arXiv:2304.00464v2 [cs.RO] 4 Apr 2023

UniDexGrasp++：通过几何感知课程学习和迭代通才-专家学习改进灵巧抓取策略学习

Weikang Wan [1]* Haoran Geng [1,3]* Yun Liu [2]
Zikang Shan [1] Yaodong Yang [1,3] Li Yi [2] He Wang [1]
†[1]北京大学 [2]清华大学[3]北京通用人工智能研究院

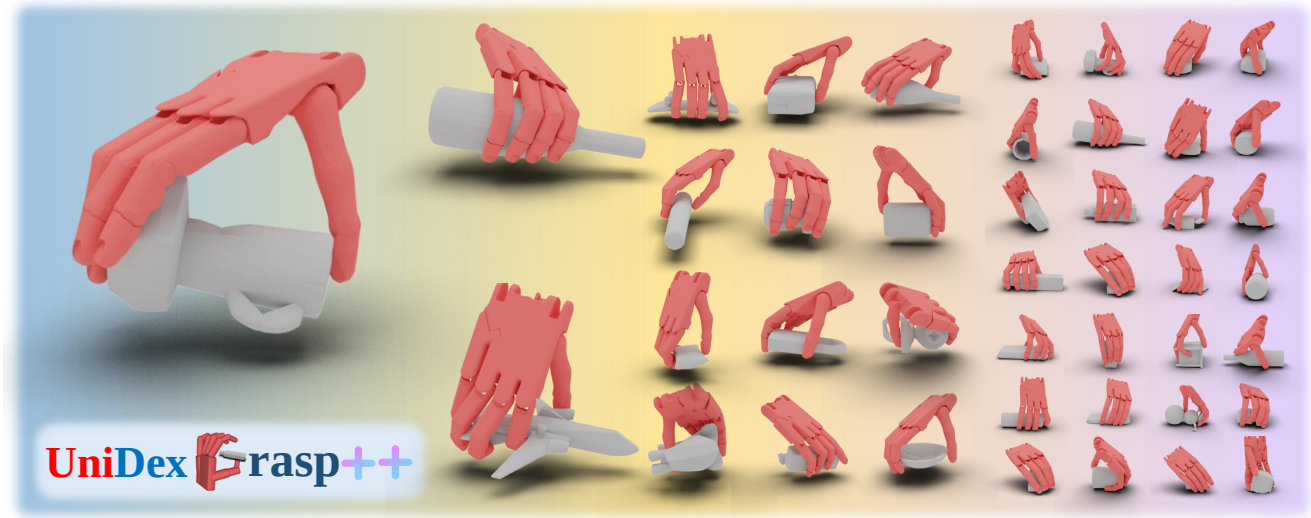


图 1：本文提出了一种新颖的灵巧抓取策略学习流程 UniDexGrasp++。与 UniDexGrasp[71], UniDexGrasp++ 在 3000+ 桌面场景下针对具有随机物体姿态的不同物体实例进行训练。其性能显著优于此前的最先进水平，在训练集和测试集上的成功率分别达到 85.4% 和 78.2%。

摘要

我们提出了一种新颖的物体无关方法，在桌面场景下利用真实点云观测和本体感知信息，学习用于灵巧物体抓取的通用策略，称为 UniDexGrasp++。为应对跨数千个物体实例学习基于视觉的策略这一挑战，我们提出了几何感知课程学习（***GeoCurriculum***）和几何感知迭代通才-专家学习（***GiGSL***），利用任务的几何特征，显著提升了泛化能力。通过所提出的技术，我们的最终策略能够对数千个物体实例实现通用灵巧抓取，在训练集和测试集上的成功率分别达到**85.4%**和**78.2%**，分别比最先进基线 UniDexGrasp 高出 11.7% 和 11.3%。

1. 引言

机器人抓取是机器人学中的一项基础且经过广泛研究的问题，近年来也受到计算机视觉领域越来越多的关注。近期的研究 [62, 6, 18, 24, 67, 17, 13] 在使用强化学习或运动规划为平行夹爪开发抓取算法方面取得了显著进展。然而，传统的平行夹爪灵活性有限，这限制了它们在日常生活中协助人类的能力。

因此，灵巧抓取正变得日益重要，因为它能提供更多样化的抓取策略，并支持更先进的操作技术。灵巧手动作空间的高维特性（例如，24 至 30 个自由度）是一项关键优势，使其具备高度的灵活性；与此同时，这也是成功执行抓取的主要难点。此外，复杂的手部关节运动会显著削弱基于运动规划的抓取

*贡献相同。
†通讯作者。

方法的效果，使强化学习成为灵巧抓取领域的主流方法。

然而，直接训练基于视觉的通用灵巧抓取策略 [37, 38, 40, 59]极具挑战性。首先，基于视觉的策略学习众所周知难度很高，因为强化学习产生的策略梯度通常噪声过大，无法有效更新视觉骨干网络。其次，此类策略学习本质上是多任务强化学习问题，存在极大的差异（例如几何形状和姿态各不相同），因而难以解决 [40, 29, 59]。尽管近期强化学习（RL） [4, 2, 37, 8, 9, 49, 41, 27, 38, 58, 69]在复杂的灵巧操作任务中取得了令人瞩目的成果，但训练出的策略难以泛化到大量对象以及未见过的对象。同时，大多数研究 [4, 2, 69, 9, 58, 49, 27] 都假设机器人掌握所有预知信息，例如对象的位置和旋转角度，这使得它们不适用于真实世界。

近期的一项工作 UniDexGrasp [70], 在其涵盖 3000 多个物体实例的基准测试上，展现了基于视觉的灵巧抓取的良好效果。其策略仅以机器人本体感知信息和逼真的点云观测作为输入。为简化策略学习，UniDexGrasp 提出了物体课程学习：从单个物体开始进行强化学习，并逐步将同一类别或相似类别中的相似物体纳入训练，从而获得基于状态的教师策略。获得该教师策略后，他们使用 DAgger [51]对其进行蒸馏，得到基于视觉的策略。最终，该策略在训练集和测试集划分上的成功率分别达到 73.7% 和 66.9%。UniDexGrasp 的一个局限是，其基于状态的教师策略在训练集上的成功率仅为 79.4%，这进一步限制了基于视觉的学生策略的性能。物体课程学习的另一个局限是未考虑物体姿态，且依赖类别标签。

为克服这些局限，我们提出了 UniDex- Grasp++，这是一种能够显著提升 UniDexGrasp 性能的新型流程。首先，为了提升基于状态的教师策略的性能，我们提出了几何感知任务课程学习（***GeoCurriculum***），该方法根据场景点云的几何特征衡量任务相似性。为了进一步提升策略的泛化能力，我们采用了通才-专家学习 [63, 39, 23, 29] 的理念：针对任务空间的子集训练一组专家，然后将其蒸馏到一个通才策略中。我们进一步提出了几何感知迭代通才-专家学习 ***GiGSL*** ，利用几何特征决定由哪个专家处理哪个任务，并迭代进行蒸馏和微调。我们的方法得到性能最佳的基于状态的策略，在训练集和测试集上的成功率分别达到**87.9%** 和**83.7%** 。随后，我们将性能最佳的专家策略蒸馏到一个基于视觉的通才策略中，并在基于视觉的策略上再次执行 GiGSL，直至性能饱和。通过完整流程，我们的最终的基于视觉的策略展现了对 3000+个物体实例的通用灵巧抓取能力，在训练集和测试集上的成功率分别达到**85.4%** 和**78.2%** ，显著优于最先进基线 UniDexGrasp，分别高出 11.7% 和 11.3%。在 Meta-World [73] 上开展的补充实验进一步证明了我们方法的有效性，其表现优于此前最先进的多任务强化学习方法。

2. 相关工作

2.1. 灵巧抓取

灵巧手因其在机器人学中实现类人操作的潜力而受到广泛关注 [53, 52,43, 14, 3, 12, 33, 32, 41, 36, 45, 35]。灵巧手潜力巨大，但由于其高度灵活性，相关研究也极具挑战。灵巧抓取是该领域备受关注的课题。一些工作 [5, 15, 3]采用解析方法对手和物体的运动学与动力学进行建模，但通常需要进行简化，例如使用简单的手指与物体几何形状，以确保规划过程可行。近年来，强化学习和模仿学习方法 [8, 9, 49, 41, 27, 4, 58, 69]取得了令人鼓舞的成果。尽管这些工作表现良好，但都假设机器人在训练和测试期间能够获取所有预知状态（例如物体位置、速度）。然而，在真实世界中，这些状态信息难以轻松且准确地获取。为缓解这一问题，一些工作 [38, 37, 46, 70] 考虑了更贴近现实的设定，将机器人本体感知与 RGB 图像或三维场景点云作为策略的输入，因为这些信息在真实世界中更容易获取。我们的工作与近期提出的 UniDexGrasp [70]关系更为密切，该工作学习了适用于 3000+不同物体的基于视觉的策略。本文提出了一种新颖的流程，显著提升了 UniDexGrasp 的性能与泛化能力，称为 UniDexGrasp++。

2.2. 基于视觉的策略学习

大量研究致力于探索如何从视觉输入中学习策略 [74, 30, 60, 72, 61,22, 21, 20]。为简化优化和训练过程，一些工作采用预训练视觉模型并冻结骨干网络，如 [57, 48, 56] 等工作所示。另一些工作，如 [69, 68], , 则采用多阶段训练。我们的工作与 [8, 7, 70], 更为相关；他们首先训练基于状态的策略，然后将其蒸馏为基于视觉的策略。此外，我们的工作充分利用视觉骨干网络的预训练，并将其融入模仿（监督）学习和强化学习的过程中，从而能够在基于视觉的设置下训练出具有泛化能力的策略。

2.3. 模仿学习中的泛化与策略蒸馏

为了在策略学习中泛化到较大的环境变化（例如物体几何形状、任务语义），以往工

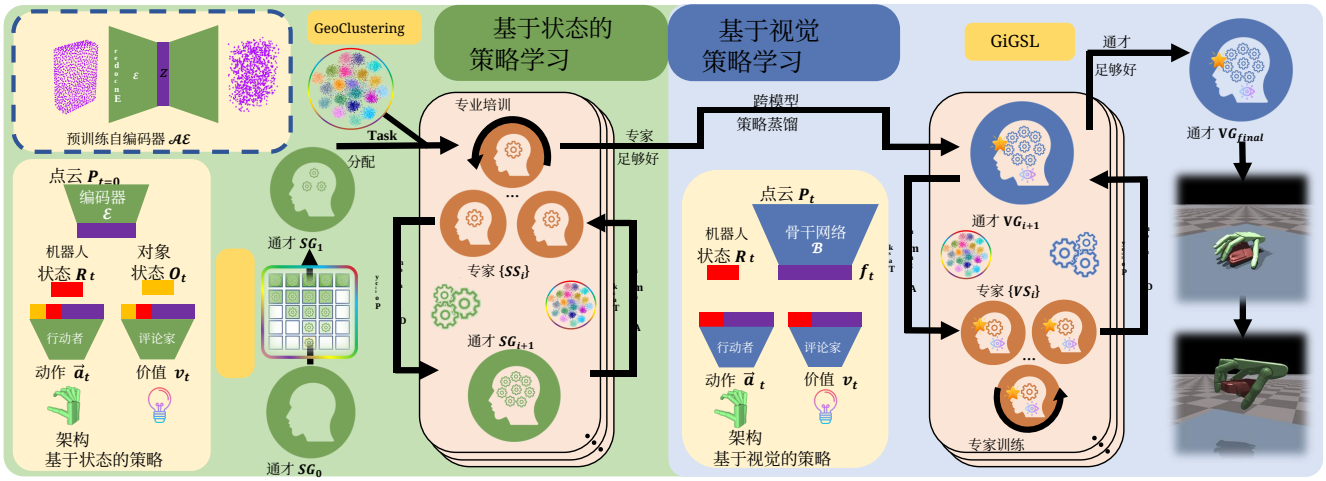


图 2：**方法概述**。 我们提出先采用基于状态的策略学习阶段，随后进入基于视觉的策略学习阶段。基于状态的策略以机器人状态 R_t、物体状态 S_t以及第一帧场景点云的几何特征 z 作为输入。我们利用几何感知任务课程（***GeoCurriculum***）来学习初始基于状态的通才策略。之后，通过在我们提出的几何感知迭代通才-专才学习（***GiGSL***）中迭代执行专家微调并将结果蒸馏回通才，进一步提升该通才策略；其中，将任务分配给哪个专家由我们的几何感知聚类（***GeoClustering***）决定。对于基于视觉的策略学习，我们首先将最终基于状态的专家策略蒸馏到初始基于视觉的通用策略中，然后对视觉通用策略执行 *GiGSL*，直至获得性能最高的最终基于视觉的通用策略。

作采用了模仿学习，包括行为克隆 [64, 31], 、通过增强示范数据进行强化学习 [49, 69, 47, 59, 16] 以及逆强化学习 [42, 1, 26, 19, 34]，以利用专家示范或策略。一些工作 [63, 39, 23, 29]采用了通才-专家学习的思路：先在任务空间的一个子集上训练一组专家（教师），然后利用上述模仿学习和策略蒸馏方法，将其蒸馏为覆盖整个任务空间的单一通才（学生）。尽管这些工作在若干基准测试中取得了显著进展 [40, 73, 28, 65, 11]，但它们要么没有认识到如何划分任务或环境变化对专家的重要性，要么关注的设定与我们的方法不同。在这项工作中，我们在划分专家时利用任务的几何特征，并采用课程学习，从而大幅提升了泛化能力。

3. 问题定义

在本研究中，我们关注如何在桌面场景下，利用真实点云观测和本体感知信息，学习灵巧物体抓取的通用策略，这与 [70, 46] 类似。

我们从多样化的抓取任务中学习这样的通用策略。抓取任务定义为 $\tau = (o,\ R)$，其中 $o \in \mathbb{O}$ 是物体数据集 $\mathbb{O}$中的一个对象实例，且 $R \in \mathbf{SO(3)}$ 是对象的初始三维旋转。为构建环境，我们随机采样一个对象 o，让它从一定高度落下，由此随机确定其初始姿态，然后将对象中心移至桌面中心。我们始终将灵巧手初始化在桌面中心上方的固定姿态。若对象与目标之间的位置差小于阈值，则任务成功。这是一种多任务策略学习设置，我们要求学习到的策略能够在多样化的抓取任务中良好泛化，例如，适用于随机初始姿态以及数千个对象，包括未见过的对象。

4. 方法

本节全面介绍我们提出的、用于解决复杂任务的方法。在第4.1节中，我们概述了我们的方法及其训练流程。我们提出的方法采用基于DAgger的蒸馏和迭代通才-专家学习（iGSL）策略，详情见第4.2节。此外，我们引入几何感知聚类，以确定由哪个专家处理哪个任务，从而实现几何感知迭代通才-专才学习（GiGSL），相关内容见第4.3节。在第4.4节中，我们介绍一种几何感知任务课程学习方法，用于训练首个基于状态的通才策略。

4.1. 方法概述

据此， [70, 8, 7], 我们可以将策略学习分为两个阶段：1）基于状态的策略学习阶段；2）基于视觉的策略学习阶段。众所周知，直接学习基于视觉的策略非常具有挑战性，因此我们首先学习一个能够获取预言信息的基于状态的策略，并利用该策略辅助和简化基于视觉的策略学习。完整流程如图 2所示。

基于状态的策略学习阶段。本阶段的目标

是获得一个通用策略，或者说我们称之为通才，它在第一帧接收来自机器人状态 R_t、物体状态 O_t和场景点云 $P_{t=0}$ 的输入。这里的对象点云由多个多视角深度相机捕获的深度点云融合而成。我们将 $P_{t=0}$纳入输入，以保留场景几何信息，并使用预训练点云自编码器的编码器提取其几何特征。请注意，此时点云编码器被冻结，以尽可能简化处理，避免干扰策略学习。我们将 P_t 的视觉处理留给基于视觉的策略。

尽管通过强化学习学习基于状态的策略比学习基于视觉的策略更容易实现，但要在如此多样化的多任务设置中取得较高的成功率，仍然非常具有挑战性。因此，我们提出一种几何感知课程学习方法***(GeoCurriculum)***，以简化多任务强化学习并提高成功率。

经过这一***GeoCurriculum***后，我们获得了首个能够处理所有任务的基于状态的通才 SG_1 。随后，我们提出一种几何感知迭代通才-专家学习策略，称为***GiGSL***，以进一步提升通才的性能。该过程通过迭代进行：学习多个基于状态的专家$\{\text{SS}_i\}$，使其专门处理特定范围的任务；再将这些专家蒸馏到通才 SG_{i+1}中，其中 i 表示迭代索引。通过这种迭代学习，整体性能持续提升，直至趋于饱和。

基于视觉的策略学习。对于基于视觉的策略，我们只允许其访问真实世界中可获取的信息，包括机器人状态 R_t 和场景点云 P_t。在这一阶段，我们需要联合学习一个视觉骨干网络 $\mathcal{B}$，以从 f_t 中提取 P_t，并与我们的策略一同学习（见图中蓝色部分。2）。这里我们采用 Point-+Transformer [40] 作为 $\mathcal{B}$，我们发现它具有更大的容量，因此性能优于 PointNet [44]。我们随机初始化首个视觉通才 VG_1的网络权重。首先进行跨模态蒸馏，将最新的基于状态的专家 $\{\text{SS}_n\}$ 蒸馏到 VG_1中。随后，我们便可开始基于视觉策略的 *GiGSL* 循环，在微调 $\{\text{VS}_i\}$ 与蒸馏到 VG_{i+1} 之间交替进行，直至基于视觉的通才性能趋于饱和。最终基于视觉的通才 VG_{final}即为我们学习得到的通用抓取策略，具有最高的性能。整个流程的伪代码请参阅补充材料。

4.2.*iGSL*：迭代通才-专家学习

回顾通才-专家学习（*GSL*）。通才-专家学习的构想源自一系列关于多任务策略学习的研究 [63, 39, 23, 29]。最新的一篇论文 [29] 提出了*GSL*，这是一种将整个任务空间划分为多个子空间，并由一位专家负责一个子空间的方法。由于每个子空间包含的任务变化较少，因此更易于学习，每位专家都能得到良好训练，并在各自的任务分布上表现出色。最后，所有专家都将被蒸馏为一个通才。

请注意，[29] 仅包含一轮专家学习和通才学习。直观来看，增加迭代轮数可能会有所帮助。在*GSL*中，蒸馏通过 GAIL [26]或 DAPG [49] 实现，但我们发现它们的表现平平。在这项工作中，我们提出了一种基于 DAgger 的更优策略蒸馏方法，通过迭代实现通才-专家学习。

基于 DAgger 的策略蒸馏。DAgger [51] 是一种在线模仿学习算法。与只需要专家示范数据的 GAIL 或 DAPG 不同，DAgger [51] 需要一个专家策略（称为教师），以及负责与环境交互的学生。当学生采取动作时，教师策略会将其动作作为监督信号，以改进学生。由于学生始终使用自己的策略与环境交互，这种模仿属于在线学习，因此不会受到行为克隆算法中常见的协变量偏移问题的影响。先前的工作，例如用于灵巧抓取的 [70]和用于手内操作的 [8, 7]，已使用 DAgger 将基于状态的教师策略蒸馏到基于视觉的学生策略；UniDexGrasp [70] 中的研究表明，在策略蒸馏方面，DAgger 优于 GAIL 和 DAPG。

然而，DAgger 的一个局限在于，它只关注策略网络，而舍弃了 PPO [55] 和 SAC [25]等常见演员-评论家强化学习方法通常包含的价值网络。在这种情况下，如果教师同时包含行动者和评论家，并将其蒸馏到学生模型中，那么学生模型将只有行动者而没有评论家，因此无法再使用演员-评论家强化学习进行微调。这使得 *GSL* 只能进行一个循环，也阻碍了通才的进一步改进。

为缓解这一问题，我们提出一种新的蒸馏方法，在使用 DAgger 学习行动者的同时联合学习评论家函数。在监督式策略蒸馏过程中，我们基于 DAgger 的蒸馏方法同时学习策略和评论家函数：策略损失是教师策略 $\pi_{teacher}$ 与学生策略 π_θ 生成的动作之间的均方误差（MSE）（与 DAgger 中相同）；评论家损失则是预测的价值函数 V_ϕ 与使用广义优势估计（GAE）[54]计算的估计回报 R_t 之间的均方误差。

$$\mathcal{L} = \frac{1}{|D_{\pi_\theta}|}\sum_{\tau\in D_{\pi_\theta}}(\pi_{teacher}(s_t)-\pi_\theta(s_t))^2 + \frac{1}{|D_{\pi_\theta}|\,T}\sum_{\tau\in D_{\pi_\theta}}\sum_{t=0}^{T}(V_\phi(s_t)-\hat{R}_t)^2 \tag{1}$$

这种基于DAgger的蒸馏方法使我们能够保留行动者和评论家，同时取得极高的性能。与仅预训练行动者的 IL AD [69]相比，

并直接微调演员-评论家强化学习（评论家网络从头开始训练）相比，我们的方法使演员-评论家强化学习能够在训练好的行动者和评论家网络上进行微调，从而提高强化学习训练的稳定性和有效性。

专家微调与通才蒸馏之间的迭代。借助我们提出的基于DAgger的蒸馏方法，可以执行以下步骤：1）从通过*GeoCurriculum*学习得到的第一个通才模型开始；2）将通才模型克隆为多个专家，并在各自的任务分布上对每个专家进行微调；3）使用基于DAgger的蒸馏方法，将所有专家蒸馏到一个通才模型中；我们可以在步骤2）和3）之间迭代，直到性能趋于饱和。

4.3. *GiGSL*：几何感知的迭代通才-专家学习

iGSL 尚有一个重要问题待解决：如何划分任务空间。在 [29], 中，任务数量有限，因此可以为每个任务分配一个专家，也可以随机分配。然而，我们的工作需要处理无限数量的任务，因为初始物体姿态可以连续变化。我们只能承担有限数量的专家，因此需要找到一种方法，将采样得到的任务分配给某个专家。我们认为，相似的任务应分配给同一个专家，因为只有当任务变化较小时，专家才能通过强化学习有效提升。为此，我们提出了***GeoClustering***，一种用于任务空间几何感知聚类的策略。

***GeoClustering* 策略**。我们将任务空间 $\mathbb{T} = \mathbb{O} \times \mathbf{SO(3)}$ 划分为 N_{clu} 个簇，每个簇中的任务C_j 由指定的专家 S_j 在专 家微调期间处理。我们首先从 $\mathbb{T}$ 中采样大量任务 $\{\tau^{(k)}\}_{k=1}^{N_{\text{sample}}}$（在我们的实现中为 $N_{\text{sample}} \approx 270,000$），并使用 K-Means 对其视觉特征进行聚类。大规模任务样本的聚类结果近似表示了整个连续任务空间的聚类情况。

我们首先使用样本任务 $\{\tau^{(k)}\}_{k=1}^{N\text{sample}}$ 中初始化对象的点云 $\{P_{t=0}^{(k)}\}_{k=1}^{N\text{sample}}$ 训练点云三维自编码器。该自编码器采用编码器-解码器结构。编码器 $\mathcal{E}$ 对 $P_{t=0}^{(k)}$ 进行编码，并输出编码后的潜在特征 $z^{(k)} = \mathcal{E}(P_{t=0}^{(k)})$。解码器 $\mathcal{D}$ 将 $z^{(k)}$ 作为输入，并生成点云 $\hat{}P_{t=0}^{(k)}$。模型使用重建损失 $\mathcal{L}_{\text{AE}}$进行训练，该损失为 $P_{t=0}^{(k)}$ 与 $\hat{}P_{t=0}^{(k)}$ 之间的倒角距离。更多详情请参阅补充材料。

在对基于状态的专家进行聚类时，我们使用预训练编码器 $\mathcal{E}$ 对任务 $\tau^{(k)}$ 的对象点云$P_{(k)\ t=0}$进行编码，并获得潜在编码 $z^{(k)}$。我们将这个几何与位姿编码的潜在编码 $z^{(k)}$用作聚类特征。然后，我们使用 K-Means 对这些采样任务 $\{z^{(k)}\}_{k=1}^{N_{\text{sample}}}$ 的特征进行聚类，并生成

算法 1*GeoClustering*

要求： 任务空间 $\mathbb{T}$，预训练 AutoEn- 的编码器 $\mathcal{E}$ 编码器或视觉策略中的骨干网络 $\mathcal{B}$。目标簇数量 簇 N_{clu}

1：从 $\mathbb{T}$ 中采样 N_{sample} 个任务 $\{\tau^{(k)}\}_{j=1}^{N_{\text{sample}}}$

2：提取特征：

基于状态：$\{z^{(k)}\}_{k=1}^{N_{\text{sample}}} \leftarrow \{\mathcal{E}(P_{t=0}^{(k)})\}_{k=1}^{N_{\text{sample}}}$

基于视觉：$\{f^{(k)}\}_{k=1}^{N_{\text{sample}}} \leftarrow \{\mathcal{B}(P_{t=0}^{(k)})\}_{k=1}^{N_{\text{sample}}}$

3：使用 K-Means 获取簇中心：

基于状态：$\{c_j\}_{j=1}^{N_{\text{clu}}} \leftarrow$K-Means($\{z^{(k)}\}$)

基于视觉：$\{c_j\}_{j=1}^{N_{\text{clu}}} \leftarrow$K-Means($\{f^{(k)}\}$)

4：**返回**簇中心 $\{c_j\}_{j=1}^{N_{\text{clu}}}$

N_{clu} 个簇及其对应的簇中心 $\{c_j\}_{j=1}^{N_{\text{clu}}}$：

对于基于视觉的专家，得益于训练好的视觉骨干网络，我们直接用它生成特征 $f^{(k)}$，以替代基于状态的设置中对应的编码特征 $z^{(k)}$。最后，专家的聚类可以表示为：

在专家微调期间，我们以在线方式将给定任务$\tau^{(k)}$ 分配给专家，以处理无限的任务空间。在微调期间，如果某个专家的中心 c_j 与特征 $z^{(k)}$或 $f^{(k)}$最近，我们就将 $\tau^{(k)}$ 分配给 SS_j或 VS_j。然后，每个专家只需在分配给自己的任务集上进行训练，并将学到的特定知识蒸馏给通才。

总结与讨论。***GeoClustering*** 策略解决了任务空间划分问题，使一位专家能够专注于集中的任务分布，从而提升各位专家的性能。GeoClustering 的伪代码请参见算法 3 。*GeoClustering*。

作为划分任务空间的一种方式，基于以下原因，我们的几何感知聚类比基于类别标签的划分更加合理且有效：1）并非每个对象实例都具有类别标签；2）考虑到较大的类别内几何变化，属于同一类别的物体未必会由同一位专家处理；3）物体姿态也会影响抓取，而这一因素在基于类别标签的划分中被完全忽略，但我们的方法能够很好地捕捉到它。

4.4.*GeoCurriculum*：几何感知任务课程学习

从零开始进行*GiGSL* 的问题 对于基于状态的策略学习，理论上我们可以从零开始进行*GiGSL* 。一种直接的方法是在整个任务空间上从零开始直接学习一个通才，然后按照G-S-G-S-...步骤对其进行改进。然而，使用强化学习在整个任务空间上直接学习这第一个通才将极具挑战性，通常会得到一个成功率不理想的通才。

另一种方法是先学习 N_{clu} 专家，将其蒸馏为通才，然后按照 S-G-S-G-... 的步骤进行。如何

算法 2*GeoCurriculum*

要求：任务空间 $\mathbb{T}$，**用于训练的** N_{train} **个任务** $\{\tau^{(k)}\}_{k=1}^{N_{\text{train}}} \subset \mathbb{T}$，**课程学习的** N_{level} **个层级，以及每个层级的** N_{sub}**个子簇；来自预训练自编码器的编码器** $\mathcal{E}$ **1：从编码器获取特征：** $\{z^{(k)}\}_{k=1}^{N_{\text{train}}}$**2：层级 0：找到特征空间的中心** $z_c \leftarrow$ **GeoClustering(**$N_{\text{clu}} = 1$**)，以及特征最接近** z_c**的任务** τ_c**，训练** $\mathcal{C}_0 = \{\tau_c\}$ **（其中** $\|\mathcal{C}_0\| = 1$**）。3：对于层级** l**，依次遍历 1,...,**$N_{\text{level}} - 1$**，执行 4：层级** l**：将层级** $l-1$ **的每个簇拆分为** N_{sub} **个子簇。找到特征最接近各子簇特征中心的** N_{sub} **个任务，并将这些任务添加到** $\mathcal{C}_l$**，训练**$\mathcal{C}_l$ **（其中** $\|\mathcal{C}_l\| = N_{\text{sub}}^l$**）。5：结束循环 6：层级** N_{level}**：训练** $\mathcal{C}_{N_{\text{level}}} = \{\tau^{(k)}\}_{k=1}^{N_{\text{train}}}$ **（其中** $\|\mathcal{C}_{N_{\text{level}}}\| = N_{\text{train}}$**）7：返回** $\{\mathcal{C}_l\}_{l=0}^{N_{\text{level}}}$

然而，这仍然远非最优。鉴于任务空间中的巨大变化，我们需要 $N_{\text{clu}} >> 1$，以便每个专家所面对的任务变化足够小，从而能够有效地学习。如此大量的专家会使训练成本高昂。此外，由于每个专家都是从头开始单独训练的，它们的策略可能彼此大相径庭，从而引发新的问题。考虑两个相似的任务，却被分配给不同的专家（它们恰好位于任务子空间的边界两侧）。由于这两个专家是独立训练的，因此无法保证它们会以相似的方式处理这两个相似的任务，也就是说，策略在子空间边界附近是不连续的。在策略蒸馏过程中，对于这些“边界任务”，通才可能会从不同专家那里获得差异显著的动作监督。因此，策略中的这种不连续性可能导致难以收敛，并损害策略对未见任务的泛化能力。

回顾 UniDexGrasp 中的物体课程遵循 UniDexGrasp [70]，我们考虑利用课程学习，使最初的通才学习更容易。[70]引入了一种物体课程：首先使用强化学习训练策略，使其抓取一个对象实例（该对象可以处于不同的初始姿态）；该策略充分训练后，通过纳入同一类别中的若干相似物体来增加物体数量，然后使用强化学习在新的物体集合上微调策略；接着纳入该类别中的所有物体，再次增加物体数量并微调策略；最后，将物体范围扩展到整个训练物体集中的所有不同类别，并完成最终微调。 [70] 表明，物体课程对其性能至关重要，使其基于状态的策略在训练集上的成功率从 31% 提升至 74%。

模型	训练(%)	测试(%)	
		未见物体已见类别	未见类别
PPO[55]	24.3	20.9	17.2
DAPG[49]	20.8	15.3	11.1
ILAD[69]	31.9	26.4	23.1
GSL[29]	57.3	54.1	50.9
UniDexGrasp[70]	79.4	74.3	70.8
我们的方法（基于状态）	**87.9**	**84.3**	**83.1**
PPO[55]+DAgger[51]	20.6	17.2	15.0
DAPG[49]+DAgger	17.9	15.2	13.9
ILAD[69]+DAgger	27.6	23.2	20.0
GSL[29]+DAgger	54.1	50.2	44.8
UniDexGrasp[70]	73.7	68.6	65.1
我们的方法（状态）+DAgger	77.4	72.6	68.8
我们的方法（基于视觉）	**85.4**	**79.6**	**76.7**

表 1：**训练集和测试集上评估对象的平均成功率**。为便于区分，我们用绿色表示基于状态的策略，用蓝色表示基于视觉的策略。

GeoCurriculum。中使用的对象课程有一个根本局限： [70]它没有考虑物体姿态，并且依赖类别标签。与我们在第4.3节讨论中的观点类似，我们提出利用几何特征来衡量任务之间的相似性，而非物体身份和类别标签。因此，我们引入*GeoCurriculum*，一种利用分层任务空间划分的几何感知任务课程。

具体而言，我们设计了一种 N_{level} 任务课程，将变化程度逐渐增加的任务分配给策略学习，以促进循序渐进的学习。如算法2所示，我们首先找到一个任务 $\tau^{(k_c)}$，其特征最接近所有采样任务的特征中心，并训练策略（第 0 级）。随后，对于每一级 l，我们根据几何特征 $z^{(k)}$，将前一级 $l-1$ 中的每个簇划分为 N_{sub} 个子簇（在我们的实现中为 30 个），并找到 N_{sub} 相应的中心。然后，我们将特征最接近这些子中心的任务添加到当前已分配的任务中 $\mathcal{C}_{i-1}$。最终，我们得到课程的分层任务组，即：

$$\{\mathcal{C}_l\}_{l=0}^{N_{\text{level}}} = \textit{GeoCurriculum}(\mathbb{T}) \tag{2}$$

在训练过程中，我们在每个指定的任务集上迭代训练策略。从只处理 $\mathcal{C}_0$中的一个任务，到处理 $\mathcal{C}_{N_{\text{level}}}$ 中的所有训练任务，策略逐步成长，其表现优于直接在所有任务上训练。

5. 实验

5.1. 实验设置

我们在具有挑战性的灵巧抓取基准测试 Uni DexGrasp [70]中评估了我们方法的有效性。该基准套件是近期提出的，旨在用于学习具有泛化能力的灵巧抓取。

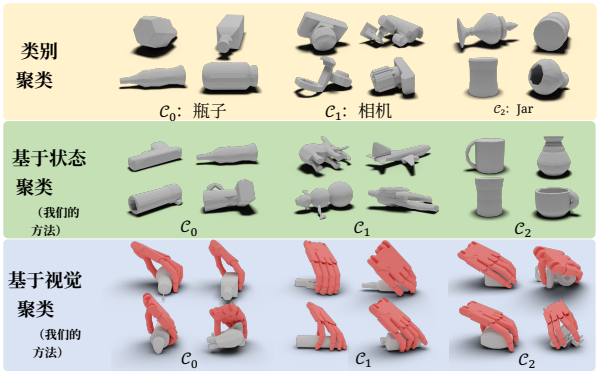


图3：**基于类别标签的聚类与我们提出的几何感知聚类之间的比较**。我们的基于状态的聚类依据预训练编码器提取的首帧点云特征进行，而基于视觉的策略则利用其视觉骨干网络提取特征进行聚类。由于基于视觉的聚类具有任务感知能力，我们还展示了第三行中灵巧手的抓取姿态。

UniDexGarsp 包含 3165 个不同的物体实例，涵盖 133 个类别。由于在 UniDexGrasp 环境中，预训练所需的真实抓取姿态生成和点云渲染过程成本极高，因此我们仅考虑 UniDexGarsp 中的非目标条件设置，该设置不指定抓取手部姿态。每个环境都会随机初始化一个对象及其初始姿态，环境由固定摄像机采集的全景三维点云 P_t 构成，用于基于视觉的策略学习。

对于网络架构，在基于状态的设置中，我们为策略网络和价值网络使用包含 4 个隐藏层（1024、1024、512、512）的多层感知机，并在基于视觉的设置中额外使用 PointNet+Transformer [40]对输入的三维场景点云进行编码。基于视觉的专家训练期间，我们冻结视觉骨干网络。实验中我们使用 $K = N_{clu} = 20$ 。其他详细超参数见补充材料。

5.2. 主要结果

我们首先在基于状态的策略学习设置中训练我们的方法，并将其与多种基线方法进行比较（表1中的绿色部分）。我们的流程采用 PPO [55]进行专家强化学习。对于这些基线方法：PPO [55] 是一种常用的强化学习方法；DAPG [50], 和 ILAD [69] 是结合强化学习进一步利用专家示范数据的模仿学习方法；GSL [29] 采用了通才-专家学习的思路：使用 PPO 进行专家学习，并使用 DAPG 整合示范数据进行通才学习，但对各个专家采用随机划分，且只进行一次策略蒸馏。UniDexGrasp [70]使用 PPO 和基于类别的物体课程学习。为将我们的方法与这些基线方法进行比较，我们将最终的基于状态的专家 $\{SS_n\}$ 蒸馏为一个基于状态的通才 SG_{n+1}

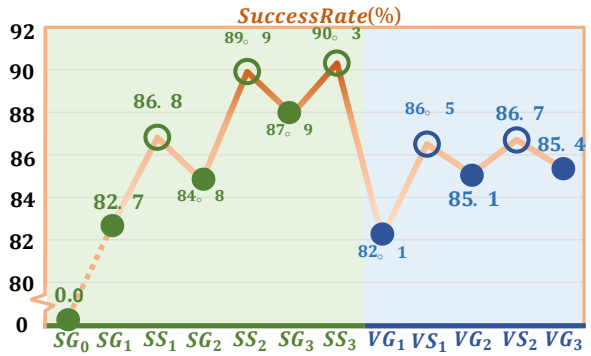


图 4：**GiGSL 训练期间的成功率。我们绘制了每个训练步骤的成功率，其中绿色代表基于状态的策略，蓝色代表基于视觉的策略，空心点代表专家策略，实心点代表通才策略。**

（尽管后续不会使用后者）。借助我们提出的技术，我们的方法在训练集和测试集上的成功率分别达到 88% 和 84%，比 **UniDexGrasp** 在基于状态的设置下**分别提升了 9% 和 11%**。

随后，我们在基于视觉的策略学习设置下，将我们的方法与基线方法（表1中的蓝色部分）进行比较。对于 PPO [55], DAPG [50], ILAD [69] 和 GSL [29], , 由于这些方法未考虑观测空间的变化（从状态到视觉），直接在视觉输入下训练会彻底失败，因此我们使用 DAgger 将基于状态的策略蒸馏为基于视觉的策略 [51] 。对于我们的方法，我们将所提出的完整流程“Ours（基于视觉）”与直接使用 DAgger 将我们的基于状态的策略蒸馏为基于视觉的策略这一变体进行比较，后者称为“Ours（state）+DAgger”。在基于视觉的设置下，我们的最终结果在训练集和测试集上分别达到 85% 和 78%，比最先进水平的基线 UniDexGrasp 分别高出**12%** 和 **11%**。

5.3. 训练过程分析

几何感知聚类有助于策略学习。我们在图中展示了一些定性结果。3。第一行展示了一种简单的聚类方式，该方式基于对象类别。但正如我们上文所分析的，这种聚类方法不包含物体几何形状信息，因此对抓取学习的帮助有限。第二行展示了我们提出的基于状态的聚类策略，该策略基于点云编码器提取的特征 $\mathcal{E}$，能够将形状相似的物体聚为一类。此外，第三行展示了我们的基于视觉的聚类策略，该策略利用包含更多任务相关信息的视觉主干网络，因此聚类得到的物体不仅形状相似，抓取姿态也相似。

我们的方法在定量性能上有所提升*GiGSL*。我们在图中展示了每个学习或微调步骤的成功率。4。无论是基于状态还是基于视觉的策略，通才-专家微调和蒸馏带来的提升都表明了其有效性——

		技术						成功率（%）		
		地理感知课程	迭代微调s	地理感知聚类	迭代微调v	端到端蒸馏	Transformer骨干网络	训练	Test 未见物体	Test 未见类别
状态-based	1							79.4	74.3	70.8
	2	✓						82.7	76.8	74.2
	3	✓	✓					84.0	77.9	74.8
	4	✓	✓	✓				**87.9**	**84.3**	**83.1**
视觉-based	5	✓	✓					77.4	72.6	68.8
	6	✓	✓			✓		78.0	72.1	69.1
	7	✓	✓		✓	✓		78.9	74.7	70.2
	8	✓	✓	✓	✓	✓		82.1	77.1	71.9
		✓	✓	✓	✓		✓	82.7	76.2	73.4
	10	✓	✓	✓		✓	✓	82.5	76.1	72.0
	11	✓	✓		✓	✓	✓	78.6	73.7	72.3
	12	✓	✓	✓	✓	✓	✓	**85.4**	**79.6**	**76.7**

表2：**消融实验**。 对于基于状态的策略（绿色）和基于视觉的策略学习（蓝色），我们将所用技术与各种消融方案进行比较。

	PPO[55]	GSL[29]	Ours
MT-10（%）	58.4±10.1	77.5±2.9	**80.3±0.5**

表 3：**Meta-World 中的补充实验**。

证明了我们的几何感知迭代通才-专才学习 ***GiGSL*** 策略设计能够提升通用灵巧抓取的最终表现。

5.4. 消融实验

消融研究结果见表2。对于**基于状态的策略学习阶段**（绿色部分），我们对消融结果分析如下。

1）（第1、2行）*GeoCurriculum*的有效性。使用我们提出的 *GeoCurriculum* （第2行）比在 [70]第1行中使用物体课程学习表现更好。

2）（第2、3行）迭代微调的有效性s。与单轮迭代相比（第2行），策略可从迭代微调过程中获益，并在训练集和测试集上取得更高的成功率（第3行）。另见图4。

3）（第3、4行）*GeoClustering* 在基于状态的设置中的有效性。借助预训练视觉特征，分配给同一专家的任务集中在相近的特征簇中，因此彼此相似。与随机将任务分配给专家相比（第3行），这显著降低了策略学习的难度，进而提升了性能（第4行）。

关于**基于视觉的策略学习阶段**（蓝色部分）的消融研究，我们默认在基于状态的策略训练中使用*GeoClustering*；本部分中*GeoClustering*旁的勾号表示是否在基于视觉的策略学习中使用它。如果“Transformer Backbone”未勾选，我们使用PointNet。 [44]

4)（第 5、6、9、12 行）端到端蒸馏的效果。我们发现，直接将最终基于状态的专家策略$\{SS_n\}$ 蒸馏到基于视觉的通才 VG_1 （第 6、12 行），效果优于先蒸馏基于状态的专家 $\{SS_n\}$ 到基于状态的通才 SG_{n+1}，再将该通才蒸馏到基于视觉的通才 VG_1 （第 5、9 行）。

5)（第6、7行和第10、12行）迭代微调的效果v。通过迭代微调过程，策略能够受益，并在训练集和测试集上取得比单阶段训练更高的成功率。另请参见图4。

6)（第7、8行和第11、12行）*GeoClustering* 在基于视觉的设置中的效果。利用通用策略视觉骨干网络提取的视觉特征来划分专家，与随机划分专家相比，最终性能显著提升（比较第11行和第12行，训练集和测试集上的提升分别为**7%** 和**5%** ）。

7)（第8、12行）Transformer骨干网络的有效性。结果表明，PointNet+Transformer骨干网络 [40] （第13行）具有更强的表达能力，相比使用PointNet [44] 骨干网络（第8行），能够提升基于DAgger的蒸馏效果。

5.5. Meta-Wo中的补充实验 rld

为进一步证明我们所提出训练策略的有效性，我们在Meta-World基准测试上开展了更多实验 [73]，，该基准测试更侧重于基于状态的多任务学习。我们采用***iGSL*** 方法处理MT10任务，表3 中的结果表明，我们的方法优于先前最先进的方法，也体现了所设计训练策略的优势：该策略支持迭代蒸馏和微调。更多细节和结果见补充材料。

6. 结论与讨论

本文提出了一种新颖的流程 UniDex- Grasp++，显著提升了 UniDexGrasp 的性能和泛化能力。我们认为，这种泛化能力对于真实机器人灵巧抓取中的仿真到现实迁移也至关重要。其局限性在于，我们仅在仿真中处理灵巧抓取任务，未来工作将扩展到真实机器人。

参考文献

[1] Pieter Abbeel 和 Andrew Y Ng. 通过逆强化学习进行学徒式学习. 载于第二十一届机器学习国际会议论文集，第 1 页，2004 年 . 3[2] Ilge Akkaya, Marcin Andrychowicz, Maciek Chociej, Ma- teusz Litwin, Bob McGrew, Arthur Petron, Alex Paino, Matthias Plappert, Glenn Powell, Raphael Ribas 等. 使用机器人手解决鲁比克魔方. arXiv 预印本 arXiv:1910.07113，2019年. 2[3] Sheldon Andrews 和 Paul G Kry. 面向目标的多指操作：控制策略与分析. Computers & Graphics，37(7):830–839，2013年. 2[4]OpenAI：Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray 等. 灵巧的手内操作学习. 《国际机器人研究期刊》，39(1):3–20，2020 年 . 2[5] Yunfei Bai 和 C. Karen Liu. 同时使用手掌和手指进行灵巧操作. 载于 2014年 IEEE 机器人与自动化国际会议（ICRA），第 1560 1565 页，2014年 2 – . [6] Michel Breyer, Jen Jen Chung, Lionel Ott, Roland Sieg- wart 和 Juan Nieto. 体积抓取网络：杂乱环境中的实时六自由度抓取检测. arXiv 预印本 arXiv:2101.01132，2021年. 1[7] Tao Chen, Megha Tippur, Siyang Wu, Vikash Kumar, Ed- ward Adelson 和 Pulkit Agrawal. 视觉灵巧性：基于深度信息的手内灵巧操作 . arXiv 预印本 arXiv:2211.11744，2022 年 . 2, 3, 4[8] Tao Chen, Jie Xu 和 Pulkit Agrawal. 一种通用手内物体重新定向系统. 机器人学习会议，2021年. 2, 3, 4[9] Sammy Christen, Muhammed Kocabas, Emre Aksan, Jemin Hwangbo, Jie Song 和 Otmar Hilliges. D-grasp：面向手物交互的物理合理动态抓取合成. 载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，2022年. 2[10] Djork-Arn´ e Clevert, Thomas Unterthiner 和 Sepp Hochre- iter. 使用指数线性单元（ELU）进行快速、准确的深度网络学习. arXiv 预印本 arXiv:1511.07289，2015年. 13[11] Karl Cobbe, Chris Hesse, Jacob Hilton 和 John Schul- man. 利用程序化生成对强化学习进行基准测试. 载于国际机器学习会议，第 2048–2056 页。PMLR，2020年. 3[12] Nikhil Chavan Dafle, Alberto Rodriguez, Robert Paolini, Bowei Tang, Siddhartha S Srinivasa, Michael Erdmann, Matthew T Mason, Ivan Lundberg, Harald Staab 和 Thomas Fuhlbrigge. 外在灵巧性：借助外力进行手内操作. 载于 2014年 IEEE 机器人与自动化国际会议（ICRA），第 1578–1585 页。IEEE，2014年. 2[13] Qiyu Dai, Yan Zhu, Yiran Geng, Ciyu Ruan, Jiazhao Zhang 和 He Wang. GraspNeRF：基于多视角的透明及镜面物体六自由度抓取检测，采用可泛化 NeRF. arXiv 预印本 arXiv:2210.06575，2022年. 1

[14] Mehmet R Dogar 和 Siddhartha S Srinivasa。使用灵巧手进行推抓：力学与一种方法。发表于2010年IEEE/RSJ智能机器人与系统国际会议，第2123–2130页。IEEE，2010年。2[15] Mehmet R. Dogar 和 Siddhartha S. Srinivasa。使用灵巧手进行推抓：力学与一种方法。发表于2010年IEEE/RSJ智能机器人与系统国际会议，第2123–2130页，2010年。2[16] Yan Duan、Xi Chen、Rein Houthooft、John Schulman 和 Pieter Abbeel。连续控制的深度强化学习基准测试。发表于国际机器学习会议，第1329–1338页。PMLR，2016年。3[17] Hongjie Fang、Hao-Shu Fang、Sheng Xu 和 Cewu Lu。Transcg：用于透明物体深度补全的大规模真实世界数据集及抓取基线。IEEE机器人与自动化快报，第1–8页，2022年。1[18] Hao-Shu Fang、Chenxi Wang、Minghao Gou 和 Cewu Lu。Graspnet-1billion：面向通用对象抓取的大规模基准测试。发表于IEEE/CVF计算机视觉与模式识别会议论文集，第11444–11453页，2020年。1[19] Justin Fu、Katie Luo 和 Sergey Levine。通过对抗逆强化学习学习鲁棒奖励。arXiv预印本 arXiv:1710.11248，2017年。3[20] Haoran Geng、Ziming Li、Yiran Geng、Jiayi Chen、Hao Dong 和 He Wang。Partmanip：从点云观测中学习跨类别可泛化部件操作策略，2023年。2[21] Haoran Geng、Helin Xu、Chengyang Zhao、Chao Xu、Li Yi、Siyuan Huang 和 He Wang。Gapartnet：通过可泛化且可操作部件实现跨类别领域泛化的对象感知与操作。arXiv预印本 arXiv:2211.05272，2022年。2[22] Yiran Geng、Boshi An、Haoran Geng、Yuanpei Chen、Yaodong Yang 和 Hao Dong。用于机器人操作的端到端可供性学习。arXiv预印本 arXiv:2209.12941，2022年。2[23] Dibya Ghosh、Avi Singh、Aravind Rajeswaran、Vikash Kumar 和 Sergey Levine。分而治之强化学习。arXiv预印本 arXiv:1711.09874，2017年。2、3、4[24] Minghao Gou、Hao-Shu Fang、Zhanda Zhu、Sheng Xu、Chenxi Wang 和 Cewu Lu。RGB很重要：在单目RGBD图像上学习7自由度抓取姿态。发表于2021年IEEE国际机器人与自动化会议（ICRA），第13459–13466页。IEEE，2021年。1[25] Tuomas Haarnoja、Aurick Zhou、Pieter Abbeel 和 Sergey Levine。软演员-评论家：具有随机演员的离策略最大熵深度强化学习。发表于国际机器学习会议，第1861–1870页。PMLR，2018年。4[26] Jonathan Ho 和 Stefano Ermon。生成对抗模仿学习。神经信息处理系统进展，第29卷，2016年。3、4、14、15[27] Wenlong Huang、Igor Mordatch、Pieter Abbeel 和 Deepak Pathak。通过几何感知多任务学习实现灵巧操作的泛化。arXiv预印本 arXiv:2111.03062，2021年。2

[28] Stephen James、Zicong Ma、David Rovick Arrojo 和 Andrew J Davison。Rlbench：机器人学习基准与学习环境。IEEE机器人与自动化快报，5(2):3019–3026，2020年。3
[29] Zhiwei Jia、Xuanlin Li、Zhan Ling、Shuang Liu、Yiran Wu 和 Hao Su。通过通才-专家学习改进策略优化。发表于国际机器学习会议，第10104–10119页。PMLR，2022年。2、3、4、5、6、7、8、12、15[30] Dmitry Kalashnikov、Alex Irpan、Peter Pastor、Julian Ibarz、Alexander Herzog、Eric Jang、Deirdre Quillen、Ethan Holly、Mrinal Kalakrishnan、Vincent Vanhoucke 等。用于基于视觉的机器人操作的可扩展深度强化学习。发表于机器人学习会议，第651–673页。PMLR，2018年。2[31] Michael Kelly、Chelsea Sidrane、Katherine Driggs-Campbell 和 Mykel J Kochenderfer。Hg-dagger：与人类专家进行的交互式模仿学习。发表于2019年国际机器人与自动化会议（ICRA），第8077–8083页。IEEE，2019年。3[32] Vikash Kumar、Abhishek Gupta、Emanuel Todorov 和 Sergey Levine。从经验和模仿中学习灵巧操作策略。arXiv预印本 arXiv:1611.05095，2016年。2[33] Vikash Kumar、Emanuel Todorov 和 Sergey Levine。基于学习得到的局部模型的最优控制：应用于灵巧操作。发表于2016年IEEE国际机器人与自动化会议（ICRA），第378–383页。IEEE，2016年。2[34] Fangchen Liu、Zhan Ling、Tongzhou Mu 和 Hao Su。基于状态对齐的模仿学习。arXiv预印本 arXiv:1911.10947，2019年。3[35] Yunze Liu、Yun Liu、Che Jiang、Kangbo Lyu、Weikang Wan、Hao Shen、Boqiang Liang、Zhoujie Fu、He Wang 和 Li Yi。Hoi4d：用于类别级人-物交互的四维第一人称视角数据集。发表于IEEE/CVF计算机视觉与模式识别会议论文集，第21013–21022页，2022年。2[36] Qingkai Lu、Kautilya Chenna、Balakumar Sundaralingam 和 Tucker Hermans。将多指抓取规划为学习得到的深度网络中的概率推断。发表于《机器人研究：第18届国际研讨会ISRR》，第455–472页。Springer，2020年。2[37] Priyanka Mandikal 和 Kristen Grauman。利用来自视频的人手姿态先验学习灵巧抓取。发表于机器人学习会议（CoRL），2021年。2[38] Priyanka Mandikal 和 Kristen Grauman。利用以物体为中心的视觉可供性学习灵巧抓取。发表于IEEE国际机器人与自动化会议（ICRA），2021年。2[39] Tongzhou Mu、Jiayuan Gu、Zhiwei Jia、Hao Tang 和 Hao Su。利用自监督物体候选区域进行面向组合泛化能力的策略重构。神经信息处理系统进展，33:8883–8894，2020年。2、3、4[40] Tongzhou Mu、Zhan Ling、Fanbo Xiang、Derek Yang、Xuanlin Li、Stone Tao、Zhiao Huang、Zhiwei Jia 和 Hao Su. Maniskill：基于大规模示范数据的通用操作技能基准。arXiv 预印本 arXiv:2107.14483，2021年。2, 3, 4, 7, 8, 14[41] Anusha Nagabandi、Kurt Konolige、Sergey Levine 和 Vikash Kumar。用于学习灵巧操作的深度动力学模型。发表于机器人学习会议，第 1101–1112 页。PMLR，2020年。2[42] Andrew Y Ng、Stuart Russell 等。逆强化学习算法。发表于国际机器学习大会，第 1 卷，第 2 页，2000年。3[43] Allison M Okamura、Niels Smaby 和 Mark R Cutkosky。灵巧操作概述。发表于《2000年 ICRA 会议论文集：千年会议。IEEE 国际机器人与自动化会议。研讨会论文集》（目录号 00CH37065），第 1 卷，第 255–262 页。IEEE，2000年。2[44] Charles R Qi、Hao Su、Kaichun Mo 和 Leonidas J Guibas。PointNet：用于三维分类与分割的点集深度学习。arXiv 预印本 arXiv:1612.00593，2016年。4, 8, 14[45] Haozhi Qi、Ashish Kumar、Roberto Calandra、Yi Ma 和 Jitendra Malik。通过快速运动适应实现手内物体旋转。arXiv 预印本 arXiv:2210.04887，2022 年。2[46] Yuzhe Qin 、Binghao Huang、Zhao-Heng Yin、Hao Su 和 Xiaolong Wang。DexPoint：用于仿真到现实灵巧操作的通用点云强化学习。arXiv 预印本 arXiv:2211.09423，2022年。2, 3[47] Ilija Radosavovic、Xiaolong Wang、Lerrel Pinto 和 Jitendra Malik。用于灵巧操作的仅基于状态的模仿学习。发表于2021年 IEEE/RSJ 智能机器人与系统国际会议（IROS），第 7865–7871 页。IEEE，2021年。3[48] Ilija Radosavovic、Tete Xiao、Stephen James、Pieter Abbeel、Jitendra Malik 和 Trevor Darrell。通过掩码视觉预训练实现真实世界机器人学习。arXiv 预印本 arXiv:2210.03109，2022年。2[49] Aravind Rajeswaran、Vikash Kumar、Abhishek Gupta、Giulia Vezzani、John Schulman、Emanuel Todorov 和 Sergey Levine。通过深度强化学习和示范数据学习复杂的灵巧操作。arXiv 预印本 arXiv:1709.10087，2017年。2, 3, 4, 6, 12[50] Aravind Rajeswaran、Vikash Kumar、Abhishek Gupta、Giulia Vezzani、John Schulman、Emanuel Todorov 和 Sergey Levine。通过深度强化学习和示范数据学习复杂的灵巧操作。arXiv 预印本 arXiv:1709.10087，2017年。7[51] St´ephane Ross、Geoffrey Gordon 和 Drew Bagnell。将模仿学习和结构化预测归约为无遗憾在线学习。发表于第十四届人工智能与统计国际会议论文集，第627–635 页。JMLR 研讨会与会议论文集，2011年。2, 4, 6, 7, 12[52] Daniela Rus。分段光滑三维物体的手内灵巧操作。The International Journal of Robotics Research，18(4):355–381，1999年。2[53] J Kenneth Salisbury 和 John J Craig。关节式机械手：力控制与运动学问题。The International Journal of Robotics Research，1(1):4–17，1982年。2

[54] John Schulman、Philipp Moritz、Sergey Levine、Michael Jordan 和 Pieter Abbeel。使用广义优势估计进行高维连续控制。arXiv 预印本 arXiv:1506.02438，2015年。4[55] John Schulman、Filip Wolski、Prafulla Dhariwal、Alec Radford 和 Oleg Klimov。近端策略优化算法。arXiv 预印本 arXiv:1707.06347，2017年。4, 6, 7, 8, 12, 15[56] Younggyo Seo、Danijar Hafner、Hao Liu、Fangchen Liu、Stephen James、Kimin Lee 和 Pieter Abbeel。用于视觉控制的掩码世界模型。arXiv 预印本 arXiv:2206.14244，2022年。2[57] Younggyo Seo、Kimin Lee、Stephen L James 和 Pieter Abbeel。利用视频进行无动作预训练的强化学习。发表于国际机器学习会议，第 19561–19579 页。PMLR，2022年。2[58] Qijin She、Ruizhen Hu、Juzhan Xu、Min Liu、Kai Xu 和 Hui Huang。通过夹爪与物体交互的动态表征学习高自由度触及与抓取。arXiv 预印本 arXiv:2204.13998，2022年。2[59] Hao Shen、Weikang Wan 和 He Wang。通过基于示范数据的生成对抗式自我模仿学习类别级通用物体操作策略。arXiv 预印本 arXiv:2203.02107，2022年。2, 3[60] Aravind Srinivas、Michael Laskin 和 Pieter Abbeel。CURL：用于强化学习的对比式无监督表征。arXiv 预印本 arXiv:2004.04136，2020年。2[61]Adam Stooke、Kimin Lee、Pieter Abbeel 和 Michael Laskin。将表征学习与强化学习解耦。发表于国际机器学习会议，第 9870–9879 页。PMLR，2021年。2[62] Martin Sundermeyer、Arsalan Mousavian、Rudolph Triebel 和 Dieter Fox。Contact-GraspNet：在杂乱场景中高效生成六自由度抓取姿态。发表于2021年 IEEE 国际机器人与自动化会议（ICRA），第 13438–13444 页。IEEE，2021年。1[63] Yee Teh、Victor Bapst、Wojciech M Czarnecki、John Quan、James Kirkpatrick、Raia Hadsell、Nicolas Heess 和 Razvan Pascanu。Distral：稳健的多任务强化学习。《神经信息处理系统进展》，第 30 卷，2017年。2, 3, 4. [64] Faraz Torabi、Garrett Warnell 和 Peter Stone。基于观测的行为克隆。arXiv 预印本 arXiv:1805 01954，2018年 3 . . [65] Yusuke Urakami、Alec Hodgkinson、Casey Carlin、Randall Leu、Luca Rigazio 和 Pieter Abbeel。DoorGym：可扩展的开门环境与基线智能体。arXiv 预印本 arXiv:1908.01887，2019 年。3[66] Ashish Vaswani 、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。注意力就是一切。《神经信息处理系统进展》，第 30 卷，2017年。14[67] Chenxi Wang、Hao-Shu Fang、Minghao Gou、Hongjie Fang、Jin Gao 和 Cewu Lu。杂乱环境中的抓取性发现：实现快速、准确的抓取检测。发表于 IEEE/CVF 国际计算机视觉会议（ICCV）论文集，第 15964–15973 页，2021年10月。1[68] Ruihai Wu、Yan Zhao、Kaichun Mo、Zizheng Guo、Yian Wang、Tianhao Wu、Qingnan Fan、Xuelin Chen、Leonidas Guibas 和 Hao Dong。VAT-mart：学习用于操作 3D 铰接物体的视觉动作轨迹提议。发表于国际学习表征会议，2022年。2[69] Yueh-Hua Wu、Jiashun Wang 和 Xiaolong Wang。从人类抓取可供性中学习可泛化的灵巧操作。arXiv 预印本 arXiv:2204.02320，2022年。2、3、4、6、7、12[70] Yinzhen Xu、Weikang Wan、Jialiang Zhang、Haoran Liu、Zikang Shan、Hao Shen、Ruicheng Wang、Haoran Geng、Yijia Weng、Jiayi Chen、Tengyu Liu、Li Yi 和 He Wang。UniDexGrasp：通过学习多样化的提议生成和目标条件策略，实现通用机器人灵巧抓取，2023年。2、3、4、6、7、8、12、13、14、15[71] Zhenjia Xu、Zhanpeng He 和 Shuran Song。面向铰接物体的通用操作策略网络。IEEE机器人与自动化快报，7(2)：2447–2454，2022年。1[72] Denis Yarats、Amy Zhang、Ilya Kostrikov、Brandon Amos、Joelle Pineau 和 Rob Fergus。提升基于图像的无模型强化学习的样本效率。发表于AAAI人工智能大会论文集，第35卷，第10674–10681页，2021年。2[73] Tianhe Yu、Deirdre Quillen、Zhanpeng He、Ryan Julian、Karol Hausman、Chelsea Finn 和 Sergey Levine。Meta-world：多任务与元强化学习的基准测试和评估。发表于机器人学习会议，第1094–1100页。PMLR，2020年。2、3、8、15[74] Fangyi Zhang、J¨urgen Leitner、Michael Milford、Ben Upcroft 和 Peter Corke。面向用于机器人运动控制的基于视觉的深度强化学习。arXiv 预印本 arXiv:1511.03791，2015年。2

A. 方法与实现细节

我们在算法 3中形式化了完整的流程方法。

A.1. 方法细节

算法 3 UniDexGrasp++

输入：任务空间 $\mathbb{T}$，K 基于状态的专家 $\{SS^j\}$，一个基于状态的通才 SG，K 基于视觉的专家 $\{VS^j\}$，一个基于视觉的通才 VG

$\{\mathcal{C}_l\} \leftarrow$ ***GeoCurriculum***$(\mathbb{T})$，用于物体课程。
几何感知任务课程学习，用于训练 SG_0。
用于 $i = 1,\ 2,\ \ldots\ldots$**执行**:
  初始化专家 $SS_i^j = SG_i$
  $\{c_j\} \leftarrow$***GeoClustering***$(\mathbb{T})$
  将最接近 c_j 的任务在线分配给专家 SS_i^j 并训练 $SS^j{}_i$ ▷ 强化学习
  如果 $\{SS_i^j\}$ 达到最优 **则中断**
  else
    蒸馏 $\{SS_i^j\}$ 至 SG_{i+1} ▷ 基于DAgger的蒸馏
  结束 if
结束 for
蒸馏 $\{SS_i^j\}$ 至 VG_0 ▷ 基于DAgger的蒸馏
对于 $i = 1、2、\ldots\ldots$**执行**:
  初始化专家 $VS_i^j = VG_i$
  $\{c_j\} \leftarrow$***GeoClustering***$(\mathbb{T})$
  将距离 c_j 最近的任务在线分配给专家 VS_i^j 并训练 $VS^j{}_i$ ▷ 强化学习
  蒸馏 $\{VS_i^j\}_{i=1}^K$ 到 VG_{i+1} ▷ 基于DAgger的蒸馏
  如果 VG_{i+1} 是最优的 **则中断**
  结束如果
结束循环

关于*GiGSL*的详细信息：在基于状态的策略学习阶段，当当前策略的成功率 SS_n 仅略高于前一策略 SS_{n-1}（不到0.5%）时，我们便终止训练。此时，我们将其蒸馏 SS_n 到基于视觉的策略。在基于视觉的策略学习阶段，当当前策略的成功率 VG_n 仅略高于前一策略 VG_{n-1}（不到0.5%）时，我们便停止训练。随后，我们将 VG_n 用作最终策略。

关于自编码器：我们使用样本任务 $\{\tau^{(k)}\}_{k=1}^{N\text{sample}}$ 中初始化物体的点云 $\{P_{t=0}^{(k)}\}_{k=1}^{N\text{sample}}$ 来训练点云三维自编码器。该自编码器采用编码器-解码器结构。编码器 $\mathcal{E}$ 对 $P_{t=0}^{(k)}$ 进行编码，并输出编码后的潜在特征 $z^{(k)} = \mathcal{E}(P_{t=0}^{(k)})$。解码器 $\mathcal{D}$ 以 $z^{(k)}$ 为输入，并生成点云 $\hat{P}_{t=0}^{(k)}$。

$$z^{(k)} = \mathcal{E}(P_{t=0}^{(k)}) \tag{3}$$

$$\hat{P}_{t=0}^{(k)} = \mathcal{D}(z^{(k)}) \tag{4}$$

模型使用重建损失 $\mathcal{L}_{\text{AE}}$进行训练，该损失是 $P_{t=0}^{(k)}$ 与 $\hat{P}^{(k)}_{t=0}$之间的倒角距离。

$$\mathcal{L}_{\text{AE}} = \text{ChamferDistance}(P_{t=0}^{(k)}, \hat{P}_{t=0}^{(k)}) \tag{5}$$

关于*GeoCurriculum*的详情：在我们的实现中，我们选择 $N_{\text{level}} = 4$ 并使用四阶段 *GeoCurriculum* 进行训练，其中任务数量为 1-300-900-N_{traino}。我们还比较了不同的 N_{level}，结果见第 C

A.2. 关于基线方法的详情

PPO 近端策略优化（PPO） [55]是一种广泛使用的无模型在策略强化学习方法。我们采用 PPO 作为强化学习基线。

DAPG 示范增强策略梯度（DAPG） [49]是一种广泛使用的模仿学习（IL）方法，通过利用专家示范来降低样本复杂度。遵循 ILAD [69], 的方法，我们使用运动规划生成示范数据。

ILAD ILAD [69] 是一种能够提升 DAPG 泛化能力的模仿学习方法。它在 DAPG 的基础上引入了一种新颖的模仿学习目标，在策略学习过程中通过对生成的示范数据进行行为克隆，共同学习对象的几何表示。此方法使用与 DAPG 相同的生成示范数据。

GSL 通才-专家学习（GSL） [29] 是一种分三个阶段的学习方法：首先，使用强化学习在所有环境变化上训练一个通才；然后，将该通才的权重克隆到大量专家，并对其进行微调，每个专家都使用强化学习训练，以掌握选定的一小部分变化。最后，GSL 使用这些专家收集示范数据，并在模仿学习部分采用 DAPG 来训练通才。为进行公平比较，我们在实现中对强化学习部分采用 PPO [55]，对模仿学习部分采用 DAPG [49]。

UniDexGrasp UniDexGrasp [70]是一种两阶段学习方法。在第一阶段的基于状态的训练中，他们提出了对象课程学习（OCL），从单个物体开始进行强化学习，并逐步将同一类别或相似类别中的相似物体纳入训练，以获得基于状态的教师策略。获得该教师策略后，他们使用 DAgger [51] 将其蒸馏为基于视觉的策略。

B. 实验细节

如第 4 节所述，我们在几何感知课程 学习阶段使用 PPO [55]，以获得第一个通才 SG_1，并

参数	描述
$\boldsymbol{q} \in \mathbb{R}^{18}$	关节位置
$\dot{\boldsymbol{q}} \in \mathbb{R}^{18}$	关节速度
$\boldsymbol{\tau}_{\text{dof}} \in \mathbb{R}^{24}$	自由度力
$x_{\text{finger}} \in \mathbb{R}^{3\times5}$	指尖位置
$\alpha_{\text{finger}} \in \mathbb{R}^{4\times5}$	指尖方向
$\dot{x}_{\text{finger}} \in \mathbb{R}^{3\times5}$	指尖线速度
$\omega_{\text{finger}} \in \mathbb{R}^{3\times5}$	指尖角速度
$F_{\text{finger}} \in \mathbb{R}^{3\times5}$	指尖力
$\tau_{\text{finger}} \in \mathbb{R}^{3\times5}$	指尖扭矩
$\boldsymbol{t} \in \mathbb{R}^{3}$	手部根节点全局平移
$R \in \mathbb{R}^{3\times3}$	手部根节点全局方向
$\boldsymbol{a} \in \mathbb{R}^{24}$	动作

表4：机器人状态定义。

在专家学习阶段 $\{\text{SS}_i\}$, $\{\text{VS}_i\}$ 训练这些专家。在通才学习阶段 $SG_i(i > 1)$ 和 VG_i，我们采用所提出的基于 DAgger 的策略蒸馏。请注意，在$\{\text{VS}_i\}$ 学习阶段，我们会冻结视觉骨干网络。

B.1. 环境设置

状态定义 基于状态的策略的完整状态记为 $\mathcal{S}_t^{\mathcal{S}} = (R_t, O_t, P_{t=0})$，而基于视觉的策略的完整状态表示为 $\mathcal{S}_t^{\mathcal{V}} = (R_t, P_t)$。机器人状态 R_r 详见表 4，物体先验状态 O_t 包括物体姿态（位置有 3 个自由度，旋转矩阵有 9 个自由度）、线速度和角速度。为加快训练过程，我们仅从场景点云中对对象和手采样 1024 个点 P_t。

动作空间 动作空间是灵巧手上 24 个执行器的电机指令。前 6 个电机控制灵巧手的全局位置和方向，其余 18 个电机控制手指。我们根据执行器规格将动作范围归一化为 $(-1,1)$。

相机设置与 UniDexGrasp 类似， [70], 我们采用由五个 RGBD 相机构成的设置，这些相机按照图5所示，布置在桌子周围和上方。系统原点位于桌子中心，相机的位置为 ([0.5, 0, 0.05], [-0.5, 0, 0.05], [0, 0.5, 0.05], [0, -0.05, 0.05], [0, 0, 0.55])，其焦点设为[0, 0, 0。05]我们将五个相机生成的局部点云融合为一个场景点云 P_t。

奖励函数： 我们使用 UniDexGrasp 中非目标条件化的奖励版本 [70], 并将其形式化为

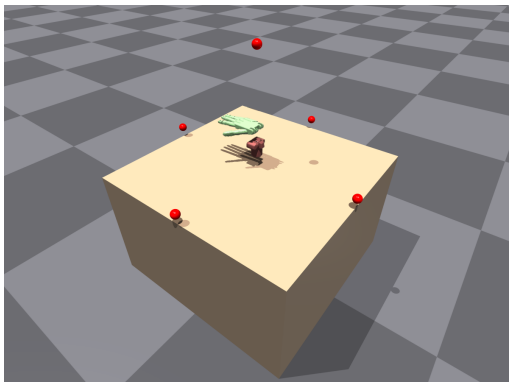


图5：**相机位置**

如下（此处所有 ω_* 均为与 UniDexGrasp 相同的超参数）：

到达奖励 r_{reach} 鼓励机器人手指伸向物体。这里，$\mathbf{x}^{\text{finger}}$ 和$\mathbf{x}^{\text{obj}}$ 分别表示每根手指和物体的位置：

$$r_{\text{reach}} = -\omega_r \sum \|\mathbf{x}_{\text{finger}} - \mathbf{x}_{\text{obj}}\|_2 \tag{6}$$

抬升奖励 r_{lift} 鼓励机器人手在手指足够靠近物体时抬起物体。f 是用于判断机器人是否达到抬升条件的标志：$f = \mathbf{Is}(\sum \|\mathbf{x}^{\text{finger}} - \mathbf{x}_{\text{obj}}\|_2 < \lambda_{f_1}) + \mathbf{Is}(d_{\text{obj}} > \lambda_0)$。这里，$d_{\text{obj}} = \|\mathbf{x}^{\text{obj}} - \mathbf{x}_{\text{target}}\|_2$，其中$\mathbf{x}^{\text{obj}}$ 和$\mathbf{x}^{\text{target}}$ 分别是物体位置和目标位置。 a_z 是沿z轴施加于手部根部的缩放力（$\omega_l > 0$）。

$$r_{\text{lift}} = \begin{cases} \omega_l * (1 + a_z) & \text{if } f = 2 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

移动奖励 r_{move} 鼓励对象伸向目标；当对象被抬起并非常接近目标时，还会给予奖励项：

$$r_{\text{move}} = \begin{cases} -\omega_m d_{\text{obj}} + \frac{1}{1+\omega_b d_{obj}} & \text{if } d_{\text{obj}} < \lambda_0 \\ -\omega_m d_{\text{obj}} & \text{otherwise} \end{cases} \tag{8}$$

最后，我们将各个组成部分相加，并将奖励函数表示如下：

$$r = r_{\text{reach}} + r_{\text{lift}} + r_{\text{move}} \tag{9}$$

B.2. 训练细节

网络架构： 状态策略 $\pi_{\mathcal{E}}$ 和基于视觉的策略 $\pi_{\mathcal{S}}$ 使用的多层感知机由 4 个隐藏层（1024、1024、512、512）组成。我们使用指数线性单元（ELU） [10] 作为激活函数。

超参数	值
环境数量（Isaac Gym，基于状态）	1024
环境数量（Isaac Gym，基于视觉）	32
环境间距（Isaac Gym）	1.5
每次策略更新的采样步数（PPO）	8
每次策略更新的采样步数（DAgger）	1
每个智能体的批次数	4
学习轮次数	5
缓冲区大小（DAgger）	2000
回合长度	200
策略迭代的饱和阈值	0.005
折扣因子	0.96
GAE 参数	0.95
熵系数	0.0
PPO 裁剪范围	0.2
学习率	0.0003
价值损失系数	1.0
最大梯度范数	1.0
初始噪声标准差	0.8
期望 KL 散度	0.16
裁剪观测值	5.0
裁剪动作	1.0
N	270,000
$N^{\text{采样}}$	3200
$N_{\text{clu}}^{\text{训练}}$	20
ω_r	0.5
ω_l	0.1
ω_m	2
ω_b	10

表5： **抓取策略的超参数**。

自编码器中 PointNet 的网络结构为（1024、512、64）。我们使用 PointNet + Transformer 骨干网络 [40]作为视觉骨干网络，并使用不同的 PointNet [44]处理具有不同分割掩码的点（机器人、对象、整个点云）。此外，还有一个额外的多层感知机，仅为机器人状态输出一个 256 维的隐藏向量。多层感知机和 PointNet 提取的所有特征都会输入 Transformer [66]。输出向量经过全局注意力池化，提取出一个 256 维的表示，然后输入一个层大小为 [256, 128, 特征维度] 的最终多层感知机，以输出视觉特征，随后将其与机器人状态拼接。

训练超参数： 我们实验中的超参数列于表5。

训练时间： 实验使用四块 NVIDIA RTX 3090 Ti 完成。训练过程包括 GeoCurriculum 第一阶段的 20,000 个环境步数，以及其他阶段的 15,000 个环境步数（针对每个单独的策略）。总共需要两天。

C. 补充结果与分析

本节包含第 5 节所述实验的扩展结果。

关于 *GeoCurriculum* 的更多消融实验。 我们进一步对 *GeoCurriculum* 的有效性进行了消融实验，结果见表6。具体而言，我们将提出的 *GeoCurriculum* 方法与不使用任何课程学习的方法，以及文献 [70]中提出的物体课程学习（OCL）方法进行了比较。我们的结果表明，在物体实例及其初始姿态存在较大差异的复杂灵巧抓取任务中，课程学习对于取得成功至关重要。此外，我们发现，我们的*GeoCurriculum* 方法考虑了不同物体和姿态之间的几何相似性，其表现优于仅考虑物体类别标签的 OCL 方法。此外，我们还对课程学习阶段的数量进行了消融实验。对于 3 阶段 *GeoCurriculum*，任务数量为 1-100-N_{train}；对于 4 阶段 *GeoCurriculum*，任务数量为 1-300-900-N_{train}；对于 5 阶段 *GeoCurriculum*，任务数量为 1-20-100-1000-N_{train}。我们比较了所有实验中 SG_1 的表现。由于 5 阶段 *GeoCurriculum* 的表现与 4 阶段 *GeoCurriculum* 相近，因此为简化起见，我们在主要实验中选择了 4 阶段方案。

模型	训练(%)	测试(%)	
		未见物体已见类别	未见类别
无课程学习	30.5	23.4	20.6
OCL[70]	79.4	74.3	70.8
GeoCurriculum (3)	81.3	75.6	73.3
GeoCurriculum (4)	82.7	**76.8**	**74.2**
GeoCurriculum (5)	**82.9**	76.4	74.0

表6：**关于*GeoCurriculum*的消融实验。** OCL 指的是文献 [70]中提出的对象课程学习。括号中的数字表示课程学习的阶段数。

关于 *iGSL* 的更多消融实验。 对于迭代通用-专用学习（*iGSL*）中使用的策略蒸馏方法，我们将基于 DAgger 的策略蒸馏与几种流行的模仿学习方法进行比较，包括行为克隆（我们还加入了价值函数学习，使整个过程能够迭代进行）、GAIL [26] 和 DAPG [70]。所有方法均采用 *GeoCurriculum*，并比较 SG_{n+1}的性能。表7展示了结果，表明我们基于 DAgger 的策略蒸馏方法显著优于其他方法。

值得注意的是，我们的方法使用教师检查点，而其他方法仅使用教师提供的示范数据。

模型	训练(%)	测试(%)	
		未见物体 已见类别	未见类别
BC + 价值	12.4	8.6	8.4
GAIL[26]	30.7	26.9	26.0
DAPG[70]	61.4	52.6	47.9
Ours	**87.9**	**84.3**	**83.1**

表 7：**策略蒸馏方法的消融实验**。

关于 *GiGSL* 的更多消融实验结果。首先，我们针对 *GiGSL* 中 GeoClustering 的簇数量 N_{clu} 进行了消融实验。我们比较了不同 VG_n 下最终基于视觉的策略 N_{clu}的性能。结果见表8 ，结果表明，增大 N_{clu} 超过某个阈值后，性能不会提升，甚至可能下降。

模型	训练(%)	测试(%)	
		未见物体 已见类别	未见类别
0（无专家）	77.4	72.6	68.8
10	80.3	74.9	75.2
20	**85.4**	**79.6**	**76.7**
50	77.2	71.2	69.9

表 8：消融实验**簇数量**。

接下来，我们将我们的 *GeoClustering* 与随机聚类和基于类别标签的聚类进行比较（为公平起见，我们将所有类别平均划分为 N_{clu} 个部分）。在基于类别类别的聚类中，我们使用所有物体及其初始姿态预训练一个分类任务。随后，我们使用倒数第二层的特征进行聚类，并在基于状态的策略学习中将该特征与机器人状态和物体状态拼接起来。我们比较了不同方法下最终的基于视觉的策略 VG_n 的性能。结果见表9。

Meta-World 上的更多结果这里展示了 Meta-World 上的更多结果 [73], 这一广受欢迎的多任务策略学习基准测试。MT-10 任务包含 10 个多样且具有挑战性的任务，例如开门或拾取物体，这些任务需要广泛的技能和能力。MT-50 任务集是 MT-10 任务集的扩展，另外包含 50 个更加复杂多样的任务。我们在表10 中的结果表明，我们的

模型	训练（%）	测试（%）	
		未见物体 已见类别	未见类别
随机	77.0	71.9	68.2
类别标签。	79.7	73.9	74.1
Ours	**85.4**	**79.6**	**76.7**

表 9：**关于预训练自编码器的消融实验**。编码器提取的特征在基于状态的设置中用于*GeoClustering* 。

所提出的方法 ***iGSL*** 在多任务 MT-10 和 MT-50 上表现良好，并优于基线方法。

	PPO[55]	GSL[29]	Ours
MT-10（%）	58.4±10.1	77.5±2.9	**80.3±0.5**
MT-50（%）	31.1±4.5	43.5±2.2	**45.9±1.7**

表 10：**Meta-World 中的附加实验**。

更多定性抓取结果 我们在图中展示了更多定性结果。6 和图7。在图6中，我们提供了更多关于我们方法的结果 *GeoClustering* 在基于视觉的策略学习阶段。基于视觉的策略 VG_1 利用其视觉骨干网络提取任务的视觉特征以进行聚类。由于基于视觉的聚类具有任务感知能力，我们还展示了这些任务的抓取姿态。图6 中的结果表明，我们的方法能够根据对象几何形状、姿态特征以及通才策略相应的抓取策略对任务进行聚类。在图7中，我们展示了不同对象在不同初始姿态下的若干抓取轨迹。

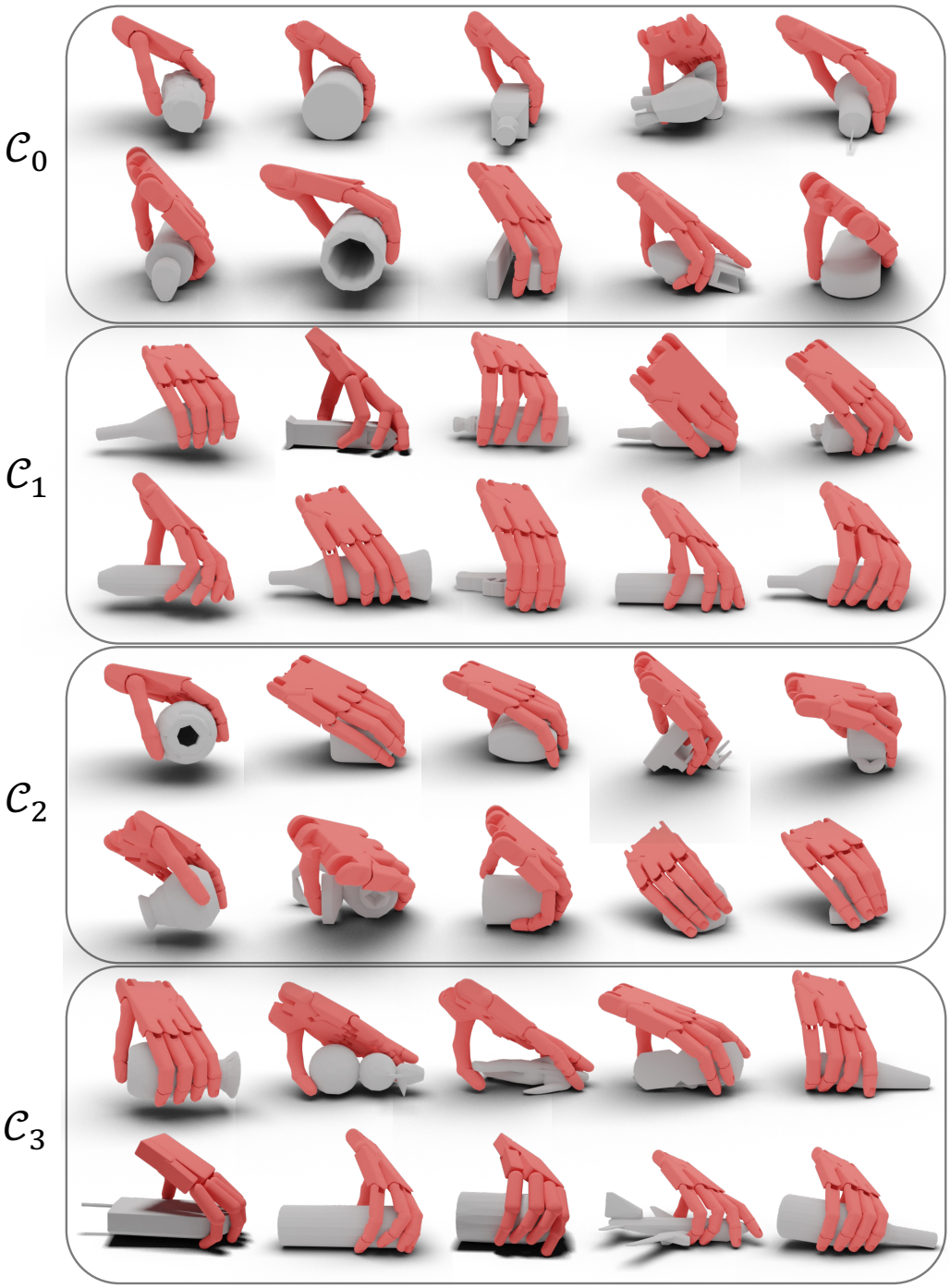


图 6：**定性抓取结果**。对于 4 个簇中的每一个，我们展示 10 个任务及策略对应的抓取姿态。这些簇由我们在基于视觉的策略学习阶段使用*GeoClustering* 生成。

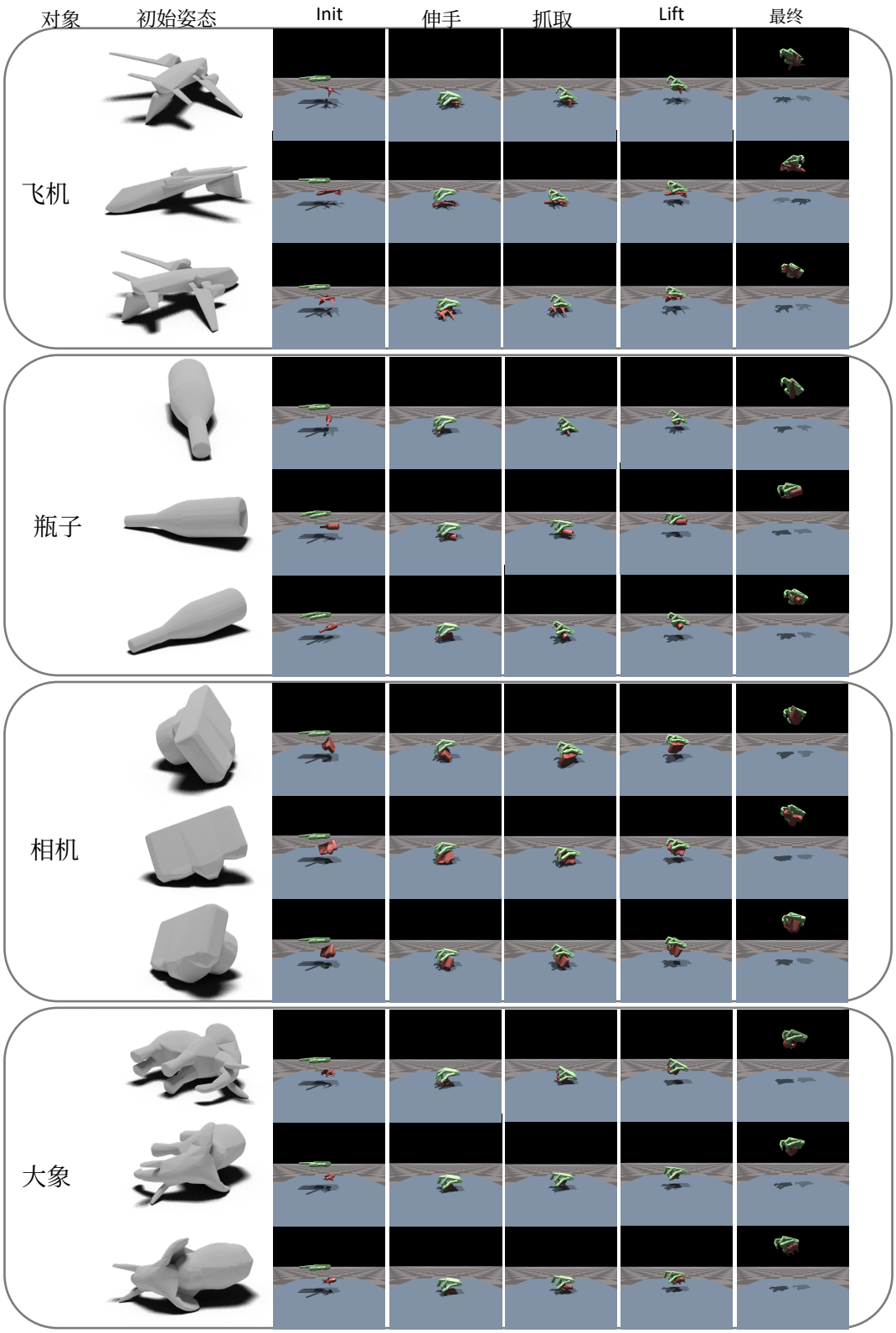


图7：**定性抓取轨迹**。我们展示了多个物体在不同初始姿态下的若干抓取轨迹。