RIT Rochester Institute of Technology
UNIVERSITY of ROCHESTER
kLab

How Efficient Are Today's Continual Learning Algorithms?

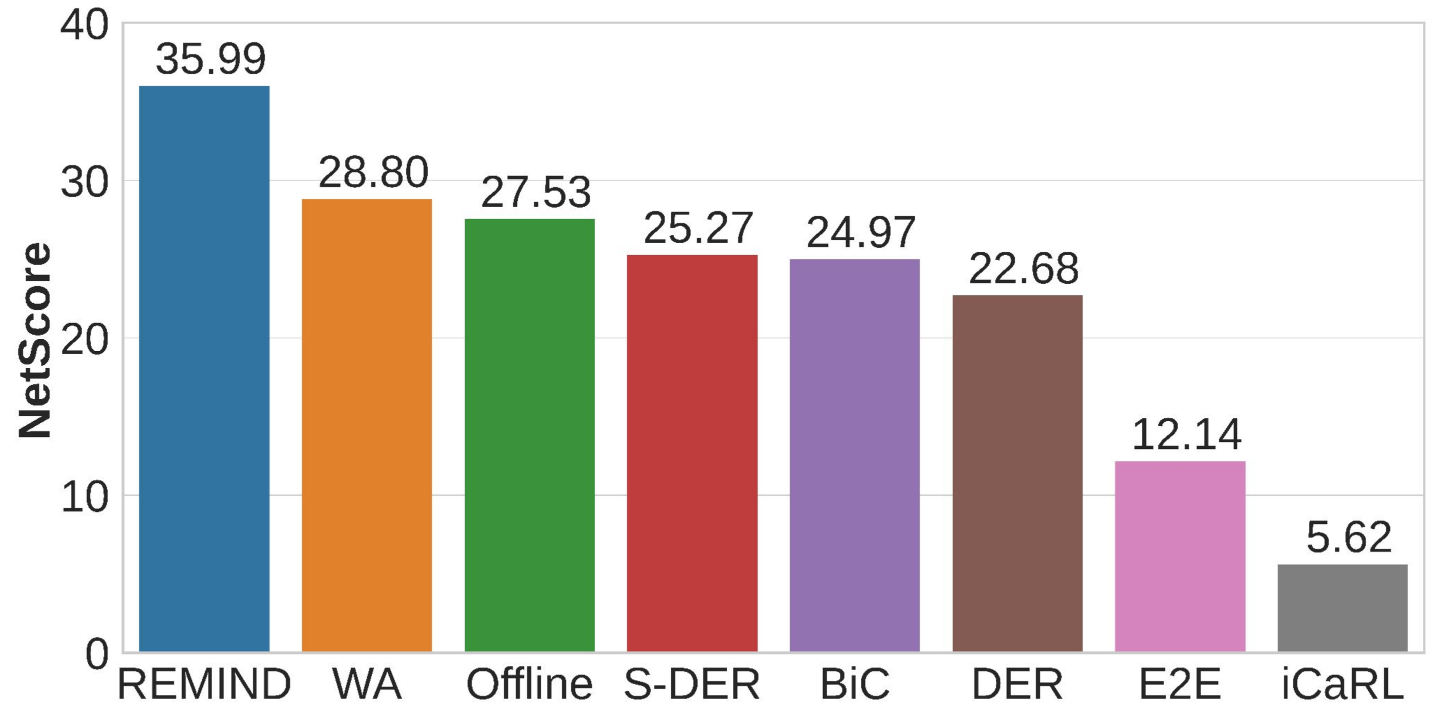
Md Yousuf Harun[1], Jhair Gallardo[1], Tyler L. Hayes[1], Christopher Kanan[2]
[1]Rochester Institute of Technology, [2]University of Rochester

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

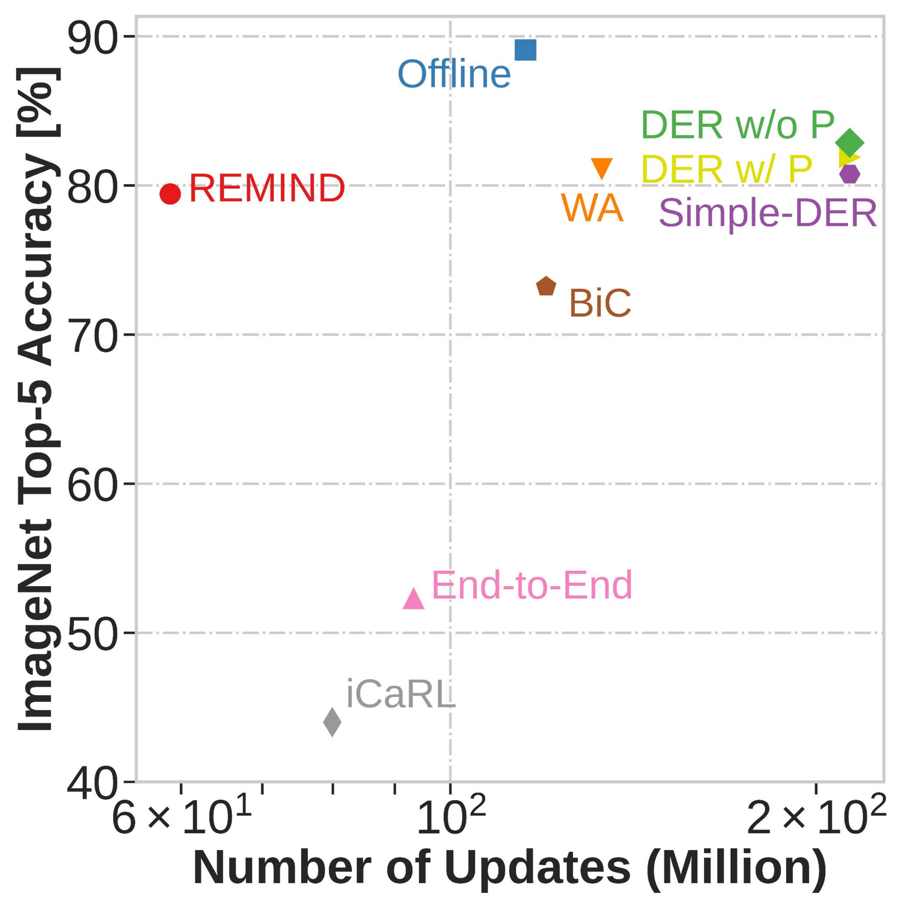
Overview

- Continual learning (CL) has focused on catastrophic forgetting, but a major motivation for CL is ***efficiently updating*** deep neural networks (DNNs) with new data, rather than retraining from scratch when dataset grows over time.
- Although catastrophic forgetting has largely been alleviated, many state-of-the-art CL methods overlook model size, computational overhead, memory usage, data efficiency and training time. For making a real-world impact, CL cannot ignore these factors.
- We study the computational efficiency of existing methods which reveals that many are as expensive as training offline models from scratch – this is alarming and defeats the efficiency aspect of CL.

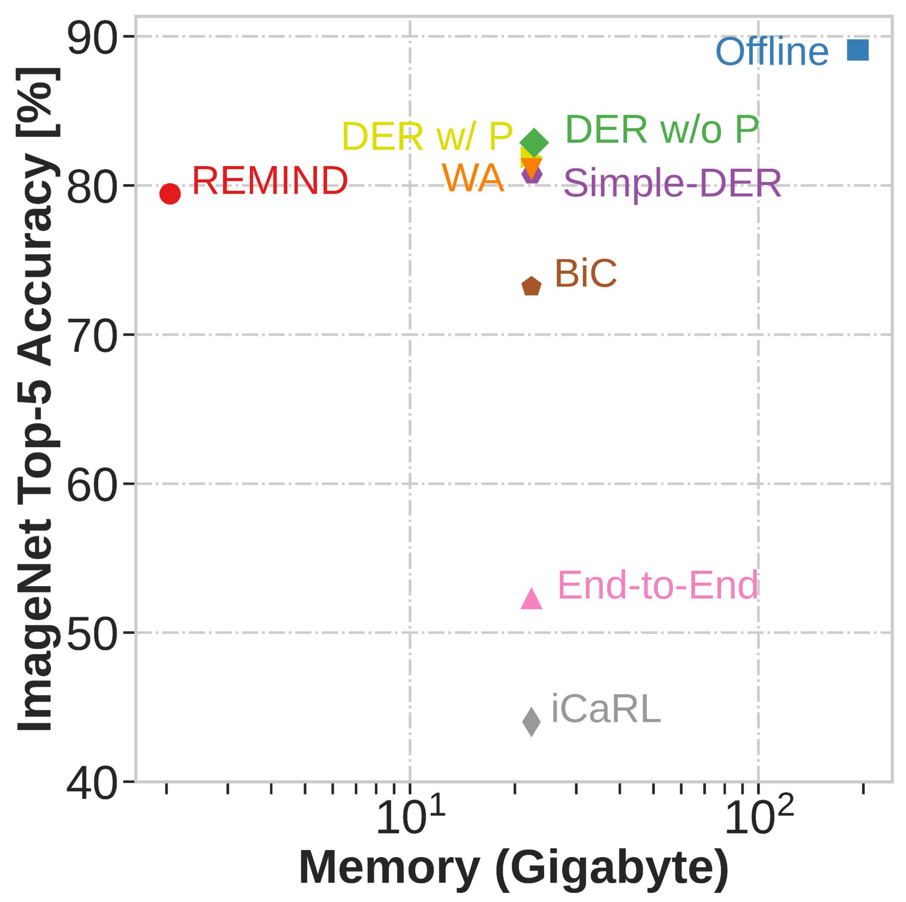
Results



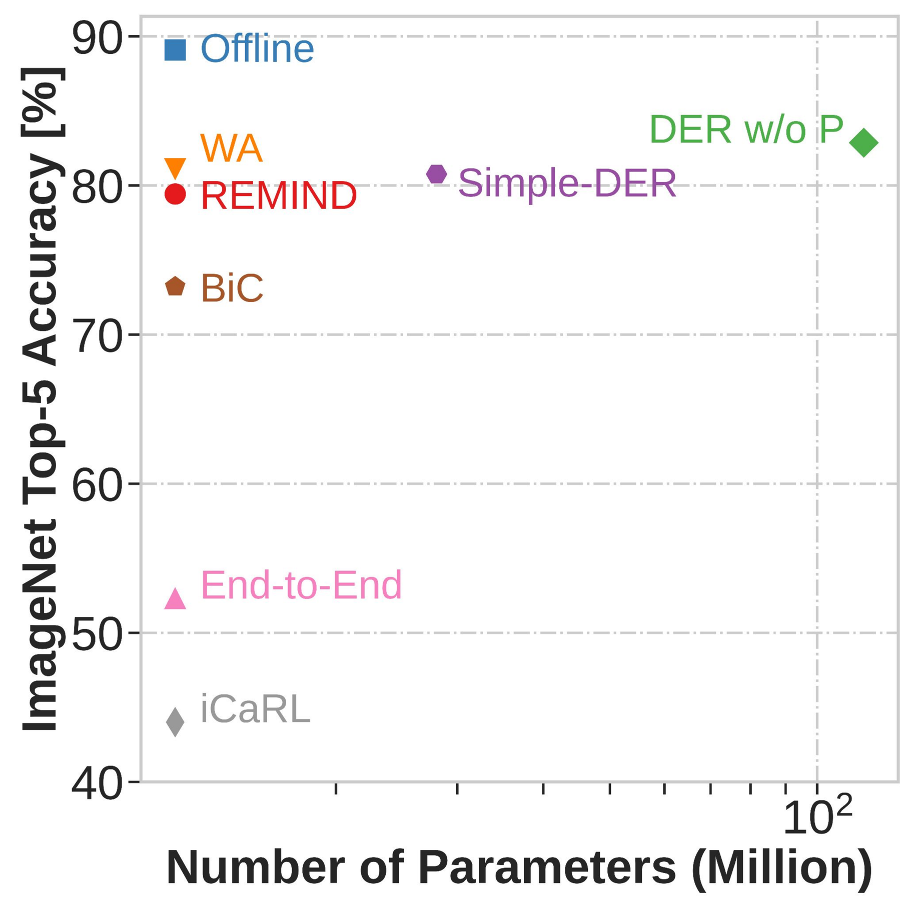
[a] NetScore Comparison



[b] Accuracy vs Backprop Updates



[c] Accuracy vs Memory



[d] Accuracy vs Parameter

Besides REMIND and WA, many CL methods become more expensive than an offline model as they have lower NetScore (fig[a]) and require more compute (updates) than an offline model (fig[b]). Moreover, they require increased memory (fig[c]) and a large number of parameters (fig[d]).

Evaluation Criteria

We propose NetScore for evaluating state-of-the-art class incremental learning methods for ImageNet-1K dataset.

NetScore $\Omega(\mathcal{G})$: assigns score in terms of four factors: accuracy $a(\mathcal{G})$, model size $p(\mathcal{G})$, compute $u(\mathcal{G})$, and memory $m(\mathcal{G})$. The coefficients α, β, γ, and ζ control the contribution of each factor.

$$\Omega(\mathcal{G}) = s\log\left(\frac{a(\mathcal{G})^{\alpha}}{p(\mathcal{G})^{\beta}u(\mathcal{G})^{\gamma}m(\mathcal{G})^{\zeta}}\right)$$

- ❖ Compute (Updates): Number of single input backpropagation steps
- ❖ Memory usage: Old data, current data, model
- ❖ Model size: Number of parameters

Criteria for Efficient CL

We propose following criteria to align CL with real-world needs:

1. Systems need computational efficiency, while avoiding forgetting.
2. We must move beyond toy datasets.
3. CL paradigms should be justified and their limitations mentioned.
4. CL methods should be routinely compared with offline models based on performance and compute.
5. CL models should work across a range of data ordering schemes in addition to extreme edge-cases, e.g., IID and class incremental.
6. Systems should allow online updates and be robust across batch sizes.

These criteria are missing in existing state-of-the-art CL algorithms as they require large number of parameters (11.68-116.89M), network updates (79.94-213.17M), increased memory (22.32-22.74GB), and work only for edge cases.

Summary

- ❖ Existing CL algorithms are not well aligned with many real-world applications, especially on-device learning.
- ❖ Given the computational expense and carbon emission involved in retraining DNNs, CL has potential to reduce economic and environmental costs of deep learning, but only if CL is computationally cheaper than offline retraining.
- ❖ We urge research community to focus on efficient CL beyond catastrophic forgetting. Besides efficiency, CL systems need to work well across multiple data orderings.
- ❖ In recent work, we propose SIESTA, a CL algorithm that meets our criteria (Harun et al., 2023).

Acknowledgements

We thank NSF for financially supporting this research.