GRASP: A Rehearsal Policy for Efficient Online Continual Learning

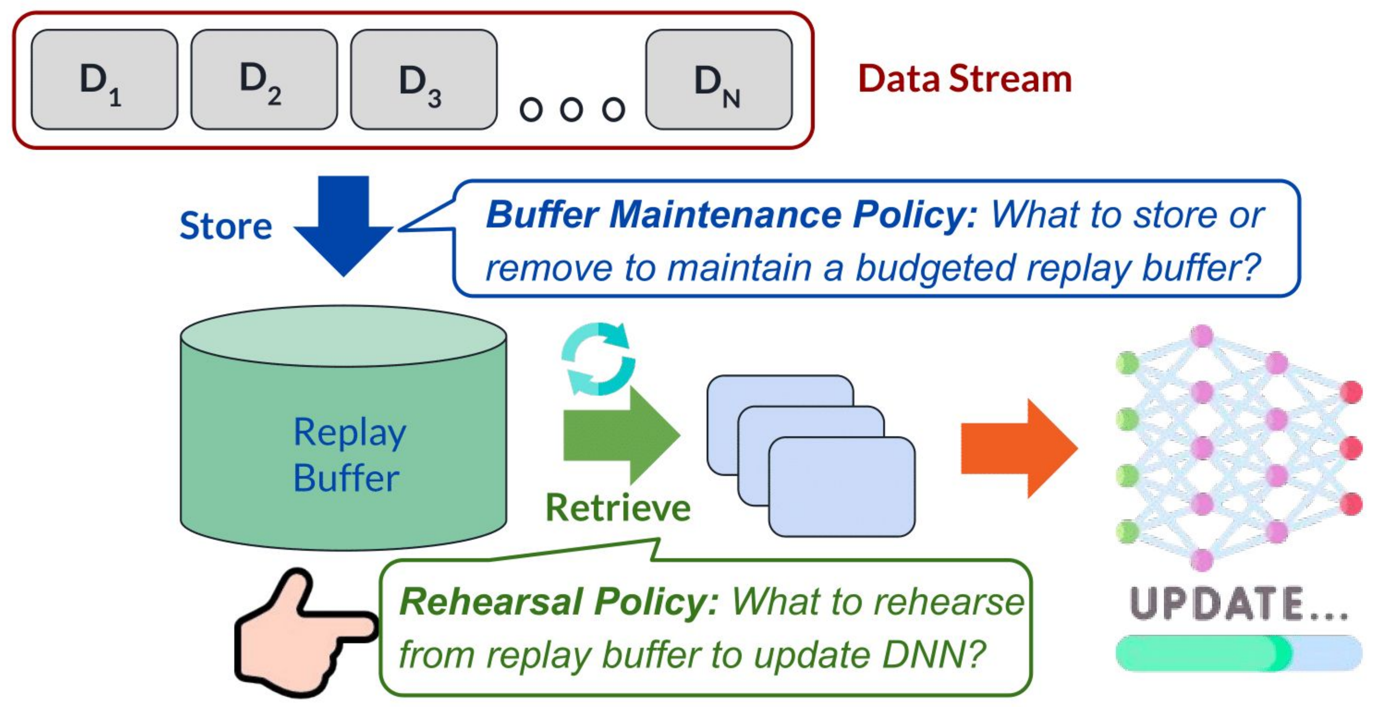
Md Yousuf Harun[1], Jhair Gallardo[1], Junyu Chen[2], Christopher Kanan[2]

[1]Rochester Institute of Technology, [2]University of Rochester

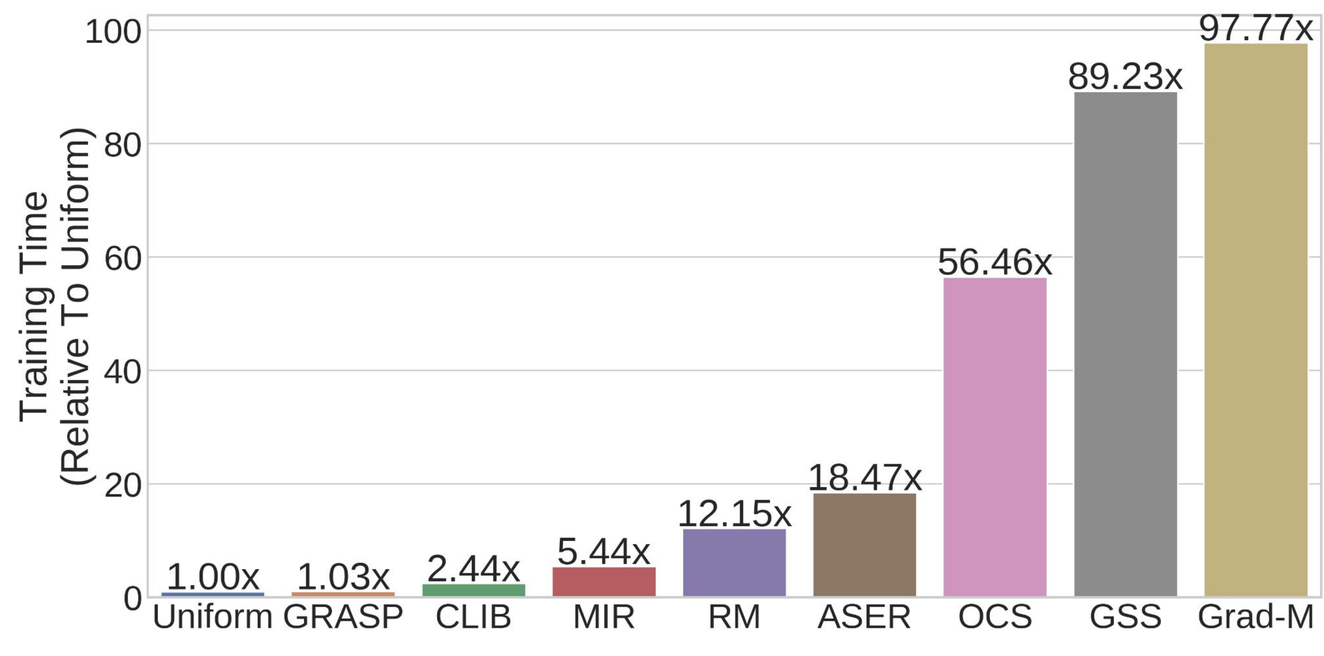
Code: https://yousuf907.github.io/graspsite

Overview

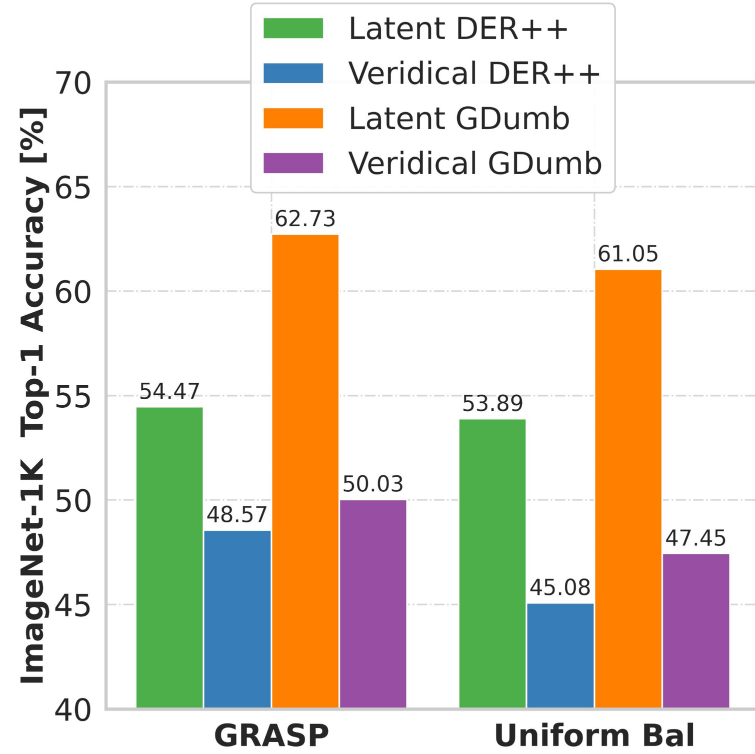
- A major challenge in continual learning (CL) is that non-stationary data streams cause **catastrophic forgetting** of previously learned abilities.
- A popular solution is **rehearsal**: storing past observations in a buffer and then sampling the buffer using a **rehearsal policy** to update the deep neural networks (DNNs).
- For large-scale CL, **uniform selection** has been shown to outperform more sophisticated policies.
- **Goal:** We need more efficient policy to reduce training time and achieve superior performance.
- We propose **GRASP**, a dynamic rehearsal policy.



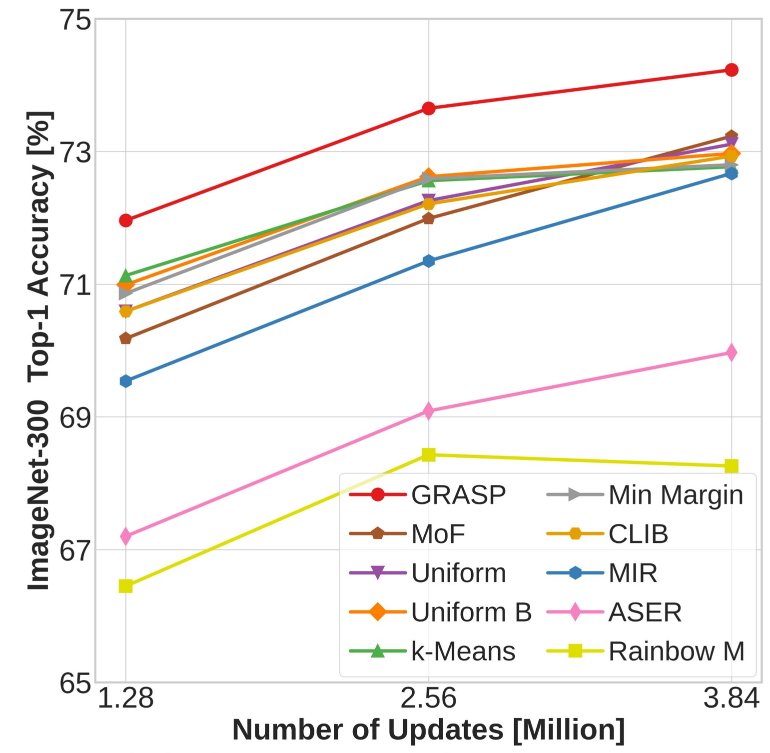
Results



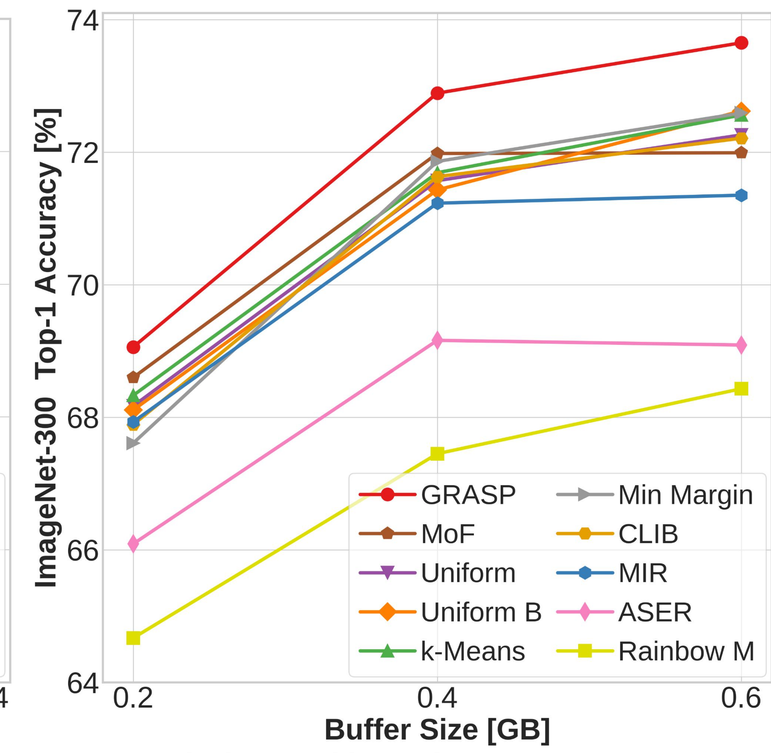
(a) GRASP is significantly faster than existing SOTAs



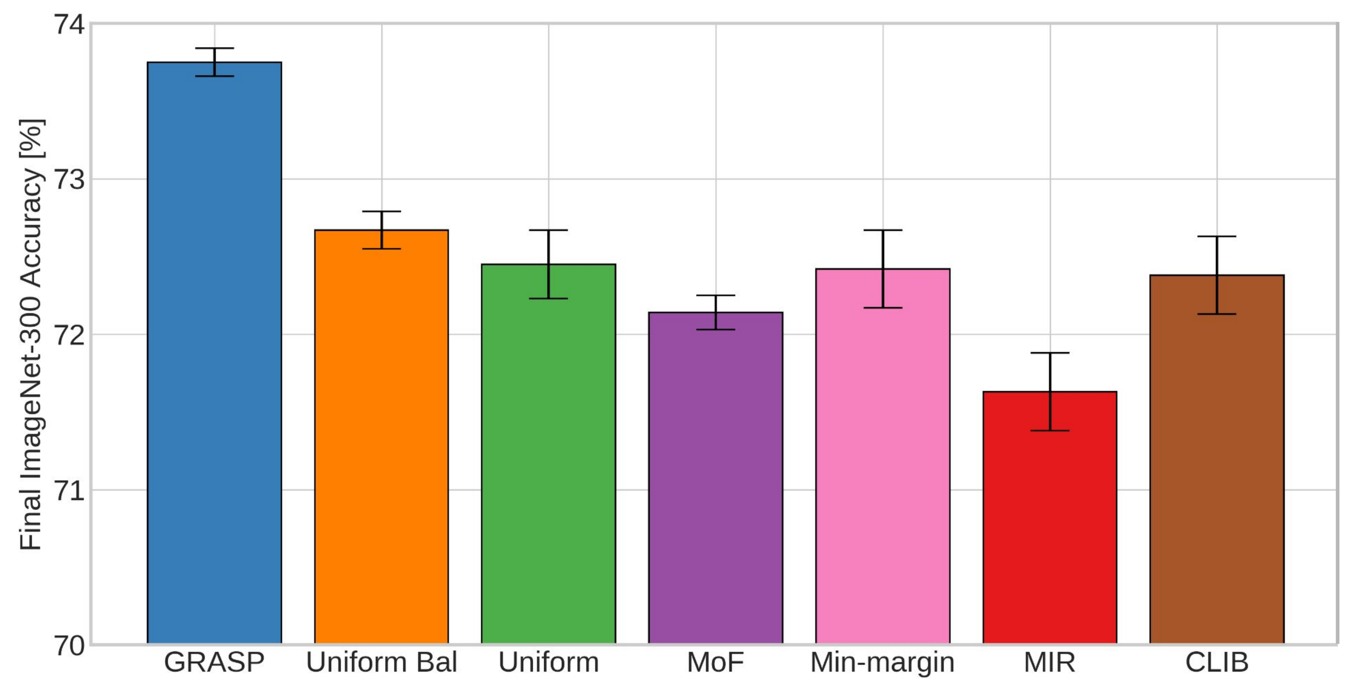
(b) CIL Performance



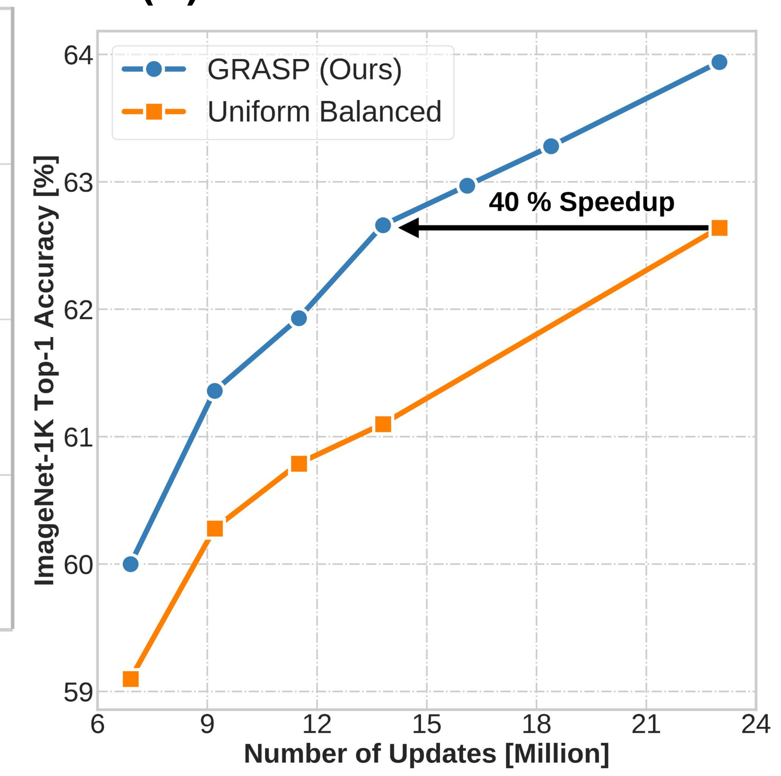
(c) Compute Constraints



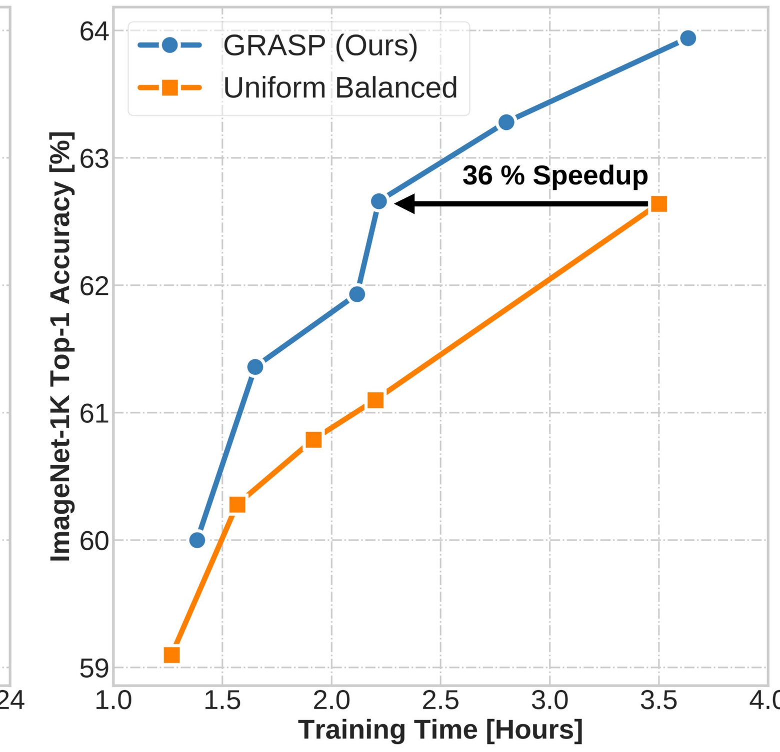
(d) Buffer Constraints



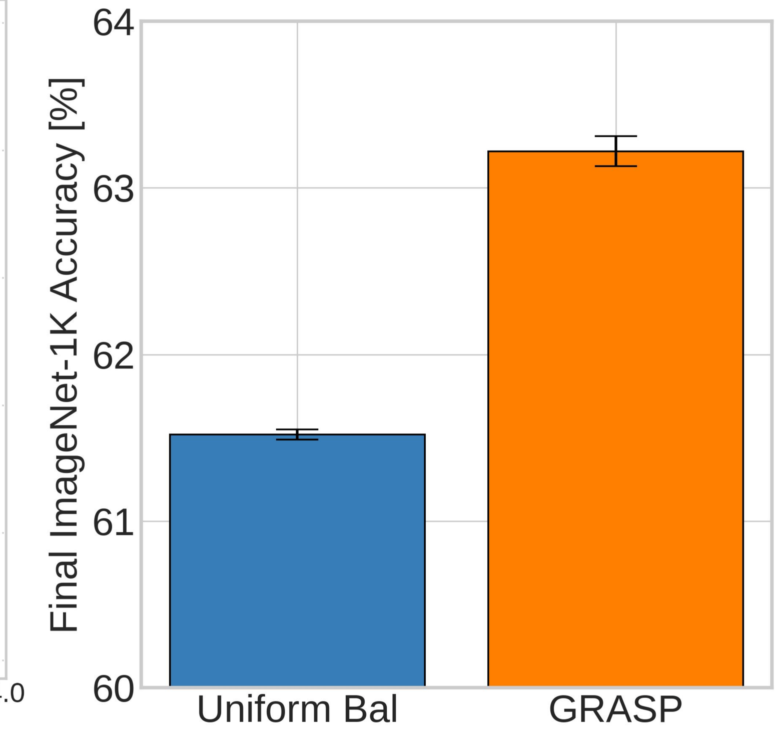
(e) GRASP outperforms 17 other rehearsal policies. Here we show the high-performing policies.



(f) Compute Efficiency



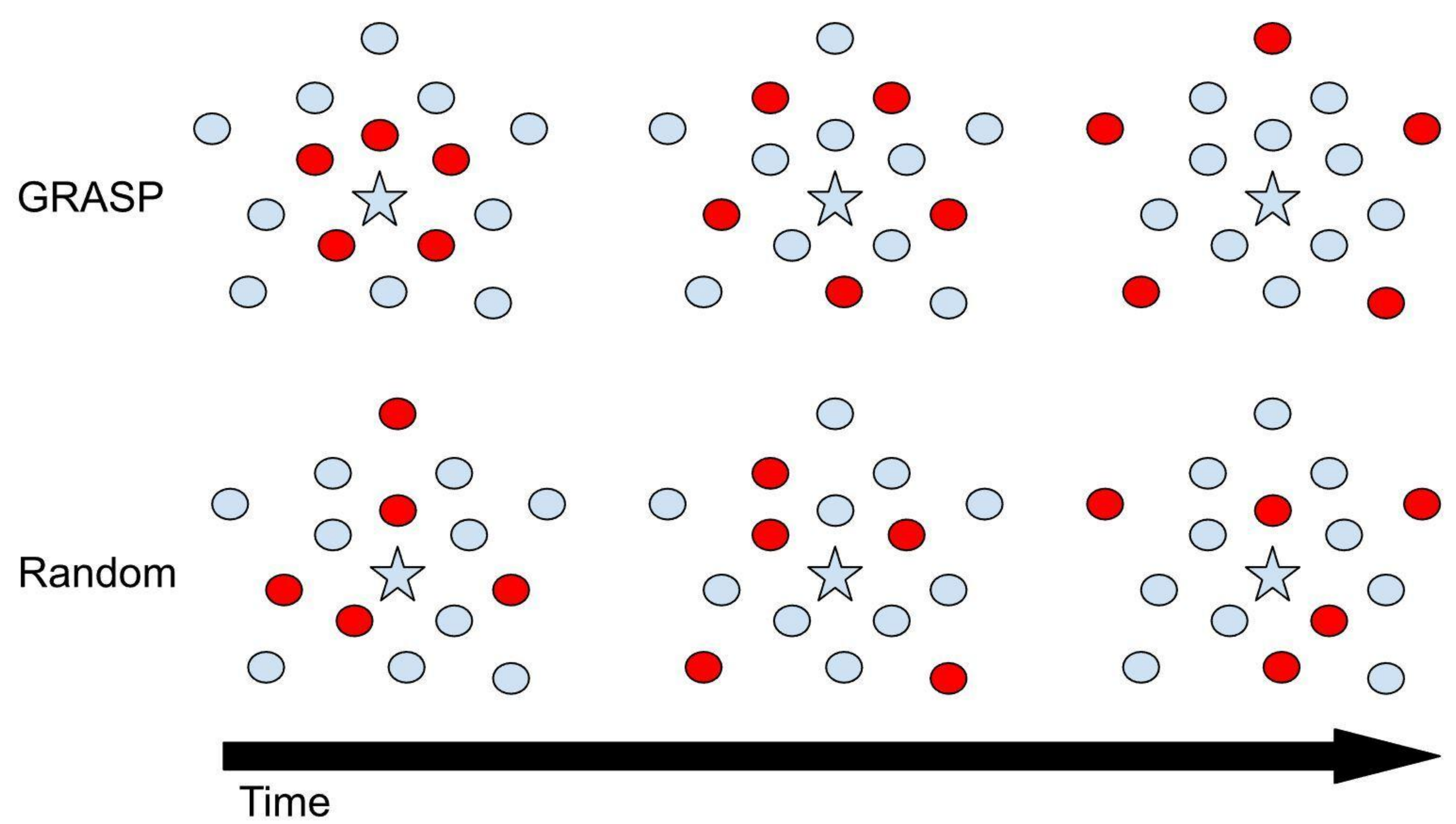
(g) Training Time Efficiency



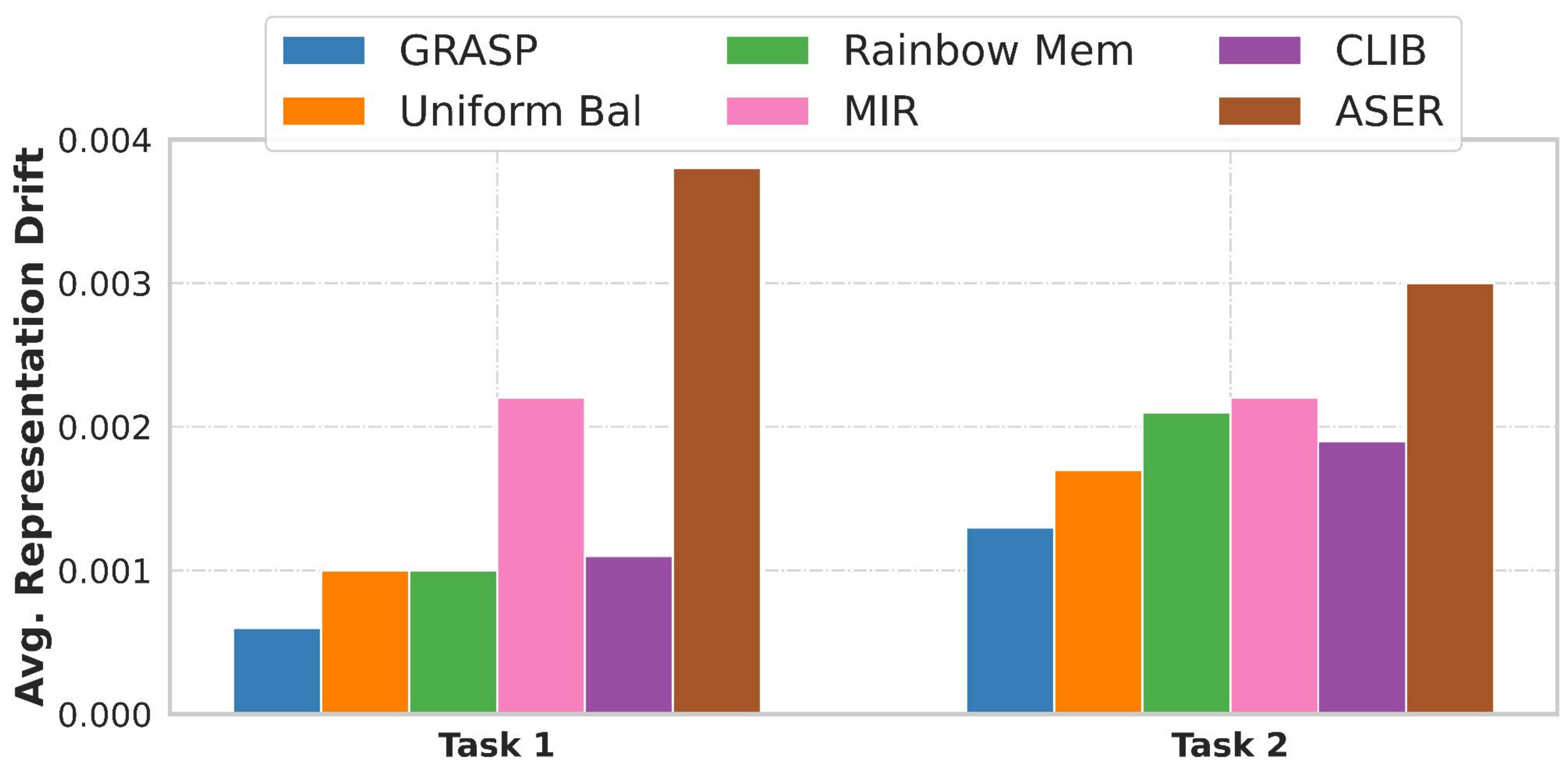
(h) IID CL Performance

How Does GRASP Work?

GRASP first selects the most prototypical (easy) samples from the rehearsal buffer and then gradually selects less prototypical (harder) samples, where easy samples are closest to the class mean and hard samples are farthest.



Why Is GRASP More Effective?



- We measure the representation drift by computing the mean squared error between the penultimate embedding vectors across consecutive training iterations.
- Using GRASP, previously learned representations change less abruptly across rehearsal updates.

Summary

- In CIL on ImageNet, GRASP outperforms 17 rehearsal policies including uniform balanced.
- GRASP is compute and memory efficient.
- GRASP outperforms uniform balanced when integrated with various CL methods e.g., SIESTA, GDumb, and DER++.
- GRASP is effective across data distributions including CIL, IID, and long-tailed distributions.
- GRASP outperforms uniform balanced in text classification on 5 benchmark datasets.
- GRASP is effective for both veridical and latent rehearsal.
- GRASP has potential to supplant expensive periodic retraining and make on-device CL more efficient.

Acknowledgements

We thank NSF for financially supporting this research.