RIT Rochester Institute of Technology

Lifelong Machine Learning with Deep Streaming Linear Discriminant Analysis

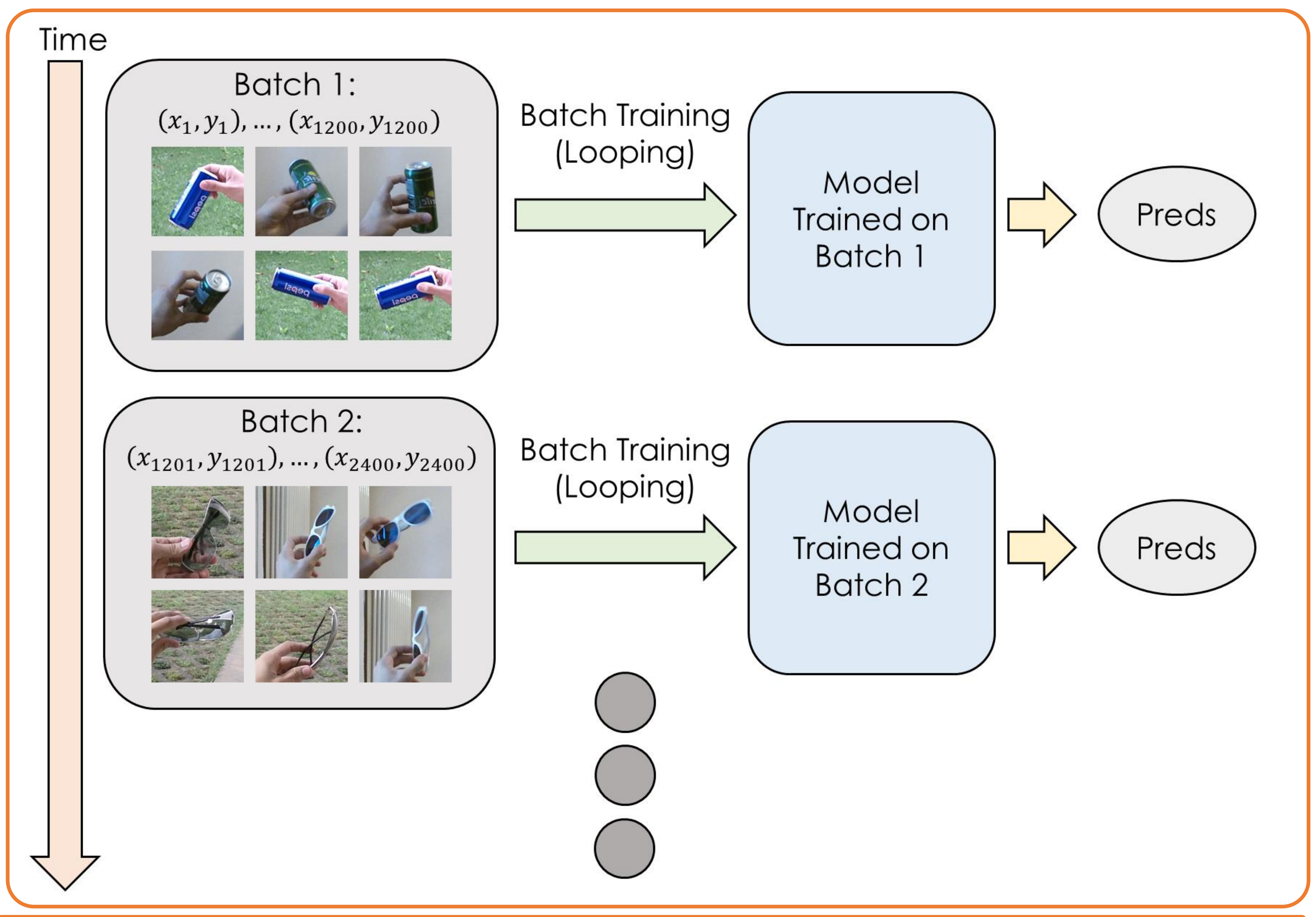
Tyler L. Hayes and Christopher Kanan
Rochester Institute of Technology, Rochester NY
tlh6792@rit.edu; kanan@rit.edu

kLab

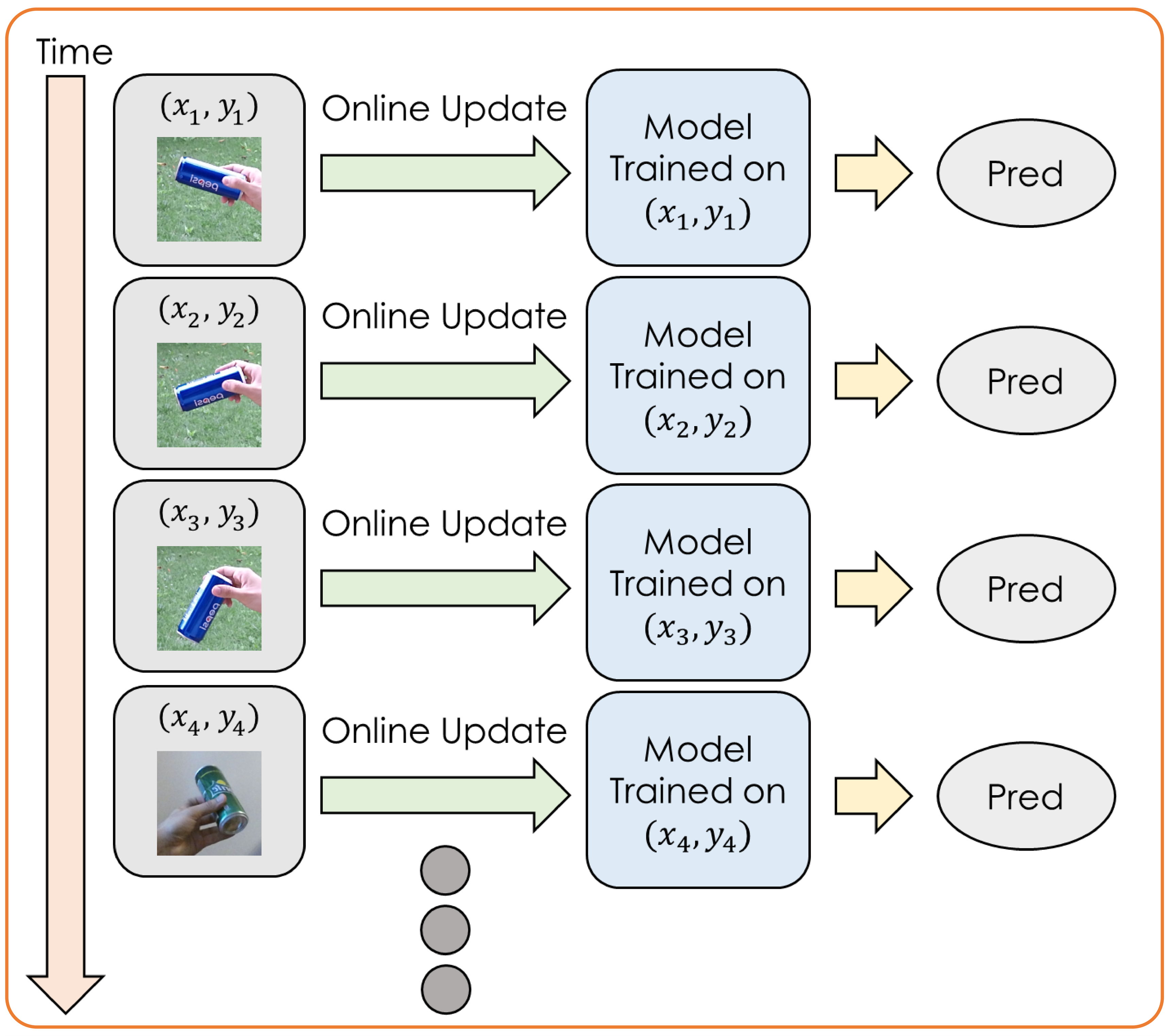
Overview

- Agents must be capable of learning and using information immediately.
- **Deep neural networks** (DNNs) are widely used for perception tasks, but if they are updated on changing data distributions, they **catastrophically forget** previous knowledge.
- **Streaming learning** requires agents to learn from non-independent and identically distributed (iid) data streams in real-time, i.e., one example at a time and a single pass through the dataset.
- **Deep Streaming Linear Discriminant Analysis (SLDA)** trains the output layer of a convolutional neural network (CNN) incrementally.
- SLDA outperforms recent incremental batch and streaming models with fewer memory and computational costs.

Incremental Batch Learning



Streaming Learning



Deep Streaming Linear Discriminant Analysis

- SLDA stores a running mean per class (μ_k) and a tied covariance matrix (Σ).
- We compute the precision matrix $\Lambda = [(1-\epsilon)\Sigma + \epsilon I]^{-1}$.
- Predictions are made by assigning to an input embedding z_t the label of the closest Gaussian in feature space using the stored means and covariance:

$$\hat{y}_t = \underset{k}{\operatorname{argmax}} \left[(\Lambda \mu_k)^T z_t - \frac{1}{2} (\mu_k \cdot \Lambda \mu_k) \right].$$

Experimental Evaluation

- **ImageNet-1K:** Popular large-scale image classification dataset (1,000 classes).
- **CORe50:** Streaming dataset containing video sequences of 10 different object categories. Temporal dependences are natural for streaming.

$$\Omega_{all} = \frac{1}{T} \sum_{t=1}^{T} \frac{\alpha_t}{\alpha_{\text{offline},t}}$$

α_t = accuracy of streaming learner at time t
$\alpha_{\text{offline},t}$ = accuracy of offline model at time t

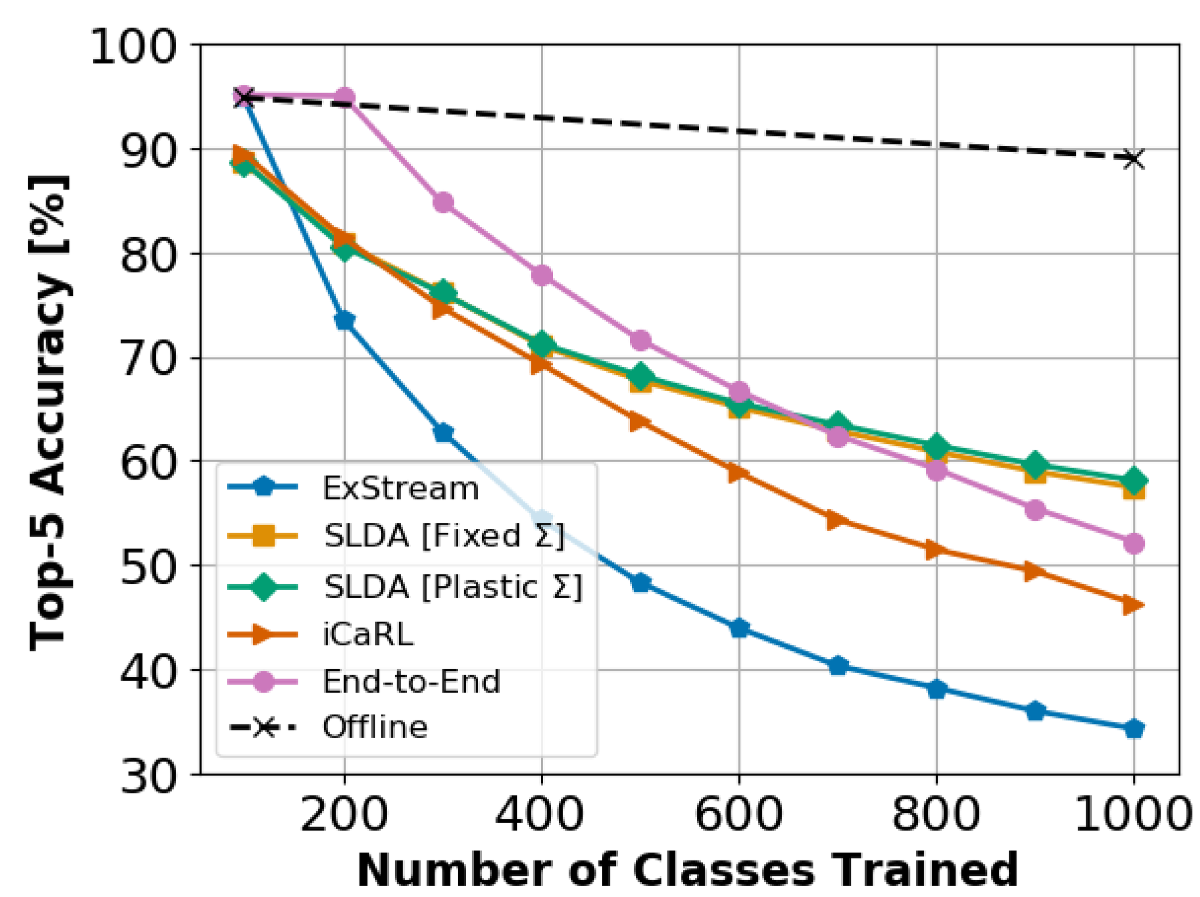
Streaming Learning Paradigms

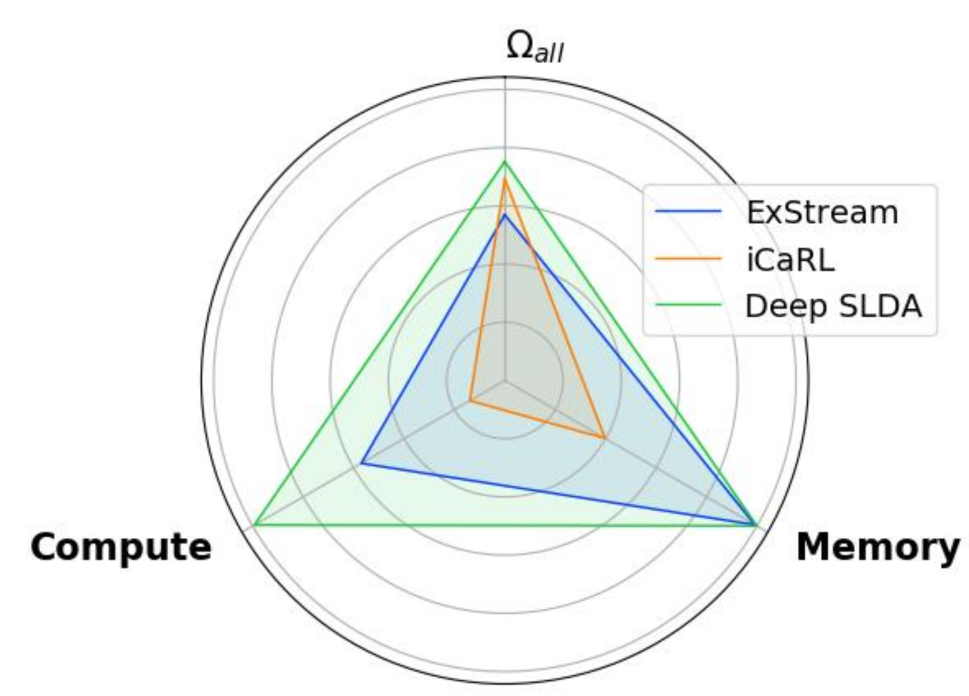
- **iid:** data stream is randomly shuffled.
- **Class iid:** data stream is organized by class.
- **Instance:** data stream is temporally ordered by object instances.
- **Class Instance:** data stream is temporally ordered by object instances by class.

Comparison Models

- We compare several models with the ResNet-18 CNN:
- **Deep SLDA:** Two variants: a fixed covariance and a plastic covariance.
- **ExStream:** Streaming learner that uses partial rehearsal and clustering.
- **iCaRL:** Popular incremental batch model that stores images for replay and uses distillation loss. Uses nearest class mean classifier.
- **End-to-End:** State-of-the-art incremental batch model on ImageNet-1K. Stores images for replay and uses distillation like iCaRL. Uses the CNN for classification and uses multiple augmentation techniques.
- **Offline:** Optimized offline learner. An upper bound on performance.

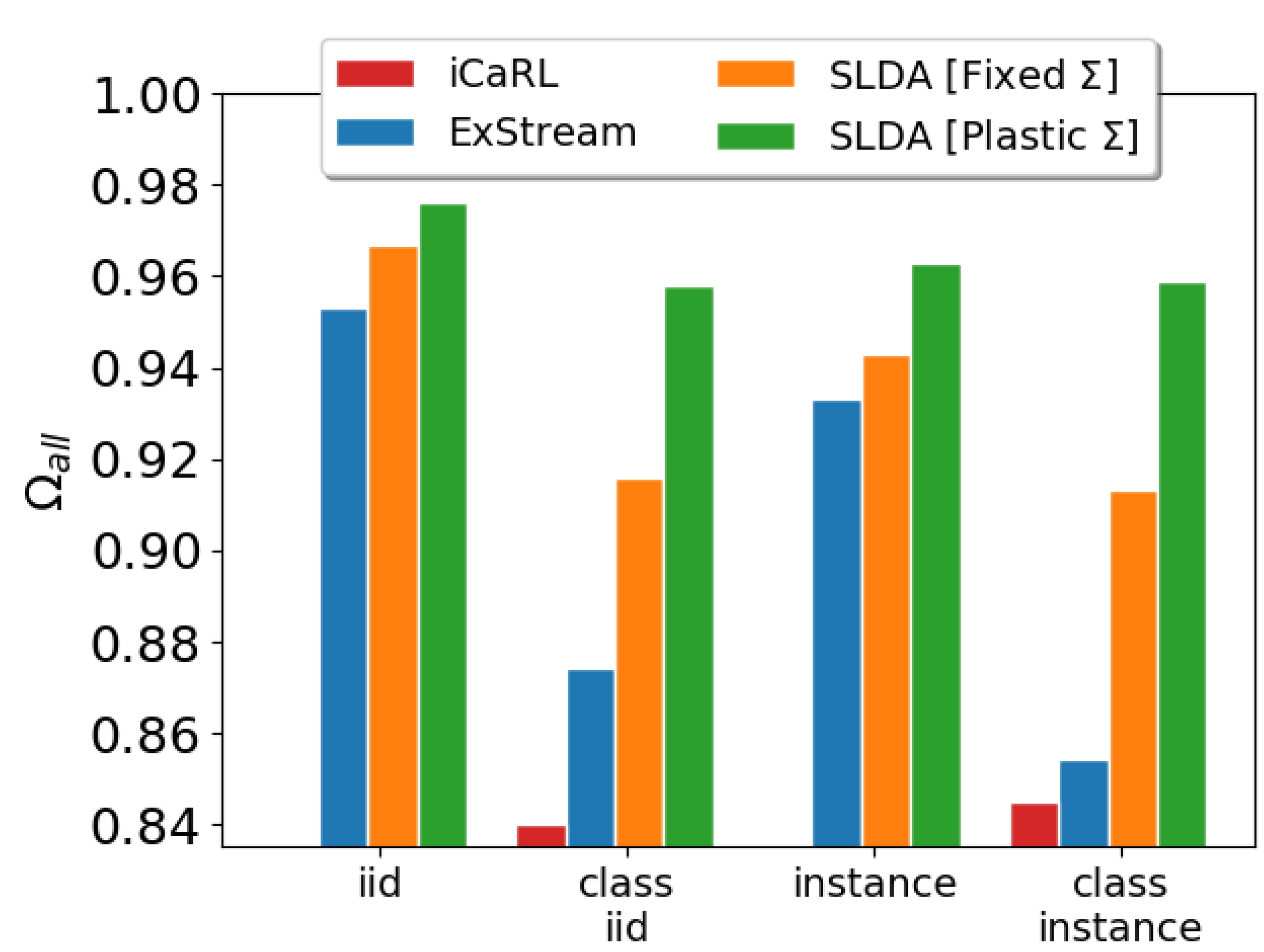
ImageNet-1K Results





SLDA runs 137x faster than iCaRL (in only 27 minutes) and saves over 1000x the memory!

CORe50 Results



Summary

- SLDA is popular in the data mining community but has not been used recently for **large classification datasets**.
- We **combine SLDA with a CNN** and exceed incremental batch learning models, while being much more **lightweight**.
- Our offline results suggest greater performance is achievable by training hidden layers, but we urge future developers to test only training the output layer to ensure gains are being realized.

Acknowledgements:
We thank DARPA L2M for financially supporting this research.