The Challenges of Continuous Self-Supervised Learning
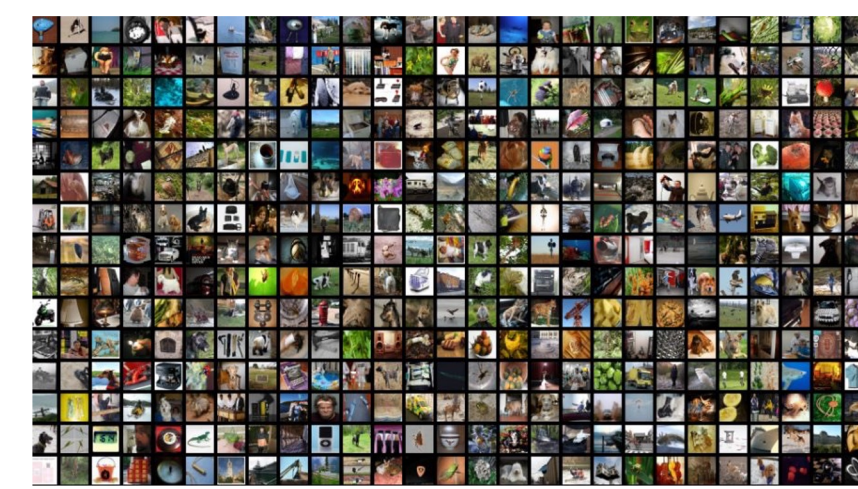
Senthil Purushwalkam*, Pedro Morgado*, Abhinav Gupta
http://www.senthilpurushwalkam.com/publication/continuoussl/
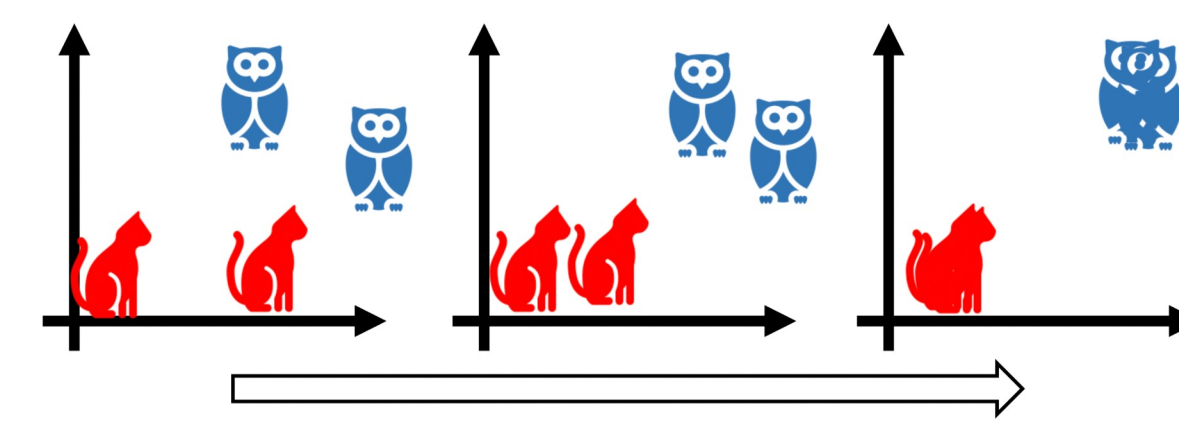
Carnegie Mellon University
The Robotics Institute

Conventional SSL

Finite Data

IID Data

Stationary Semantics
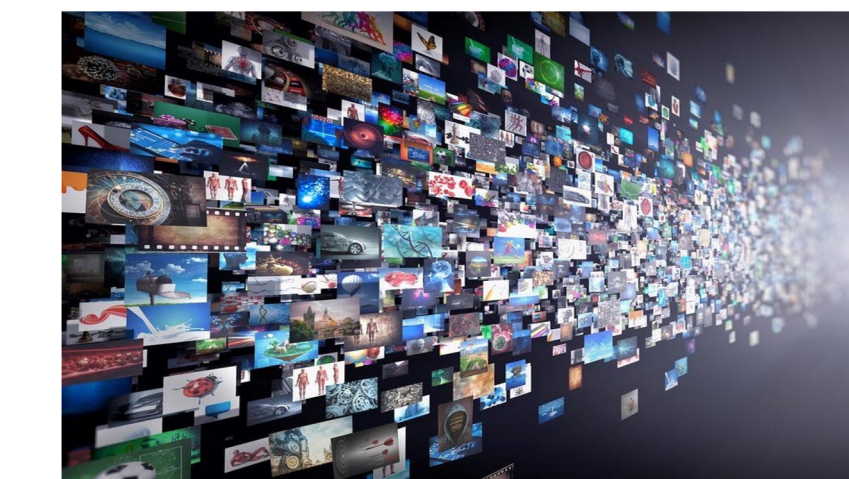
Continuous SSL

Infinite Data

Non-IID Data

Non-stationary Semantics
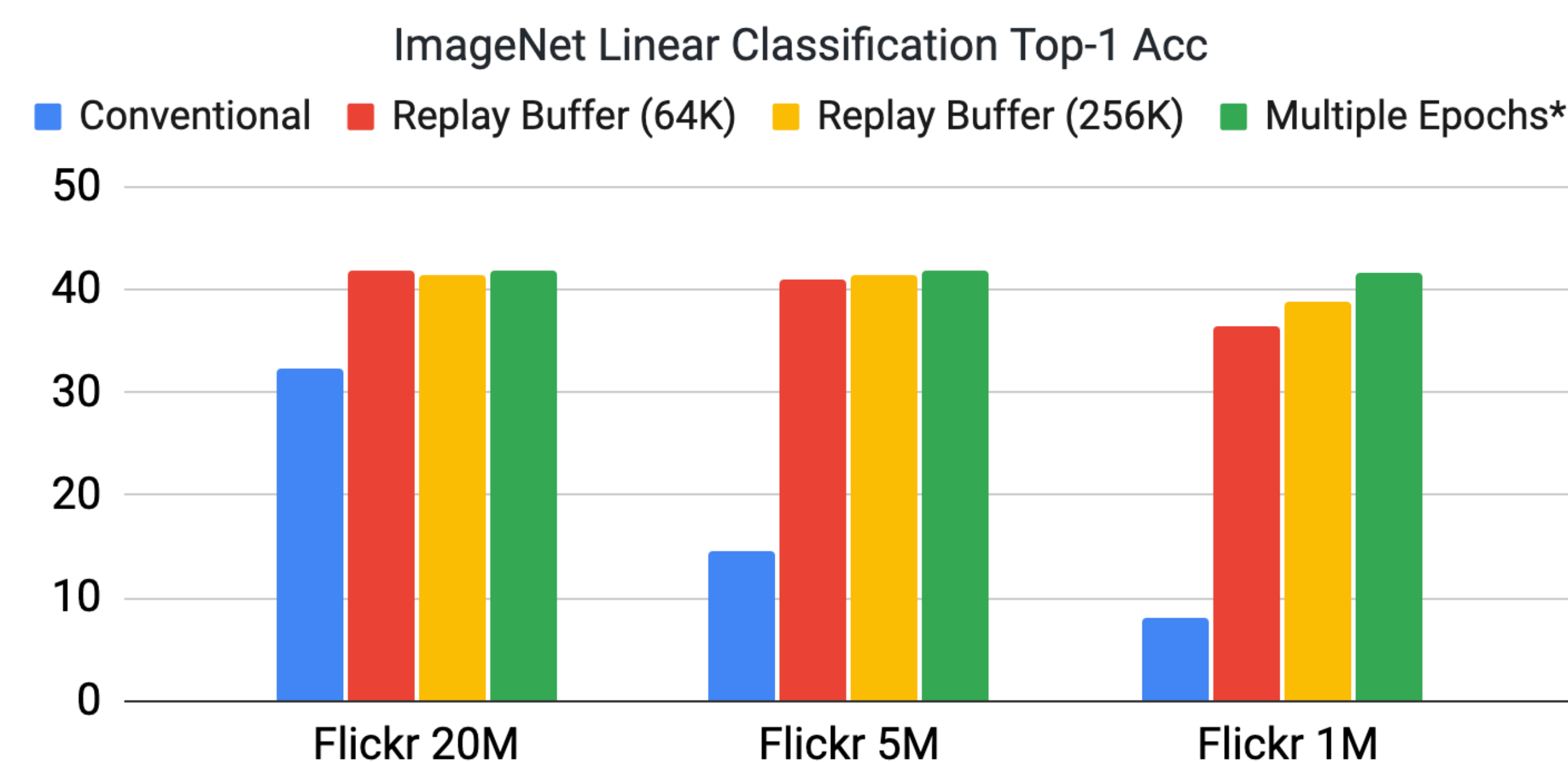

Infinite Data
⇒Single Pass Training

Data: Images sampled from *flickr*

Solution: Replay Buffers
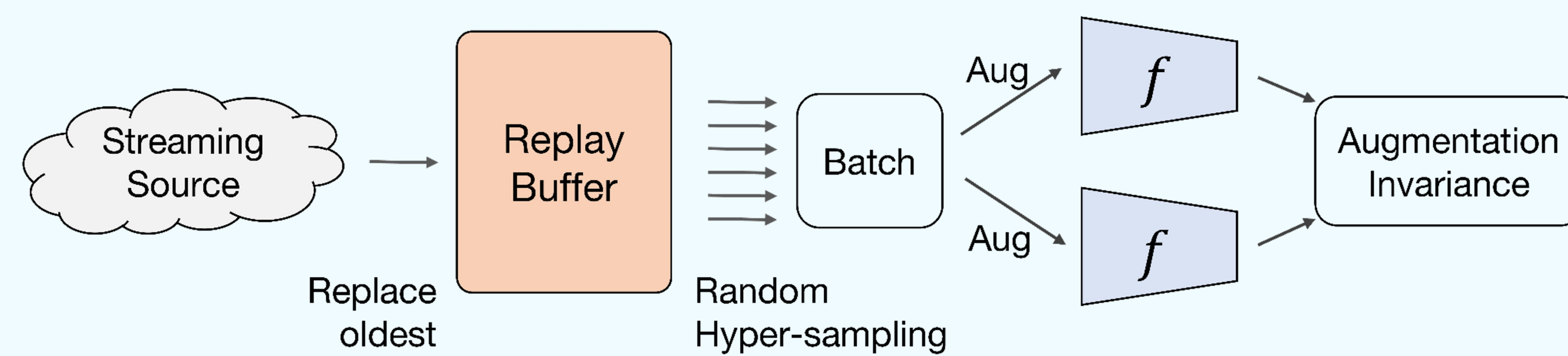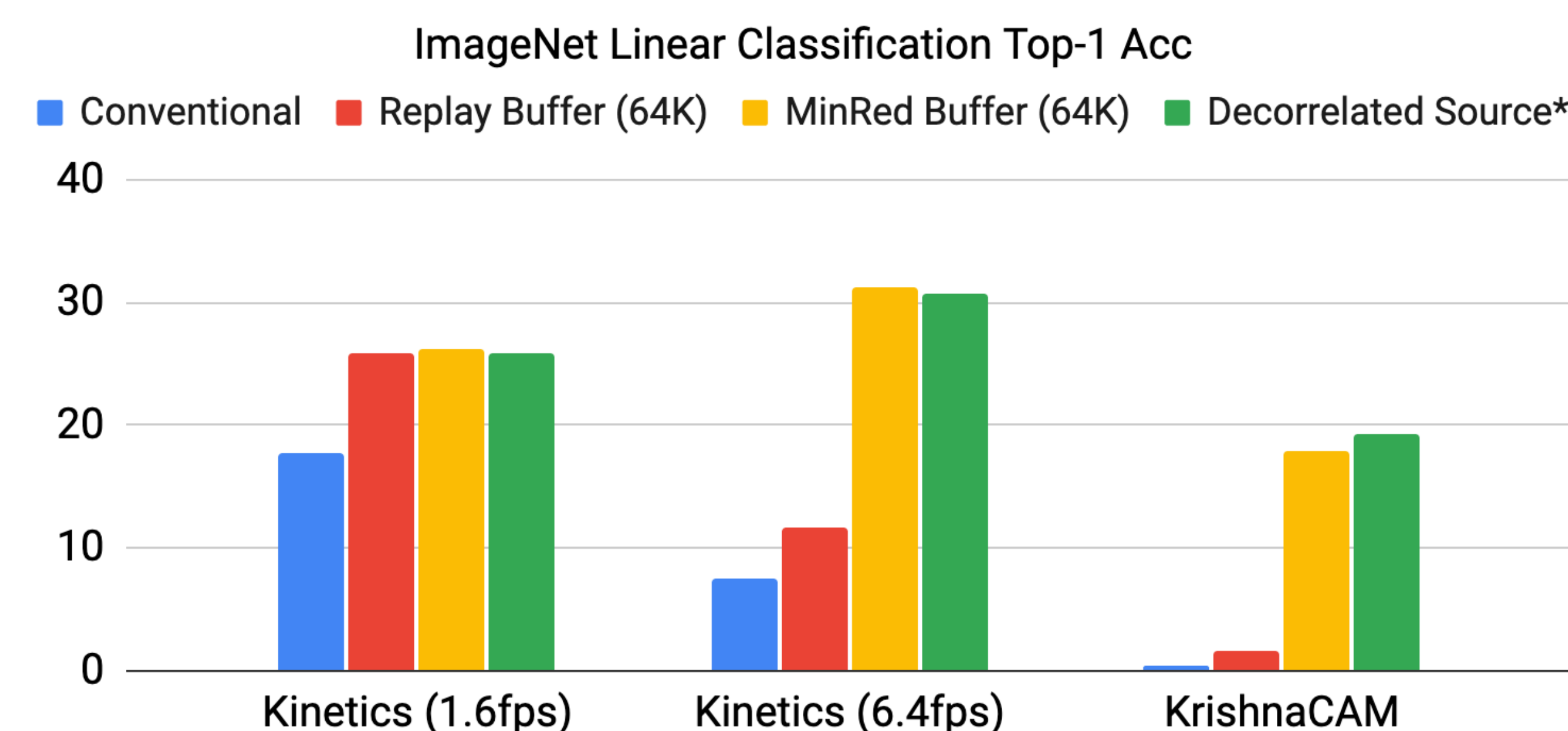

Allows re-using samples from streaming source

ImageNet Linear Classification Top-1 Acc
(Conventional, Replay Buffer (64K), Replay Buffer (256K), Multiple Epochs*)
Flickr 20M, Flickr 5M, Flickr 1M

Non-IID Data
⇒Poor Optimization

Data: Sequential frames from videos

Kinetics [1]

Krishna CAM [2]
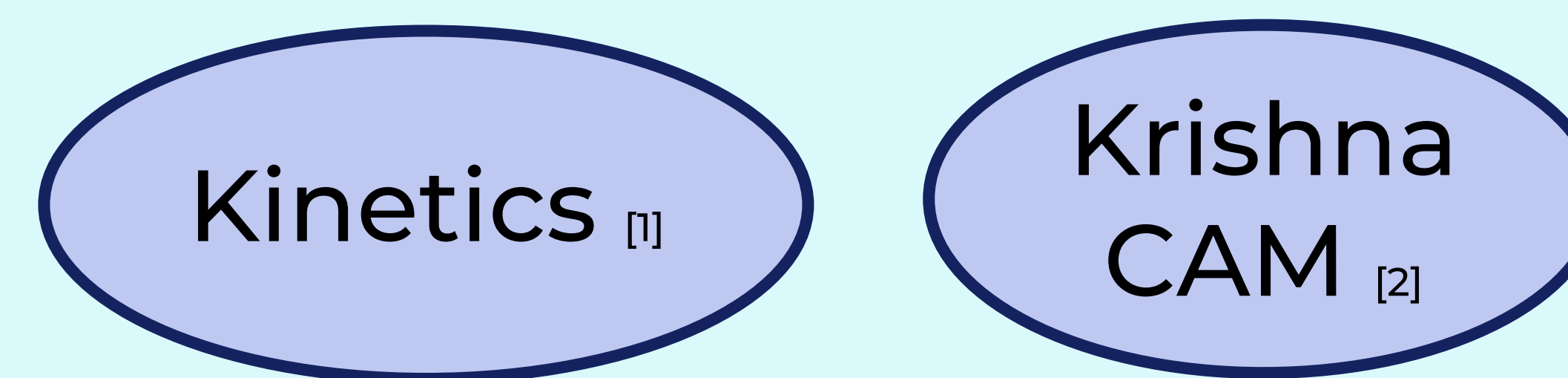
Solution: Minimum Redundancy Buffers

Replace Redundant Samples

$$\mathcal{B} \leftarrow \mathcal{B} \setminus i^*$$

$$i^* = \arg\min_{i \in \mathcal{B}} \min_{j \in \mathcal{B}} d_{\cos}(\bar{z}_i, \bar{z}_j)$$

ImageNet Linear Classification Top-1 Acc
(Conventional, Replay Buffer (64K), MinRed Buffer (64K), Decorrelated Source*)
Kinetics (1.6fps), Kinetics (6.4fps), KrishnaCAM
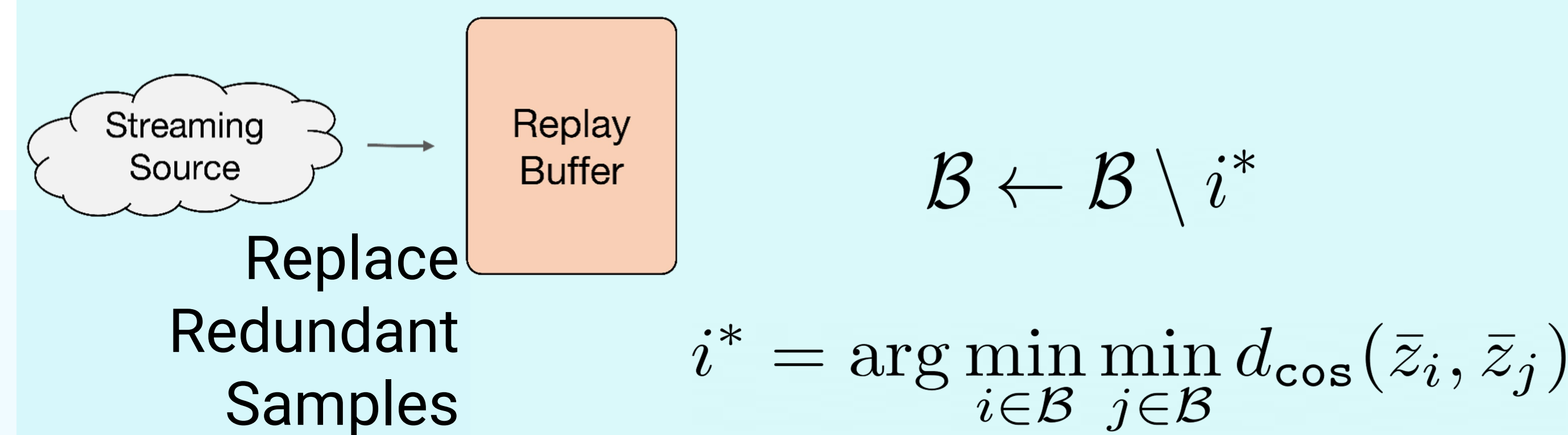
Non-stationary Semantics
⇒Catastrophic Forgetting

Data:

Divide ImageNet-21K classes

D_1 D_2 D_3 D_4

Train sequentially

D_1 D_2 D_3 D_4

Training Progress

Conventional, Replay Buffer*, MinRed Buffer*

Accuracy Change on D0 Classes
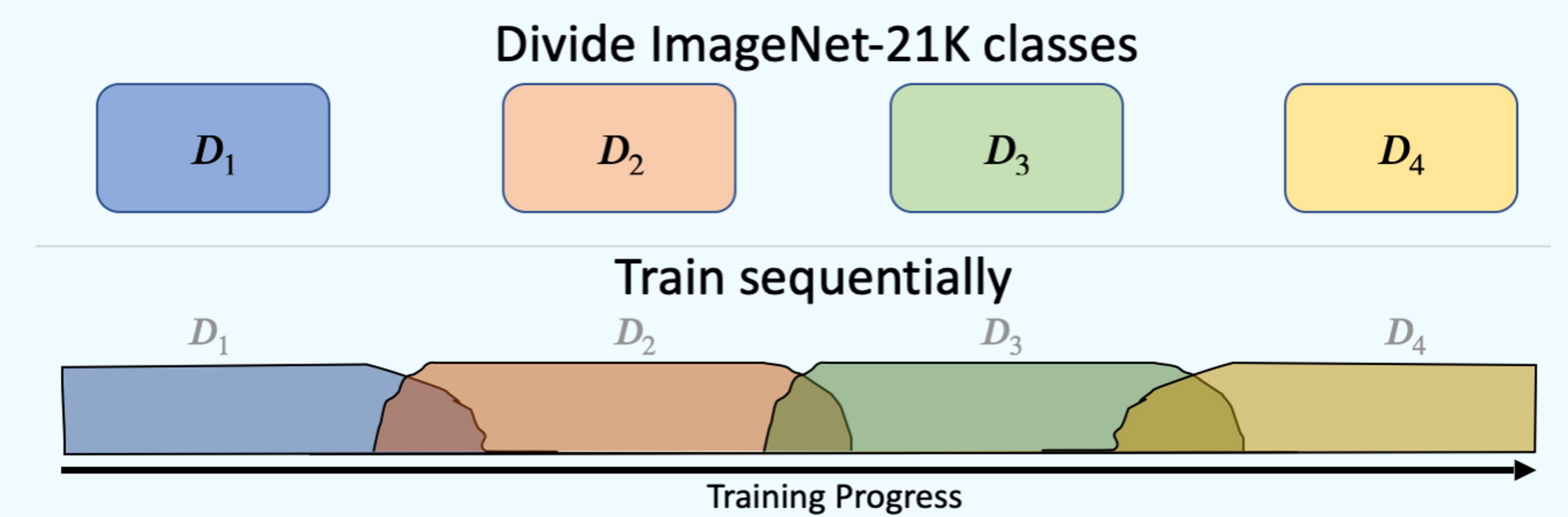After D1, After D2, After D3

Accuracy Change on D1 Classes
After D2, After D3

Accuracy Change on D2 Classes
After D3

---
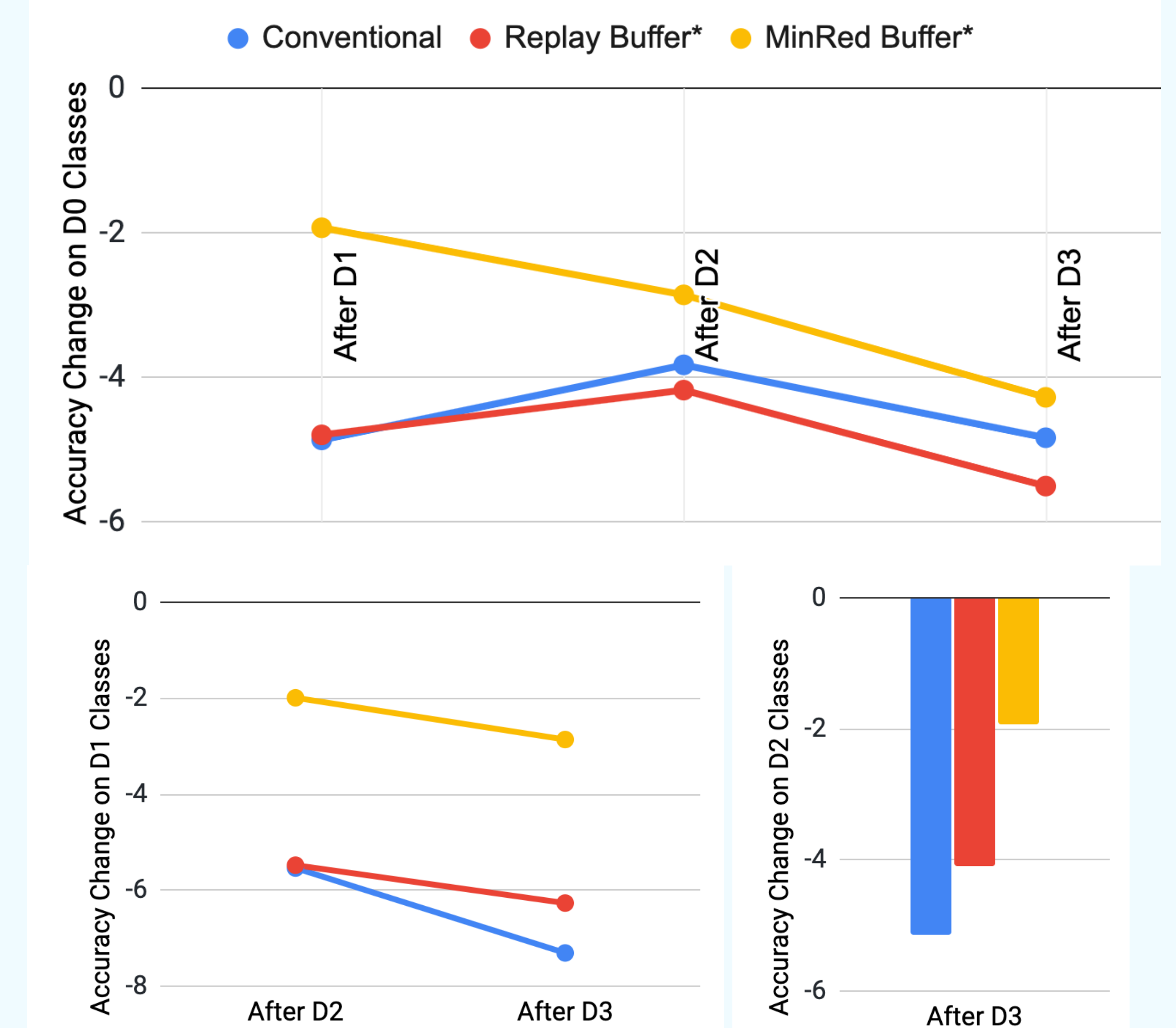
1. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6299–6308 (2017)
2. Singh, K.K., Fatahalian, K., Efros, A.A.: Krishnacam: Using a longitudinal, single-person, egocentric dataset for scene understanding tasks. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1–9. IEEE (2016)
3. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15750–15758 (2021)