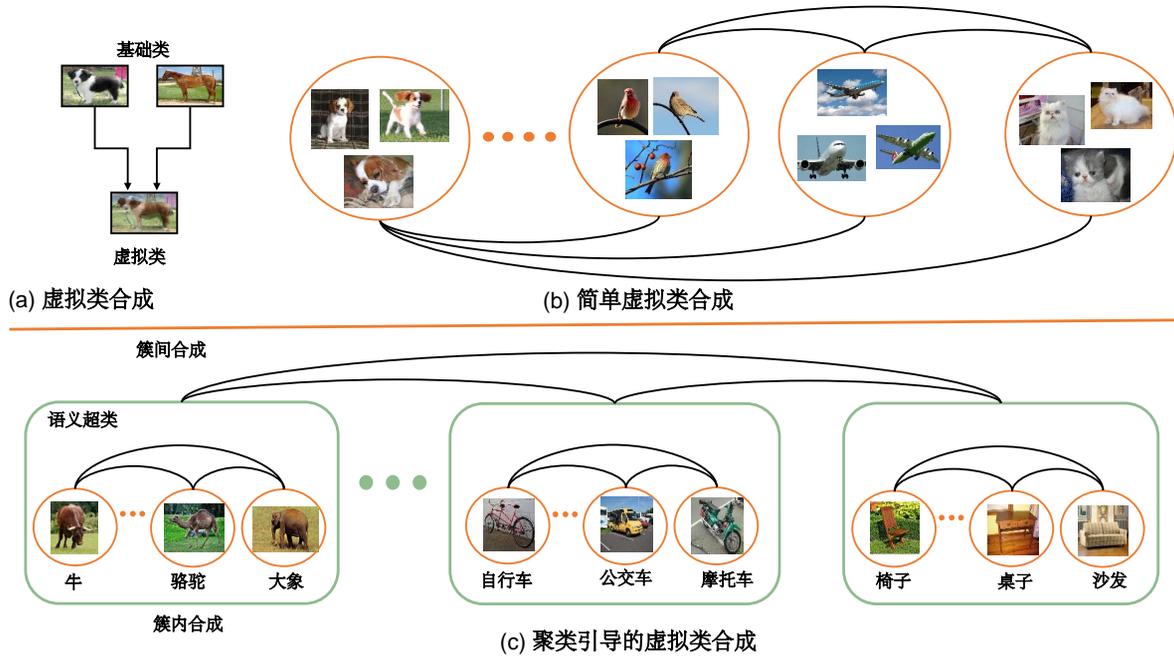
为了从基类中学习视觉先验知识帮助识别新类目标，当前大多数 NCD 方法都采用了两阶段框架 [4–9]。在第一阶段，采用预训练在有标签数据集  $D^l$  和无标签数据集  $D^u$ （如果有必要）上学习判别性特征表示。第二阶段，使用无监督聚类为无标签数据集  $D^u$  进行聚类分配。

对于 NCD 任务，使用随机初始化的模型对无标签数据进行聚类是很困难的，因此，模型需要从基础类别中获得足够的视觉先验知识。一个常见的策略是执行监督预训练，模型  $\Phi$  在有标签数据集  $D^l$  上通过标准的交叉熵损失函数进行学习：

$$L_{ce} = -\frac{1}{N} \sum_{j=1}^N \log \left( \frac{e^{s_j}}{\sum_{i=1}^{C^l} e^{s_i}} \right) \quad (1)$$

其中， $N$  是有标签数据集  $D^l$  中图像的数量，对于第  $j$  个图像  $x_j$ ， $s_j = \Phi(x_j)$  表示该图像的真实类别的分类分数。由于  $D^l$  只包含基础类别，模型  $\Phi$  学到的表示不可避免地会偏向这些类别。

为了解决这个问题，一些研究 [7, 9, 52] 在监督训练前，用有标签数据集  $D^l$  和无标签数据集  $D^u$  进行自监督学习 [15]，以便模型能够学习一些额外的与标签无关的内在特征表示。尽管在预训练阶段引入额外的自监督学习步骤已经被证明在 NCD 任务中是有效的，但预训练阶段的有标签数据集仅包含有限数量的视觉类别，缺



**Fig. 2** 针对 NCD 提出的虚拟类合成的示意图。(a) 通过对来自不同基础类的两张图像进行线性插值, 并约束权重参数, 合成辅助虚拟类。(b) 基础虚拟类合成会以二次增长的方式增加虚拟类的数量。(c) 基于聚类引导的虚拟类合成可以有效减少虚拟类的数量并提升 NCD 的性能。通过  $k$ -means, 形成语义相似的超类, 从而在超类之间进行我们提出的类间合成 (inter-cluster synthesis) 以及在每个超类内进行类内合成 (intra-cluster synthesis)。

乏先验知识使得模型学习到的表示对未来可能出现的未知类别的辨别能力较弱。

在无监督聚类阶段, 大多数 NCD 方法 [5–9, 16] 基于伪标签, 通过自训练策略训练模型。

在 NCD 任务中, 由于有标签数据集  $D^l$  和无标签数据集  $D^u$  属于不同的标签空间, 因此在第一阶段预训练模型后, 自训练过程进行了如下修改: 首先, 模型通过伪标签分配方法生成  $D^u$  的伪标签。然后, 模型在  $D^u$  上使用伪标签以及在  $D^l$  上使用真实标签进行联合训练, 以对  $D^u$  进行聚类。前者用于学习未知类别的特征表示, 后者则防止模型遗忘在有标签数据集中学到的先验知识。

通过在  $D^l$  和  $D^u$  上进行自训练, 期望模型能为新类别学习到区分度高的表示。然而, 由于已知类别标签的强监督, 模型学习的表示可能会严重偏向已知类别, 这样模型为  $D^u$  分配的伪标签可能会存在很多噪音。由于模型在含有噪音的伪标签的  $D^u$  上训练, 错误预测的置信度增加, 导致在随后的自训练中倾向于学习错误的信息, 产生较差的聚类结果。

### 3.2 利用辅助虚拟类进行学习

为了减少伪标签的噪音, 学习较少偏向的特征表示至关重要。一个自然的想法是学习更丰富的视觉先验知识, 以使所学到的表示更加多样化和可迁移。为此, 我们提出在预训练阶段学习更多的视觉类别。为了简单且高效

地获得大量额外视觉类别的样本, 合成辅助虚拟类别是一个有效的解决方案。受 Mixup [10] 和 IL2A [51] 的启发, 我们可以通过线性插值两类不同类别的样本来合成额外的虚拟类别样本。具体来说, 在预训练阶段, 给定有标签数据集  $D^l$ , 我们可以通过组合来自  $D^l$  的两个训练图像  $x_p$  和  $x_q$  来生成一个新的辅助虚拟类别图像  $x_{new}$ :

$$x_{new} = \lambda x_p + (1 - \lambda)x_q \quad (2)$$

其中,  $p$  和  $q$  表示不同的类别。  $x_p$  和  $x_q$  分别从类  $p$  和类  $q$  随机采样。为了使生成的虚拟类别位于类别  $p$  和类别  $q$  之间的低密度区域,  $\lambda$  的值在 [0.4, 0.6] 之间采样 [51], 这减少了虚拟类别与原始基础类别之间的重叠。如果预训练阶段的有标签数据集中包含  $C$  个类别, 那么生成的虚拟类别的数量将为  $(C \times (C - 1)/2)$ 。在预训练阶段, 模型  $\Phi$  在合成的虚拟类别和  $D^l$  上联合训练, 使用标准的交叉熵损失进行训练, 虚拟类别的分类头将在训练结束时被移除。尽管基于改进 Mixup 的样本成对插值可以生成大量的虚拟类别, 但虚拟类别和基础类别的数量呈二次关系。这将导致在高维分类器中消耗过多的计算资源。在下一节中, 我们提出了聚类引导的虚拟类别合成策略来解决这个问题。

### 3.3 聚类引导的虚拟类合成

为了有效的减少合成的虚拟类的数量，消除语义相似的虚拟类是一种直观的方法。一般来说，如果两个虚拟类在语义上相似，那么用于合成这些虚拟类的基础类也应该是相似的。因此，如果我们能在  $D^l$  中找到语义相似的基础类，并在合成虚拟图像时将它们合并，就可以显著减少合成的虚拟类的数量。

为此，我们提出了一种基于聚类引导的虚拟类合成策略。具体来说，我们首先通过优化特征表示与分类器权重之间的余弦相似性，并使用标准交叉熵损失，在  $D^l$  上训练模型  $\Phi$ 。结果是得到经过  $l_2$  归一化的分类器权重，这些权重可以被解释为特征空间中类簇的代表性特征向量（即中心）。为了快速对基础类进行聚类，我们使用  $k$ -means [23, 53–57] 对  $l_2$  归一化的分类器进行聚类，将其聚为  $K$  个簇，每个簇可以看作一个语义上的超类。

对于这  $K$  个基础类的语义超类，在基于 Mixup 的虚拟类合成过程中，属于同一超类的所有基础类将被赋予相同的标签。因此，生成的虚拟类的数量将减少到  $(K \times (K - 1)/2)$ ，其中  $K \ll C$ 。使得虚拟类的数量大大减少。我们将这种虚拟类合成称为簇间合成。

由于插值图像属于不同的超类，簇间压缩仅生成与已知类在语义上非常不同的虚拟类。因此，我们还尝试通过混合超类内的基础类来合成虚拟类。假设第  $i$  个超类中基础类的数量为  $Y_i$ ，则生成的细粒度虚拟类的数量将为  $\sum_i^K \mathcal{Y}_i \times (\mathcal{Y}_i - 1)/2$ 。我们将这种虚拟类合成称为簇内合成。我们将在实验部分讨论它们的性能。

## 4 实验

### 4.1 实验设置

#### 4.1.1 数据集

我们在实验中使用了四个基准数据集：CIFAR-10、CIFAR-100、Tiny-ImageNet 和 ImageNet。训练数据集被划分为有标签数据集和无标签数据集。无标签数据集中新类别的数量是预先已知的。这些数据集的划分详情见表 2。值得注意的是，以往文献中常用的数据集要么太小 (CIFAR-10、CIFAR-100)，要么太大 (ImageNet)。前者存在视觉概念有限的问题，而后者则存在基础类与新类别之间严重不平衡的问题。为了涵盖更广泛的场景，我们进一步引入了 Tiny-ImageNet 数据集，作为新的中等规模的 NCD 基准数据集，并在实验中使用了 200-100 的划分方式。该数据集比 CIFAR 数据集包含更丰富的视觉概念，相较于 ImageNet 数据集，其基础类别和新类别更加平衡。

#### 4.1.2 评估设置和指标

根据 [8]，我们主要在任务感知 (task-aware) 和任务无关 (task-agnostic) 设置中评估我们提出的 RPK 模块。在任务感知设置中，测试样本属于基础类或新类是已知

的。在执行新类发现时，基础类的测试样本将仅被分配到基础类，并且性能仅在在有标签数据集的测试部分报告 (表 1)。同时，新类的测试样本将仅被分配到新类，并且性能分别在无标签数据集的训练部分 (表 7) 和测试部分 (表 1) 报告。在任务无关设置中，当执行新类发现时，所有测试样本将被分配到基础类和新类。性能将在有标签和无标签数据集的测试部分上显示 (表 1 和表 9)。对于任务感知和任务无关设置，我们使用分类准确率指标来评估有标签数据集的表现，并使用聚类准确率指标来评估无标签数据集的表现。聚类准确率定义如下：

$$ClusterACC = \max_{p \in P(Y_u)} \frac{1}{M} \sum_{i=1}^M \mathbf{1}\{y_i = p(\hat{y}_i)\} \quad (3)$$

其中， $M$  是无标签数据集  $D^u$  中测试样本的数量。 $y_i$  和  $\hat{y}_i$  分别表示测试样本  $x_i \in D^u$  的真实标签和聚类分配标签。 $P(Y_u)$  是  $Y_u$  上所有排列的集合。用匈牙利算法 [58] 获得最优排列。

#### 4.1.3 实现细节

ResNet18 作为我们所有实验的骨干网络。在 UNOv2 基准 [8] 中，网络首先在有标签数据集上以监督方式训练 200 个 epoch，然后在有标签数据集和无标签数据集上进行 500 个 epoch 的无监督聚类训练。按照 UNOv2 的细节，改进的增强方法和多裁剪方法也被采用。在 CIFAR 和 Tiny-ImageNet 数据集上的所有实验中，批量大小设置为 512。网络使用随机梯度下降 (SGD) 进行优化，动量为 0.9。学习率首先在 10 个 epoch 内线性预热至 0.1，然后使用余弦退火算法衰减至 0.001。对于带有余弦相似度的 softmax 交叉熵损失函数，温度  $\tau$  设置为 0.1。对于聚类引导的虚拟类别合成，语义超类的数量  $K$  在 CIFAR100、Tiny-ImageNet 和 ImageNet 数据集上分别设置为 20、40 和 100。在上述所有数据集上，我们在每次迭代中生成了  $4 \times$  批量大小的辅助虚拟类别样本，并将这些样本与原始训练数据一起训练。

### 4.2 消融实验

在本节中，我们在 CIFAR-100 和 Tiny-ImageNet 数据集上进行了消融实验，以验证我们提出的 RPK 模块中各个部分的有效性：不同的虚拟类合成变体、语义超类的数量、不同的采样间隔和不同的空间数据增强方法。我们采用了 UNOv2[8] 作为基线方法，它是 UNO [8] 的改进版本，并且在训练时不使用辅助的虚拟类。

#### 4.2.1 不同的虚拟类合成变体

我们首先验证了以下几种不同虚拟类合成变体的有效性：

- 简单的虚拟类合成：通过从基础类中随机采样两个不同类别的样本来合成虚拟类样本。

**Table 1** 消融实验。使用不同的虚拟类合成策略评估所提方法的准确率。“All”代表简单的虚拟类合成，“Intra”代表类内虚拟类合成，“Inter”代表类间虚拟类合成。“Dum”代表合成的虚拟类数量。所有结果均在测试集上进行测量。

| 方法           | CIFAR100-50  |              |              |      |             |             |             |             |             |             | Tiny-ImageNet-100 |             |             |             |             |             |             |
|--------------|--------------|--------------|--------------|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------------|-------------|-------------|-------------|-------------|-------------|-------------|
|              | All          | Intra        | Inter        | Dum  | 任务感知        |             |             | 任务无关        |             |             | Dum               | 任务感知        |             |             | 任务无关        |             |             |
|              |              |              |              |      | 基类          | 新类          | 所有类         | 基类          | 新类          | 所有类         |                   | 基类          | 新类          | 所有类         | 基类          | 新类          | 所有类         |
| UNOv2        | $\times$     | $\times$     | $\times$     | 0    | <b>82.3</b> | 60.2        | 71.3        | 75.5        | 57.8        | 66.8        | 0                 | 66.7        | 31.4        | 49.1        | 55.1        | 29.9        | 42.5        |
| + RPK (Ours) | $\checkmark$ | $\times$     | $\times$     | 1225 | 82.2        | 62.9        | 72.6        | 75.2        | 60.5        | 67.9        | 4950              | 68.8        | 31.9        | 50.4        | 56.8        | 30.6        | 43.7        |
|              | $\times$     | $\checkmark$ | $\checkmark$ | 235  | 82.1        | 63.6        | 72.9        | <b>75.5</b> | 60.7        | 68.1        | 876               | 68.5        | 31.7        | 50.1        | 56.9        | 30.4        | 43.7        |
|              | $\times$     | $\checkmark$ | $\times$     | 45   | 81.7        | 62.3        | 72.0        | 74.9        | 59.6        | 67.3        | 96                | 66.9        | 31.6        | 49.3        | 55.2        | 29.8        | 42.5        |
|              | $\times$     | $\times$     | $\checkmark$ | 190  | 82.1        | <b>64.0</b> | <b>73.1</b> | <b>75.5</b> | <b>61.2</b> | <b>68.4</b> | 780               | <b>68.9</b> | <b>32.9</b> | <b>50.9</b> | <b>57.0</b> | <b>31.5</b> | <b>44.3</b> |

- 簇内虚拟类合成：利用基础类的语义超类簇，从同一语义超类簇中的不同类别中随机采样两个样本来合成虚拟类。
- 簇间虚拟类合成：在虚拟类合成过程中，将同一语义超类簇中的所有样本分配相同的标签，虚拟类通过从不同的语义超类簇中随机采样两个样本合成。
- 簇内和簇间虚拟类合成的组合。

实验在 CIFAR100-50 和 Tiny-ImageNet-100 数据集的有标签和无标签数据的测试集上进行，采用了任务感知和任务无关的评估设置。比较结果如表 1 所示。根据这些实验结果，所有虚拟类合成变体的讨论如下：

(1) 与 UNOv2 相比，将简单的虚拟类合成应用于 UNOv2 可以提高 CIFAR100-50 和 Tiny-ImageNet-100 上的性能，这表明通过使用辅助虚拟类别进行学习，所提出的方法从更丰富的先验视觉知识中取得性能提升。然而，简单虚拟类合成的结果显著低于簇间虚拟类合成的结果，这表明它并不是学习多样化和可迁移特征表示的最佳策略。我们推测原因是训练过程中合成的虚拟类过于冗余（例如，Tiny-ImageNet-100 有 4950 个类），这可能会使模型学习产生混淆，导致对新类的区分度降低。此外，这种方法在处理大量有标签数据集时计算成本高且不可行。

(2) 与 (1) 类似，将簇内虚拟类合成应用于 UNOv2 可以提升 CIFAR100-50 的性能，而在 Tiny-ImageNet-100 上则表现一般。这是因为 CIFAR100 包含有限的视觉概念，而通过合成的簇内虚拟类进行训练，可以使模型为与基础类语义相似的新类学习到更具区分性的特征表示。然而，Tiny-ImageNet 包含更多类别且视觉概念更广泛，因此在使用少量合成的簇内虚拟类（例如 Tiny-ImageNet-100 中的 96 个）时，未观察到性能提升是可以理解的。

(3) 将簇间虚拟类合成应用于 UNOv2 几乎在所有评估指标上都取得了最佳性能，这表明学习多样化和可迁移特征的有效性。相比于简单的虚拟类合成，它还显著减少了合成的虚拟类数量，从而提高了训练效率。通过将同一语义超类簇中的样本分配相同的标签，将语义相似的基础类整合为一类，这使得合成的虚拟类之间的差异更大，且减少对模型起负面影响的易于混淆的类别。

**Table 2** 我们实验中使用的每个数据集的有标签集和无标签集的信息。

| 数据集                  | 有标签集  |     | 无标签集 |     |
|----------------------|-------|-----|------|-----|
|                      | 图像数量  | 类数量 | 图像数量 | 类数量 |
| CIFAR10              | 25K   | 5   | 25K  | 5   |
| CIFAR100-20          | 40K   | 80  | 10K  | 20  |
| CIFAR100-50          | 25K   | 50  | 25K  | 50  |
| Tiny-ImageNet200-100 | 50K   | 100 | 50K  | 100 |
| ImageNet             | 1.25M | 882 | ≈30K | 30  |

**Table 3** 对语义超级类簇数量的敏感性分析，“Dum”表示合成的虚拟类别的数量。

| 方法          | $K$ | Dum  | CIFAR100-20 | CIFAR100-50 |
|-------------|-----|------|-------------|-------------|
| UNOv2 + RPK | 1   | 0    | 90.2        | 60.9        |
|             | 5   | 10   | 91.0        | 62.5        |
|             | 10  | 45   | 91.1        | 63.4        |
|             | 20  | 190  | <b>91.5</b> | <b>64.5</b> |
|             | 50  | 1225 | 90.9        | 63.5        |

(4) 我们还探索了将簇内和簇间虚拟类合成结合应用于 UNOv2 的效果，结果显示其性能略低于簇间虚拟类合成，但仍取得了次佳性能。这并不令人意外，因为对于簇内虚拟类合成，我们将同一语义超类簇中的基础类视为不同类，而在簇间虚拟类合成中，我们将这些基础类视为同一类。对这两种类型的虚拟类进行联合学习会阻碍优化过程，从而导致较差的结果。

总之，(1) 展示了使用辅助虚拟类学习的潜在有效性以及简单虚拟类合成的缺点。(2) 说明簇内虚拟类合成只能合成有限数量的虚拟类，这些类在语义上与基础类相似，因此在处理广泛视觉概念的场景中较为困难。(3) 证明通过使用簇间虚拟类合成，用得到的新语义且冗余的虚拟类来进行学习，能够对新类发现起重要作用。(4) 表明簇内虚拟类合成和簇间虚拟类合成由于虚拟类类型的冲突而不具有互补性。

#### 4.2.2 语义超类的数量

我们在 CIFAR100-20 和 CIFAR100-50 上使用标准的 NCD 设置评估语义超类簇数  $K$  的影响。表 3 显示了应用了簇间虚拟类合成的 UNOv2 的结果。随着  $K$  的增加，合成的虚拟类数量也会增加。可以观察到，当  $K$  的值为 20 时，模型达到了最佳效果。这是因为  $K$  值过小会导致合成的语义新虚拟类过少，使模型难以学习更

**Table 4** UNOv2+RPK 在 CIFAR100-20 和 CIFAR100-50 基准测试中使用不同聚类方法的结果。

| 方法          | 聚类方法                          | CIFAR100-20 | CIFAR100-50 |
|-------------|-------------------------------|-------------|-------------|
| UNOv2       | 无                             | 90.2        | 60.9        |
| UNOv2 + RPK | Agglomerative Clustering [54] | 91.4        | 64.3        |
|             | Spectral Clustering [55]      | <b>91.6</b> | <b>64.5</b> |
|             | <i>k</i> -means [53]          | 91.5        | <b>64.5</b> |

**Table 5** UNOv2+RPK 在 CIFAR100-20 和 CIFAR100-50 基准测试中使用不同采样范围的结果。

| 方法          | 采样区间      | CIFAR100-20 | CIFAR100-50 |
|-------------|-----------|-------------|-------------|
| UNOv2 + RPK | [0.1,0.9] | 90.9        | 62.4        |
|             | [0.2,0.8] | 90.9        | 62.8        |
|             | [0.3,0.7] | 91.1        | 62.8        |
|             | [0.4,0.6] | <b>91.5</b> | <b>64.5</b> |

丰富的视觉先验知识。相反， $K$  值过大会导致冗余的辅助虚拟类和巨大的计算成本，这与简单的虚拟类合成的情形相近。考虑到我们的方法在  $K$  值大于 1 时能够持续改善性能，因此当对数据集的语义超类没有先验知识时，我们可以选择一个不会显著增加计算成本的合适  $K$  值。在本文中，对于 CIFAR-100 和 Tiny-ImageNet 数据集，我们简单地将  $K$  值设置为每个数据集类别总数的五分之一（即 CIFAR-100 为 20，Tiny-ImageNet 为 40）。对于 ImageNet 数据集，我们将  $K$  值简单地设置为数据集类别总数的十分之一（即  $K=100$ ）。

#### 4.2.3 不同的聚类方法

我们评估了不同聚类方法（包括凝聚聚类 [54]、谱聚类 [55] 和  $k$ -means [53]）在 CIFAR100-20 和 CIFAR100-50 上使用标准 NCD 设置的效果。表 4 显示了使用不同聚类方法的 UNOv2+RPK 的结果。可以看出，我们的方法在使用不同的聚类方法时取得了相似的结果，这表明我们的方法对聚类方法不敏感。

#### 4.2.4 不同的采样区间

我们在 CIFAR100-20 和 CIFAR100-50 上使用标准 NCD 设置，评估了  $\lambda$  采样区间的选择。表 5 显示了 UNOv2+RPK 在不同  $\lambda$  采样区间下的结果。采样的范围越大，虚拟类与基类重叠的概率就越高。可以观察到，随着采样范围的减小，模型的性能逐渐提升。这是因为与单个基类的重叠过多会导致合成的虚拟类在语义上与该基类非常相似，使得学习到的表示不够多样化。

#### 4.2.5 不同的空间数据增强方法

我们评估了两种流行的空间数据增强方法（例如 Mixup 和 Cutmix [48]）在 CIFAR100-20 和 CIFAR100-50 数据集上进行标准 NCD 设定下的效果。表 6 显示了 UNOv2+RPK 使用不同空间数据增强方法的结果。可以观察到，UNOv2+RPK 在两种空间数据增强方法下的表现都优于 UNOv2，证明了学习多样化和可迁移表示的有效性。此外，Mixup 在两个基准数据集上的表现均优于 Cutmix。这是因为 Cutmix 在块级对两张图像进行混

**Table 6** UNOv2+RPK 在 CIFAR100-20 和 CIFAR100-50 基准测试中使用不同空间数据增强方法的结果。

| 方法          | Mixup 方法 | CIFAR100-20 | CIFAR100-50 |
|-------------|----------|-------------|-------------|
| UNOv2       | 无        | 90.2        | 60.9        |
| UNOv2 + RPK | Cutmix   | 90.6        | 63.7        |
|             | Mixup    | <b>91.5</b> | <b>64.5</b> |

合，而不是对整个图像进行线性插值，这样混合后的图像可能仅包含一张原始图像的物体，甚至可能没有物体（因为原始图像块可能仅包含图像的背景），这将导致生成的图像无法准确表示合成的虚拟类别。

### 4.3 与最先进方法的比较

我们将 RPK 应用于 UNOv2，与最先进的方法进行比较，包括  $k$ -means [53], KCL [11], MCL [12], DTC [4], RS [5], RS+ [5]（带有增量学习方案的 RS），NCL [7], NCL + HNG [7]（NCL 与困难负样本生成），UNO [8] 和 ComEx [17]。标准 NCD 结果（在无标签集的训练集上的聚类准确率）报告在表 7（用于 CIFAR10、CIFAR100 和 Tiny-ImageNet 数据集）和表 8（用于 ImageNet 数据集）中。任务无关设置的结果（在有标签和无标签集的测试集上的分类准确率）报告在表 9。

在表 7 中可以看到，UNOv2 + RPK 优于所有最先进的方法。这些结果验证了为 NCD 学习更丰富的先验知识的有效性。具体来说，与 UNOv2 相比，CIFAR10 的提升相对较小，主要原因是 CIFAR10 包含的基础类别太少，导致生成的虚拟类数量非常有限。对于更具实际意义的 CIFAR100-20 和 CIFAR100-50 设置，UNOv2 + RPK 自然比 UNOv2 分别提升了 1.3% 和 3.6%，这证明了在 NCD 任务中通过辅助虚拟类进行训练的好处。最后，对于极具挑战性的 Tiny-ImageNet-100 数据集，我们的方法仍然显著提高了性能，证明了 RPK 模块在更广泛语义概念场景中的有效性。

在表 9 中，我们可以观察到 UNOv2 + RPK 在任务无关设置下表现最佳或类似于所有方法。尽管 UNOv2 + RPK 在 CIFAR10 和 CIFAR100-20 上的改进有限，主要原因是这两个数据集的结果已经趋于饱和，尤其是在 CIFAR100-20 上，无标签分类头的准确率甚至超过了有标签分类头的准确率。相比之下，UNOv2 + RPK 在更具挑战性的 CIFAR100-50 数据集上表现出色，特别是在无标签分类头的准确率上，这证明了通过采用生成的虚拟类来学习更多样化和可迁移的表示可以在更实际的 NCD 设置中提升性能。在 Tiny-ImageNet-100 数据集中，与最佳方法 UNOv2 相比，它在所有类别上平均提升了 1.8% 的性能。

#### 4.3.1 即插即用不同方法

所提出的 RPK 模块具有即插即用的特性，可以应用于主流的两阶段方法。在表 7 和表 9，我们还报告了将

**Table 7** 在 CIFAR-10、CIFAR-100 和 Tiny-ImageNet 数据集下的 NCD 设置中，与最新方法的对比。报告了无标签数据集训练部分的聚类准确率。“RS+”是带有增量学习方案的 RS，“NCL + HNG”是带有困难负样本生成的 NCL。UNOv2 是 UNO 的改进版本，具有更好的性能。

| 方法                   | 期刊会议       | CIFAR10         | CIFAR100-20     | CIFAR100-50     | Tiny-ImageNet-100 |
|----------------------|------------|-----------------|-----------------|-----------------|-------------------|
| <i>k</i> -means [53] | Classic    | 72.5±0.0        | 56.3±1.7        | 28.3±0.7        | -                 |
| KCL [11]             | ICLR'18    | 72.3±0.2        | 42.1±1.8        | -               | -                 |
| MCL [12]             | ICLR'19    | 70.9±0.1        | 21.5±2.3        | -               | -                 |
| DTC [4]              | ICCV'19    | 88.7±0.3        | 67.3±1.2        | 35.9±1.0        | -                 |
| OpenMix [6]          | CVPR'21    | <b>95.3</b>     | 87.2            | -               | -                 |
| Joint [16]           | ICCV'21    | 93.4±0.6        | 76.4±2.8        | -               | -                 |
| UNO [8]              | ICCV'21    | 92.6±0.5        | 85.0±0.6        | 52.9±1.4        | -                 |
| DualRank [9]         | NeurIPS'21 | 91.6±0.6        | 75.3±2.3        | -               | -                 |
| ComEx [17]           | CVPR'22    | 93.6±0.3        | 85.7±0.7        | 53.4±1.3        | -                 |
| RS [5]               | ICLR'20    | 90.4±0.5        | 73.2±2.1        | 39.2±2.3        | 13.9±0.4          |
| + RPK (Ours)         | -          | <b>91.3±0.3</b> | <b>76.8±0.6</b> | <b>43.8±0.5</b> | <b>15.2±0.6</b>   |
| RS+ [5]              | ICLR'20    | 91.7±0.9        | 75.2±4.2        | 44.1±3.7        | 16.5±0.1          |
| + RPK (Ours)         | -          | <b>92.1±0.3</b> | <b>77.9±2.5</b> | <b>49.3±2.6</b> | <b>17.7±0.5</b>   |
| NCL [7]              | CVPR'21    | 93.4±0.2        | 82.3±2.6        | 48.9±0.5        | 22.9±0.2          |
| + RPK (Ours)         | -          | <b>93.9±0.1</b> | <b>84.5±2.1</b> | <b>52.2±0.5</b> | <b>24.1±0.2</b>   |
| NCL + HNG [7]        | CVPR'21    | <b>93.4±0.1</b> | 86.6±0.4        | 52.5±2.0        | 22.3±0.2          |
| + RPK (Ours)         | -          | 93.1±0.2        | <b>87.3±0.1</b> | <b>54.8±0.3</b> | <b>23.4±0.3</b>   |
| UNOv2 [8]            | ICCV'21    | 93.6±0.5        | 90.2±0.4        | 60.9±0.7        | 31.0 ± 1.1        |
| + RPK (Ours)         | -          | <b>93.8±0.2</b> | <b>91.5±0.2</b> | <b>64.5±1.7</b> | <b>32.4 ± 0.7</b> |

**Table 8** 在 ImageNet 数据集设置下与最新方法的对比。

| 方法           | 期刊会议       | ImageNet    |
|--------------|------------|-------------|
| RS [5]       | ICLR'20    | 85.7        |
| RS+ [5]      | ICLR'20    | 85.7        |
| OpenMix [6]  | CVPR'21    | 85.7        |
| NCL [7]      | CVPR'21    | 89.5        |
| NCL+HNG [7]  | CVPR'21    | 90.7        |
| Joint [16]   | ICCV'21    | 86.7        |
| UNO [8]      | ICCV'21    | 90.6        |
| DualRank [9] | NeurIPS'21 | 88.9        |
| ComEx [17]   | CVPR'22    | 90.9        |
| UNOv2 [8]    | ICCV'21    | 90.7        |
| UNOv2+RPK    | -          | <b>91.1</b> |

RPK 应用于其他流行的 NCD 方法（例如，RS [5], NCL [7]）的结果。在表 7 中，我们可以观察到，RPK 提高了 RS、NCL 及其变体的性能，除了在 CIFAR10 上的“NCL + HNG”。在表 9 中，将 RPK 应用于 RS+，在几乎所有性能指标上都显著提升了性能。值得注意的是，与“UNOv2 + RPK”相比，“RS+ + RPK”显著提高了有标签分类头的性能，这表明通过合成虚拟类来学习多样化和可迁移的特征对基础类的分类也是有益的。对于 UNOv2，他们提出的统一目标使有标签分类头能够达到非常高的分类准确率。因此，将 RPK 应用于 UNOv2 难以进一步提高有标签分类头的性能。总而言之，在其他主流方法中的性能提升表明了我们方法的显著优势。

#### 4.3.2 分布无关设置的结果

为了进一步验证我们方法的通用性，我们还评估了 RPK 在最近提出的分布无关 NCD 设置 [23] 下的性能，该设置假设新类别样本来自任意未知的类别分布。我们将 RPK

应用于 BYOP [23]，这是一种最近提出的分布无关的两阶段 NCD 方法，并将其与当前最先进的方法进行比较，包括 RS、RS+、NCL、UNO 和 UNO+BYOP [23]。在表 10 和表 11 中，报告了 CIFAR10 和 CIFAR100-20 在不同的不平衡比例下的比较结果。可以看出，UNO+BYOP+RPK 在大多数指标上取得了最佳结果，并且显著超过了 UNO+BYOP，充分验证了 RPK 在这一更具挑战性的环境中的有效性。

#### 4.3.3 与 OpenMix 的对比

通过合成辅助的虚拟类别，RPK 使模型能够在 NCD 方法的第一阶段学习更多可迁移的表示，而 OpenMix 主要通过 NCD 方法的第二阶段混合新旧类别样本来减少训练样本的不确定性。与 OpenMix 相比，我们的 RPK 模块可以应用于任何两阶段的 NCD 方法，并且易于实现。由于 OpenMix 是一种不修改第一阶段训练的两阶段 NCD 方法，因此 RPK 可以自然地与 OpenMix 结合。表 12 显示了 OpenMix+RPK 和 OpenMix 的对比结果。可以看出，OpenMix+RPK 在 CIFAR100 和 Tiny-ImageNet 数据集上实现了性能提升。

#### 4.3.4 大规模数据集的结果

我们在大规模的 ImageNet 数据集上评估了所提出的 RPK 模块，比较结果如表 8 所示。尽管基础类的数量相对于新类数量已经足够（882 和 30），UNOv2+RPK 仍然实现了性能提升，这证明了我们所提出模块的有效性。通过学习额外合成的虚拟类，能够获得更丰富的先验知识，进一步提升性能。

**Table 9** 在任务无关设置下与最新方法在 CIFAR-10、CIFAR-100 和 Tiny-ImageNet 数据集上的对比。测试集上评估了有标签集的分类准确率和无标签集的聚类准确率。

| 方法           | CIFAR10     |             |             | CIFAR100-20 |             |             | CIFAR100-50 |             |             | Tiny-ImageNet-100 |             |             |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------------|-------------|-------------|
|              | 基类          | 新类          | 所有类         | 基类          | 新类          | 所有类         | 基类          | 新类          | 所有类         | 基类                | 新类          | 所有类         |
| KCL [11]     | 79.4        | 60.1        | 69.8        | 23.4        | 29.4        | 24.6        | -           | -           | -           | -                 | -           | -           |
| MCL [12]     | 81.4        | 64.8        | 73.1        | 18.2        | 18.0        | 18.2        | -           | -           | -           | -                 | -           | -           |
| DTC [4]      | 58.7        | 78.6        | 68.7        | 47.6        | 49.1        | 47.9        | 30.2        | 34.7        | 32.5        | -                 | -           | -           |
| UNO [8]      | 93.5        | 93.3        | 93.4        | 73.2        | 73.1        | 73.2        | 71.5        | 50.7        | 61.1        | -                 | -           | -           |
| ComEx [17]   | 95.0        | 92.6        | 93.8        | 75.2        | 77.3        | 75.6        | 75.3        | 53.5        | 64.4        | -                 | -           | -           |
| RS+ [5]      | 90.6        | <b>88.8</b> | 89.7        | 71.2        | 56.8        | 68.3        | 69.7        | 40.9        | 55.3        | 54.9              | 16.5        | 33.0        |
| + RPK (Ours) | <b>94.2</b> | 87.9        | <b>91.0</b> | <b>76.4</b> | <b>65.0</b> | <b>74.1</b> | <b>74.1</b> | <b>46.5</b> | <b>60.2</b> | <b>60.2</b>       | <b>19.4</b> | <b>37.2</b> |
| UNOV2        | 96.2        | <b>91.2</b> | 93.7        | 76.3        | 77.6        | 76.5        | <b>75.5</b> | 57.8        | 66.8        | 55.7              | 31.1        | 43.4        |
| + RPK (Ours) | <b>96.4</b> | <b>91.2</b> | <b>93.8</b> | <b>76.4</b> | <b>78.4</b> | <b>76.8</b> | <b>75.5</b> | <b>61.2</b> | <b>68.4</b> | <b>57.8</b>       | <b>32.6</b> | <b>45.2</b> |

**Table 10** 在不同不平衡比例下与最新方法在 CIFAR10 数据集上的对比。“Trad.” 代表 “Traditional”，“Nov.” 代表 “Novel”。

| 数据集 →<br>规则 →<br>方法 ↓ | CIFAR10 (不平衡比例: 100) |             |             |             |             |             |             | CIFAR10 (不平衡比例: 10) |             |             |             |             |             |             |
|-----------------------|----------------------|-------------|-------------|-------------|-------------|-------------|-------------|---------------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                       | 传统                   |             | 任务感知        |             |             | 任务无关        |             | 传统                  |             | 任务感知        |             |             | 任务无关        |             |
|                       | 新类                   | 基类          | 新类          | 所有类         | 基类          | 新类          | 所有类         | 新类                  | 基类          | 新类          | All         | 基类          | 新类          | 所有类         |
| RS [5]                | 46.3                 | 71.8        | 43.2        | 57.5        | -           | -           | -           | 69.7                | 87.4        | 63.6        | 75.5        | -           | -           | -           |
| RS+ [5]               | 45.3                 | 64.4        | 50.1        | 57.3        | <b>64.4</b> | 55.5        | 60.0        | 66.5                | 77.3        | 63.3        | 70.3        | 77.3        | 62.3        | 69.8        |
| NCL [7]               | 47.2                 | 71.6        | 43.1        | 57.4        | -           | -           | -           | 62.6                | 86.9        | 56.9        | 71.9        | -           | -           | -           |
| UNO [8]               | 43.9                 | 69.6        | 52.2        | 60.9        | 56.0        | 55.6        | 55.8        | 59.6                | 88.1        | 59.1        | 73.6        | 78.2        | 58.8        | 68.5        |
| UNO+BYOP [23]         | 59.3                 | 70.1        | 53.3        | 61.7        | 56.6        | 56.6        | 56.6        | 63.2                | 88.5        | 61.7        | 75.1        | <b>78.4</b> | 61.0        | 69.7        |
| UNO+BYOP+RPK          | <b>67.7</b>          | <b>72.2</b> | <b>59.0</b> | <b>65.6</b> | 60.0        | <b>63.9</b> | <b>62.0</b> | <b>71.8</b>         | <b>89.6</b> | <b>68.4</b> | <b>79.0</b> | 75.3        | <b>65.9</b> | <b>70.6</b> |

**Table 11** 在不同不平衡比例下与最新方法在 CIFAR100-20 数据集上的对比。“Trad.” 代表 “Traditional”，“Nov.” 代表 “Novel”。

| Dataset →<br>Protocol →<br>方法 ↓ | CIFAR100-20 (不平衡比例: 100) |             |             |             |             |             |             | CIFAR100-20 (不平衡比例: 10) |             |             |             |             |             |             |
|---------------------------------|--------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                                 | 传统                       |             | 任务感知        |             |             | 任务无关        |             | 传统                      |             | 任务感知        |             |             | 任务无关        |             |
|                                 | 新类                       | 基类          | 新类          | 所有类         | 基类          | 新类          | 所有类         | Nov.                    | 基类          | 新类          | 所有类         | 基类          | 新类          | 所有类         |
| RS [5]                          | 36.5                     | 40.0        | 36.2        | 39.2        | -           | -           | -           | 47.6                    | 58.8        | 47.8        | 56.6        | -           | -           | -           |
| RS+ [5]                         | 35.3                     | 38.2        | 32.9        | 37.1        | 38.2        | 24.7        | 35.5        | 48.8                    | 55.5        | 45.2        | 53.4        | 55.5        | 35.6        | 51.5        |
| NCL [7]                         | 35.5                     | 39.1        | 28.7        | 37.0        | -           | -           | -           | 50.3                    | 57.1        | 45.6        | 54.8        | -           | -           | -           |
| UNO [8]                         | 35.2                     | 43.9        | 32.5        | 41.6        | 40.7        | 29.9        | 38.5        | 46.9                    | 60.9        | 46.6        | 58.0        | 57.9        | 38.4        | 54.0        |
| UNO + BYOP [23]                 | 50.3                     | 44.3        | 35.3        | 42.5        | 41.3        | 33.0        | 39.6        | 54.5                    | 61.4        | 48.3        | 58.8        | 58.2        | 40.6        | 54.7        |
| UNO + BYOP + RPK                | <b>52.6</b>              | <b>48.9</b> | <b>39.2</b> | <b>47.0</b> | <b>46.0</b> | <b>37.5</b> | <b>44.3</b> | <b>58.8</b>             | <b>66.4</b> | <b>51.3</b> | <b>63.4</b> | <b>62.9</b> | <b>45.2</b> | <b>59.4</b> |

**Table 12** 在 NCD 设置下，与 OpenMix 在 CIFAR-100 和 Tiny-ImageNet 数据集上的对比。

| 方法           | CIFAR100-20     | CIFAR100-50     | Tiny-ImageNet-100 |
|--------------|-----------------|-----------------|-------------------|
| OpenMix [6]  | 87.1±0.4        | 55.0±1.6        | 23.4±1.2          |
| + RPK (Ours) | <b>89.2±0.5</b> | <b>59.6±0.9</b> | <b>25.8±1.6</b>   |

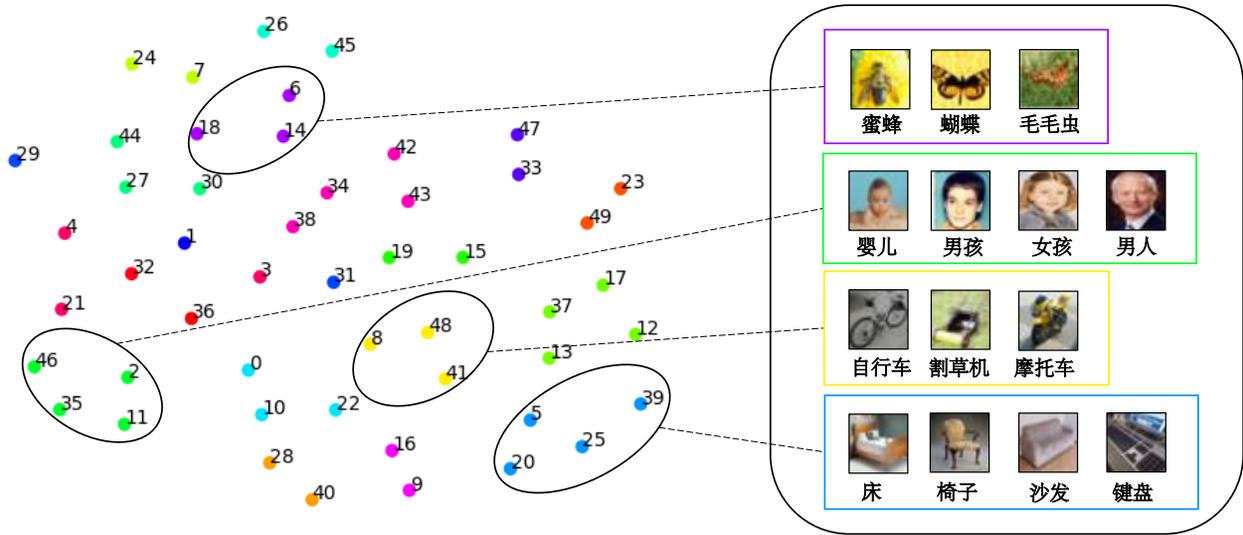
**Table 13** 在 NCD 设置下，与 SOTA 方法在 Stanford Cars、CUB 和 FGVC-Aircraft 数据集上的对比。

| 方法           | Stanford Cars   | CUB             | FGVC-Aircraft   |
|--------------|-----------------|-----------------|-----------------|
| UNO          | 49.8±1.4        | 59.2±0.4        | 52.1±0.7        |
| GCD [18]     | 42.6±0.4        | 56.4±0.3        | 49.5±1.0        |
| rKD [56]     | 53.5±0.8        | 65.7±0.6        | 55.8±0.9        |
| + RPK (Ours) | <b>54.7±0.2</b> | <b>66.1±0.8</b> | <b>57.3±0.8</b> |

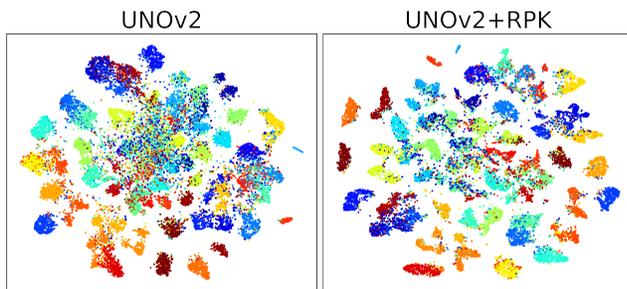
#### 4.3.5 细粒度数据集的结果

为了评估 RPK 在细粒度数据集上的性能，我们将 RPK 应用于 rKD [56]，这是一种最近提出的两阶段 NCD 方法，在细粒度基准上进行评估，与当前最先进的方法进行比较，包括 UNO、GCD[18] 和 rKD [56]。表 13 报告了在 Stanford Cars、CUB 和 FGVC-Aircraft 数据集上的比较结果。可以看出，尽管我们的方法取得了最佳结果，

但相比 rKD，性能提升相对有限。这是因为在细粒度数据集中，不同类别之间的视觉差异较小，这使得合成的虚拟类别与基础类之间的视觉差异也较小，从而使模型难以通过额外的虚拟类别学习到更多多样化的特征。



**Fig. 3** CIFAR100-50 基类的广义语义超类可视化。语义超类的数量  $K$  为 20，每个点代表一个基类，颜色相同的点表示它们属于同一个超类。通过对基类分类器的  $l_2$  标准化权重进行聚类，我们的方法能够很好地将语义相似的基类划分到同一个超类。例如，语义相似的类“bicycle”（自行车）、“mower”（割草机）和“motorcycle”（摩托车）实际上被分配到同一个超类。同样，“baby”（婴儿）、“boy”（男孩）、“girl”（女孩）和“man”（男人）也自然地划分到同一个超类。通过使用生成的超类簇进行超类间虚拟类的合成，合成的虚拟类之间的差异将被放大，同时减少了琐碎虚拟类的合成。



**Fig. 4** CIFAR100-50 无标签训练数据的 t-SNE 可视化。

#### 4.4 生成语义超类的可视化

为了验证聚类引导虚拟类合成的有效性，我们可视化了生成的语义超类。具体来说，我们对 CIFAR100-50 的数据集的  $l_2$  归一化基础类分类器权重进行了 t-SNE 可视化，结果如图 3 所示。我们发现，当使用 k-means 对基础类分类器的权重向量进行聚类时，语义相似的类别几乎都被划分到相同的语义超类中。例如，“自行车”、“割草机”和“摩托车”都可以被认为属于“交通工具”超类。同样，“婴儿”、“男孩”、“女孩”和“男人”都可以被归为“人”超类。通过为同一语义超类中的样本分配相同的标签，我们认为不同集群的合成虚拟类之间的差异将会很大，从而生成的类会更加有意义。

#### 4.5 学习到的特征空间的可视化

我们对 CIFAR100-50 的设置进行了定性分析。在图 4 中，我们使用 t-SNE 对学习到的特征空间进行了可视化。从图中可以看出，原始的 UNOV2 在表示空间中使大量不同的新类样本交织在一起，导致模型难以正确识别它

们。相比之下，UNOV2+RPK 为新类样本生成了更加紧凑的每类表示，并在表示空间中保持了更好的分离度。

## 5 结论

在这项工作中，我们通过学习合成的辅助虚拟类为 NCD 任务提出了一个更丰富的先验知识 (RPK) 模块。我们的核心思想是在未来可能遇到的新类上提前学习多样化和判别性强的表示。为实现这一目标，我们使用带有约束权重参数的 Mixup 来合成虚拟类，模拟新类的分布，并在有监督的预训练中学习虚拟类。为了有效学习多样化和可迁移的特征并提高性能，我们开发了一种聚类引导的虚拟类合成策略，以消除冗余的虚拟类。四个数据集上的实验结果表明，将我们的 RPK 应用于主流的 NCD 方法可以显著提高性能。

## 6 声明

### 6.1 资助

本工作由 NSFC 资助 (编号: 62225604 和 62206135)。

### 6.2 作者贡献

各作者的贡献如下。彭志茂提出了本文的主要思想，完成了大部分实验，并撰写了论文的主体部分。王恩光参与了实验设计，并与彭志茂共同撰写了实验部分。刘夏雷改进了思想并修订了论文的摘要和方法部分，参与了大部分讨论。程明明参与了大部分关于完善思想的讨论，并修订了论文的引言和相关工作部分。

### 6.3 利益冲突声明

作者声明与本手稿相关的利益冲突不存在。  
没有其他声明。

### References

- [1] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 770–778.
- [2] He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2017, 2961–2969.
- [3] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 39(6): 1137–1149.
- [4] Han K, Vedaldi A, Zisserman A. Learning to discover novel visual categories via deep transfer clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, 8401–8409.
- [5] Han K, Rebuffi SA, Ehrhardt S, Vedaldi A, Zisserman A. Automatically Discovering and Learning New Visual Categories with Ranking Statistics. In *International Conference on Learning Representations (ICLR)*, 2020, 1–13.
- [6] Zhong Z, Zhu L, Luo Z, Li S, Yang Y, Sebe N. 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*, 2021, 9462–9470.
- [7] Zhong Z, Fini E, Roy S, Luo Z, Ricci E, Sebe N. Neighborhood Contrastive Learning for Novel Class Discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 10867–10875.
- [8] Fini E, Sangineto E, Lathuilière S, Zhong Z, Nabi M, Ricci E. A unified objective for novel class discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 9284–9292.
- [9] Zhao B, Han K. Novel Visual Category Discovery with Dual Ranking Statistics and Mutual Knowledge Distillation. In *Advances in Neural Information Processing Systems*, 2021, 22982–22994.
- [10] Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. mixup: Beyond Empirical Risk Minimization. In *International Conference on Learning Representations (ICLR)*, 2018, 1–13.
- [11] Hsu YC, Lv Z, Kira Z. Learning to cluster in order to transfer across domains and tasks. In *International Conference on Learning Representations (ICLR)*, 2018, 1–20.
- [12] Hsu YC, Lv Z, Schlosser J, Odom P, Kira Z. Multi-class classification without multi-class labels. In *International Conference on Learning Representations (ICLR)*, 2019, 1–16.
- [13] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 2009, 22(10): 1345–1359.
- [14] Wang M, Wang S, Yang X, Yuan J, Zhang W. Equity in Unsupervised Domain Adaptation by Nuclear Norm Maximization. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.
- [15] Gidaris S, Singh P, Komodakis N. Unsupervised representation learning by predicting image rotations. In *International Conference on Learning Representations (ICLR)*, 2018, 1–16.
- [16] Jia X, Han K, Zhu Y, Green B. Joint Representation Learning and Novel Category Discovery on Single- and Multi-modal Data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 610–619.
- [17] Yang M, Zhu Y, Yu J, Wu A, Deng C. Divide and Conquer: Compositional Experts for Generalized Novel Class Discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 14268–14277.
- [18] Vaze S, Han K, Vedaldi A, Zisserman A. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 7492–7501.
- [19] Caron M, Touvron H, Misra I, Jégou H, Mairal J, Bojanowski P, Joulin A. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, 9650–9660.
- [20] Guo LZ, Zhang YG, Wu ZF, Shao JJ, Li YF. Robust semi-supervised learning when not all classes have labels. *Advances in Neural Information Processing Systems*, 2022, 35: 3305–3317.
- [21] Joseph K, Paul S, Aggarwal G, Biswas S, Rai P, Han K, Balasubramanian VN. Novel Class Discovery without Forgetting. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIV*, 2022, 570–586.
- [22] Roy S, Liu M, Zhong Z, Sebe N, Ricci E. Class-incremental Novel Class Discovery. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII*, 2022, 317–333.
- [23] Yang M, Wang L, Deng C, Zhang H. Bootstrap your own prior: Towards distribution-agnostic novel class discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, 3459–3468.
- [24] Yu Q, Ikami D, Irie G, Aizawa K. Self-labeling framework for novel category discovery over domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2022, 3161–3169.
- [25] An W, Tian F, Chen P, Tang S, Zheng Q, Wang Q. Fine-grained category discovery under coarse-grained supervision with hierarchical weighted self-contrastive learning. *arXiv preprint arXiv:2210.07733*, 2022.
- [26] Van Engelen JE, Hoos HH. A survey on semi-supervised learning. *Machine Learning*, 2020: 373–440.
- [27] Rasmus A, Berglund M, Honkala M, Valpola H, Raiko T. Semi-supervised learning with ladder networks. In *Advances in neural information processing systems*, 2015, 3546–3554.

- [28] Sajjadi M, Javanmardi M, Tasdizen T. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In *Advances in neural information processing systems*, 2016, 1171–1179.
- [29] Laine S, Aila T. Temporal ensembling for semi-supervised learning. In *International Conference on Learning Representations (ICLR)*, 2017, 1–13.
- [30] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in neural information processing systems*, 2017, 1195–1204.
- [31] Xie Q, Dai Z, Hovy E, Luong T, Le Q. Unsupervised data augmentation for consistency training. In *Advances in neural information processing systems*, 2020, 6256–6268.
- [32] Grandvalet Y, Bengio Y. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems*, 2004.
- [33] Lee DH, et al.. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, 2013, 896.
- [34] Wu H, Prasad S. Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Transactions on Image Processing*, 2017: 1259–1270.
- [35] Xie Q, Luong MT, Hovy E, Le QV. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, 10687–10698.
- [36] Wang X, Kihara D, Luo J, Qi GJ. EnAET: A self-trained framework for semi-supervised and supervised learning with ensemble transformations. *IEEE Transactions on Image Processing*, 2020: 1639–1647.
- [37] Chen T, Kornblith S, Swersky K, Norouzi M, Hinton GE. Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, 2020: 22243–22255.
- [38] Wang L, Liu Y, Di H, Qin C, Sun G, Fu Y. Semi-supervised dual relation learning for multi-label classification. *IEEE Transactions on Image Processing*, 2021: 9125–9135.
- [39] Ke R, Aviles-Rivero AI, Pandey S, Reddy S, Schönlieb CB. A three-stage self-training framework for semi-supervised semantic segmentation. *IEEE Transactions on Image Processing*, 2022: 1805–1815.
- [40] of the IEEE/CVF conference on computer vision P, pattern recognition. Autoaugment: Learning augmentation strategies from data. In *CVPR*, 2019, 113–123.
- [41] Cubuk ED, Zoph B, Shlens J, Le QV. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, 702–703.
- [42] Berthelot D, Carlini N, Goodfellow I, Papernot N, Oliver A, Raffel CA. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 2019.
- [43] Sohn K, Berthelot D, Carlini N, Zhang Z, Zhang H, Raffel CA, Cubuk ED, Kurakin A, Li CL. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 2020: 596–608.
- [44] Zhang B, Wang Y, Hou W, Wu H, Wang J, Okumura M, Shinzaki T. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in neural information processing systems*, 2021: 18408–18419.
- [45] Wang Y, Chen H, Heng Q, Hou W, Savvides M, Shinzaki T, Raj B, Wu Z, Wang J. Freematch: Self-adaptive thresholding for semi-supervised learning. *International Conference on Learning Representations (ICLR)*, 2023.
- [46] Chen H, Tao R, Fan Y, Wang Y, Wang J, Schiele B, Xie X, Raj B, Savvides M. SoftMatch: Addressing the Quantity-Quality Trade-off in Semi-supervised Learning. *International Conference on Learning Representations (ICLR)*, 2023.
- [47] Verma V, Lamb A, Beckham C, Najafi A, Mitliagkas I, Lopez-Paz D, Bengio Y. Manifold mixup: Better representations by interpolating hidden states. In *International conference on machine learning*, 2019, 6438–6447.
- [48] Yun S, Han D, Oh SJ, Chun S, Choe J, Yoo Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, 6023–6032.
- [49] Kim JH, Choo W, Song HO. Puzzle mix: Exploiting saliency and local statistics for optimal mixup. In *International Conference on Machine Learning*, 2020, 5275–5285.
- [50] Zhong Z, Cui J, Liu S, Jia J. Improving calibration for long-tailed recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, 16489–16498.
- [51] Zhu F, Cheng Z, Zhang XY, Liu CI. Class-Incremental Learning via Dual Augmentation. In *Advances in Neural Information Processing Systems*, 2021, 14306–14318.
- [52] Han K, Rebuffi S, Ehrhardt S, Vedaldi A, Zisserman A. AutoNovel: Automatically Discovering and Learning Novel Visual Categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021: 6767–6781.
- [53] MacQueen J, et al.. Some methods for classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, 1967, 281–297.
- [54] Müllner D. Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*, 2011.
- [55] Ng A, Jordan M, Weiss Y. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2001, 14.
- [56] Gu P, Zhang C, Xu R, He X. Class-relation Knowledge Distillation for Novel Class Discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, 16474–16483.

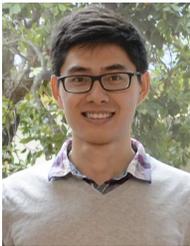
- [57] Vaze S, Han K, Vedaldi A, Zisserman A. Generalized Category Discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, 7492–7501.
- [58] Kuhn HW. The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 1955: 83–97.



**Zhimao Peng** is currently a Ph.D. candidate in the College of Computer Science, Nankai University, under the supervision of Prof. Ming-Ming Cheng. His research interests include deep learning, image classification, and self/semi-supervised learning.



**Enguang Wang** is currently a master student at Nankai University, under the supervision of Xialei Liu. His research interests include applications in computer vision, self-supervised learning and open world tasks.



**Xialei Liu** is currently an associate professor at Nankai University. Before that, He was a postdoctoral researcher at the University of Edinburgh. He received his Ph.D. degrees from the Autonomous University of Barcelona in 2019, supervised by Prof. Joost van de Weijer and Prof. Andrew D. Bagdanov. He works in the field of computer vision and machine learning. His research interests include continual learning, self-supervised learning, and few-shot learning.



**Ming-Ming Cheng** received his Ph.D. degree from Tsinghua University in 2012, and then worked with Prof. Philip Torr in Oxford for 2 years. He is now a professor at Nankai University, leading the Media Computing Lab. His research interests include computer vision and computer graphics. He is a senior member of the IEEE and on the editorial boards of IEEE TPAMI and IEEE TIP.