基于不确定性引导的智能强化主动学习图像分类方法

酒明远[1,2,3]，吴国伟[1]，宋旭光[1]，李书攀[1,2,3]，徐明亮[1,2,3]

(1. 郑州大学计算机与人工智能学院，河南 郑州 450001；
2. 郑州大学智能集群系统教育部工程研究中心，河南 郑州 450001；
3. 国家超级计算郑州中心，河南 郑州 450001)

摘　　要：随着深度学习技术的快速发展，其在图像分类等任务中取得了显著成果。然而，这些模型的成功往往依赖于大量高质量的标注数据，而在实际应用中，标注数据通常稀缺，人工标注过程又极为耗时、费力，限制了模型的推广与应用。近年来，主动学习因其能够在有限标注预算下提升模型性能而受到广泛关注，其核心思想是根据样本的不确定性、多样性或代表性等指标，挑选最有价值的数据进行标注。针对传统主动学习方法多依赖手动设计的启发式采样策略，难以适应不同任务场景，且选择策略难以动态优化等问题，提出一种基于智能强化主动学习(SRAL)的图像分类方法，通过将样本选择过程建模为马尔科夫决策过程，利用强化学习的自适应策略优化能力，引导模型从未标注样本中动态挑选最具价值的样本用于标注。其中，状态由未标注样本提取的特征构成，动作表示是否选择样本进行标注，奖励函数则定义为当前样本加入训练集后模型准确率的变化差值。采用演员-评论家(Actor-Critic)算法进行策略优化，并引入不确定性启发式排序作为辅助信息以提升学习效率。实验结果表明，在 CIFAR-10，SVHN 和 FASHION-MNIST 等数据集上，所提出的 SRAL 方法在相同标注预算下，相比于其他主动学习方法，能够显著提高分类准确率，且在各数据集上均展现出较好的稳定性和泛化能力，验证了 SRAL 方法在提高图像分类模型性能方面的有效性与优势。

关　键　词：深度学习；强化学习；主动学习；图像分类；策略优化

中图分类号：TP 391.41；TP 18　　**DOI：**10.11996/JGj.2095-302X.2026010047

文献标识码：A　　**文 章 编 号：**2095-302X(2026)01-0047-10

Image classification method based on uncertainty-driven smart reinforcement active learning

JIU Mingyuan[1,2,3], WU Guowei[1], SONG Xuguang[1], LI Shupan[1,2,3], XU Mingliang[1,2,3]

(1. School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou Henan 450001, China;
2. Engineering Research Center of Intelligent Swarm Systems, Ministry of Education, Zhengzhou University, Zhengzhou Henan 450001, China;
3. National Supercomputing Center in Zhengzhou, Zhengzhou Henan 450001, China)

Abstract: With the rapid development of deep learning, remarkable achievements have been made in image classification and related tasks. However, the success of these models heavily relies on large amounts of high-quality labeled data. In real-world applications, labeled data is often scarce, and manual annotation is time-consuming, labor-intensive, and costly, which limits the scalability and deployment of deep learning models. In recent years, active learning has gained significant attention due to its ability to improve model performance under limited annotation budgets. The core idea of active learning is to select the most valuable data for labeling based on certain

收稿日期：2025-06-13；**定稿日期：**2025-10-10；**通信作者：**徐明亮，E-mail：iexumingliang@zzu.edu.cn
Received：13 June，2025；**Finalized：**10 October，2025；**Corresponding author：**XU Mingliang, E-mail：iexumingliang@zzu.edu.cn
基金项目：国家自然科学基金(62272422, U22B2051，62325602)；河南省优秀青年基金(252300421225)；郑州大学有组织科研团队培育项目(35220549)
Foundation items：National Natural Science Foundation of China (62272422, U22B2051, 62325602); Natural Science Foundation of Henan Province (252300421225) and Organized Young Scientific Research Team Cultivation Foundation of Zhengzhou University (35220549)

criteria such as uncertainty, diversity, or representativeness. To address the limitations of traditional active learning methods, which often rely on manually designed heuristic sampling strategies that struggle to adapt to different task scenarios and are difficult to dynamically optimize, a Smart Reinforcement Active Learning (SRAL) approach for image classification is proposed. The sample selection process is modeled as a MARKOV DECISION PRocess (MDP), leveraging reinforcement learning's adaptive strategy optimization ability to guide the model in dynamically selecting the most valuable samples from the unlabeled data for labeling. In this framework, the state is represented by features extracted from the unlabeled samples, the action indicates whether a sample should be selected for labeling, and the reward function is defined as the change in model accuracy after incorporating the selected sample into the training set. The Actor-Critic algorithm is adopted to optimize the sampling policy, and uncertainty-based heuristic ranking is incorporated as auxiliary information to improve the learning efficiency. Experimental results demonstrate that the proposed SRAL method significantly improves classification accuracy under the same labeling budget compared to other active learning approaches on datasets such as CIFAR-10, SVHN, and FASHION-MNIST. Furthermore, SRAL exhibits robust stability and strong generalization ability across these datasets. This confirms the effectiveness and advantages of SRAL in enhancing the performance of image classification models.

Keywords: deep learning; reinforcement learning; active learning; image classification; policy optimization

随着深度学习的兴起，图像分类的性能已达到前所未有的水平。深度神经网络，包括卷积神经网络[1] (Convolutional Neural Network，CNN)和Transformer[2]，通过自动学习图像中的判别性特征，在图像分类任务上取得了重大突破。然而，传统的深度学习方法严重依赖于大规模标注数据集，而手动标注这些数据通常耗时且成本高昂。因此，在保证模型精度的同时，如何减少标注数据的需求已成为深度学习领域的一大研究趋势。

主动学习(Active learning)作为一种高效的数据利用范式，旨在通过智能化策略从海量未标注数据中挑选出最具代表性和信息价值的样本[3]，从而在减少人工标注成本的同时，仍能训练出性能优异的模型。该方法通过迭代地选择信息量丰富的样本，交由专家进行精确标注，并将其加入训练集中，不断优化分类器的性能。与传统全监督学习相比，主动学习在标注样本数量大幅减少的情况下，依然能够达到相当甚至更优的性能表现，特别适用于标注成本高、数据获取困难的实际场景。因此，如何设计合理且具有泛化能力的样本选择策略，成为研究主动学习的核心问题之一。一个高效的样本选择机制不仅关系到模型性能的提升，还直接决定了标注资源的利用效率与系统的整体性能。早期的主动学习策略主要基于不确定性[4-6]、代表性[7-8]以及多样性[9]，但这些策略大多是人工设计的。近年来，越来越多的研究开始探索通过优化精心设计的损失函数[10-11]来改进样本选择策略。然而，由于训练环境可能会不断变化，上述策略可能无法始终选择最具信息量的样本，导致标注预算的低效利用，进而影响最终的分类性能。

针对上述问题，本文提出了一种新的基于智能强化主动学习(Smart Reinforcement Active Learning，SRAL)模型。该模型将主动学习问题建模为一个马尔可夫决策过程(Markov Decision Process, MDP)，旨在通过学习一个动态策略，智能地选择最具信息量的未标注样本，从而在保证模型性能的前提下，尽可能减少对人工标注的依赖。在每轮样本选择过程中，首先采用基于不确定性的准则对所有未标注样本进行初步排序，以过滤出潜在的关键样本。随后，系统结合专家的反馈信息，通过强化学习框架对样本选择策略进行动态优化。与传统依赖人工规则设计的样本选择方法相比，该策略具备更强的自适应性和决策能力。

为了进一步提升策略的稳定性与性能，本文在算法中引入了 Actor-Critic 结构，将演员网络(Actor)与评论家网络(Critic)相结合，实现策略更新与价值评估的协同优化。此外，该模型采用深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)算法，以稳定强化学习过程中在高维连续空间下的训练效果。实验结果表明，该方法不仅能够有效提升主动学习效率，还具备良好的泛化能力和收敛性能。

1 相关工作

在现有文献中，主动学习策略已被广泛研究，其核心在于如何设计高效的样本选择机制，以最大

化有限标注资源的利用效率。这些方法大致可分为基于启发式规则与基于优化目标函数 2 类，前者依赖经验直觉，后者则具备更为坚实的理论基础与可解释性。其中，基于不确定性采样的方法是最为经典的一类策略，可通过分析模型对样本的预测概率分布来评估其不确定性[12]，并优先选择预测结果最模糊或最接近决策边界的样本，从而促使模型更快速地学习难以判别的样本特征。基于代表性的方法则关注样本在特征空间中对整体未标注数据分布的代表能力[13]，通常选取那些与当前已标注样本差异较大、或处于未标注样本聚簇中心[14-15]的样本，以提升样本集的多样性与覆盖性。这类策略能够有效避免样本冗余，提高模型的泛化能力。基于多样性[16]的主动学习方法旨在选取与已标注样本差异较大、能够覆盖更广泛特征空间的样本，从而提升模型的泛化能力。并通过衡量样本间的相似性或分布距离，优先选择在特征空间中彼此差异较大、分布分散的样本，避免冗余信息，确保训练数据具有代表性与全面性，常用于补充不确定性采样的局限性。此外，近年来也有研究将不确定性、代表性与多样性等因素进行联合建模[17]，提出了基于约束优化的主动学习方法，通过构建多目标或加权损失函数，在样本选择过程中实现各类准则的权衡与协同，进一步提高样本选择的合理性和稳定性。这些方法通常采用迭代优化的方式，从整体上提升了主动学习的智能水平与实际应用效果。

随着神经网络的发展，强化学习在学习策略方面表现出了极强的能力，其核心思想是通过与环境的交互学习最优的行为策略，旨在最大化智能体在长期执行任务中的累积奖励。强化学习的研究经历了多个阶段，从最早的经典控制问题到如今在复杂任务中的广泛应用，研究方法和应用范围不断拓展。最初，强化学习主要集中在有限状态空间和简单问题上，如 MDP 中的基本问题求解。随着算法和计算能力的进步，研究者们逐渐将强化学习扩展到更复杂的环境中，包括高维空间和连续控制问题。在这方面，2013 年 MNIH 等[18]提出的深度 Q 网络首次将 CNN 与 Q 学习相结合，是强化学习领域的重要突破。2016 年，VAN HASSELT 等[19]提出双重 DQN (Double Deep Q-Network)，通过分离动作选择和动作评估，减轻 Q 学习中的过估计问题，提高了算法的稳定性和性能。此外，传统的强化学习方法通常面临着收敛速度慢、样本利用率低等问题，为了提高样本效率，SCHAUL 等[20]提出优先经验回放机制，根据经验的重要性进行采样，提升了学习效率。随后，FANG 等[21]将启发式主动学习算法重新定义为一个强化学习问题。HAUSSMANN 等[22]提出强化主动学习(Reinforced Active Learning，RAL)算法，并使用贝叶斯神经网络作为样本选择策略的学习预测器。LIU 等[23]提出的深度强化主动学习(Deep RAL，DRAL)采用了类似的思想。在此基础上，SUN 和 GONG[24]将深度卷积神经网络提取图像的特征作为强化学习算法的“状态”，并使用深度 Q-Learning 算法来训练一个 Q 网络，根据 Q 网络的输出来决定是否对数据进行标注，但该方法不适合处理连续的动作空间且训练过程也不稳定。WANG 等[25]将主动学习建模为 MDP，并基于 Actor-Critic 架构的强化学习算法，使用 DDPG 算法来训练模型。该方法虽然稳定了训练并能更快地找到更优的策略，但是内部对于“状态”的表达会随着样本集的变大而变得复杂，且对于“奖励”的设置没有充分考虑到数据分布对训练结果的影响。

本文方法与之前的方法相比主要由以下几点不同：

1) SRAL 首先在样本选择的初期阶段引入不确定性排序机制，使用基于不确定度量的排序方法筛选具有较高不确定性的样本。然后，将不确定性高的样本作为输入进入强化学习过程，使用 Actor-Critic 结构进一步优化。通过这种方式，SRAL 将不确定性排序与强化学习策略紧密结合，为强化学习提供了一个经过初步筛选的、具有较高信息量的样本集。DRAL 和 RAL 方法虽然也使用强化学习框架解决主动学习中的样本选择问题，但其更多依赖于固定的 Q-Learning 或 DDPG 等方法，在选择样本时，强化学习直接从未标注样本中选择，且不显著地引入不确定性排序机制。在这些方法中，样本选择的策略较为简单，主要依靠强化学习模型的训练过程来不断优化；

2) SRAL 方法提出一种动态中间奖励机制，即在每次样本选择后，动态地计算当前样本对模型准确率的影响，并计算实时反馈。具体来说，奖励是基于选择样本前后的模型准确率差值来确定的。且能够使强化学习在每一步都得到及时反馈，从而加速模型的学习和收敛。DRAL 和 RAL 方法在奖励的计算上往往存在一定的延迟，通常需要等到一轮样本选择完成后，才会根据最终的模型效果来进行奖励评估；

3) SRAL 方法采用了 Actor-Critic 结构来进行

样本选择策略的优化，Actor 网络负责从当前状态中选择最具信息量的样本进行标注，而 Critic 网络负责评估当前的动作(即选择的样本)是否有效，进而提供奖励信号来指导 Actor 网络的更新。DRAL 和 RAL 方法也采用了强化学习框架，由于强化学习的结构可能更加简单，通常依赖于 Q-Learning 等传统方法；

4) SRAL 方法在样本选择过程中不仅考虑样本的价值，还设定了小预算，每次选择一定数量的样本进行标注，直至达到总标注预算。在每个预算周期结束时，模型会基于当前已标注数据集重新训练，并更新样本选择策略。DRAL 和 RAL 方法的样本选择过程通常是全局性的，并没有像 SRAL 那样通过小预算分阶段进行动态调整。

2 方法框架

2.1 模型结构

本文提出的 SRAL 算法将主动学习框架建模为一种新的 MDP。整个过程由状态、动作、奖励、状态转换和预算组成，其总体结构如图 1 所示。首先，对于所有未标注的样本集，可根据不确定度量对其进行一个基础的排序，然后选择前 n 个样本，并使用 ResNet 模型提取这 n 个样本的特征作为“状态”。之后将这个“状态”的特征矩阵值馈送到演员网络，其输出的值 0 或者 1 定义为 Action，用来指示专家是否应该标记这些样本。如果演员网络的输出为 1，则表示应该手动标注这 n 个样本，如果为 0 则不对其进行标注。奖励则是由选择样本前后的模型准确率差决定，如果差值为正，表明所选择的样本有积极的影响，则将所选的样本添加到已标注的集合中进行进一步的模型学习，否则返回到未标记的集合中进行未来的选择。

该模型在开始训练之前，会将未标记样本经过排序挑选出前 n 个样本交由专家标记，并用这些个样本对分类器进行预训练。该方法认为经过排序的样本已具备良好的价值，然后根据初步价值的高低进行挑选可以达到事半功倍的效果(本文中采用基于不确定性采样的方法进行排序)。在达到挑选样本的总预算之前会设定多个阈值，每当挑选的样本数达到阈值之后，就对其进行标记并加入到已标注数据集，然后用最新的已标注数据集对分类器进行训练，经过训练后的分类器再对未标记样本进行新的排序用于下个阈值到达前的样本挑选。在本文提出的 SRAL 框架中，演员-评论家(Actor-Critic)网络是核心组件，其中演员网络根据当前状态和学习策略从未标记的集合中选择信息量最大的样本，而评论家网络则决定这些行为是否提高了模型的性能。

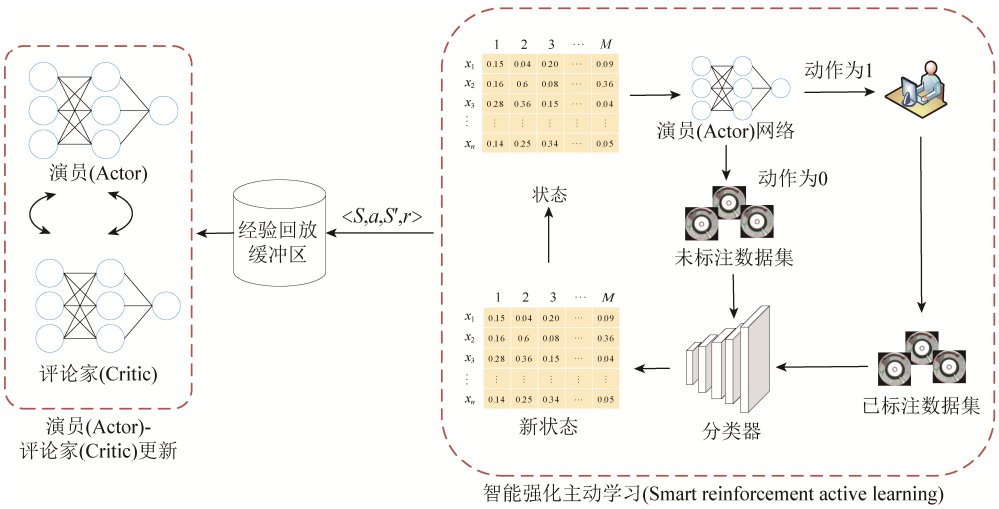

图 1 智能强化主动学习网络示意图

Fig. 1 Smart reinforcement active learning framework

2.2 框架细节

本文提出的 SRAL 以 Actor-Critic 网络作为模型的核心组件。其中，Actor 网络根据当前状态 $S_t \in S$ 和已学习的策略(即动作)从未标注样本集中选择最具信息量的样本，其下标 t 代表当前的主动学习迭代轮次。Critic 网络负责评估所选动作是否能够提升模型性能。两者共同在深度强化学习框架下训练，从而不断优化样本选择策略，最终实现更优的主动学习效果，这一点将在后续实验中进一步验证。

1) 状态。为了更快地选择出相对有价值的样本进行标注，同时避免在所有样本中进行盲目的选择，在考虑分类器具有一定分类能力的基础上，将所有未标注样本输入到分类器中进行初步分类，本文中采用主动学习中基于熵的分类方法并且将得到的结果按照从大到小的顺序进行排序。每个未标注样本通过输入到预训练的 ResNet-18 分类器中提取特征。具体来说，特征向量来自于 ResNet-18 网络的倒数第 2 层(即全连接层之前的输出)。

得到新的未标注样本序列之后，取当前的 n 个样本作为此刻的状态。并将每批的前 n 个样本的选择定义为一种状态，即

$$S = X_i^{u_n} \tag{1}$$

式中：$X_i^{u_n}$ 表示 n 个未标注样本。

2) 动作。将前面提到的 S 输入到演员网络中来计算每个样本的选择动作。为了更好地获得最终的动作，该方法采用了 2 个卷积层和一个池化层的网络，然后是 3 个完全连接的线性层，最后使用非线性函数将结果映射到(–1,1)。如果该值小于 0，则将对应的动作设置为 0，表示丢弃该样本，否则选择该样本。然后将当前的 n 个样本提交给专家进行注释，并添加到标记集中。$\{(x_i,y_i)\}_{i=1}^{n}$ 表示被选中并标注的样本，因此更新后的新标注集为：$(X_l,Y_l):=\left((X_l,Y_l)\cup\{(x_i,y_i)\}_{i=1}^{n}\right)$。假设评论家网络的参数写成$\theta_a$，则将上述操作基于状态 S 生成动作 a 的学习策略$\pi(S;\theta_a)$。

3) 状态转换。该步骤的目的是获得下 n 个样本的特征，定义为 $S'=X_{i+1}^{u_n}$。需要说明的是，此举并不是一次性获得 b 个样本，然后提交给专家进行标注(其中 b 可以理解为触发需要专家标注的数字，是一个固定值)。因此，在达到 b 之前，会利用分类器提供的排序样本序列以提取图像的特征。一旦达到 b，专家将对样本进行标注并输入分类器进行训练，以得到新的样本序列和特征。

4) 奖励。奖励函数在强化学习算法中的作用至关重要，其不仅评估行为的选择质量，还指导演员网络做出有价值的决策。然而，在许多强化学习算法中，奖励通常是在整个动作序列结束后才给出的，这种设计导致了反馈的延迟。具体而言，智能体通常无法在每个动作之后即时获得奖励，从而使得其在学习过程中难以迅速、准确地将奖励信号与其先前的决策关联起来。反馈的延迟使得学习过程变得不够高效，且智能体往往需要更多的时间才能找到最佳的策略。为了应对这一问题，本文提出了一种新的方法，通过每一步计算模型的验证精度变化，来动态地计算中间奖励。该做法使得智能体可以更早、更准确地收到来自环境的反馈，提升学习效率，从而加速整个训练过程的收敛，其中奖励函数可表示为

$$r(S,a) = Acc(\phi_i) - Acc(\phi_{i-1}) \tag{2}$$

式中：Acc 表示预测精度；ϕ_i 表示经过动作 a 发生后的学习模型。在强化学习中，该部分的目标是通过与环境的交互最大化奖励，并被定义为 Q-Value 函数。用 Bellman 方程建立状态-动作函数之间的关系，即

$$Q(S,a;\theta_a) = E\left[\gamma Q(S',\pi(S';\theta_a);\theta_c) + r(S,a)\right] \tag{3}$$

式中：θ_c 表示近似 Q 值函数的评论家网络；γ表示延迟参数。基于这些关系，该方法执行价值函数估计、策略改进和优化操作，以便使智能体能够在环境中学习并做出决策。受到深度 Q-Learning[26]的启发，该方法通过解决最大化问题来学习演员网络的贪婪策略，即

$$\max_{\theta_a} Q\left(S,\pi(S;\theta_a);\theta_a\right) \tag{4}$$

该方法定义了 $\tilde{Q}(S,a;\theta_c) = \gamma Q\left(S',\pi(S';\theta_a);\theta_c\right) + r(S,a)$，并通过求解最小化问题学习评论家网络，即

$$\min_{\theta_c}\left(\tilde{Q}(S,a;\theta_c) - Q(S,a;\theta_c)\right)^2 \tag{5}$$

5) 预算。该方法把总预算设为 B 作为最终达到标注要求的数字。当达到总预算时，代表一个训练轮次结束，当前数据集成为模型训练的最终数据集。特别是，该方法通过一个小预算 b 来迭代地实现总预算。每当选择的样本数量达到小预算时，该部分标注样本就与之前标注的样本相结合，并用于训练分类器以进行下一次样本选择。

6) 训练目标网络。为了获得更好的性能，该方法遵循[27]使用单独的目标网络来计算 $\tilde{Q}(S,a;\theta_c)$。根据式(5)，$\pi(S';\theta_a)$ 取决演员网络的输出，$\tilde{Q}(S,a;\theta_c)$ 取决于下一个状态 S'，$\left(S',\pi(S';\theta_a)\right)$ 则由评论家网络评估。本文采用参数为 $\theta_{a'}$ 的独立演员网络和参数为 $\theta_{c'}$ 的评论家网络来计算 $\tilde{Q}(S,a;\theta_c)$。因此，可将式(5)改写为

$$\min_{\theta_c}\left(\gamma Q'(S',\pi'(S';\theta_{a'});\theta_{c'}) + r(S,a) - Q(S,a;\theta_c)\right)^2 \tag{6}$$

式中：$\pi'(\cdot,\theta_{a'})$ 表示来自目标 Actor 网络的目标策略，$Q'(.,\cdot,\theta_{c'})$ 表示目标评论家网络的函数。应用 DDPG 对式(6)中的最小化问题进行优化。在每个迭

代轮次结束时，目标的演员网络和评论家网络被更新为

$$\theta_{a'} := \lambda\theta_a + (1-\lambda)\theta_{a'};\quad \theta_{c'} := \lambda\theta_c + (1-\lambda)\theta_{c'} \tag{7}$$

式中：$\lambda \in (0,1)$ 表示权衡参数。

在训练过程中，该方法将每个获得的转换 (S,a,S',r) 存储在回放缓冲区中，使智能体能够反复利用这些经验进行学习，从而有效抑制数据之间的相关性，提高学习的效率和稳定性。随后，从缓冲区中随机选择小批量的过渡样本，用来更新演员网络和评论家网络的参数。

3 数据集

本文方法在 CIFAR-10，SVHN 和 Fashion-MNIST 的 3 个数据集上进行了图像分类任务的验证，这些数据集被广泛用于评估主动学习在图像分类任务中的性能。CIFAR-10 数据集包含 10 个类别的彩色图像，共 60 000 张，其中 50 000 张用于训练，10 000 张用于测试，每张图像分辨率为 32×32 像素，类别包括飞机、汽车、鸟、猫、鹿、狗、青蛙、马、船和卡车。该数据集具有多样性和适中的难度，被广泛应用于深度学习模型的训练与性能评估。SVHN 数据集是一个大规模的数字识别数据集，用于机器学习和深度学习研究，该数据集从谷歌街景图像中提取，包含 73 257 张训练图像、26 032 张测试图像和 531 131 张额外图像，每张图像为 32×32 像素，标注为 0~9 的数字类别。SVHN 数据集具有实际场景的复杂性，是评估模型在数字识别任务中性能的重要基准。类似地，Fashion-MNIST 数据集包含 10 个服饰类别，共 70 000 张灰度图像，每张图像大小为 28×28 像素。其中 60 000 张用于训练，10 000 张用于测试。类别包括 T 恤、裤子、鞋子和包等，旨在替代经典的 MNIST 数据集，提供更具挑战性和实用性的图像分类任务基准，广泛应用于深度学习模型的训练和性能评估。

4 实验

4.1 实验参数

在图像分类任务中，考虑到 ResNet-18[28]在准确率和训练稳定性方面的优势，本文选择 ResNet-18 作为主学习器，并在最终选定的样本上进行模型评估。在样本选择过程中用于分类的模型采用传统的 CNN，初始学习率设置为 0.01，批次大小为 128，优化器采用随机梯度下降(Stochastic Gradient Descent，SGD)，权重衰减为 5×10^{-4}，动量设置为 0.9[29]。训练过程的最大迭代次数为 100 次，所有实验均在 PyTorch[30]平台上完成。该方法采用了结构相同的演员和评论家网络，均由 5 个完全连接的层组成。优化过程使用 Adam 优化器，并将学习率和延迟因子 γ 分别设置为 0.000 1 和 0.99。权衡参数设定为 λ=0.01，回放缓冲区的容量为 3 000，且仅当缓冲区存储的经验长度超过 128 时才进行随机抽样。每次随机抽取的样本大小为 64，小预算 b 的大小设定为 300。

4.2 实验对比分析

对于图像分类任务，本文方法与几种现有的主动学习方法进行了比较，包括随机选择(Random)、熵(Entropy，EN)[31]、最小置信度(Least Confidence，LC)[32]、边际抽样(Margin Sampling，MS)[33]和 DRAL[25]。对于每种方法，首先从未标记样本集 D_u 中随机选择 1 000 张图像作为初始样本集，并构成初始标记数据集 D_l，用于训练初始分类器。在训练过程中，每当选择的样本数量达到预算 1 000 时便将这些样本添加到 D_l 中，形成一个新的数据集 $D_{l+\text{new}}$，在新的数据集中重新训练分类器。这个过程会不断重复，直到总预算 B 用尽。

图 2 展示了在 CIFAR-10 数据集上对小预算参数 b 的消融实验结果，样本规模设置为 3 000。从图中可以看出随着预算 b 增大，性能整体呈先上升后略微下降的趋势。当 b=300 时有最高均值，b=400 时方差更小、更稳定。后续实验将预算 b 设置为 300。

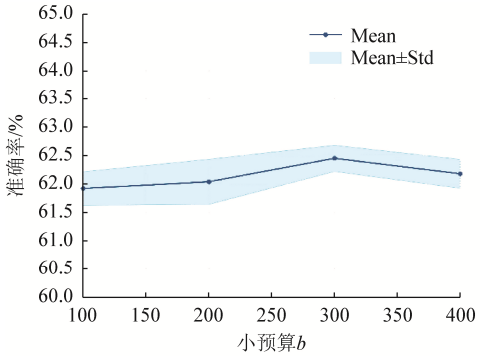

图 2 SRAL 在小预算 b 不同时的实验结果(CIFAR-1 数据集，样本数量为 3 000)

Fig. 2 Experimental results of SRAL under different small budget b (CIFAR-10 dataset, number 3 000)

表 1~3 分别展示了本文方法与其他几种常见的主动学习方法在 CIFAR-10 数据集、Fashion-MNIST 数据集和 SVHN 数据集上分类准确率的结果，并通过消融实验验证了该方法的鲁棒性。本文

还对每个实验设置进行了多次重复实验，得到每种方法的性能均值和方差。消融实验(SRAL+Random 方法)是指在该方法框架的基础上只按照随机的样本排列方式进行挑选，并未融合基于主动学习的启发式排序方法。消融实验的结果说明本方法的优越性建立在与启发式排序方法结合的基础之上。从表中结果可以看出，在选择相同数量的样本进行模型训练时，本文方法明显优于其他主动学习方法，这验证了其动态样本选择策略的有效性和适用性。

表 1　在 CIFA-10 数据集上的实验结果对比(准确率均值±方差)

Table 1　Comparison of experimental results on the CIFAR-10 dataset (Accuracy mean ± Standard deviation)

方法	样本数量						
	1 000	20 00	3 000	4 000	5 000	6 000	7 000
Random	49.44±0.86	56.90±0.49	61.60±0.28	64.73±0.20	69.96±0.39	67.76±0.32	69.25±0.15
EN[31]	50.02±0.19	57.29±0.19	61.58±0.17	64.55±0.20	68.99±0.33	67.39±0.20	68.80±0.19
LC[32]	49.50±0.43	57.77±0.26	61.35±0.28	64.61±0.47	69.72±0.20	67.75±0.31	69.19±0.24
MS[33]	49.94±0.39	57.23±0.64	61.42±0.40	64.50±0.23	70.25±0.07	68.42±0.21	69.72±0.12
DRAL[25]	50.21±0.19	57.77±0.40	61.64±0.42	65.18±0.010	70.10±0.22	68.54±0.17	69.75±0.15
SRAL+Random	50.40±0.04	57.83±0.36	61.83±0.07	65.60±0.49	69.90±0.67	68.62±0.17	70.00±0.11
SRAL	**50.67±0.10**	**58.45±0.11**	**62.36±0.55**	**65.77±0.44**	**70.49±0.19**	**68.98±0.26**	**70.26±0.15**

注：加粗数据表示最优值。

表 2　在 SVHN 数据集上的实验结果对比(准确率均值±方差)

Table 2　Comparison of experimental results on the SVHN dataset (Accuracy mean ± Standard deviation)

方法	样本数量						
	1 000	2 000	3 000	4 000	5 000	6 000	7 000
Random	73.65±0.17	85.06±0.70	87.44±0.05	88.92±0.13	89.46±0.58	90.23±0.13	90.63±0.41
EN[31]	73.62±1.07	84.55±0.48	88.55±0.14	89.99±0.46	91.24±0.14	92.50±0.09	92.93±0.21
LC[32]	75.02±0.53	85.40±0.26	88.14±0.19	90.08±0.28	91.32±0.06	92.16±0.29	93.11±0.16
MS[33]	72.60±0.45	85.28±0.90	88.51±0.31	90.55±0.19	91.52±0.19	92.41±0.14	93.13±0.05
DRAL[25]	74.63±0.93	86.23±0.40	89.19±0.19	91.00±0.18	91.56±0.12	92.92±0.10	93.44±0.08
SRAL+Random	75.21±0.16	86.58±0.37	89.40±0.14	90.96±0.23	91.36±0.20	92.72±0.29	93.15±0.37
SRAL	**75.45±0.46**	**86.94±0.16**	**89.66±0.22**	**91.30±0.13**	**91.80±0.13**	**93.05±0.10**	**93.60±0.10**

注：加粗数据表示最优值。

表 3　在 FASHION-MNIST 数据集上的实验结果对比(准确率均值±方差)

Table 3　Comparison of experimental results on the FASHION-MNIST dataset (Accuracy mean ± Standard deviation)

方法	样本数量						
	1 000	2 000	3 000	4 000	5 000	6 000	7 000
Random	76.28±0.81	78.30±0.60	79.64±0.82	79.84±0.95	81.64±0.68	81.71±0.71	82.84±0.07
EN[31]	77.51±1.85	82.77±0.43	84.26±0.67	86.16±0.51	87.14±0.18	87.78±0.55	88.17±0.46
LC[32]	77.22±0.94	82.96±0.27	84.90±0.18	86.38±0.07	86.95±0.29	87.97±0.42	88.72±0.38
MS[33]	78.33±1.32	83.54±0.54	88.56±0.28	86.84±0.40	87.35±0.53	88.10±0.43	88.79±0.23
DRAL[25]	79.81±0.31	84.79±0.29	86.25±0.17	86.97±0.23	87.55±0.43	88.67±0.15	88.93±0.24
SRAL+Random	80.07±0.25	84.44±0.17	86.24±0.24	86.78±0.26	87.67±0.16	88.25±0.26	88.89±0.24
SRAL	**80.47±0.18**	**85.01±0.13**	**86.71±0.20**	**87.40±0.17**	**87.95±0.20**	**88.83±0.14**	**89.25±0.20**

注：加粗数据表示最优值。

同时，本文在 SVHN 数据集上进行了 SRAL 和 DRAL 的 t-test 显著性测试，SRAL 方法在 2 个 3 000 和 6 000 样本数量上与 DRAL 有着明显差异(p 值分别为 0.042 5 和 0.000 6)，而在 1 000，2 000，4 000 和 7 000 样本数量下差异比较显著(p 值最高为 0.237 4)，当样本数量增多时，SRAL 和 DRAL 方法的性能趋近一致，主要原因在于当样本足够多时，随机抽取的样本已经能够反映出数据的分布形

式，此时主动学习的优势不明显。

在 CIFAR-10 数据集数据集上，本文方法与当前图像分类最新方法进行了对比。图 3 为 SRAL 与最新方法 F-SAM[34]在样本数量为 1 000~2 000 时的实验结果。从图中可以看出，当样本数量低于 1 500 时，SRAL 的准确率更高，而在超过 1 500 后，F-SAM 的准确率则更具优势。由此可见，相较于 F-SAM 和 SRAL 方法在样本数量较少时展现出更优的样本选择效果，在样本增多时与最新方法有着一定差距，主要原因在于 F-SAM 设计了专门针对梯度噪声与对抗扰动分解的模型，而本方法侧重于样本选择策略，所使用的分类模型相比于 F-SAM 比较简单，因此在样本数量较多的情况下，性能明显劣于 F-SAM 方法。而在样本较少的情况下，由于 SRAL 能够挑选出对分类模型有益的困难样本，能够显著提升分类性能。

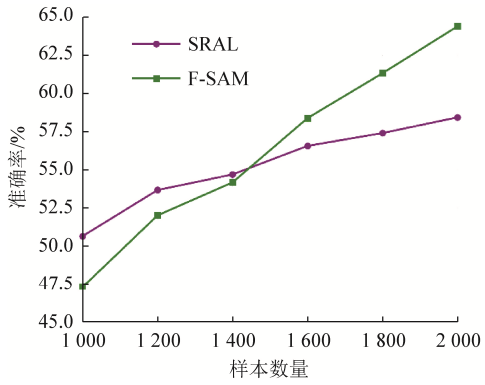

图 3 SRAL 与 F-SAM 在不同样本数量下的对比结果 (CIFAR-10 数据集)

Fig. 3 Experimental results of SRAL and F-SAM under different sample sizes (CIFAR-10 dataset)

为了进一步观察 SRAL 每次挑选样本的分布，本文使用 t-SNE 方法对 CIFAR-10 数据集上不同主动学习方法生成的选定样本进行了可视化，如图 4 所示。t-SNE[35]被用来将每个样本的高维特征映射到二维空间，从而方便地观察样本在特征空间中的分布情况。图 4 中的每个子图显示了在不同迭代次数下，采用不同方法(如随机选择、EN、LC、MS、DRAL 和 SRAL)挑选样本的分布。每次迭代中，不同方法选择 1 000 张新图像(以黑色点表示)，并交由专家进行手动标注。黑色点代表了通过主动学习策略选择的新样本，而彩色点则表示在模型训练后，通过不同方法挑选过的样本。不同彩色点对应于不同类别的样本，其展示了通过当前选择策略对模型训练的影响。从可视化结果中可以明显看到，在 SRAL 方法的前期迭代中，选择的样本分布更加均匀，且不同类别之间的边界变得更加清晰，随着迭代次数的增加，SRAL 方法进一步优化了样本选择，使类别之间的边界更加明显，且各类别样本在特征空间中的分布逐渐变得更加分离。与其他方法相比，SRAL 方法在每次迭代中通过动态优化样本选择策略，能够更加迅速地选择到具有代表性且分布均匀的样本，进一步提高了模型的泛化能力和鲁棒性。

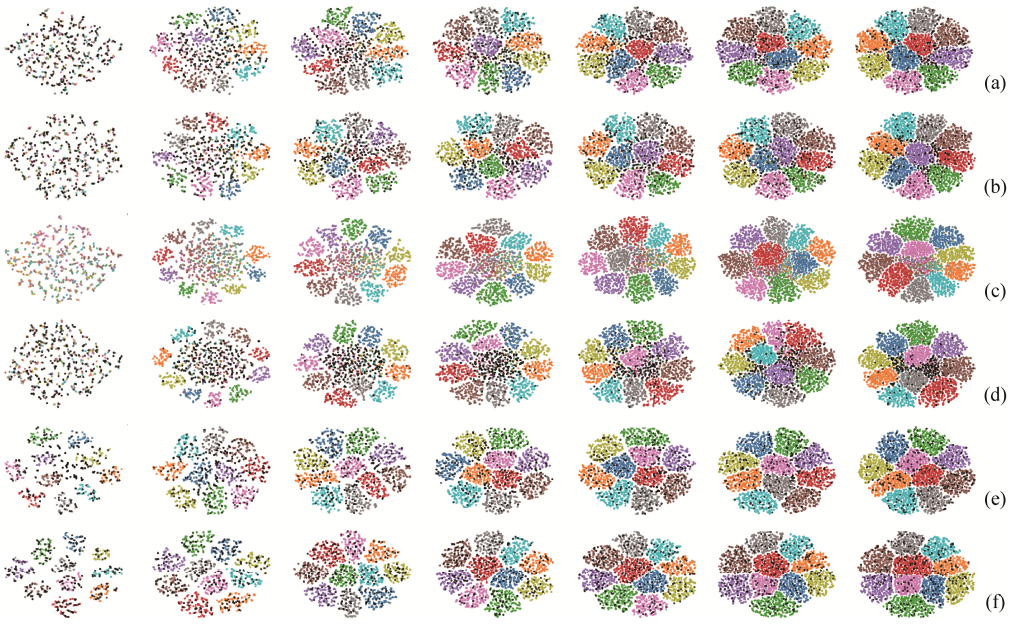

图 4 在不同迭代次数下挑选样本的二维分布 t-SNE 可视化图(每轮迭代挑选 1 000 张图片)

Fig. 4 2D distribution t-SNE visualization comparison chart (Each iteration select 1 000 images) ((a) Random; (b) EN; (c) LS; (d) MS; (e) DRAL; (f) SRAL)

图 5 展示了使用 SRAL 方法在多个数据集上挑选的部分样本例图。被 SRAL 挑选出来的样本意味着在标注和训练过程中具有更高的价值。在图 5 中，可以观察到以下几点：

图 5　SRAL 在多个数据集上挑选出的样本例图

Fig. 5　Image instances selected by the SRAL on multiple datasets ((a) CIFAR-10; (b) SVHN; (c) FASHION-MNIST)

1) 图中的选取样本通常具有较大的类别多样性，即可代表数据集中的不同类别；

2) SRAL 特别关注模型不确定性较高的样本。这些困难样本往往处于类别边界或模型的“盲区”，因此其对模型的训练价值更高。

5　结论

本文提出一种新颖的 SRAL 方法，将样本选择建模为 MDP。其核心优势在于克服传统主动学习中广泛采用的人工设计样本选择策略的局限性，通过 Actor-Critic 网络动态优化样本选择策略，使其更适应复杂且不断变化的环境。同时，SRAL 利用 DDPG 算法对模型进行训练，以提升选择策略的有效性。通过在多个分类数据集上的验证结果表明，在挑选相同数量样本进行模型训练的情况下，该方法在提升模型分类性能方面表现更为显著。

尽管 SRAL 方法在多个数据集上的实验结果表明其优越性，但在样本量较大的情况下，其性能逐渐与 F-SAM 等方法接近，甚至存在一定差距。同时，SRAL 方法依赖于强化学习的策略优化，虽然该方法能够有效提升样本选择效果，但其计算复杂度较高，尤其在大规模数据集上，可能导致训练过程的时间成本和计算资源消耗过大。因此，未来的研究可以探索更加高效的动态样本选择策略，进一步优化学习过程，考虑结合跨模态学习和迁移学习的技术以应对更复杂的任务和数据集，为主动学习领域带来更多的突破和创新。

参考文献 (References)

[1] CHEN L Y, LI S B, BAI Q, et al. Review of image classification algorithms based on convolutional neural networks[J]. Remote Sensing, 2021, 13(22): 4712.

[2] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. [2025-04-13]. https://dl.acm.org/doi/10.5555/3295222.3295349.

[3] WANG M, MIN F, ZHANG Z H, et al. Active learning through density clustering[J]. Expert Systems with Applications, 2017, 85: 305-317.

[4] LEWIS D D, GALE W A. A sequential algorithm for training text classifiers[M]//CROFT B W, RIJSBERGEN C J. SIGIR’94. London: Springer, 1994: 3-12.

[5] KHAN A, HAQ I U, HUSSAIN T, et al. PMAL: a proxy model active learning approach for vision based industrial applications[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2022, 18(2s): 123.

[6] TANG S G, YU X Y, CHEANG C F, et al. Transformer-based multi-task learning for classification and segmentation of gastrointestinal tract endoscopic images[J]. Computers in Biology and Medicine, 2023, 157: 106723.

[7] DEMIR B, PERSELLO C, BRUZZONE L. Batch-mode active-learning methods for the interactive classification of remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(3): 1014-1031.

[8] YANG Y Z, LOOG M. Single shot active learning using pseudo annotators[J]. Pattern Recognition, 2019, 89: 22-31.

[9] NGUYEN H T, SMEULDERS A. Active learning using pre-clustering[C]//The 21st International Conference on Machine Learning. New York: ACM, 2004: 79.

[10] DASGUPTA S. Analysis of a greedy active learning strategy[EB/OL]. [2025-04-13]. https://dl.acm.org/doi/abs/10.5555/2976040.2976083.

[11] DESCHAMPS S, SAHBI H. Reinforcement-based display selection for frugal learning[C]//2022 26th International Conference on Pattern Recognition . New York: IEEE Press, 2022: 1186-1193.

[12] CARAMALAU R, BHATTARAI B, KIM T K. Visual transformer for task-aware active learning[EB/OL]. [2025-04-13]. https://arxiv.org/abs/2106.03801.

[13] GISSIN D, SHALEV-SHWARTZ S. Discriminative active learning[EB/OL]. [2025-04-13]. https://arxiv.org/abs/1907.06347.

[14] HUANG S J, JIN R, ZHOU Z H. Active learning by querying informative and representative examples[EB/OL]. [2025-04-13]. https://dl.acm.org/doi/10.5555/2997189.2997289.

[15] YIN C Y, CHEN S S, YIN Z C. Clustering-based active learning classification towards data stream[J]. ACM Transactions on Intelligent Systems and Technology, 2023, 14(2): 38.

[16] SHEN W J, LI Y H, CHEN L, et al. Multiple-boundary clustering and prioritization to promote neural network retraining[C]//The 35th IEEE/ACM International Conference on Automated Software Engineering. New York: IEEE Press, 2020: 410-422.

[17] DESCHAMPS S, SAHBI H. Reinforcement-based frugal learning for interactive satellite image change detection[C]// IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. New York: IEEE Press, 2022: 627-630.

[18] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing Atari with deep reinforcement learning[EB/OL]. [2025-04-13]. https://arxiv.org/abs/1312.5602.

[19] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning[C]//The 30th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2016: 2094-2100.

[20] SCHAUL T, QUAN J, ANTONOGLOU I, et al. Prioritized experience replay[EB/OL]. [2025-04-13]. https://arxiv.org/abs/1511.05952.

[21] FANG M, LI Y, COHN T. Learning how to active learn: a deep reinforcement learning approach[EB/OL]. [2025-04-13]. https://aclanthology.org/D17-1063/.

[22] HAUSSMANN M, HAMPRECHT F, KANDEMIR M. Deep active learning with adaptive acquisition[EB/OL]. [2025-04-13]. https://dl.acm.org/doi/10.5555/3367243.3367383.

[23] LIU Z M, WANG J Y, GONG S G, et al. Deep reinforcement active learning for human-in-the-loop person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2019: 6121-6130.

[24] SUN L, GONG Y H. Active learning for image classification: a deep reinforcement learning approach[C]//The 2nd China Symposium on Cognitive Computing and Hybrid Intelligence . New York: IEEE Press, 2019: 71-76.

[25] WANG J W, YAN Y G, ZHANG Y B, et al. Deep reinforcement active learning for medical image classification[C]//The 23rd International Conference on Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Cham: Springer, 2020: 33-42.

[26] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[27] TANG X, WU S, CHEN G, et al. Learning to label with active learning and reinforcement learning[C]//The 26th International Conference on Database Systems for Advanced Applications. Cham: Springer, 2021: 549-557.

[28] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 770-778.

[29] CARAMALAU R, BHATTARAI B, KIM T K. Sequential graph convolutional network for active learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2021: 9578-9587.

[30] PASZKE A, GROSS S, MASSA F, et al. PyTorch: an imperative style, high-performance deep learning library[EB/OL]. [2025-04-13]. https://dl.acm.org/doi/10.5555/3454287.3455008.

[31] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27(3): 379-423.

[32] SETTLES B. Active learning literature survey[R]. Madison: University of Wisconsin-Madison, 2009.

[33] SCHEFFER T, DECOMAIN C, WROBEL S. Active hidden Markov models for information extraction[C]//The 4th International Conference on Advances in Intelligent Data Analysis. Cham: Springer, 2001: 309-318.

[34] LI T, ZHOU P, HE Z B, et al. Friendly sharpness-aware minimization[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2024: 5631-5640.

[35] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9: 2579-2605.