中图法分类号:TP181;TP39　　**文献标识码**:A　　**文章编号**:1006-8961(2026)11-4005-16

论文引用格式:Liu W H, Zhang Z M, Zhang Y and Guo Y L. 2026. Curriculum-driven dynamic data pruning for self-supervised learning. Journal of Image and Graphics, 31(11):4005-4020(刘文昊, 张钊铭, 张晔, 郭裕兰. 2026. 面向自监督学习的课程式动态数据剪枝方法. 中国图象图形学报, 31(11):4005-4020)[DOI:10.11834/jig.250542]

面向自监督学习的课程式动态数据剪枝方法

刘文昊[1],张钊铭[2],张晔[1],郭裕兰[1*]

1. 中山大学电子与通信工程学院,深圳　518000; 2. 西安现代控制技术研究所,西安　710000

摘　要:**目的**　动态数据剪枝是提升自监督预训练效率的重要手段,但现有方法多基于瞬时损失进行剪枝,难以应对自监督训练中损失波动剧烈、优化不稳定的问题,易造成关键样本被误剪和模型性能退化。针对上述缺陷,本文提出了面向自监督学习的课程式动态数据剪枝方法。**方法**　首先设计指数损失重要性表征,通过动量系数对样本历史损失做指数移动平均,获取时序平滑且对短期损失波动鲁棒的样本价值表征,避免因瞬时损失扰动而误判关键样本;在此基础上,进一步提出课程式动态剪枝策略。该策略通过一维聚类进行样本分组,形成从易到难的课程集合,进而引入滑动窗口机制来动态控制各课程的训练进度,引导模型渐进式学习以有效保障训练过程的稳定性。**结果**　在8个二维图像与三维点云基准数据集上,采用掩码自编码器(masked autoencoder,MAE)、动量对比学习(momentum contrast,MoCo)、点云掩码自编码器(point cloud masked autoencoder,PointMAE)等主流预训练框架,结合视觉Transformer(vision transformer,ViT)、残差网络(residual network,ResNet)等骨干网络开展了5组系统的对比实验。实验结果显示,在更高的数据剪枝率下,本文方法的性能全面优于现有的动态剪枝方法。同时,与全量数据训练策略相比,本文方法在下游任务上取得了近乎无损的性能。消融实验也验证了本文方法中各核心组件的有效性。**结论**　本文方法通过指数损失驱动的样本重要性度量与课程式动态剪枝策略,有效突破了现有数据剪枝方法在自监督学习场景的应用局限,在性能近乎无损的前提下大幅提升了自监督预训练效率(在ImageNet-1K数据集上最高可剪枝30.6%的训练数据),为视觉模型的高效自监督训练提供了可靠的技术支撑。

关键词:自监督学习(SSL);动态数据剪枝;动态课程学习;指数损失表征;高效预训练

Curriculum-driven dynamic data pruning for self-supervised learning

Liu Wenhao[1], Zhang Zhaoming[2], Zhang Ye[1], Guo Yulan[1*]

1. School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen　518000, China;

2. Xi'an Modern Control Technology Research Institute, Xi'an　710000, China

Abstract: **Objective** Self-supervised learning (SSL) has become a cornerstone of modern representation learning, allowing models to extract powerful features from vast unlabeled data. SSL-based methods have achieved performance comparable to or even exceeding fully supervised paradigms across various downstream tasks. This success has driven the development of large-scale vision models but often comes at the cost of immense computational resources, posing substantial barriers to research and practical applications. Data pruning, which aims to enhance training efficiency by removing less informative samples, provides a viable solution to this efficiency issue. Although dynamic data pruning methods are conceptually suited for the label-free nature of SSL, they face a fundamental challenge. Specifically, these methods typically

收稿日期:2025-11-07;**修回日期**:2026-03-20;**预印本日期**:2026-03-27

*通信作者:郭裕兰　guoyulan@sysu.edu.cn

基金项目:国家自然科学基金项目(62372491);航空科学基金项目(2025L0150M1001)

Supported by: National Natural Science Foundation of China(62372491); Aviation Science Foundation of China (2025L0150M1001)

assume that samples with low loss values are "simple" and can be safely pruned. However, this assumption does not hold reliably in the context of SSL. Owing to the complex and inherently stochastic characteristics of SSL tasks, such as masked image modeling with high masking ratios and contrastive learning, the loss associated with individual samples can exhibit extreme and frequent volatility throughout training. Consequently, critical samples might momentarily yield low loss values due to stochastic optimization dynamics and be erroneously and permanently removed from the training set, resulting in irreversible degradation in model performance. Furthermore, prematurely removing genuinely simple samples at the beginning of training can destabilize the learning process, as these samples are crucial for the model to learn fundamental low-level features. This work addresses the inherent conflict between existing dynamic pruning techniques and the intrinsic properties of SSL by introducing a novel pruning framework to ensure a stable, efficient, and lossless pre-training process. **Method** Aiming to address these challenges, curriculum-driven dynamic data pruning (CDDP), a novel framework redesigning the core dynamic pruning pipeline for SSL, is introduced. The methodology is based on two key innovations: a robust sample importance metric and a curriculum-based data scheduling strategy. First, the exponential loss importance (ELI) metric is introduced to address the unreliability of instantaneous loss. Instead of relying on a single, noisy loss value at a specific time step, ELI provides a temporally smoothed and stable assessment of sample importance by calculating the exponential moving average of its historical loss values. This approach effectively removes high-frequency fluctuations and captures the long-term contribution of the sample to learning, thereby preventing important samples from being mistakenly pruned during transient periods of low loss. ELI is highly efficient and introduces minimal computational and memory overhead. Second, inspired by curriculum learning principles, the dynamic curriculum pruning (DCP) strategy is developed. Rather than making simple binary decisions to retain or discard samples, DCP reframes data pruning as a dynamic data scheduling problem, guiding the model from simple to complex samples through a structured learning path. At the start of each epoch, DCP initially performs two steps: 1) applying an efficient 1D k-means clustering algorithm to partition the entire dataset into a predefined number of "courses" based on current ELI scores; and 2) sorting these courses by difficulty and defining a curriculum window that specifies the subset of courses to be used for training in the current epoch. As training progresses, this window gradually shifts from the simplest courses to the most difficult ones, ensuring that the model builds a solid foundation of basic features before transitioning to more complex examples. Aiming to further control the pruning ratio precisely over the training process, the framework incorporates a random pruning mechanism to achieve the target pruning rate. Additionally, an annealing strategy is introduced to deactivate pruning in the final training stages, allowing the model to converge on the full dataset and mitigating potential gradient bias. **Result** Comprehensive evaluations validate the effectiveness, generalizability, and robustness of CDDP across multiple data modalities and learning paradigms. The experiments encompassed 2D image datasets (CIFAR-10, ImageNet-1K) and 3D point cloud datasets (ShapeNet), utilizing various self-supervised pre-training schemes, including masked autoencoder (MAE) with vision transformer (ViT) backbones, momentum contrast (MoCo) with residual network (ResNet) backbones, and point cloud masked autoencoder (PointMAE) for 3D data. Performance was assessed across a wide range of downstream tasks, such as image classification on CIFAR-100, semantic segmentation on ADE20K, 3D object classification on ModelNet40 and ScanObjectNN, and 3D part segmentation on the ShapeNetPart dataset. The proposed CDDP framework was compared against the full-dataset baseline and state-of-the-art dynamic data pruning methods, such as InfoBatch and SeTa. Results unequivocally demonstrate the superiority of the proposed approach. CDDP achieved lossless performance compared with the full-dataset baseline, markedly outperforming competitors on challenging transfer learning and large-scale 2D and 3D tasks. Notably, on the large-scale ImageNet-1K dataset, CDDP matched baseline performance across all downstream tasks while pruning training data by 30.6%. This superior performance extended to the 3D domain, where the framework achieved virtually identical results to the baseline across all downstream tasks while pruning over 32% of the data. Furthermore, detailed ablation studies confirmed the individual contributions of the ELI metric and the DCP strategy, demonstrating that their synergistic combination is crucial for achieving lossless and stable pruned training. **Conclusion** This paper introduces a novel CDDP framework specifically designed for self-supervised learning. Experiment results demonstrate that the proposed method notably improves training efficiency while achieving lossless performance compared with the full-dataset baseline, outperforming existing state-of-the-art pruning methods.

Key words: self-supervised learning (SSL); dynamic data pruning; dynamic curriculum learning; exponential loss metric; efficient pre-training

0 引 言

自监督学习(self-supervised learning,SSL)已成为主流的视觉数据表征学习范式之一(Gui 等,2024)。该技术通过设计预训练任务,从海量无标签数据中挖掘内在结构与语义信息,有效提升模型的通用表征能力(耿伟峰 等,2023;盛振涛 等,2023)。经预训练的模型仅需在下游任务上进行少量样本微调,即可在多个应用场景中达到甚至超越全监督方法的性能(张亚茹 等,2025)。常用的自监督预训练任务包括预测图像的缺失部分(Bao等,2022;He等,2022)、判断图像块的相对位置(Gidaris等,2018)、在特征空间中进行实例对比学习(Chen 等,2020;He 等,2020)等。其中,以掩码自编码器(masked autoencoder,MAE)(He等,2022)为代表的掩码重建方法和以动量对比学习(momentum contrast,MoCo)(He 等,2020)为代表的对比学习方法,已在多项二维视觉任务中取得优于传统有监督训练的性能。自监督学习范式的成功已延伸至三维视觉领域,点云掩码自编码器(point cloud masked autoencoder,Point-MAE)(Pang等,2022)等研究证实了其在点云表征学习中的有效性。

近年来,自监督学习依托互联网海量无标签数据,有力推动了视觉基础模型(vision foundation models)的涌现与发展(Girdhar 等,2023;Kirillov 等,2023)。然而,基于大规模数据的自监督预训练往往伴随极高的计算资源消耗,对硬件设施、训练时长与能源成本提出严苛要求,严重制约了相关学术研究的深入推进与技术成果的落地应用。例如,尽管 DeepSeek-V3 采取了混合专家模型(mixture of experts,MoE)等高效训练策略,其预训练成本仍高达 532.8 万美元(Liu 等,2024)。因此,面向自监督学习开展数据剪枝技术研究,有效提升大规模预训练的效率,具有重要的理论价值与应用意义。

数据剪枝旨在识别并剔除数据集中对模型训练贡献较低的冗余或噪声样本,从而在保持甚至提升模型性能的前提下提升训练效率。现有数据剪枝方法多面向有监督学习任务,主要可分为静态剪枝与动态剪枝两类。静态数据剪枝方法,也称为核心集选择方法,在训练开始前依据几何分布(Sener 和 Savarese,2018)或不确定性(Gal 等,2017)等特定度量,一次性筛选出固定的数据子集。近期研究进一步将核心集优化与数据蒸馏相融合(吴钦浩 等,2026),有效缓解了高剪枝率下的性能损失。然而,此类方法所筛选的数据往往与特定模型架构、数据集和训练策略强耦合,泛化性能受限。相比之下,动态数据剪枝在训练过程中实时评估样本的重要性,并动态调整筛选的数据子集。该类方法将数据筛选与模型优化过程相结合,使剪枝决策能够自适应于模型在训练期间的动态变化。例如,InfoBatch(Qin 等,2023)认为损失值较低的样本对模型更新的贡献较小,可视为“简单样本”,在每个训练步骤中动态计算所有样本损失值的均值作为剪枝阈值,并随机剔除部分损失低于阈值的样本,以 O(1)复杂度实现高效剪枝。该方法在多种模型架构、数据集及有监督训练任务中均取得了近乎无损的剪枝效果,显著降低了训练成本。在此基础上,SeTa(scale efficient training for large dataset)(Zhou 等,2025)同样以损失值作为样本表征,结合随机剪枝与动态滑窗选择算法,在保障性能的前提下实现了更优的剪枝率。

尽管动态数据剪枝已在有监督学习任务中取得显著突破,但将现有方法直接迁移至自监督学习场景仍面临诸多挑战。具体而言,图像重建、对比学习等自监督预训练任务通常具有较高的复杂性和不确定性(Tian 等,2021),导致样本损失值在训练过程中呈现剧烈且频繁的波动。然而,现有动态剪枝方法多直接将低损失样本判定为冗余样本予以剔除,使得部分关键样本因训练过程中的偶然低损失被误剪,进而对模型表征性能造成不可逆损害。此外,现有方法在训练初始阶段即大量剪枝低损失值样本,这种做法可能进一步加剧自监督训练的不稳定性,导致模型收敛困难。综上,InfoBatch 等依赖瞬时损失的方法容易将关键样本误判为简单样本,虽然 SeTa 有效地优化了剪枝率,但仍然难以解决损失波动问题。因此,现有的动态剪枝方法难以直接适配自监督学习范式,亟需探索适配自监督训练特性的数据剪枝思路。

课程学习(curriculum learning, CL)作为一种模拟人类认知过程的训练策略,为解决上述挑战提供了有效思路。该方法由Bengio等人(2009)首次提出,核心思想是按照从易到难的顺序组织和呈现训练数据,从而加速模型收敛并提升泛化性能。在此基础上,自步学习(self-paced learning, SPL)进一步根据模型自身的学习状态动态调整课程难度(Kumar等,2010)。为融合课程学习的先验知识与自步学习的自适应性,后续研究提出了自步课程学习(self-paced curriculum learning, SPCL)方法,通过引入正则化项将样本难度权重与模型损失共同纳入优化过程,实现了对学习进程的稳健控制(Jiang等,2015)。鉴于此,针对训练难度较高的自监督学习任务,引入"由易到难"的自步课程学习思想,对数据实施结构化、动态化管理而非简单的二元剪枝决策,是提升自监督学习中数据剪枝效率与模型表征性能的有效途径。然而,课程学习仅提供"由易到难"的训练理念,缺乏与动态剪枝机制的具体适配方案。为此,本文将该思想具象化为课程式动态剪枝机制:基于样本难度动态划分课程集合,并通过滑动窗口控制课程的参与顺序,实现从易到难的渐进式学习。

针对自监督训练中损失波动剧烈及初期训练不稳定导致的剪枝失效问题,本文提出一种面向自监督学习的动态数据剪枝方法(curriculum-driven dynamic data pruning, CDDP)。首先,为稳健评估样本重要性,提出了指数损失重要性(exponential loss importance, ELI)表征,采用损失的指数移动平均替代瞬时损失,有效滤除短期噪声,为剪枝决策提供可靠依据。其次,基于上述课程学习思想,构建课程式动态剪枝策略(dynamic curriculum pruning, DCP),根据ELI指标将全部样本动态地划分为不同难度的课程,并在训练过程中引导模型从易到难渐进学习,实现数据剪枝与自监督预训练的紧密耦合。本文在涵盖二维图像和三维点云的8个基准数据集上,结合多种预训练策略与骨干网络开展实验验证。结果表明,本文方法在显著提升自监督预训练效率的同时,各项任务性能均与全量数据训练近乎持平,为自监督学习的高效数据剪枝提供了新方案。

1 方法介绍

本文所提出的课程式动态数据剪枝方法CDDP的整体流程如图1所示。首先,将数据输入模型计算各样本损失,并将其转换为ELI表征。随后,对样本进行课程式动态数据分组,并在训练过程中通过由易到难的课程调度,实现课程式动态数据剪枝。下面从指数损失重要性表征、课程式动态剪枝策略以及剪枝率调控与退火策略3个方面对本文方法进行详细阐述。

1.1 指数损失重要性(ELI)表征

1.1.1 训练损失作为数据价值表征的局限性

现有动态数据剪枝算法为兼顾剪枝效率与动态

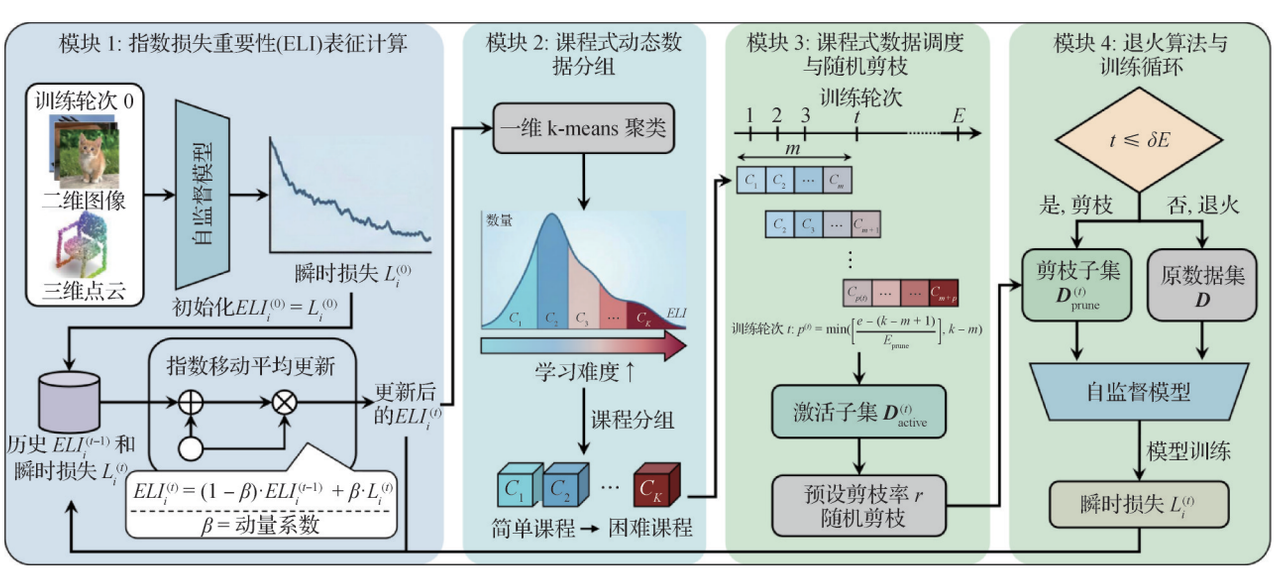


图1 CDDP数据剪枝方法整体流程

Fig. 1 The overall pipeline of the CDDP data pruning framework

适应性，通常采用样本的瞬时训练损失作为数据价值的度量表征，其基本假设是样本损失值可有效反映其学习难度。在有监督学习任务中，优化目标明确且训练过程相对稳定，整体损失曲线呈现平滑且单调下降的趋势，因此样本的瞬时损失能够相对可靠地表征其在当前阶段的冗余程度。而在自监督学习中，预训练任务的复杂性较高（如高比例掩码重建、困难负样本挖掘等），导致优化过程稳定性差，尤其是单个样本的损失在训练过程中呈现高频且剧烈的波动特征，导致训练损失作为数据价值表征的可靠性与稳定性大幅下降。

为直观验证该问题，图 2 展示了分别采用有监督训练与自监督训练框架在 ImageNet-1K 数据集上训练 ResNet50（residual network）模型得到的损失曲线对比结果。可以观察到，与有监督训练相比，自监督训练的损失曲线表现出更频繁且剧烈的波动特征。这种训练过程的不稳定性，导致样本在任意训练轮次的瞬时损失值无法可靠表征其真实学习价值。对模型表征学习至关重要的样本，可能因优化过程的随机性而偶然出现低损失值，进而被误判为“简单样本”而错误剔除，最终造成模型整体性能的不可逆退化。因此，亟需设计一种鲁棒的样本价值度量表征，实现对自监督学习中样本学习价值的稳定、准确评估。

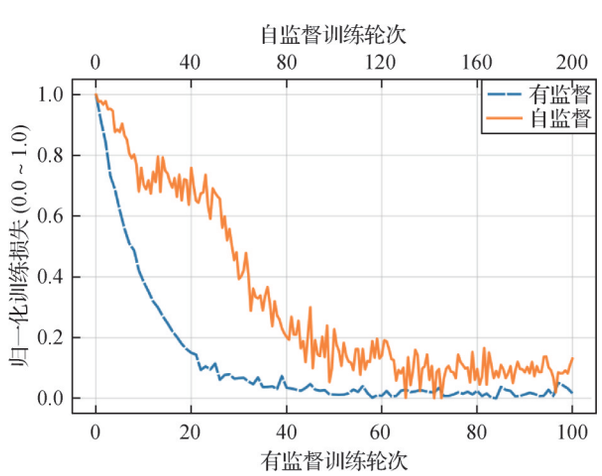


图 2　有监督与自监督训练的损失曲线对比

Fig. 2　Comparison of loss curves between supervised and self-supervised training

1.1.2　用于评估样本自监督学习价值的 ELI 表征

为在自监督学习中稳健度量样本重要性，本文提出指数损失重要性表征 ELI。其核心思想是引入时序信息，通过计算样本损失的指数移动平均值，构建对高频波动与瞬时扰动具有鲁棒性的样本价值表征。ELI 利用时序平滑特性滤除损失曲线中的短期噪声，从而更准确地反映样本在整个训练过程中的长期学习贡献，提升样本学习价值评估的可靠性。

对于数据集 $S=\{s_1,s_2,\cdots,s_N\}$ 中的任意样本 s_i，其在训练迭代步骤 t 的 ELI 值，记为 $ELI_i^{(t)}$，计算为

$$ELI_i^{(t)}=(1-\beta)\cdot L_i^{(t)}+\beta\cdot ELI_i^{(t-1)} \tag{1}$$

式中，$L_i^{(t)}$ 为样本 s_i 在第 t 次迭代的损失值；β 为动量系数，是一个取值于区间 $[0,1)$ 的超参数。β 的值决定了历史信息对当前 ELI 值的影响程度。β 越接近 1，历史信息的权重越大，ELI 曲线越平滑，但对损失变化的响应越慢；β 越接近 0，ELI 值越接近当前损失值。特别地，在初始步骤 $t=0$ 时，将 ELI 初始化为初始损失值，即

$$ELI_i^{(0)}=L_i^{(0)} \tag{2}$$

对于被剪枝的样本 s_j，因当前轮次无法获取其最新损失值，本文参考 InfoBatch 方法的损失值继承策略，直接继承上一轮次的 ELI 值，即

$$ELI_j^{(t)}=ELI_j^{(t-1)} \tag{3}$$

与瞬时损失相比，ELI 表征在样本重要性评估方面具有显著优势。首先，ELI 通过引入历史信息，为样本重要性评估提供一个更长的时间窗口，有效规避单次迭代损失抖动带来的干扰；其次，ELI 的计算效率较高，对于每个样本仅需存储上一轮 ELI 值，并执行基本的加法和乘法运算。与 SeTa 等直接依赖瞬时损失的方法不同，ELI 避免了因偶然的低损失数值而误剪关键样本。

1.2　课程式动态剪枝策略（DCP）

1.2.1　阈值剪枝策略在自监督训练中的适配问题

自监督学习存在收敛速度慢、优化难度大的固有问题，且该问题随模型规模增大而愈发显著（Tian 等，2021）。在此背景下，若沿用有监督学习中的阈值剪枝策略，将低损失样本简单视为“冗余”并在早期进行剔除，将进一步加剧训练的不稳定性。这是由于训练初期的低损失样本通常包含清晰的边缘、纹理等基础结构信息，是模型学习通用特征表示、构建层次化语义理解的重要基石。过早筛除这些样本会破坏基础特征的学习过程，导致模型收敛困难甚至性能下降（Wang 等，2024）。

针对上述问题，受课程学习的核心思想启发

(Bengio等,2009),本文提出了课程式动态剪枝策略DCP。该策略根据样本学习难度与模型实时学习状态,自适应地将样本划分为不同难度的课程,引导模型遵循从易到难的顺序完成课程式学习。与InfoBatch等方法的阈值剪枝不同,DCP摒弃了静态的"保留—剔除"二元剪枝决策,将数据剪枝重构为动态调度问题,在保留基础样本的同时逐渐地引入困难样本,避免因早期过度剪枝引发的训练不稳定。

1.2.2 DCP策略构建:动态分组与课程调度

DCP策略主要包含两个核心步骤:基于ELI的动态数据分层与时序课程调度。该策略在每个训练轮次开始时动态执行,以适应模型的最新学习状态。

1)基于ELI的动态数据分层。在每个训练轮次e开始时,首先计算数据集中所有N个样本的最新ELI值,得到集合$\boldsymbol{ELIs}$,即

$$\boldsymbol{ELIs}=\{ELI_0^{(e)},ELI_1^{(e)},\cdots,ELI_N^{(e)}\} \tag{4}$$

为了根据难度对数据进行分组,采用一维k-means聚类算法,将所有样本按其ELI值划分为k个簇,即

$$\boldsymbol{C}_1^{(e)},\boldsymbol{C}_2^{(e)},\cdots,\boldsymbol{C}_k^{(e)}=f_{\text{k-means}}(\boldsymbol{ELIs},k) \tag{5}$$

式中,k为超参数,具体设置见2.1节。

随后,根据各簇中心点对这k个簇进行升序排序,使得簇$\boldsymbol{C}_j^{(e)}$中心点值小于$\boldsymbol{C}_{j+1}^{(e)}$中心点值。由此得到按学习难度从易到难排列的"课程列表",其中$\boldsymbol{C}_1^{(e)}$为当前最简单的课程,$\boldsymbol{C}_k^{(e)}$为最困难的课程。

2)时序课程调度。获得有序课程列表后,DCP并非在所有训练阶段使用全部课程,而是定义了一个关于课程窗口大小的超参数$m(1\leqslant m\leqslant k)$,表示任意时刻模型需要学习的课程数量。随着训练的进行,该窗口将从最简单的课程开始,逐步向更困难的课程滑动。

具体而言,设总剪枝轮次数为E_{prune},则课程窗口的起始索引$p^{(e)}$在训练轮次e时的计算方法为

$$p^{(e)}=\min\left(\left\lfloor\frac{e\cdot(k-m+1)}{E_{\text{prune}}}\right\rfloor,k-m\right) \tag{6}$$

式(6)确保了在训练初期(e较小时),$p^{(e)}$为0,模型学习最简单的课程;随着e增长,$p^{(e)}$逐步增加,窗口向右滑动;训练末期,$p^{(e)}$达到最大值$k-m$,模型专注于学习最困难的m个课程。

确定起始索引$p^{(e)}$后,当前轮次激活训练的数据子集$\boldsymbol{D}_{\text{active}}^{(e)}$由窗口内的所有课程构成,具体为

$$\boldsymbol{D}_{\text{active}}^{(e)}=\bigcup_{j=p^{(e)}+1}^{p^{(e)}+m}\boldsymbol{C}_j^{(e)} \tag{7}$$

1.3 剪枝率调控与退火策略

在DCP完成宏观数据调度的基础上,CDDP进一步引入两种辅助机制,实现对训练过程的精细化控制。

首先,为精确控制剪枝率,CDDP借鉴SeTa的随机剪枝策略,对经课程调度筛选出的激活数据子集$\boldsymbol{D}_{\text{active}}^{(e)}$按预设比例$r$随机剔除部分样本,以满足预期的剪枝率要求。本文参考SeTa方法,定义动态剪枝算法的全局剪枝率$ratio$为

$$ratio=1-\frac{\sum_{i=1}^{E}M^{(i)}}{N\cdot E} \tag{8}$$

式中,N为数据集样本总数,E为总训练轮次数,$M^{(i)}$为第i轮剪枝后实际参与训练的样本数量。

其次,为消除剪枝在训练末期可能引入的梯度偏差(Qin等,2023),CDDP引入退火策略,参考InfoBatch的做法,将动态剪枝过程(包括课程调度与随机剪枝)限定在训练的前$\delta\cdot E$个轮次内执行,其中δ为退火系数。此后,所有剪枝机制暂停,模型恢复在全量数据集上训练直至收敛。上述两种机制共同保障CDDP在提升训练效率的同时,维持模型的最终性能与训练稳定性。

2 实验结果与分析

2.1 实验设置

为全面验证本文所提CDDP数据剪枝方法的有效性与泛化能力,本文在涵盖二维图像与三维点云的8个基准数据集上开展了系列评估实验。

在预训练阶段,采用的二维数据集为CIFAR-10(Canadian institute for advanced research-10)(Krizhevsky,2009)与ImageNet-1K(Deng等,2009),三维数据集为ShapeNet(Chang等,2015)。针对二维数据,分别采用MAE(He等,2022)框架结合视觉Transformer(vision Transformer,ViT)(Dosovitskiy等,2021)骨干网络,以及动量对此学习(momentum contrast,MoCo)(He等,2020)框架结合ResNet(He等,2016)骨干网络进行自监督预训练;针对三维数据,则采用基于重建的PointMAE(Pang等,2022)自监督学习框架进行预训练。

在下游任务评估阶段，在预训练的数据集上进行微调，并在图像分类数据集 CIFAR-100（Krizhevsky，2009）、图像语义分割数据集 ADE20K（Zhou 等，2017）、三维物体分类数据集 ModelNet40（Wu 等，2015）与 ScanObjectNN（Uy 等，2019）、三维部件分割数据集 ShapeNetPart（Yi 等，2016）上进行多任务迁移学习微调，以全面衡量预训练模型的表征能力。此外，为验证 CDDP 方法的鲁棒性与泛化性，在 CIFAR-10 和 CIFAR-100 数据集上使用 ResNet-18 骨干网络进行了有监督条件下的数据剪枝实验。

对比方法包括不进行任何剪枝的全量数据训练基线（baseline），以及当前最先进的动态数据剪枝方法 InfoBatch（Qin 等，2023）与 SeTa（Zhou 等，2025）。由于本文方法与对比方法均为动态剪枝策略，实际剪枝率动态变化且难以精确保持一致。为确保对比公平，在所有无损剪枝率实验中，对 InfoBatch 和 SeTa 的超参数进行精细调优，使其剪枝率略低于本文方法，即其参与训练的数据量略多于本文方法，以在更严苛条件下验证本文方法的性能优势。所有动态数据剪枝方法仅作为即插即用模块在预训练阶段使用。在 ImageNet-1K 和 ShapeNet 数据集上，预训练参数细节（如优化器、学习率等）均严格遵循 MAE、MoCo 与 PointMAE 的官方设定。由于在 CIFAR-10 上 MoCo 和 MAE 预训练策略缺少官方实现代码，本文根据数据集规模选择合适的模型规模、优化器与学习率等设定。此外，ImageNet-1K 的相关实验在一台配备 8 块 NVIDIA 3090 显卡的服务器上进行，其余实验则在配备单块 NVIDIA 3090 显卡的计算机上完成。

本文方法涉及的超参数主要包括 ELI 的动量系数 β，课程数量 k、课程窗口 m、随机剪枝的剪枝率 r 以及退火系数 δ。所有实验中取 $\beta = 0.1$，$\delta = 0.875$。自监督实验中取 $k = 10$ 和 $m = 8$；有监督实验中取 $k = 5$ 和 $m = 4$。剪枝率 r 根据不同实验需求进行调整。

2.2 对比实验

为系统评估 CDDP 策略的综合性能，从 3 个维度设计对比实验：1）在固定剪枝率下验证其性能优势；2）在保持性能近乎无损的剪枝条件下验证其效率与性能的平衡能力；3）将有监督学习作为泛化性测试场景，验证其跨任务适用性。实验涵盖 2D/3D 数据、小/大规模数据集、分类/分割等多种任务类型，并将 CDDP 与全量数据训练基线（baseline）及当前可适配自监督场景的 InfoBatch、SeTa 进行对比。

对比方法选择说明：现有静态剪枝方法通常依赖几何分布或不确定性度量，剪枝结果与特定模型架构、数据集强耦合，难以在多种数据模态和网络结构间泛化（Qin 等，2023）。动态剪枝方法中，UCB（upper confidence bound）（Raju 等，2021）、Kakurenbo（Nguyen 等，2023）、Dyn-Unc（dynamic-uncertainty）（He 等，2024）等多数方法均需利用标签信息计算样本重要性指标（如梯度范数、置信度等），无法直接迁移至无标签的自监督学习场景。相比之下，InfoBatch（Qin 等，2023）和 SeTa（Zhou 等，2025）仅依赖样本损失值作为重要性度量，不涉及标签信息，可有效适配不同数据模态、网络架构及自监督预训练任务。因此，本文选取 InfoBatch 和 SeTa 作为主要对比方法，以验证 CDDP 在自监督剪枝场景下的有效性。

2.2.1 固定剪枝率下不同剪枝方法的自监督训练实验结果对比

首先在小型数据集 CIFAR-10 上进行预训练，参考 SeTa 的实验设定，在 30%、50% 和 70% 三种剪枝率下开展对比实验。上述剪枝率设置为数据剪枝领域的常用选择，旨在全面评估不同方法在中等及极端剪枝强度下的综合性能，为综合对比提供可靠依据。考虑到数据集规模较小，本文采用了 MAE 框架结合 ViT-Tiny 骨干网络，以及 MoCo 框架结合 ResNet-18 骨干网络进行自监督预训练。其中，MAE 框架预训练 800 轮，MoCo 框架预训练 200 轮。由于动态数据剪枝方法的实际剪枝率难以精确设定（如式（8）所示），本文对 CDDP 及对比方法的剪枝率进行精细调节以匹配预期剪枝率。下游任务评估中，本文参考 InfoBatch 的设置，对预训练模型进行 8 次微调，并汇报分类精度的均值和标准差。表 1 和表 2 分别展示了在 CIFAR-10 和 CIFAR-100 数据集上各方法的预训练模型的下游任务实验结果。

在 CIFAR-10 数据集上进行 100 轮微调。实验结果表明，在 30% 的剪枝率下，CDDP 在下游分类任务上取得与 baseline 相当的性能（93.9%/95.4% vs. 93.8%/95.4%），而 InfoBatch 和 SeTa 均出现一定性能损失（93.3%/95.1%，93.5%/95.2% vs. 93.8%/95.4%）。在 50% 的剪枝率下，CDDP 仍保持与 baseline 相当的性能（93.7%/95.3% vs. 93.8%/95.4%），InfoBatch 和 SeTa 的性能损失进一步扩大（92.4%/

94.9%, 93.3%/95.1% vs. 93.8%/95.4%)。在70%的剪枝率下,虽然所有方法相较baseline均出现性能下降,但CDDP仍优于对比方法(93.0%/95.3% vs. 91.8%/94.8%, 92.5%/95.0%)。

在CIFAR-100数据集上进行100轮迁移学习微调。在30%的剪枝率下,CDDP实现近乎无损性能(74.9%/77.8% vs. 75.0%/77.7%),对比方法则出现明显性能下降(74.3%/77.1%, 74.2%/77.2% vs. 75.0%/77.7%)。在更具挑战的50%和70%的剪枝率下,虽然所有剪枝方法均出现性能损失,但CDDP仍显著优于对比方法(50%剪枝率:74.7%/77.3% vs. 72.2%/76.6%, 73.8%/77.0%; 70%剪枝率:73.5%/77.0% vs. 70.9%/76.3%, 72.9%/76.6%)。上述结果表明,CDDP能够有效实现自监督预训练的数据剪枝,在提升训练效率的同时保持良好的模型表征能力与迁移能力。

表1 在CIFAR-10数据集上不同数据剪枝方法的预训练模型的准确率微调结果

Table 1 Fine-tuning accuracy on the CIFAR-10 dataset for models pre-trained with different data pruning methods

/%

方法	预训练框架MAE			预训练框架MoCo		
	剪枝率=30%	剪枝率=50%	剪枝率=70%	剪枝率=30%	剪枝率=50%	剪枝率=70%
baseline(无剪枝)	–	93.8±0.1	–	–	95.4±0.1	–
InfoBatch(Qin等,2023)	93.3±0.1	92.4±0.2	91.8±0.2	95.1±0.2	94.9±0.2	94.8±0.3
SeTa(Zhou等,2025)	93.5±0.1	93.3±0.1	92.5±0.2	95.2±0.2	95.1±0.1	95.0±0.1
CDDP(本文)	**93.9**±0.1	**93.7**±0.2	**93.0**±0.1	**95.4**±0.1	**95.3**±0.1	**95.3**±0.1

注:加粗字体表示各列除去baseline的最优结果;baseline方法汇报了使用全部数据进行预训练的微调结果,"–"表示无结果。

表2 在CIFAR-100数据集上不同数据剪枝方法的预训练模型的准确率微调结果

Table 2 Fine-tuning accuracy on the CIFAR-100 dataset for models pre-trained with different data pruning methods

/%

方法	预训练框架MAE			预训练框架MoCo		
	剪枝率=30%	剪枝率=50%	剪枝率=70%	剪枝率=30%	剪枝率=50%	剪枝率=70%
baseline(无剪枝)	–	75.0±0.1	–	–	77.7±0.2	–
InfoBatch(Qin等,2023)	74.3±0.1	72.2±0.2	70.9±0.4	77.1±0.2	76.6±0.3	76.3±0.3
SeTa(Zhou等,2025)	74.2±0.1	73.8±0.3	72.9±0.1	77.2±0.1	77.0±0.2	76.6±0.4
CDDP(本文)	**74.9**±0.1	**74.7**±0.2	**73.5**±0.1	**77.8**±0.2	**77.3**±0.2	**77.0**±0.3

注:加粗字体表示各列除去baseline的最优结果;baseline方法汇报了使用全部数据进行预训练的微调结果,"–"表示无结果。

2.2.2 无损剪枝条件下不同剪枝方法的自监督训练实验结果对比

除固定剪枝率实验外,本文进一步在无损剪枝条件下对各方法进行对比。由于动态剪枝算法的实际剪枝率受超参数影响而难以精确预设,本文参考InfoBatch和SeTa等工作的实验设置,以0.05为步长对剪枝率超参数r进行网格搜索,在保持下游性能近乎无损前提下为CDDP方法选取最大的剪枝率。无损性能是数据剪枝研究的核心追求,在此条件下对比不同方法的性能表现,能够更客观地评估各方法在效率与精度平衡方面的实际能力。基于上述设定,本文在多个数据集上开展实验以验证CDDP的性能优势。

1)CIFAR-10数据集上预训练实验结果。与2.2.1节相同,本文首先在CIFAR-10数据集上进行预训练,除剪枝率超参数r外,其余配置完全一致。其中,MAE框架取$r=0.1$,MoCo框架取$r=0.2$。为确保对比公平,实验中使CDDP的实际剪枝率始终略高于InfoBatch和SeTa,即CDDP在训练数据更少的条件下进行模型训练。表3展示了两种预训练策略在下游任务上的实验结果。

在CIFAR-10数据集上的评估结果表明,与

表 3　在 CIFAR-10 和 CIFAR-100 数据集上不同数据剪枝方法的预训练模型的准确率微调结果

Table 3　Fine-tuning accuracy on the CIFAR-10 and CIFAR100 datasets for models pre-trained with different data pruning methods

方法	预训练框架 MAE/ViT-Tiny			预训练框架 MoCo/ResNet-18		
	剪枝率/%	CIFAR-10	CIFAR-100	剪枝率/%	CIFAR-10	CIFAR-100
baseline（无剪枝）	0	93.8±0.1	75.0±0.1	0	95.4±0.1	77.7±0.2
InfoBatch（Qin 等，2023）	25.8	93.4±0.1	74.3±0.1	36.5	95.1±0.2	77.1±0.2
SeTa（Zhou 等，2025）	26.0	93.5±0.1	74.2±0.1	36.4	95.2±0.2	77.2±0.1
CDDP（本文）	26.2	**93.9**±0.1	**75.0**±0.1	36.7	**95.4**±0.1	**77.8**±0.2

注：加粗字体表示各列除去 Baseline 的最优结果。

baseline 相比，CDDP 在下游分类任务上达到性能无损（93.9%/95.4% vs. 93.8%/95.4%），而 InfoBatch 和 SeTa 均出现一定性能损失（93.4%/95.1%，93.5%/95.2% vs. 93.8%/95.4%）。这说明即使在小型数据集上，CDDP 也能有效剔除冗余数据，同时保持模型的表征能力。

在更具挑战性的 CIFAR-100 迁移学习分类任务上，CDDP 同样实现近乎无损性能（75.0%/77.8% vs. 75.0%/77.7%），而对比方法在相近剪枝率下均出现明显精度下降（74.3%/77.1%，74.2%/77.2% vs. 75.0%/77.7%）。上述结果进一步表明，在追求无损剪枝的严苛条件下，CDDP 仍能稳定发挥优势，在提升自监督预训练效率的同时，有效保持了模型性能。

2）ImageNet-1K 数据集上预训练的实验结果。为验证 CDDP 策略在大规模数据集上的可扩展性，本文在 ImageNet-1K 上开展自监督预训练实验。具体采用 MAE 框架结合 ViT-Base 骨干网络，以及 MoCo 框架结合 ResNet-50 骨干网络进行预训练。MAE 框架遵循原文设置预训练 800 轮，剪枝率超参数 $r = 0.0$；MoCo 框架同样遵循原文设置预训练 200 轮，剪枝率 $r = 0.15$。与前述实验一致，实验中使 CDDP 的实际剪枝率略高于其他对比方法。

表 4 展示了两种预训练框架下，不同剪枝方法在 ImageNet-1K 数据集上多个下游任务进行 100 轮全参数微调的结果。可以看出，CDDP 取得与 baseline 近乎无损的性能（83.3%/76.9% vs. 83.3%/77.0%），且全面优于 InfoBatch 和 SeTa（83.3%/76.9% vs. 83.1%/76.5%，83.1%/76.3%）。此外，本文通过冻结预训练参数仅微调线性分类层的方式评估模型表征能力（Chen 等，2020）。经过 100 轮线性分类层微调，CDDP 取得与 baseline 相当的结果（65.1%/60.2% vs. 65.4%/60.6%），并显著领先 InfoBatch 和 SeTa（65.1%/60.2% vs. 64.9%/59.5%，64.7%/59.1%）。上述结果表明，CDDP 能够适配大规模数据的自监督剪枝需求，有效保持预训练模型的表征能力。

为进一步评估预训练模型的泛化能力，本文基于 mmsegmentation 框架，采用 UperNet（Xiao 等，2018）分割模型，在 ADE20K 数据集上进行图像语义分割迁移学习实验。CDDP 在该任务上的平均交并比（mean intersection-over-union，mIoU）指标超越所有对比方法，甚至略优于全量数据预训练的 baseline 模型（46.4%/41.1% vs. 46.4%/40.9%）。这进一步验证了 CDDP 训练所得模型具有良好的泛化性与可迁移性。

3）ShapeNet 数据集上的预训练实验结果。为验证本文方法在不同数据模态上的泛化能力，本文使用 PointMAE 框架在三维点云数据集 ShapeNet 上进行了 300 轮预训练。CDDP 的剪枝率超参数设置为 $r = 0.2$。CDDP 与各对比方法均作为即插即用模块集成至预训练阶段，其余设置严格遵循 PointMAE 原文，实验结果如表 5 所示。

本文在 ModelNet40 和 ScanObjectNN 数据集上进行 300 轮微调。其中，ModelNet40 实验使用 1 024 个点的坐标信息进行微调；ScanObjectNN 选取最困难的 PB-T50-RS 模式进行评估。由表 5 可知，在剪枝率略高于 InfoBatch 和 SeTa 的条件下，CDDP 在两个分类下游任务上均展现出明显的性能优势。在 ModelNet40 上，CDDP 达到与全量数据训练基线近乎无损的分类性能（93.7% vs. 93.8%），且优于 InfoBatch 和 Seta 两种对比方法（93.7% vs. 93.0%，

表4　在ImageNet-1K数据集上不同数据剪枝方法的预训练模型的微调实验结果

Table 4　Fine-tuning results on the ImageNet-1K dataset for models pre-trained with different data pruning methods

/%

方法	MAE/ViT-Base框架				MoCo/ResNet-50框架			
	剪枝率/%	ACC IN1K微调	ACC IN1K冻结	mIoU ADE20K	剪枝率/%	ACC IN1K微调	ACC IN1K冻结	mIoU ADE20K
baseline(无剪枝)	0	83.3	65.4	46.4	0	77.0	60.6	40.9
InfoBatch(Qin等,2023)	17.2	83.1	64.9	46.2	29.5	76.5	59.5	40.5
SeTa(Zhou等,2025)	17.6	83.1	64.7	46.2	30.1	76.3	59.1	40.6
CDDP(本文)	17.6	**83.3**	**65.1**	**46.4**	**30.6**	**76.9**	**60.2**	**41.1**

注:加粗字体表示各列除去baseline的最优结果。

表5　在ShapeNet数据集上不同数据剪枝方法基于PointMAE框架得到的预训练模型的微调实验结果对比

Table 5　Comparison of fine-tuning results on the ShapeNet dataset for models pre-trained with different data pruning methods under the PointMAE framework

方法	剪枝率/%	ACC/% ModelNet40	ACC/% ScanObjectNN	mIoU/% shapeNetPart
baseline	0	93.8	85.18	86.1
InfoBatch(Qin等,2023)	32.3	93.0	84.77	**86.1**
SeTa(Zhou等,2025)	32.2	93.4	84.51	**86.1**
CDDP(本文)	32.5	**93.7**	**85.10**	**86.1**

注:加粗字体表示各列除去baseline的最优结果。

93.4%);在更具挑战性的ScanObjectNN上,CDDP同样与基线性能基本持平(85.10% vs. 85.18%),而InfoBatch和SeTa则出现明显性能下降(84.77%,84.51% vs. 85.18%)。上述结果表明,CDDP在三维点云数据的自监督学习场景下同样具备良好的泛化能力,能够在多种迁移学习任务中实现近乎无损的剪枝。

为进一步验证预训练模型在其他下游任务上的性能,本文在ShapeNetPart数据集上进行300轮部件分割微调。结果显示,CDDP训练模型的实例mIoU达到86.1%,与基线方法完全一致。这进一步证明,经CDDP剪枝预训练的自监督模型在不同三维下游任务中能够保持稳定的表征能力与近乎无损的性能。

2.2.3　固定剪枝率下不同剪枝方法的有监督训练实验结果对比

为进一步验证CDDP的鲁棒性与泛化能力,本文将其拓展至有监督学习场景进行测试。参考InfoBatch的实验设定,使用ResNet-18模型在CIFAR-10数据集上进行200轮有监督训练,并在30%、50%和70%三种固定剪枝率下开展对比实验,进而评估不同剪枝方法所训练模型的分类性能。该实验旨在证明CDDP虽然针对自监督场景设计,但在有监督任务中同样具备良好的适用性与有效性。

由表6可知,CDDP在有监督数据剪枝任务中仍表现出良好的适应性与泛化能力。在30%剪枝率下,CDDP取得95.7%的分类准确率,与全量数据训练基线(95.6%)性能相当,且优于或与其他对比方法持平。当剪枝率提升至50%和70%时,尽管模型性能出现一定下降,CDDP仍优于或持平动态随机剪枝策略及其他对比方法。尤其在70%高剪枝率下,CDDP训练模型的准确率为95.0%,与InfoBatch(94.7%)和SeTa(95.0%)基本持平。上述结果表明,CDDP不仅在自监督学习中有效,在有监督训练场景下同样具备良好的通用性与鲁棒性,验证了其作为通用数据调度方法的潜力。

表 6　在 CIFAR-10 数据集上不同数据剪枝方法基于 ResNet-18 的有监督训练实验结果对比

Table 6　Comparison of supervised training results on the CIFAR-10 dataset using ResNet-18 with different data pruning methods

方法	ACC/%		
	剪枝率 = 30%	剪枝率 = 50%	剪枝率 = 70%
baseline（无剪枝）	–	95.6	–
动态随机剪枝	94.8	94.5	93.0
ϵ-greedy（Raju 等，2021）	95.2	94.9	94.1
UCB（Raju 等，2021）	95.3	94.7	93.9
InfoBatch（Qin 等，2023）	95.6	95.1	94.7
SeTa（Zhou 等，2025）	**95.7**	**95.3**	**95.0**
CDDP（本文）	**95.7**	**95.3**	**95.0**

注：加粗字体表示各列除去 baseline 的最优结果；baseline 方法汇报了使用全部数据进行预训练的微调结果，“–”表示无结果。

2.3　消融实验与参数分析

2.3.1　实验设置

为验证 CDDP 中各组件的有效性及关键超参数的影响，本文设计了两部分实验。实验设置与 2.2.1 节一致，均在 CIFAR-10 数据集上基于 MAE 框架对 ViT-Tiny 骨干网络进行自监督预训练。为验证 ELI 表征与 DCP 策略各自的独立贡献，在 CDDP 方法基础上，首先将 ELI 表征替换为瞬时损失进行数据剪枝，构成“损失值 + DCP”变体，以验证 ELI 表征的有效性；其次保留 ELI 表征，将 DCP 策略替换为阈值剪枝策略（Qin 等，2023），构成“ELI + 阈值剪枝”变体，以验证 DCP 策略的有效性。实验结果如表 7 所示。此外，为探究动量系数 β、课程数量 k、课程窗口 m 及退火系数 δ 对剪枝率与下游任务精度的影响，进行敏感性分析，相关结果见 2.3.4 节。

2.3.2　ELI 表征的消融实验分析

与将 ELI 表征替换为瞬时损失的“损失值 + DCP”变体相比，完整 CDDP 方法在下游任务上的性能提升 0.2%/0.9%。这一结果表明，通过时序平滑获取稳健的样本价值表征，能够有效提升自监督模型的剪枝效果。如 2.5.1 节可视化分析所示，ELI 能有效平滑训练过程中的损失抖动，避免样本因瞬时损失下降而被误剪。上述结果表明，在训练稳定性较差的自监督学习中，ELI 能够为样本提供一致且鲁棒的学习价值评估，为数据剪枝提供可靠依据，有效降低因损失波动导致的关键样本误剪风险。

表 7　CDDP 方法的 ELI 表征和 DCP 策略的消融实验结果

Table 7　Ablation results of the ELI metric and DCP strategy in the CDDP method

方法	剪枝率/%	ACC/%	
		CIFAR-10	CIFAR-100
baseline	0	93.8	75.0
InfoBatch（Qin 等，2023）	49.9	92.4	72.2
损失值 + DCP	50.1	93.5	73.8
ELI + 阈值剪枝	50.0	93.2	73.1
完整 CDDP	50.1	93.7	74.7

2.3.3　动态课程剪枝策略的消融实验分析

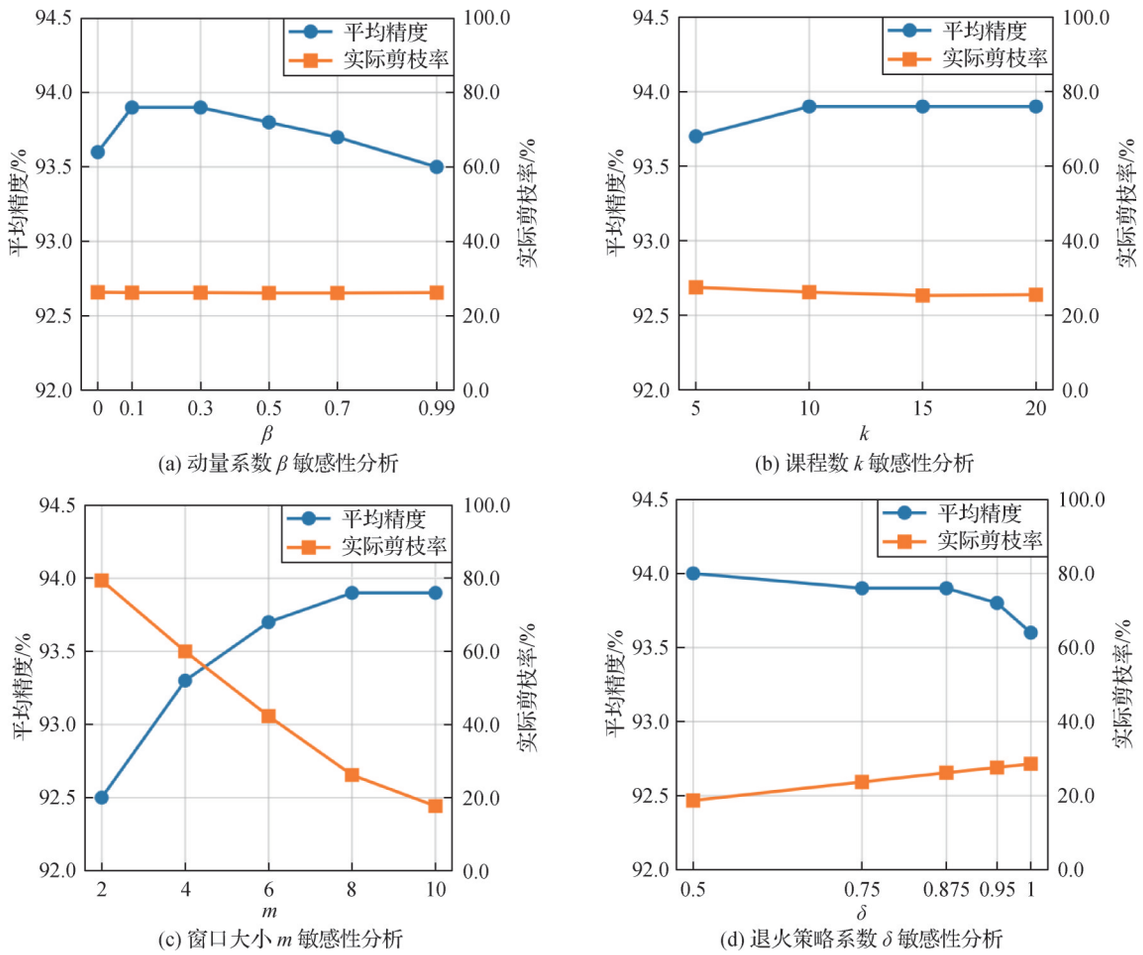
与保留 ELI 表征但采用 InfoBatch 阈值剪枝策略的“ELI + 阈值剪枝”变体相比，完整 CDDP 方法展现出更显著的性能优势，精度提升达 0.5%/1.6%。这一结果表明，相较于简单剔除低损失样本，采用从易到难的课程学习范式对训练数据进行动态调度，对于维持自监督学习中的模型稳定性与表征质量具有更为关键的作用。如 2.5.2 节可视化分析所示，DCP 依据样本 ELI 值实现的有效难度分层，为上述渐进式学习路径提供了直观佐证。

进一步与基于瞬时损失和阈值剪枝的 InfoBatch 方法相比，尽管单独使用 ELI 表征或单独使用 DCP 课程式动态剪枝策略均能带来一定性能提升，但其结果与全量数据训练的基线水平仍存在差距。而当 ELI 表征与 DCP 策略结合使用时，完整 CDDP 方法的性能不仅远超所有消融变体，且在 50.1% 剪枝率下，下游任务精度与全量数据训练的 baseline 基本持平。这一结果表明，ELI 表征与 DCP 策略之间存在协同增效作用，二者的有机结合是实现高效且无损自监督预训练剪枝的关键。

2.3.4　关键超参数敏感性分析

为探究关键超参数对 CDDP 性能的影响，在 CIFAR-10 的下游分类任务上进行敏感性分析，考察动量系数 β、课程数量 k、课程窗口 m 及退火系数 δ 对剪枝率与分类精度的影响，实验结果如图 3 所示。

1）动量系数 β。动量系数 β 控制 ELI 对历史信息的依赖程度。如图 3（a）所示，β 对剪枝率的影响



(a) 动量系数 β 敏感性分析　(b) 课程数 k 敏感性分析

(c) 窗口大小 m 敏感性分析　(d) 退火策略系数 δ 敏感性分析

图3　CDDP关键超参数敏感性分析

Fig. 3　Sensitivity analysis of key hyperparameters in CDDP((a)sensitivity analysis of momentum coefficient β;(b)sensitivity analysis of the number of curricula k;(c)sensitivity analysis of window size m;(d)sensitivity analysis of annealing coefficient δ)

较小,但对下游精度具有显著影响:β过小(< 0.1)时ELI易受瞬时波动干扰,导致精度下降;β过大(> 0.3)时对损失变化响应迟缓,同样引起精度损失。实验表明,β取值在[0.1,0.3]范围内可使精度与baseline基本持平。

2)课程数量k与窗口大小m。课程数量k决定课程划分粒度。如图3(b)所示,固定窗口与课程数比例m/k = 0.8以保持学习密度不变时,剪枝率未受明显影响。但当k较小(≤ 5)时,难以有效区分样本难度层级,导致精度下降;而k较大(≥ 15)时,虽然精度保持稳定,但会增加一维k-means聚类的计算负担。

窗口大小m控制每轮参与训练的课程数。如图3(c)所示,固定k = 10时,m的变化显著影响剪枝率,进而影响分类精度。当m较小(≤ 6)时,剪枝率提升带来训练效率增益,但造成不可忽略的性能损失;当m过大(> 8)时,剪枝率降低且训练效率下降,性能未见提升。综合权衡训练效率与精度后,本文选取k = 10,m = 8。

3)退火系数δ。退火系数δ控制剪枝机制退出的时机。如图3(d)所示,减小δ会提前终止剪枝,虽然未带来性能增益,但降低了训练效率;增大δ则削弱退火效应,虽然提升了训练效率,却导致精度明显下降。本文选取δ = 0.875,以在效率与精度间取得平衡。

2.4　训练效率与资源消耗分析

为量化评估CDDP在训练效率方面的提升,本文对比了不同方法在MoCo/ResNet-18框架下于ImageNet-1K数据集上的实际训练耗时及算法开销。实验均在配备8块NVIDIA 3090显卡的服务器上开展,记录端到端训练时间(含数据加载、前反向传播、剪枝调度等全部环节),并单独统计CDDP中一维k-

means 聚类的单次计算开销。表 8 提供了训练效率的量化结果。从表中可以观察到，CDDP 在 36.7% 的剪枝率下取得 1.51 倍加速比，实现了近似 O(1) 复杂度的数据剪枝。训练耗时较 baseline 减少 24.9 h，与 InfoBatch 和 SeTa 基本持平。统计显示，CDDP 引入的一维 k-means 聚类单轮平均耗时 0.21 s，200 轮总开销约 42 s，占整体训练时间的 0.016%，其额外计算成本远低于数据剪枝带来的时间节省，表明 CDDP 在效率与开销之间取得了良好平衡。

表 8　不同方法在 MoCo/ResNet-18 框架下于 ImageNet-1K 上的训练效率对比

Table 8　Comparison of training efficiency on ImageNet-1K under the MoCo/ResNet-18 framework with different methods

方法	剪枝率/%	训练时间/h	加速比	1D k-means 开销/ s
baseline	0	73.8	–	–
InfoBatch(Qin 等，2023)	36.5	48.8	1.51 ×	–
SeTa(Zhou 等，2025)	36.4	49.3	1.50 ×	–
CDDP(本文)	36.7	48.9	1.51 ×	0.21

注："–"表示无数据。

2.5　可视化分析

为直观展示 CDDP 在自监督预训练中的有效性，本文基于 MoCo/ResNet-50 框架在 ImageNet-1K 上的训练过程，针对 ELI 与 DCP 两个模块分别进行可视化分析。

2.5.1　ELI 表征有效性分析

1)样本训练轨迹的可视化分析。首先，本文从单一样本训练轨迹层面验证 ELI 指标的鲁棒性。实验选取两个具有代表性的样本进行分析，如图 4 所示。在训练趋于稳定后(第 50 至 200 轮)，样本 A 与样本 B 的平均损失值十分接近，表明二者在整体学习难度和重要性上大致相当。然而，样本 A 的损失曲线呈现剧烈波动，尤其在第 50 至 100 轮期间多次跌入局部极小值；而样本 B 的损失则保持相对平稳。在基于瞬时损失的动态剪枝策略下，当样本 A 的损失偶然降至波谷时，易被误判为"简单样本"而提前剔除，导致其潜在学习价值未能充分挖掘。采用 ELI 表征评估样本重要性时，样本 A 与样本 B 的 ELI 曲线更为接近。即使样本 A 的损失频繁波动或陷入极小值，其 ELI 值仍保持稳定并与样本 B 相近。这表明 ELI 能有效平滑训练过程中的损失抖动，准确反映样本的真实难度，从而避免样本因瞬时损失下降而被误剪。

2)样本分布演化的可视化分析。为进一步从整体数据分布层面揭示 ELI 的鲁棒性，本文采用核密度估计分析训练前、中、后期(Epoch 10、50、180)全量样本的度量分布演化，结果如图 5 所示。图 5(a)展示了瞬时损失 $L_i^{(t)}$ 的分布变化。可以观察到，由于自监督学习任务的高复杂度与优化过程的不确定性，瞬时损失分布在各个阶段均呈现长尾的非平稳形态。即使在训练后期(Epoch 180)，分布曲线仍伴随剧烈波动，且分布峰值较高。这表明个别样本极易因优化随机性而偶然获得极低损失，若据此进行剪枝将面

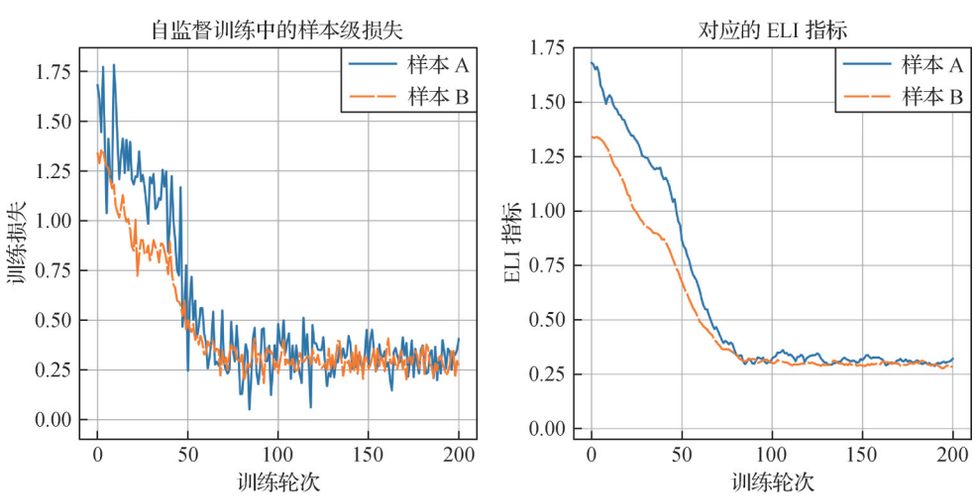
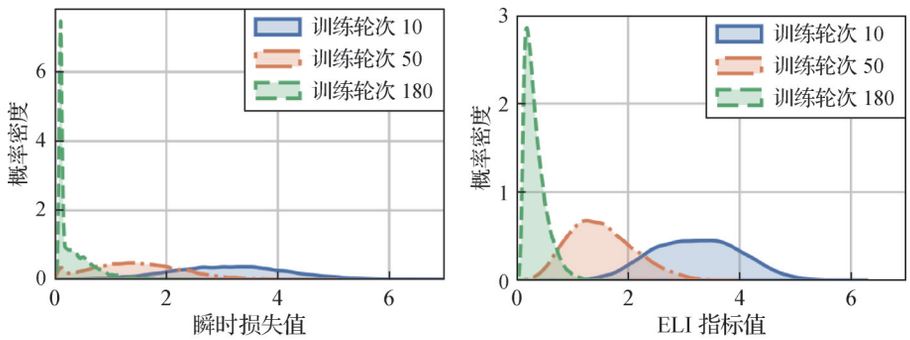


图 4　自监督训练中两个不同样本的损失值及其 ELI 表征

Fig. 4　Loss and corresponding ELI metric for two samples in self-supervised training



(a) 瞬时损失分布的核密度估计　(a) ELI 值分布的核密度估计

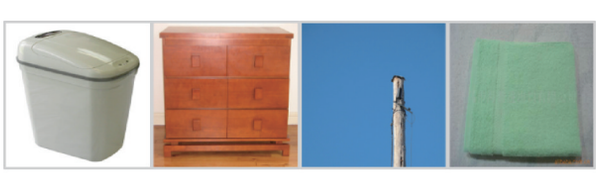
图5　不同训练阶段瞬时损失与ELI值的核密度估计对比

Fig. 5　Comparison of kernel density estimation between instantaneous loss and ELI at different training stages ((a)kernel density estimation of instantaneous loss distribution;(b)kernel density estimation of ELI value distribution)

临较高的误判风险。相比之下,图5(b)展示了ELI值的分布演化。得益于指数移动平均对历史信息的引入,ELI有效滤除了高频噪声与突发极值,各阶段分布始终保持平滑且紧凑的单峰结构,并呈现稳定收敛趋势。这一对比从整体分布层面验证了ELI在度量样本学习价值时的时序稳定性与抗噪鲁棒性,为课程式动态剪枝提供了可靠的调度依据。

2.5.2　DCP策略的可视化分析

为直观验证DCP策略的效果,对训练初期与后期被划分为"最简单课程"与"最困难课程"的样本进行可视化分析,如图6所示。图6(a)展示了训练初期(Epoch 10)被划分为最简单课程的样本,主要为物体主体突出、背景简洁以及姿态标准的图像,这类样本有助于模型建立稳定的底层特征表示。图6(b)展示了训练后期(Epoch 180)被划分为最困难课程的样本,普遍存在物体遮挡、纹理多变以及与背景混淆等复杂情况,对模型的判别能力提出更高要求。两组样本在视觉复杂度和识别难度上的明显差异表明,DCP能够依据样本ELI值实现有效的难度分层,使模型遵循从易到难的渐进式学习路径,与课程学习的核心理念高度契合。



(a) 训练初期简单课程示例 (训练轮次 10)



(b) 训练后期困难课程示例 (训练轮次 180)

图6　DCP策略中不同难度课程的样本可视化

Fig. 6　Visualization of samples from different difficulty courses in DCP strategy((a) example of easy curriculum in early training (epoch 10); (b) example of hard curriculum in late training (epoch 180))

3　结　论

针对现有动态数据剪枝方法依赖样本瞬时损失设定剪枝阈值,难以适配自监督训练损失剧烈波动、优化过程不稳定的固有特性,且训练初期简单样本的过早剔除可能导致模型优化困难等问题,本文提出了一种面向自监督学习的课程式动态数据剪枝方法(CDDP)。CDDP采用指数损失重要性表征(ELI)替代瞬时损失作为样本价值度量。ELI通过动量系数对历史损失进行指数移动平均,有效滤除了训练过程中的短期波动噪声,得到时序平滑、对短期扰动鲁棒性强的样本长期价值表征,实现了对样本重要性的可靠度量。同时,本文提出了动态课程剪枝策略(DCP),DCP首先基于样本的ELI值,采用一维k-means聚类对样本的学习难度进行动态分组,然后引入滑动课程窗口,实现从易到难的渐进式训练调度,进而通过剪枝率调控与退火策略,完成对训练样本的结构化动态学习。上述机制可有效缓解因自监督训练损失波动导致的关键样本误剪问题,同时解决了初期优化困难引发的训练不稳定问题。在涵盖二维图像与三维点云的8个基准数据集上,通过结合5种不同的预训练方法及多种骨干网络进行的实验表明,CDDP显著优于当前最先进的数据剪枝方法。本文方法在大幅提升自监督预训练效率的同

时，能够保持与全量数据训练近乎无损的性能。消融实验和可视化分析进一步验证了ELI表征对样本重要性鲁棒度量、DCP策略对训练过程稳定调度的有效性。未来工作将聚焦于超参数的自适应优化，特别是无损性能所对应剪枝率的自动判定，并将面向自监督学习的数据剪枝技术推广至自然语言处理、信号处理等更广泛的领域，以进一步降低基础模型的训练成本与能源消耗。

参考文献（References）

Bao H B, Dong L, Piao S H and Wei F R. 2022. BEiT: BERT pre-training of image transformers//Proceedings of the 10th International Conference on Learning Representations. [s.l.]: OpenReview.net [DOI: 10.48550/arxiv.2106.08254]

Bengio Y, Louradour J, Collobert R and Weston J. 2009. Curriculum learning//Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Canada: ACM: 41-48 [DOI: 10.1145/1553374.1553380]

Chang A X, Funkhouser T, Guibas L, Hanrahan P, Huang Q, Li Z, et al. 2015. ShapeNet: an information-rich 3D model repository [EB/OL]. [2025-10-28]. https://arxiv.org/pdf/1512.03012.pdf

Chen T, Kornblith S, Norouzi M and Hinton G E. 2020. A simple framework for contrastive learning of visual representations//Proceedings of the 37th International Conference on Machine Learning. [s.l.]: PMLR: 1597-1607

Deng J, Dong W, Socher R, Li L J, Li K and Li F F. 2009. ImageNet: a large-scale hierarchical image database//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 248-255 [DOI: 10.1109/CVPR.2009.5206848]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. 2021. An image is worth 16 × 16 words: transformers for image recognition at scale//Proceedings of the 9th International Conference on Learning Representations. [s.l.]: OpenReview.net

Gal Y, Islam R and Ghahramani Z. 2017. Deep Bayesian active learning with image data//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR.org: 1183-1192

Geng W F, Wang X, Jing L P and Yu J. 2023. Consensus graph learning-based self-supervised ensemble clustering. Journal of Image and Graphics, 28(4): 1069-1078 (耿伟峰, 王翔, 景丽萍, 于剑. 2023. 共识图学习驱动的自监督集成聚类. 中国图象图形学报, 28(4): 1069-1078) [DOI: 10.11834/jig.210947]

Gidaris S, Singh P and Komodakis N. 2018. Unsupervised representation learning by predicting image rotations//Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada: OpenReview.net

Girdhar R, El-Nouby A, Liu Z, Singh M, Alwala K V, Joulin A, et al. 2023. ImageBind one embedding space to bind them all//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 15180-15190 [DOI: 10.1109/cvpr52729.2023.01457]

Gui J, Chen T, Zhang J, Cao Q, Sun Z N, Luo H, et al. 2024. A survey on self-supervised learning: algorithms, applications, and future trends. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46 (12): 9052-9071 [DOI: 10.1109/tpami.2024.3415112]

He K M, Chen X L, Xie S N, Li Y H, Dollar P and Girshick R. 2022. Masked autoencoders are scalable vision learners//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 15979-15988 [DOI: 10.1109/cvpr52688.2022.01553]

He K M, Fan H Q, Wu Y X, Xie S N and Girshick R. 2020. Momentum contrast for unsupervised visual representation learning//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 9726-9735 [DOI: 10.1109/cvpr42600.2020.00975]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/cvpr.2016.90]

He M Y, Yang S, Huang T J and Zhao B. 2024. Large-scale dataset pruning with dynamic uncertainty//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, USA: IEEE: 7713-7722 [DOI: 10.1109/CVPRW63382.2024.00767]

Jiang L, Meng D Y, Zhao Q, Shan S G and Hauptmann A. 2015. Self-paced curriculum learning//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA: AAAI Press [DOI: 10.1609/aaai.v29i1.9608]

Kirillov A, Mintun E, Ravi N, Mao H Z, Rolland C, Gustafson L, et al. 2023. Segment anything//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 3992-4003 [DOI: 10.1109/iccv51070.2023.00371]

Krizhevsky A. 2009. Learning multiple layers of features from tiny images [EB/OL]. [2025-10-28]. http://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf

Kumar M P, Packer B and Koller D. 2010. Self-paced learning for latent variable models//Proceedings of the 24th International Conference on Neural Information Processing Systems — Volume 1. Vancouver, Canada: Curran Associates Inc.: 1189-1197

Liu A X, Feng B, Xue B, Wang B X, Wu B C, Lu C D, et al. 2024. DeepSeek-V3 technical report [EB/OL]. [2025-10-28]. https://arxiv.org/pdf/2412.19437.pdf

Nguyen T T, Gerofi B, Martinez-Noriega E J, Trahay F and Wahib M. 2023. KAKURENBO: adaptively hiding samples in deep neural

network training//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: 37900-37922 [DOI: 10.52202/075280-1650]

Pang Y T, Wang W X, Tay F E H, Liu W, Tian Y H and Yuan L. 2022. Masked autoencoders for point cloud self-supervised learning//Proceedings of the 17th European Conference on Computer Vision — ECCV 2022. Tel Aviv, Israel: Springer: 604-621 [DOI: 10.1007/978-3-031-20086-1_35]

Qin Z H, Wang K, Zheng Z W, Gu J Y, Peng X Y, Xu Z P, et al. 2023. InfoBatch: lossless training speed up by unbiased dynamic data pruning//Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net

Raju R S, Daruwalla K and Lipasti M. 2021. Accelerating deep learning with dynamic data pruning [EB/OL]. [2025-10-28]. https://arxiv.org/pdf/2111.12621.pdf

Sener O and Savarese S. 2018. Active learning for convolutional neural networks: a core-set approach//Proceedings of the 6th International Conference on Learning Representations. [s.l.]: OpenReview.net

Sheng Z T, Chen Y X and Qi G J. 2023. Audio-visual adversarial contrastive learning-based multi-modal self-supervised feature fusion. Journal of Image and Graphics, 28(1): 317-332 (盛振涛, 陈雁翔, 齐国君. 2023. 面向多模态自监督特征融合的音视频对抗对比学习. 中国图象图形学报, 28(1): 317-332) [DOI: 10.11834/jig.220168]

Tian Y D, Chen X L and Ganguli S. 2021. Understanding self-supervised learning dynamics without contrastive pairs//Proceedings of the 38th International Conference on Machine Learning. [s.l.]: PMLR: 10268-10278

Uy M A, Pham Q, Hua B S, Nguyen D T and Yeung S K. 2019. Revisiting point cloud classification: a new benchmark dataset and classification model on real-world data//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 1588-1597 [DOI: 10.1109/iccv.2019.00167]

Wang Y P, Chen Y F, Zhou W J, Fang A, Yan W D, Jamieson K, et al. 2024. CLIPloss and norm-based data selection methods for multimodal contrastive learning//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #480 [DOI: 10.52202/079017-0480]

Wu Q H, Xu M, Bao X Y, Zheng W L, Yu H H, Zou H T, et al. 2026. Dataset distillation via coreset optimization and prototype-augmented diffusion. Journal of Image and Graphics, 31(4): 1201-1215 (吴钦浩, 徐明, 包先雨, 郑文丽, 喻泓浩, 邹海涛, 等. 2026. 融合核心集优化与原型增强扩散的数据集蒸馏. 中国图象图形学报, 31(4): 1201-1215) [DOI: 10.11834/jig.250387]

Wu Z R, Song S R, Khosla A, Yu F, Zhang L G, Tang X O, et al. 2015. 3D ShapeNets: a deep representation for volumetric shapes//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 1912-1920 [DOI: 10.1109/cvpr.2015.7298801]

Xiao T T, Liu Y C, Zhou B L, Jiang Y N and Sun J. 2018. Unified perceptual parsing for scene understanding//Proceedings of the 15th European Conference on Computer Vision — ECCV 2018. Munich, Germany: Springer: 418-434 [DOI: 10.1007/978-3-030-01228-1_26]

Yi L, Kim V G, Ceylan D, Shen I C, Yan M Y, Su H, et al. 2016. A scalable active framework for region annotation in 3D shape collections. ACM Transactions on Graphics, 35(6): 1-12 [DOI: 10.1145/2980179.2980238]

Zhang Y R, Liu J T, Xu G Q and Hao D Y. 2025. Self-supervised coal mine image denoising with adaptive masking. Journal of Image and Graphics, 30(12): 3884-3899 (张亚茹, 刘健涛, 许国庆, 郝定溢. 2025. 基于自适应掩码的自监督矿井图像去噪. 中国图象图形学报, 30(12): 3884-3899) [DOI: 10.11834/jig.240624]

Zhou B L, Zhao H, Puig X, Fidler S, Barriuso A and Torralba A. 2017. Scene parsing through ADE20K dataset//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5122-5130 [DOI: 10.1109/cvpr.2017.544]

Zhou Q, Gao J Y and Wang Q. 2025. Scale efficient training for large datasets//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 20458-20467 [DOI: 10.1109/cvpr52734.2025.01905]

作者简介

刘文昊,男,硕士研究生,主要研究方向为动态数据剪枝。
E-mail: liuwh26@mail2.sysu.edu.cn
郭裕兰,通信作者,男,教授,博士生导师,主要研究方向为空间智能与三维视觉。E-mail: guoyulan@sysu.edu.cn
张钊铭,男,高级工程师,主要研究方向为雷达信息处理。
E-mail: zmzhang_sx@163.com
张晔,男,副研究员,主要研究方向为空间智能和三维视觉。
E-mail: zhangy2658@mail.sysu.edu.cn