

文章编号:1007-2780(2023)11-1542-12

面向多姿态点云目标的在线类增量学习

张润江^{1,2}, 郭杰龙^{2,3}, 俞辉^{2,3}, 兰海², 王希豪², 魏宪^{2,3*}

(1. 福州大学 电气工程与自动化学院, 福建 福州 350108;

2. 中国科学院 福建物质结构研究所, 福建 福州 350002;

3. 中国科学院 海西研究院 泉州装备制造研究中心, 福建 泉州 362000)

摘要:针对目前增量学习中所面向目标都是固定姿态这一现象,本文考虑了更严格的设定,即面向多姿态目标的在线类增量学习,并提出了无视姿态重放方法来缓解在线类增量学习中面对多姿态目标时的灾难性遗忘。首先,将2D/3D目标进行点云化处理,以方便提取目标的有效几何信息;其次,基于 $SE(d)$ ($d=2,3$)群对网络进行平移旋转等变性改进,使网络能够提取更丰富的几何信息,从而降低模型在每个任务中受目标姿态的影响;最后,根据损失变化采样特定样本用于重放来缓解灾难性遗忘。实验结果表明,在面对固定姿态目标MNIST、CIFAR-10时,本文方法的最终平均精度分别达到了88%和42.6%,与对比方法结果相近,但最终平均遗忘率明显优于对比方法,分别降低了约3%和15%。在面对多姿态目标RotMNIST、trCIFAR-10时,本文方法依旧能很好地保持在固定姿态目标中的表现,基本不受目标姿态的影响。此外,在3D数据集ModelNet40中的表现也依旧稳定。本文所提方法在在线类增量学习中能够不受目标姿态的影响,同时能缓解灾难性遗忘,具有很好的稳定性和可塑性。

关键词:在线类增量学习;灾难性遗忘;无视姿态重放;等变性;点云分类

中图分类号:TP391.4 文献标识码:A doi:10.37188/CJLCD.2022-0419

Online class incremental learning for multi-pose point cloud targets

ZHANG Run-jiang^{1,2}, GUO Jie-long^{2,3}, YU Hui^{2,3}, LAN Hai², WANG Xi-hao², WEI Xian^{2,3*}

(1. College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China;

2. Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences,
Fuzhou 350108, China;

3. Quanzhou Institute of Equipment Manufacturing, Haixi Institutes, Chinese Academy of Sciences,
Quanzhou 362000, China)

Abstract: In response to current phenomenon that all targets in incremental learning are fixed pose, this paper considers a more rigorous setting, *i. e.* online class incremental learning for multi-pose targets, which innovatively proposes an ignoring pose replay method to alleviate the catastrophic forgetting in facing multi-pose targets in online class incremental learning. Firstly, 2D/3D targets are point-clouded to facilitate the extraction of useful geometric information. Secondly, the network modifies for equivariance

收稿日期:2022-12-21;修订日期:2023-01-30.

基金项目:福建省科技计划(No. 2021T3003);泉州市科技计划(No. 2021C065L)

Supported by Fujian Science and Technology Plan (No. 2021T3003); Quanzhou Science and Technology Plan (No. 2021C065L)

*通信联系人, E-mail: xian.wei@fjirsm.ac.cn

based on the $SE(d)$ ($d = 2, 3$) group to enable the network to extract richer geometric information, thus reducing the impact of target poses on the model in each task. Finally, specific samples are sampled for replay to mitigate catastrophic forgetting based on loss variation. Experimental results show that when facing fixed posture targets MNIST and CIFAR-10, final average accuracy reaches to 88% and 42.6% respectively, which is comparable to the comparison method, and final average forgetting is significantly better than the comparison method, with a reduction of about 3% and 15% respectively. In the case of the multi-pose target RotMNIST and trCIFAR-10, the proposed method continues to perform well in fixed-pose targets, largely independent of target pose. In addition, the performance in 3D datasets ModelNet40 and trModelNet40 remains stable. The method proposed is able to be independent of the target pose in online class incremental learning, while achieving catastrophic forgetting mitigation, with excellent stability and plasticity.

Key words: online class-incremental learning; catastrophic forgetting; ignoring pose replay; equivariance; point cloud classification

1 引 言

深度神经网络(Deep Neural Networks, DNN)已经在许多机器学习任务中取得了巨大的成功^[1-3],这些任务都是基于独立同分布(Independent Identically Distributed, IID)数据。这一设定违反了实际应用中所面临的非平稳数据分布,如自动驾驶、智能对话系统、智慧医疗和其他实时应用。面对非 IID 数据,当 DNN 对新任务进行再训练时,神经网络在过去任务中的性能表现会迅速下降,这被称为灾难性遗忘^[4]。增量学习(Incremental Learning, IL)^[5]的出现使 DNN 能够在学习新知识的同时保留先前获得的知识,使模型具有可塑性-稳定性^[6]。

传统的增量学习大都是以离线的方式训练模型,即通过重复训练多批次当前任务的数据来增加模型的拟合效果。然而,由于隐私问题或者内存限制,离线的设定不再适用。本文考虑了一项具有挑战性的单次数据流任务,即在线类增量学习(Online Class-Incremental Learning, OCIL)^[7]。OCIL 限制每个训练任务的样本流只能看到一次,并且是非 IID 的。

以往 IL 和 OCIL 的研究对象都是具有固定姿态的样本流,即每个样本都预先进行了矫正。这种设定对自动驾驶等实时应用是不负责任的,它们在真实情况中所面临的数据流是各个姿态的,即样本特征存在平移、旋转对称变换^[8]。常用的数据扩充方法^[9]在面对 OCIL 这种单次数据流

时也变得不再适用,特别是 3D 目标姿态的高复杂性对网络提出了更严格的要求。传统的神经网络只具有平移等变性^[10]而不具备旋转等变性。在 IL 任务中,除了新任务带来的影响外,造成遗忘的另一个重要原因是网络没能提取到目标足够的几何信息,而更多地关注到了一些无关特征,如固定的位置信息而不是目标本身之间的结构信息。保证遗忘率尽可能低首先要保证能够学到更丰富的几何特征,为此,我们基于李群引入旋转等变机制^[11-13]降低网络受目标姿态的影响。同时,使用点云数据以坐标和特征值的形式出现,即 $\{(x_i, f_i)\}_{i=1}^N$, 此设定可以方便网络提取目标的几何信息,且通用于 2D 和 3D 的数据。

常见的类增量学习方法大致分为 3 大类:参数隔离方法^[14-16]、正则化方法^[17-19]和记忆重放方法^[20-22]。参数隔离方法是通过扩展网络模型为代价提高网络的可塑性,以适应新的学习任务。随着新类别的增加,该方法将会导致参数量线性增长,从而变得不可持续。正则化方法通过添加正则项,在学习新的别类时约束参数的更新方向来避免灾难性遗忘。但正则化通常以牺牲可塑性为代价,不仅会降低对新知识的接纳程度,还增加了先前知识的遗忘率,往往表现不佳^[23]。相比之下,记忆重放的方法已经比较成熟,将旧任务数据流中的少量样本存储在存储器中,并在新任务的训练中重放它们。该方法在许多具有挑战性的场景中都具有最优的表现,在 IL 中发挥着关键的作用。本文在研究中采用基于记忆重放的

方法。

基于以上问题,本文提出无视姿态重放方法,主要工作有:

(1)考虑更符合现实情况的复杂设定,即面向多姿态目标的在线类增量学习。

(2)提出一个通用于 2D 和 3D 数据的 OCIL 模型。基于李群引入旋转平移等变机制,使网络能够提取更丰富的几何信息,减弱目标姿态的影响,从而缓解灾难性遗忘。

(3)提出基于损失变化的记忆重放策略,在图像分类基准数据集 MNIST、RotMNIST、CIFAR-10 和 trCIFAR-10, 3D 点云基准数据集 ModelNet40 和 trModelNet40 上进行了实验。本文方法在目标多姿态的设定下显著优于对比方法。

2 相关基础

2.1 类增量学习

一般情况下,类增量学习是涉及 T 个任务的顺序学习,这些任务由不同的类集组成,在不同学习阶段逐渐增加,且须在任何一个训练任务 t 中对前 $t-1$ 个任务中所看到过的类能够准确分类^[24-27],可以按以下方式设定:

$$\begin{cases} C = \{c_1, c_2, \dots, c_N\}, \\ T_t = \{c | \Psi(c) = t\}, \\ \mathcal{D}_c^c = \{x_1^c, x_2^c, \dots, x_{M_c}^c\}, \\ \mathcal{D}_t^T = \{\mathcal{D}_c^c | c \in T_t\}, \end{cases} \quad (1)$$

其中: C 表示所有类别的集合; T_t 表示 C 中分配给任务 t 的子集,其由分配函数 Ψ 决定; \mathcal{D}_c^c 表示类别 c 的样本集; \mathcal{D}_t^T 则表示任务 t 的样本集,且 $\mathcal{D}_i^T \cap \mathcal{D}_j^T = \emptyset$ 。值得注意的是,不同类别 c 的样本集数量 M_c 可以相同也可以不同,本文考虑 M_c 相同情况下的场景。

2.2 等变性和不变性

等变性:当一个函数 $f: X \rightarrow Y$ 的定义域 X 被对称群 G 作用,然后计算函数 f 所得到的结果等同于先计算函数 f ,然后计算应用群 G 作用得到的结果一样时,我们称函数 f 关于变换 G 是等变的,如公式(2)所示:

$$f(g \circ x) = g \circ f(x), \quad (2)$$

其中, $g \in G$ 是对称群 G 的一个群元素。

不变性:同理,当一个函数 $f: X \rightarrow Y$ 的定义域 X 被对称群 G 作用,然后计算函数 f 所得到的结果同直接计算函数 f 得到的结果一样时,我们称函数 f 关于变换 G 是不变的,如公式(3)所示:

$$f(g \circ x) = f(x). \quad (3)$$

在分类任务中,我们期待的是对输入进行任何对称变换,最终得到的结果不发生变化,即不变性。通常的解决办法是将具有不变性的函数 f 与等变函数 f_k 组合在一起,最终达到不变性。证明过程如式(4)所示^[28]:

$$\begin{aligned} f \circ f_K \circ \dots \circ f_1 \circ \pi_0 &= f \circ f_K \circ \dots \circ \pi_1 \circ f_1 \\ &\vdots \\ &= f \circ \pi_K \circ f_K \circ \dots \circ f_1 \\ &= f \circ f_K \circ \dots \circ f_1, \end{aligned} \quad (4)$$

其中:群表示 π_1, \dots, π_K 分别作用于函数 f_1, \dots, f_K 上,且 f_k 关于 π_k, π_{k-1} 等变,即 $f_k \circ \pi_{k-1} = \pi_k \circ f_k$, $k \in \{1, 2, \dots, K\}$, 函数 f 关于 π_k 不变,即 $f \circ \pi_k = f$, 因此 $f \circ f_K \circ \dots \circ f_1$ 是不变函数。

群等变:深度学习中最常见的等变是卷积层的平移等变性,即 $L_t f(x) = f(x+t)$,然而却不具备旋转等变性,即 $L_r f(x) \neq f(r \circ x)$ 。为了能处理多姿态数据,最初的办法是数据扩充^[29-30]。2016年,文献[8]基于 p4 群和 p4m 群首次提出了具有旋转对称性的群等变卷积神经网络。从此,关于群等变性的研究吸引了许多学者的注意。例如文献[11, 31-32]提出基于 SO(3) 的三维旋转群,文献[33-34]提出基于 SE(d) 的平移旋转群,文献[35]提出基于 $\mathbb{R}^* \times T(2)$ 的平移放缩群等。

2.3 点卷积

本文使用通用于图像和点云的点卷积网络^[36-37],其定义如公式(5)所示:

$$h(x) = (g_\theta * f)(x) = \int g_\theta(x-y) f(y) dy, \quad (5)$$

其中: $g_\theta: \mathbb{R}^d \rightarrow \mathbb{R}^{c_m \times c_{out}}$ 为卷积滤波器, $f(\cdot): \mathbb{R}^d \rightarrow \mathbb{R}^{c_m}$ 为输入特征图, $h(\cdot): \mathbb{R}^d \rightarrow \mathbb{R}^{c_{out}}$ 为输出。

离散化后如公式(6)所示:

$$h(x_i) = \left(\frac{V}{n}\right) \sum_j g_\theta(x_i - x_j) f(x_j), \quad (6)$$

其中: V 为积分空间体积, n 为正交点的数量。在例如图像的 3×3 卷积层中, g_θ 对每个落在 3×3 网格上的点: $(-1, -1), (-1, 0), \dots, (1, 1)$ 都有独立的参数。

3 无视姿态重放的在线类增量学习

我们考虑一个单次数据流的类增量场景,它模拟了一个实际的设置,模型必须对每个传入的示例执行在线更新,而无需重复多次训练。在每个任务 t 到来时,系统从数据流 \mathcal{D}_t^T 中接收一组新的样本: $(\{x_i, y_i\}_{i=0}^{bs})$, 其中, bs 为每组的样本个数, x_i 为原输入样本, y_i 为样本的标签。考虑到目标的多姿态时,其输入表示为 $T_g x_i$, 意为原样本发生几何变换(旋转、平移)后的结果; g 为 $SE(d)$ ($d=2, 3$) 群的一个群元素, 是对应几何变换的群表示。

模型结构如图 1 所示, 主要包含两部分: (1) 可以提取丰富的几何特征的分类器 θ ; (2) 基于损失变化的记忆重放。

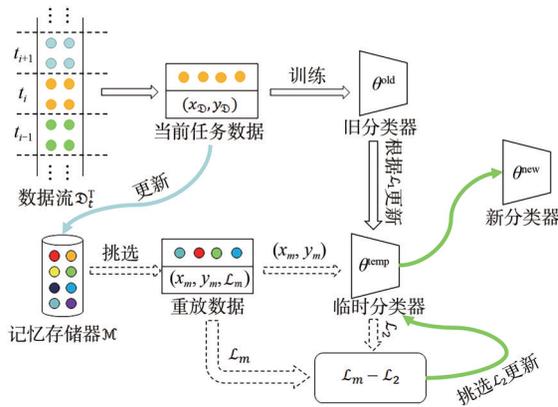


图 1 模型结构图

Fig. 1 Model structure diagram

3.1 具有多姿态识别能力的分类器 θ

为了降低 OCIL 任务中网络受目标姿态的影响, 同时提取更丰富的几何特征, 减少灾难性遗忘, 我们提出具有抗旋转平移几何变换能力的分类器, 即图 1 中的 θ 。分类器为在骨干网络 PointNet++^[38] 上进行的改造, 主要加入 3 部分: 点云化处理、群等变点卷积层和群全局池化层。

点云化处理负责将原输入转换成点云数据, 其过程如公式(7)所示:

$$x \in \mathbb{R}^d \mapsto y = \{(x_i, f_i)\}_{i=1}^n, \quad (7)$$

其中: x 是原始输入, 当输入是图像数据时 $d=2$, 当输入是 3D 数据时 $d=3$; y 是原始输入映射到高维空间后的点云化结果, 即 $\{(x_i, f_i)\}_{i=1}^n$, x_i 是点

云坐标, 其原点为输入样本的几何中心, f_i 为每个点所对应的特征值。例如, 在 CIFAR-10 数据集中, x_i 为 $(-16, 16), (-15, 16), \dots, (16, -16)$, 对应原样本图片从左上角到右下角的坐标; f_i 是 $\{r_i, g_i, b_i\}_{i=1}^{32 \times 32}$, 为原样本图片从左上角到右下角每个像素点所对应的特征值。点云化处理后的高维信息同时包含了几何位置和每个点的特征信息, 使得数据特征更加丰富, 从而能够更好地提取几何特征。

群等变点卷积层负责对网络进行平移旋转等变性改进。将点云化信息映射到更高维的空间以提取特征, 是使网络抗旋转平移几何变换的关键, 其定义由公式(8)和公式(9)给出:

$$h(g) = [\Psi^* f(x_i)](g) = \frac{1}{n_i} \sum_{j \in \text{nbhd}(i)} \Psi(g^{-1} x_j) f(x_j), \quad (8)$$

$$h(g) = [\Psi^* f(u_i)](g) = \frac{1}{n_i} \sum_{j \in \text{nbhd}(i)} \Psi(g^{-1} u_j) f(u_j), \quad (9)$$

其中: $g \in SE(d)$, $d=2$ 适用于 $x_i \in \mathbb{R}^2$, $d=3$ 适用于 $x_i \in \mathbb{R}^3$, $\{u_i \in SE(d): u_0 = x\}$, $n_i = |\text{nbhd}(i)|$ 为每个点邻域中的点数。公式(8)只适用于点卷积的第一层, 其将输入从欧式空间映射到了 $SE(d)$ 所在的李代数空间, 即: $\mathbb{Z}^2 \mapsto SE(d)$; 公式(9)适用于除第一层以外的所有点卷积层, 其在李代数空间进行映射, 即 $SE(d) \mapsto SE(d)$ 。值得注意的是, 在考虑姿态的在线类增量学习设定下, x_i 在不同时刻表现为不同的 $T_g x_i$, 即同一个样本每次出现都会表现为不同的姿态(发生了不同的平移和旋转)。另外, 由于 $SE(d)$ 群是连续群, 并不能穷举 g 的所有情况, 我们使用哈尔测度 μ 进行均匀采样。

点云化处理后, 显式地蕴含了位置信息和特征信息。而群等变点卷积层则能够将输入映射到 $SE(d)$ 所在的李代数空间, 该空间融合了样本不同位置和角度的特征, 能够使网络不受目标姿态的影响。同时, 点卷积能够使每个点邻域范围内的特征进行聚合, 表征一定范围内的几何信息, 从而使网络能够提取更丰富的几何特征。

群等变点卷积层的具体实现算法流程如算法 1 所示, 其关于旋转和平移变换是等变的, 证明过程如式(10)所示:

$$[\Psi^*L_t f(x)](g) = \frac{1}{n_i} \sum_j \Psi(g^{-1}x) f(t^{-1}x) = \frac{1}{n_i} \sum_j \Psi(g^{-1}tx) f(x) = \frac{1}{n_i} \sum_j \Psi((gt^{-1})^{-1}x) f(x) = [L_t[\Psi^*f(x)]](g), \quad (10)$$

其中: $L_t f$ 表示对输入进行 t (旋转平移) 变换, 第一行到第二行令 $x = tx$ 。证明结果满足公式(2)中等变性的定义。

算法 1 群等变点卷积

输入: $\{(u_i, f_i)\}_{i=1}^{N^*m}$, ($u_i \in T_{g_i} x_i$, m 为群元素的个数, 第一层群等变点卷积中 $m=1$)
 输出: $\{h_i\}_{i=1}^m$
 初始化: $\{g_i\}_{i=1}^m \sim \mu$
 1) FOR $i=1, \dots, N^*m$
 2) 计算 u_i 的逆元 u_i^{-1}
 3) 找到 u_i 的邻域内点集 $\text{nbhd}(i)$
 4) 计算 u_i 邻域内点的个数 $n_i = |\text{nbhd}(i)|$
 5) 初始化 $h_i = 0$
 6) FOR $j=1, \dots, N^*m$
 7) $h_i += \sum_{j \in \text{nbhd}(i)} \Psi(u_i^{-1}u_j) f(u_j)$
 8) END
 9) $h_i = \frac{1}{n_i} h_i$
 10) END

群全局池化层与普通的池化层类似, 包括最大池化和平均池化等, 本文使用了全局最大池化(GlobalMaxPooling, GP), 其定义如公式(11)所示:

$$\text{GP}f(g) = \max_{k \in gU} f(k), \quad (11)$$

其中, gU 是 $\text{SE}(d)$ 的子群 U 上的一个 g 变换。对 GP 层的输入进行对称变化, 其输出总是不变的, 即 GP 具有不变性, 满足公式(3), 放在网络的最后用来使模型整体达到不变性的效果, 相当于公式(4)中的 f , 其示意图如图 2 所示, $k_i \in gU$ 。

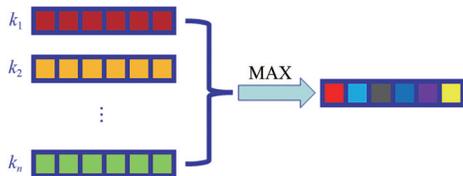


图 2 全局最大池化层

Fig. 2 Global maximum pooling layer

3.2 基于损失变化的记忆重放

为了应对 OCIL 挑战, 基于记忆重放的方法在固定内存中储存少量访问过的数据, 并在未来的任务中进行回放, 都取得了很好的成效^[7,39-40]。我们分配一个固定内存大小的记忆存储器 \mathcal{M} (容量为 M), 当样本流到达时, 使用随机采样^[41] 来保证记忆的多样性, 同时, 储存了每个样本最近一次的损失 \mathcal{L} , 即存储器中为 $\{(x_m, y_m, \mathcal{L}_m)\}_{m=1}^M$ 。为了保证 \mathcal{M} 中的样本均衡, 在每个新类到达时舍弃一部分旧样本来储存新样本, 并保证每个类别的数量相同。

记忆重放涉及的关键一步是回放样本的选择, 我们采用与^[7,42]类似的假设, 即模型应该优先回放被忘记的样本, 以减少对早期任务类别的灾难性遗忘。算法 2 描述了具体的重放过程。在时刻 t , 旧模型 θ^{old} 从数据流 \mathcal{D} 中接收一批数据 $\{(x_{\mathcal{D}}^i, y_{\mathcal{D}}^i)\}_{i=1}^{n_1}$, 根据损失 \mathcal{L}_1 执行更新产生临时模型 θ^{temp} , 如公式(12)和公式(13)所示:

$$\mathcal{L}_1 = \ell(x_{\mathcal{D}}, y_{\mathcal{D}}; \theta^{\text{old}}), \quad (12)$$

$$\theta^{\text{temp}} = \theta^{\text{old}} - \alpha \nabla_{\theta^{\text{old}}} \mathcal{L}_1, \quad (13)$$

其中: ℓ 为交叉熵损失函数, α 为学习率。接着从存储器中抽取 $n_2 \geq n_1$ 组数据 $\{(x_m^i, y_m^i, \mathcal{L}_m^i)\}_{i=1}^{n_2}$, 根据临时模型 θ^{temp} 计算损失 \mathcal{L}_2 , 并与 \mathcal{L}_m 进行比较, 挑出损失变化最大的 n_1 组 \mathcal{L}_2 用于更新 θ^{temp} , 从而产生新模型 θ^{new} , 如公式(14)和公式(15)所示:

$$\mathcal{L}_2 = \ell(x_m, y_m; \theta^{\text{temp}}), \quad (14)$$

算法 2 基于损失变化的记忆重放

输入: 数据流 $(x_{\mathcal{D}}, y_{\mathcal{D}})$;

输出: θ^{new} 。

初始化: 记忆存储器 \mathcal{M} , 学习率 α , 模型 θ^{old} 。

1) FOR $t=1, \dots, T$;

2) 通过公式(12)计算 \mathcal{L}_1 ;

3) 通过公式(13)更新临时模型 $\theta^{\text{temp}} \leftarrow \text{SGD}(\mathcal{L}_1, \theta^{\text{old}}, \alpha)$;

4) $(x_m, y_m, \mathcal{L}_m) \in \mathcal{M}$;

5) 通过公式(14)计算 \mathcal{L}_2 ;

6) 根据损失变化选择要回传的损失 $\{\mathcal{L}_2^i\}_{i=1}^{n_2} \leftarrow \text{sort}(\mathcal{L}_m - \mathcal{L}_2)$;

7) 通过公式(15)更新模型 $\theta^{\text{new}} \leftarrow \text{SGD}(\{\mathcal{L}_2^i\}_{i=1}^{n_2}, \theta^{\text{temp}}, \alpha)$;

8) 更新存储器 $\mathcal{M} \leftarrow \text{update}(\mathcal{M}, x_{\mathcal{D}}, y_{\mathcal{D}}, \mathcal{L}_1, \mathcal{L}_2)$;

9) END。

$$\theta^{\text{new}} = \theta^{\text{temp}} - \alpha \nabla_{\theta^{\text{temp}}} \left\{ \mathcal{L}_2^i \mid \text{sort}(\mathcal{L}_m - \mathcal{L}_2)_i \right\}_{i=1}^{n_1}. \quad (15)$$

最后,同步更新存储器 \mathcal{M} 中 n_2 组数据的损失,即 $\{\mathcal{L}_m^i\}_{i=1}^{n_2} = \{\mathcal{L}_2\}_{i=1}^{n_2}$,并从该 n_1 组新数据中随机采样来更新存储器 \mathcal{M} 。

4 实验结果与分析

4.1 实验数据集

我们使用 MNIST^[43]、RotMNIST^[44]、CIFAR-10^[45]、trCIFAR-10、ModelNet40^[46] 和 trModelNet40 数据集来评估本文提出的方法。其中, RotMNIST 数据集由 62 000 个随机旋转的 MNIST 数字组成,旋转角度从 $SO(2)$ 中均匀采样。trCIFAR-10 和 trModelNet40 为我们对 CIFAR-10 和 ModelNet40 进行随机的平移和旋转的几何变换,模拟真实多姿态场景。MNIST、CIFAR-10、ModelNet40 用于固定姿态目标实验,如图 3(a) 所示。RotMNIST、trCIFAR-10、trModelNet40 用于多姿态目标的实验,如图 3(b) 所示。我们列出了每个数据集的实验设置:

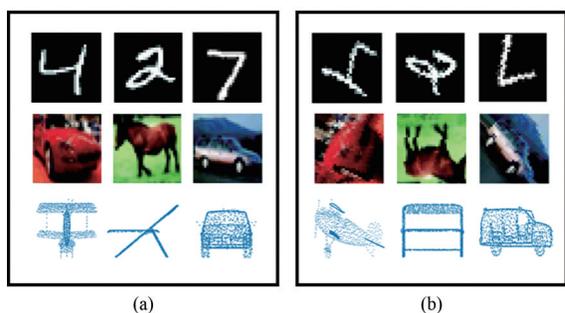


图 3 固定姿态目标(a)与多姿态目标(b)

Fig. 3 Fixed posture target (a) and multi-posture target (b)

(1) MNIST 和 RotMNIST:按照每 2 类为一个任务,分为 5 个不同的任务;将图片转为特征值为像素值的 2D 点云输入,即 $\{(x_i, f_i)\}_{i=1}^{28 \times 28}$,其中 $x_i \in \mathbb{R}^2, f_i \in \mathbb{R}$ 。遵从文献[7]的设定,每项任务分配 1 000 个样本用于训练。

(2) CIFAR-10 和 trCIFAR-10:按照每 2 类为一个任务,分为 5 个不同的任务。将图片转为特征值为像素值的 2D 点云输入,即 $\{(x_i, f_i)\}_{i=1}^{32 \times 32}$,其中 $x_i \in \mathbb{R}^2, f_i \in \mathbb{R}^3$ 。遵从文献[7]的设定,每项任务分配 9 500 个样本用于训练。

(3) ModelNet40 和 trModelNet40:按照每 4 类为一个任务,分为 10 个不同的任务;点云输入为 $\{(x_i, f_i)\}_{i=1}^{1024}$,其中 $x_i \in \mathbb{R}^3, f_i = 1$ 。由于 ModelNet40 数据集不同类的样本数量不同,故取每类的 80% 用于训练,即共有 9 843 个数据流用于训练。

4.2 实验环境

本文所有实验都在同一学习环境下进行,如表 1 所示。

表 1 实验环境

Tab. 1 Experimental environment

| 实验环境 | 环境配置 |
|--------|---|
| 编程语言 | Python3. 8. 13 |
| 开发工具 | Pycharm11. 0. 11 |
| 深度学习框架 | Pytorch1. 11. 0 |
| CUDA | 10. 2 |
| GPU | GTX TITAN xp |
| CPU | Intel(R) Xeon(R) Silver 4210 CPU @ 2. 20 GHz |

4.3 基线和评价指标

对于固定姿态目标和多姿态目标的实验,我们考虑了以下设定作为评价基线:

- (1) fine-tuning:微调,在新任务到达时连续训练,不采用任何防遗忘策略,作为实验对照下限;
- (2) iid online:所有任务数据同时出现并训练一次;
- (3) iid offline:允许数据多次出现重复训练,作为实验的上限。

这些基线所使用的分类器都是 3.1 节中具有多姿态识别能力的分类器。为了评估不同方法的效果,我们引入两个评价指标:最终平均准确率(AvgACC)和平均遗忘率(AvgF)^[41]。

4.4 实验分析

我们在 2D 数据集上进行了评估,即 MNIST、RotMNIST、CIFAR-10 和 trCIFAR-10。除了 4.3 节中提到的基线外,本节还使用了其他两种基于记忆重放的类增量学习方法作为面向 2D 数据的对照实验:ER^[41]和 ER-MIR^[7]。我们统一了记忆存储器的容量 $M=500$,且每批数据只迭代 1 次。每组实验进行了 20 次并取平均值,结果如表 2 所示。

表 2 2D 图像数据上的实验结果
Tab. 2 Experimental results on 2D image data

| 方法 | 固定姿态目标 | | | |
|-------------|----------|----------|------------|----------|
| | MNIST | | CIFAR-10 | |
| | AvgACC/% | AvgF/% | AvgACC/% | AvgF/% |
| fine-tuning | 19.8±0.5 | 98.9±0.4 | 17.2±0.6 | 84.3±0.4 |
| idd online | 88.6±0.9 | N/A | 58.3±1.7 | N/A |
| idd offline | 99.2±0.2 | N/A | 80.8±0.6 | N/A |
| ER | 82.1±1.5 | 15.0±2.1 | 33.1±1.7 | 35.4±2.0 |
| ER-MIR | 87.6±0.7 | 7.0±0.9 | 46.3±2.5 | 34.2±1.5 |
| Ours | 88.0±2.8 | 4.1±1.8 | 42.6±2.3 | 19.3±4.6 |
| 方法 | 多姿态目标 | | | |
| | RotMNIST | | trCIFAR-10 | |
| | AvgACC/% | AvgF/% | AvgACC/% | AvgF/% |
| fine-tuning | 19.7±0.3 | 97.9±1.3 | 17.4±0.5 | 85.6±1.1 |
| idd online | 86.7±0.5 | N/A | 56.4±2.4 | N/A |
| idd offline | 98.6±0.1 | N/A | 81.9±3.5 | N/A |
| ER | 23.1±1.0 | 39.6±2.3 | 15.8±2.6 | 34.2±3.5 |
| ER-MIR | 30.7±0.6 | 30.4±3.4 | 22.3±1.3 | 31.4±9.6 |
| Ours | 86.1±2.1 | 4.5±1.4 | 40.4±0.9 | 20.3±3.4 |

可以看到,当面对固定姿态目标 MNIST 时,本文方法的 AvgACC 和 AvgF 分别为 88.0% 和 4.1%,均优于 ER 和 ER-MIR 方法,且 AvgACC 与基线 iid online 接近,说明本文方法可以有效缓解灾难性遗忘,实现任务之间的稳定性。当面对固定姿态目标 CIFAR-10 时,本文方法的 AvgACC 明显优于 ER,与 ER-MIR 相比也仅差 3.7%,但 AvgF 明显优于 ER 和 ER-MIR,证明本文方法可以很大程度缓解灾难性遗忘。

而当面对多目标姿态 RotMNIST 时,ER 和 ER-MIR 与它们在 MNIST 中的表现相比具有显著的变化,AvgACC 分别下降了 59.0% 和 56.9%,AvgF 也分别增加了 24.6% 和 23.4%,说明 ER 和 ER-MIR 在面对多姿态目标时均不能学习到有效信息,且遗忘率明显增加,并不适用于真实场景。而本文方法与在 MNIST 中的表现相比,AvgACC 只下降了约 1.9%,AvgF 只增加了 0.4%。在 20 次的实验中,本文方法的 AvgACC 均保持在 85% 以上,证明本文方法可以有效抵抗目标姿态所带来的影响,且对灾难遗忘有明显的缓解作用。当面对多目标姿态 trCIFAR-10 时,本文方法的表现与上面所述一致,受目标姿态影响很小,AvgACC 只降低了约 2.2%,AvgF 只增加了 1%。而 ER 和 ER-MIR 在面对多姿态目标 trCIFAR-10

时的 AvgACC 不及它们在固定姿态目标 CIFAR-10 中表现的 1/2。虽然 AvgF 并没有很大变化,甚至 ER 在 trCIFAR-10 中的 AvgF 要略低于在 CIFAR-10 中的 AvgF,但这是由于 ER 本身在学习中的最高准确率很低所导致。

图 4 中列出了本文方法和 ER-MIR 方法中每个任务在不同时期的分类精度对比。如图 4(a~d) 所示,当面对数据集 MNIST 时,本文方法和 ER-MIR 的表现不相上下,每类任务的最终准确率都与最初的准确率相比并没有太大变化,说明本文方法和 ER-MIR 在面对固定姿态目标时都能提取有效的特征,并对灾难性遗忘有很好的抵抗力,面对 OCIL 这种严格的设定能保证稳定性。而当面对 RotMNIST 时,ER-MIR 变得不再稳定,在后续任务中发生了灾难性遗忘,每类任务的最终准确率与初始准确率相比有明显的差距,下降到了 20%~45%,远不及它在 MNIST 中的表现。相比之下,本文方法在面对 RotMNIST 时仍能保证最终准确率与初始准确率很小的差距,维持在 76%~97%,且与在 MNIST 中的表现相差无几,证明本文方法并不受目标姿态的影响,在 OCIL 任务中有效缓解灾难性遗忘,具有很好的稳定性。

如图 4(e~h) 所示,当面对 CIFAR-10 这种

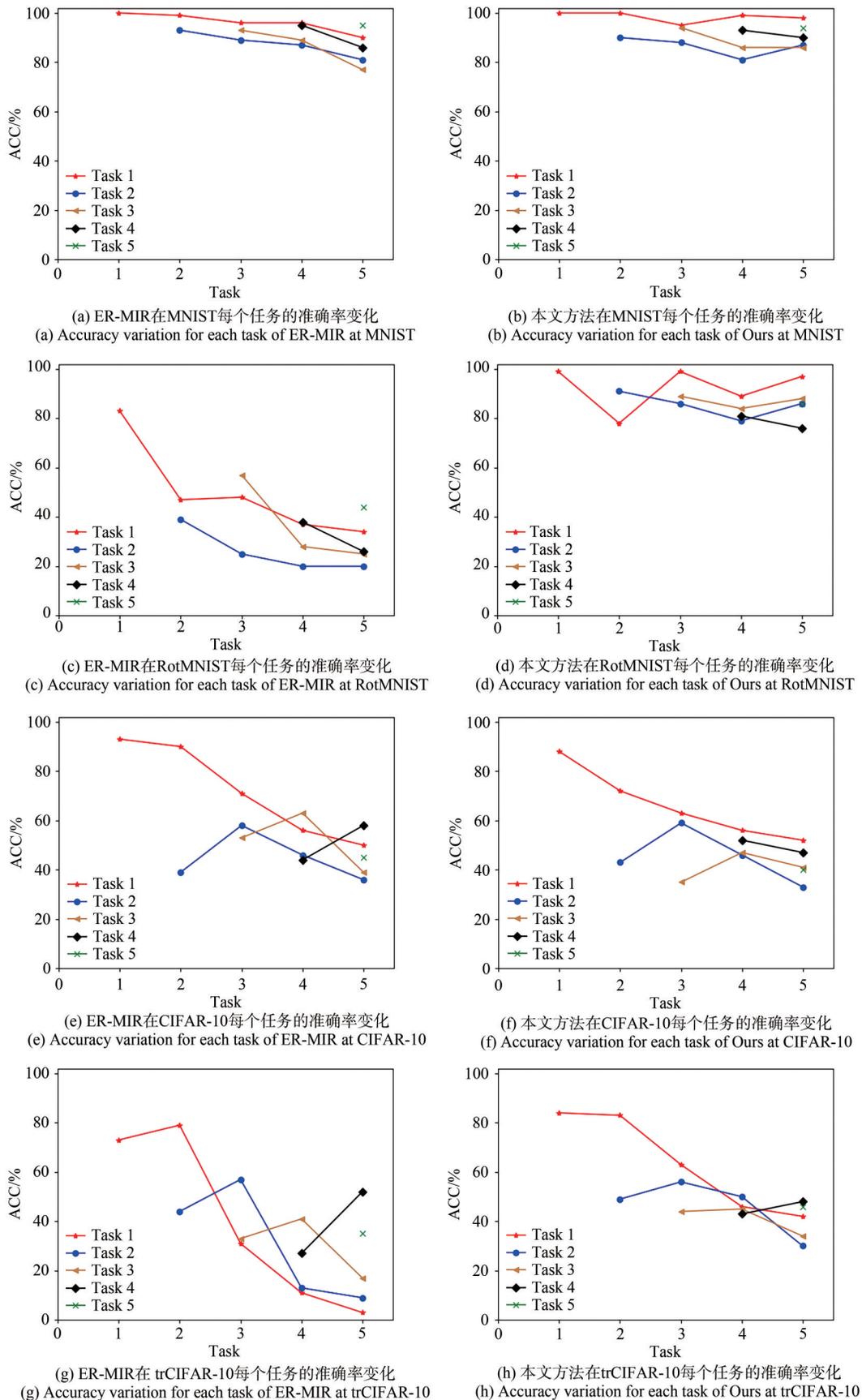


图 4 每个任务在不同时期的准确率

Fig. 4 Accuracy of each task at different times

更有挑战的数据集时,本文方法和ER-MIR虽然都有一定的遗忘,但它们的最终准确率也都能保持在33%以上。而当面对trCIFAR-10时,ER-MIR的第一个任务和第二个任务的最终准确率下降至10%以下,第三个任务的准确率也下降到了20%以下,均发生了严重的遗忘。相比之下,本文方法依旧能和在CIFAR-10中的表现一样,有效地消除了目标姿态所带来的影响,具有相对较好的稳定性。

从图4中还可以看到,本文方法在这4种数据集中每类任务的最终准确率相差不超过22%(分别为12%,22%,19%,18%),并没有出现由于单任务准确率高而提高平均准确率的现象,说明本文方法在OCIL任务中不仅不受目标姿态的影响,还有很好的平衡性。

为了验证本文方法性能,我们进一步在3D数据集ModelNet40上进行了实验。在我们的了解中,并没有发现面向3D数据的OCIL的相关研究,因此,本部分的对照实验只采用3.3节中所提到的基线实验。同时,设定记忆存储器的容量 $M=4\ 000$,即每个任务存储200个样本,且每批数据依旧只迭代1次来保持online的设定。每组实验进行了20次并取平均值,结果如表3所示。可以看到,本文方法在面对多姿态目标trModelNet40时的性能与固定姿态目标ModelNet40的AvgACC和AvgF都相差不大,分别约为4%和2%,且AvgACC超过了基线idd online的结果,证明本文方法在面对3D目标时也可以做到无视目标的姿态,同时对灾难性遗忘有很大程度的缓解。

表3 3D数据上的实验结果对比

Tab. 3 Experimental results on 3D data

| 方法 | ModelNet40 | | trModelNet40 | |
|-------------|------------|----------|--------------|----------|
| | AvgACC/% | AvgF/% | AvgACC/% | AvgF/% |
| fine-tuning | 7.3±0.9 | 67.1±2.2 | 6.7±0.9 | 58.8±2.5 |
| idd online | 51.0±1.2 | N/A | 47.9±0.8 | N/A |
| idd offline | 90.9±1.1 | N/A | 85.0±1.6 | N/A |
| Ours | 52.8±3.7 | 22.0±2.9 | 48.8±3.9 | 24.0±2.2 |

为了评估不同姿态对OCIL任务的影响,我们在MNIST数据集上进行了补充实验,如图5所示。其中,横轴表示样本的姿态变化范围,如“60”表示数据流在 $[-60^\circ, 60^\circ]$ 范围内随机旋转,以此类推,来模拟样本姿态的丰富程度。从图5可以明显看到,随着角度变化范围的增大,ER方法和

ER-MIR方法的AvgACC逐渐下降,尤其在 60° 后有显著的下降,并最终下降到基线fine-tuning附近,可见样本姿态对传统方法有较大的影响。相比之下,本文方法的AvgACC随着角度变化范围的增大只有较小的波动,波动范围保持在3%以内,证明本文方法可以在很大程度上降低目标姿态的影响,有很好的稳定性。

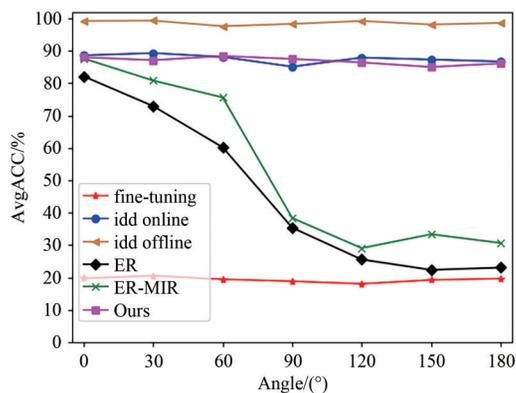


图5 MNIST不同姿态的结果

Fig. 5 Results of MNIST with different postures

5 结 论

与传统增量学习不同,本文考虑了更切合实际的复杂场景,即面向多姿态目标的在线类增量学习,该设定加剧了灾难性遗忘。为了解决这个问题,本文提出了适用于2D和3D数据的在线类增量学习方法。该算法的网络框架基于SE(d)李群引入旋转平移等变机制,使网络可以更好地提取目标的几何信息,从而使模型不受目标姿态的影响,增加模型的可塑性。本文还提出了基于

损失变化的记忆重放方法,能够配合我们的分类器缓解灾难性遗忘,在稳定性和可塑性直接得到很好的权衡。本文方法在 MNIST、RotMNIST、CIFAR-10、trCIFAR-10、ModelNet40 和 trModelNet40 数据集上进行多次实验并获得了有竞争力的结果,其中 AvgACC 除了在 CIFAR-10 表现略

低于 ER-MIR 外,其余都取得了最好的结果,分别为 88.0%, 86.1%, 42.6%, 40.4%, 52.8%, 48.8%; AvgF 则所有情况下都为最优,分别为 4.1%, 4.5%, 19.3%, 20.3%, 22.0%, 24.0%。实验结果验证了本文方法的有效性,能够同时做到不受目标姿态的影响并缓解灾难性遗忘。

参 考 文 献:

- [1] 龚勋,张志莹,刘璐,等. 人物交互检测研究进展综述[J]. 西南交通大学学报, 2022, 57(4): 693-704.
GONG X, ZHANG Z Y, LIU L, *et al.* A survey of human-object interaction detection [J]. *Journal of Southwest Jiaotong University*, 2022, 57(4): 693-704. (in Chinese)
- [2] 刘飞,高红艳,卫泽刚,等. 基于 Res-Net 深度特征的 SAR 图像目标识别方法[J]. 液晶与显示, 2021, 36(4): 624-631.
LIU F, GAO H Y, WEI Z G, *et al.* SAR image target recognition method using based on Res-Net deep features [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(4): 624-631. (in Chinese)
- [3] 李国友,李晨光,王维江,等. 基于单样本学习的多特征人体姿态模型识别研究[J]. 光电工程, 2021, 48(2): 200099.
LI G Y, LI C G, WANG W J, *et al.* Research on multi-feature human pose model recognition based on one-shot learning [J]. *Opto-Electronic Engineering*, 2021, 48(2): 200099. (in Chinese)
- [4] MCCLOSKEY M, COHEN N J. Catastrophic interference in connectionist networks: the sequential learning problem [J]. *Psychology of Learning and Motivation*, 1989, 24: 109-165.
- [5] LI Z Z, HOIEM D. Learning without forgetting [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(12): 2935-2947.
- [6] MERMILLOD M, BUGAJSKA A, BONIN P. The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects [J]. *Frontiers in Psychology*, 2013, 4: 504.
- [7] ALJUNDI R, CACCIA L, BELILOVSKY E, *et al.* Online continual learning with maximally interfered retrieval [J/OL]. *arXiv*, 2019: 1908.04742.
- [8] COHEN T, WELLING M. Group equivariant convolutional networks [C]//*Proceedings of the 33rd International Conference on Machine Learning*. New York: ICML, 2016: 2990-2999.
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [10] LECUN Y, BENGIO Y. Convolutional networks for images, speech, and time-series [M]//ARBIB M A. *The Handbook of Brain Theory and Neural Networks*. Cambridge: MIT Press, 1995.
- [11] ESTEVES C, ALLEN-BLANCHETTE C, MAKADIA A, *et al.* Learning SO(3) equivariant representations with spherical CNNs [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018: 54-70.
- [12] RAVANBAKHS S, SCHNEIDER J G, PÓCZOS B. Deep learning with sets and point clouds [C]. 5th International Conference on Learning Representations. Toulon: ICLR, 2017.
- [13] WANG Y, SUN Y B, LIU Z W, *et al.* Dynamic graph CNN for learning on point clouds [J]. *ACM Transactions On Graphics*, 2019, 38(5): 146.
- [14] CAI S S, XU Z W, HUANG Z C, *et al.* Enhancing CNN incremental learning capability with an expanded network [C]//2018 IEEE International Conference on Multimedia and Expo (ICME). San Diego: IEEE, 2018: 1-6.
- [15] MALLYA A, DAVIS D, LAZEBNIK S. Piggyback: Adapting a single network to multiple tasks by learning to mask weights [C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018: 67-82.

- [16] MALLYA A, LAZEBNIK S. PackNet: Adding multiple tasks to a single network by iterative pruning [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 7765-7773.
- [17] KIRKPATRICK J, PASCANU R, RABINOWITZ N, *et al.* Overcoming catastrophic forgetting in neural networks [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(13): 3521-3526.
- [18] CHAUDHRY A, DOKANIA P K, AJANTHAN T, *et al.* Riemannian walk for incremental learning: Understanding forgetting and intransigence [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018: 556-572.
- [19] ZENKE F, POOLE B, GANGULI S. Continual learning through synaptic intelligence [C]//*Proceedings of the 34th International Conference on Machine Learning*. Sydney: ACM, 2017: 3987-3995.
- [20] REBUFFI S A, KOLESNIKOV A, SPERL G, *et al.* iCaRL: incremental classifier and representation learning [C]//*Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 5533-5542.
- [21] 王伟, 张志莹, 郭杰龙, 等. 基于脑启发的类增量学习[J]. *计算机应用研究*, 2023, 40(3): 671-675, 688.
WANG W, ZHANG Z Y, GUO J L, *et al.* Brain-inspired class incremental learning [J]. *Application Research of Computers*, 2023, 40(3): 671-675, 688. (in Chinese)
- [22] SHIN H, LEE J K, KIM J, *et al.* Continual learning with deep generative replay [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: ACM, 2017: 2994-3003.
- [23] RIOS A, ITTI L. Closed-loop memory GAN for continual learning [C]//*Proceedings of the 28th International Joint Conference on Artificial Intelligence*. Macao, China: ACM, 2019: 3332-3338.
- [24] DOUILLARD A, CORD M, OLLION C, *et al.* PODNet: pooled outputs distillation for small-tasks incremental learning [C]. 16th European Conference on Computer Vision. Glasgow: Springer, 2020: 86-102.
- [25] HU X T, TANG K H, MIAO C Y, *et al.* Distilling causal effect of data in class-incremental learning [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 3956-3965.
- [26] BANG J, KIM H, YOO Y J, *et al.* Rainbow memory: continual learning with a memory of diverse samples [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 8214-8223.
- [27] ZHU F, CHENG Z, ZHANG X Y, *et al.* Class-incremental learning via dual augmentation [C]//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. Online: NeurIPS, 2021: 14306-14318.
- [28] HUTCHINSON M J, LE LAN C, ZAIDI S, *et al.* Lietransformer: Equivariant self-attention for lie groups [C]//*Proceedings of the 38th International Conference on Machine Learning*. Online: ICML, 2021: 4533-4543.
- [29] FASEL B, GATICA-PEREZ D. Rotation-invariant neoperceptron [C]. 18th International Conference on Pattern Recognition (ICPR'06). Hong Kong, China: IEEE, 2006: 336-339.
- [30] LAPTEV D, SAVINOV N, BUHMANN J M, *et al.* TI-POOLING: transformation-invariant pooling for feature learning in convolutional neural networks [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 289-297.
- [31] COHEN T S, GEIGER M, KÖHLER J, *et al.* Spherical CNNs [C]. 6th International Conference on Learning Representations. Vancouver: ICLR, 2018.
- [32] DENG C Y, LITANY O, DUAN Y Q, *et al.* Vector neurons: a general framework for $SO(3)$ -equivariant networks [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021: 12180-12189.
- [33] FUCHS F, WORRALL D, FISCHER V, *et al.* Se(3)-transformers: 3d roto-translation equivariant attention networks [C]//*Proceedings of the 34th International Conference on Neural Information Processing Systems*. Online: NeurIPS, 2020: 1970-1981.
- [34] FINZI M, STANTON S, IZMAILOV P, *et al.* Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data [C]//*Proceedings of the 37th International Conference on Machine Learning*.

- Virtual Event: ACM, 2020: 297.
- [35] SOSNOVIK I, SZMAJA M, SMEULDERS A. Scale-equivariant steerable networks [C]. 8th International Conference on Learning Representations. Addis Ababa: ICLR, 2019.
- [36] DAI J F, QI H Z, XIONG Y W, *et al.* Deformable convolutional networks [C]//*Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE, 2017: 764-773.
- [37] WU W X, QI Z A, FUXIN L. PointConv: deep convolutional networks on 3d point clouds [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 9613-9622.
- [38] QI C R, YI L, SU H, *et al.* PointNet++: Deep hierarchical feature learning on point sets in a metric space [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: ACM, 2017: 5105-5114.
- [39] ROBINS A. Catastrophic forgetting, rehearsal and pseudorehearsal [J]. *Connection Science*, 1995, 7(2): 123-146.
- [40] JIN X S, SADHU A, DU J Y, *et al.* Gradient-based editing of memory examples for online task-free continual learning [C]//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. Online: NeurIPS, 2021: 29193-29205.
- [41] CHAUDHRY A, ROHRBACH M, ELHOSEINY M, *et al.* Continual learning with tiny episodic memories [J/OL]. *arXiv*, 2019:1902.10486v1.
- [42] TONEVA M, SORDONI A, DES COMBES R T, *et al.* An empirical study of example forgetting during deep neural network learning [C]. 7th International Conference on Learning Representations. New Orleans: ICLR, 2019.
- [43] LECUN Y, BOTTOU L, BENGIO Y, *et al.* Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [44] LAROCHELLE H, ERHAN D, COURVILLE A, *et al.* An empirical evaluation of deep architectures on problems with many factors of variation [C]//*Proceedings of the 24th International Conference on Machine Learning*. Corvalis: ACM, 2007: 473-480.
- [45] KRIZHEVSKY A. Learning multiple layers of features from tiny images [D]. Toronto: University of Toronto, 2009.
- [46] WU Z R, SONG S R, KHOSLA A, *et al.* 3D ShapeNets: a deep representation for volumetric shapes [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 1912-1920.

作者简介:



张润江(1998—),男,山西运城人,硕士研究生,2020年于福建工程学院获得学士学位,主要从事计算机视觉及持续学习方面的研究。E-mail: 1069707145@qq.com



魏 宪(1986—),男,河南沁阳人,博士,研究员,2017年于慕尼黑工业大学获得博士学位,主要从事机器学习、几何优化方面的研究。E-mail: xian.wei@fjirsm.ac.cn