Imbalanced Continual Learning with **P**artitioning **R**eservoir **S**ampling

ECCV 2020


Chris Dongjoo Kim

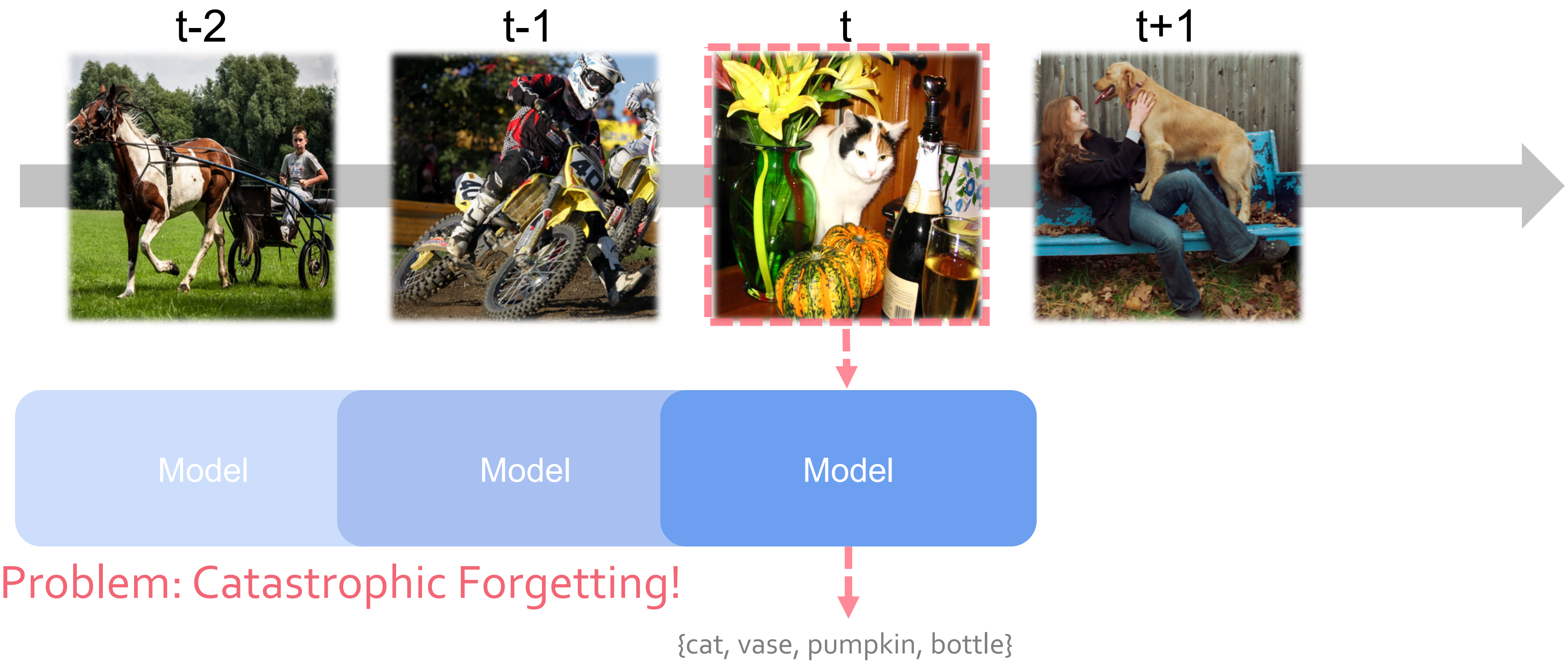

Jinseo Jeong


Gunhee Kim

SEOUL NATIONAL UNIV.
VISION & LEARNING

Continual Learning with Multi-Labels

✓ First work to explore this



Problem: Catastrophic Forgetting!

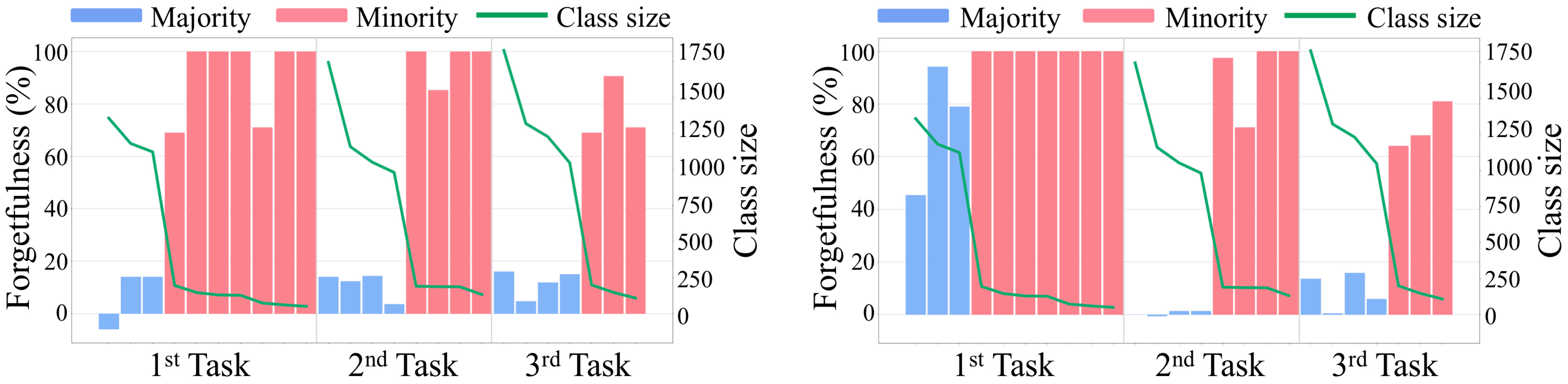
Natural Imbalance in Multi-labeled Datasets

Problem: Long-tailed distribution

Lin et al., *Microsoft COCO: Common objects in Context*, ECCV, 2014

Motivation: Fatal Forgetting on the Tail

- Forgetting measured on the earlier tasks after training



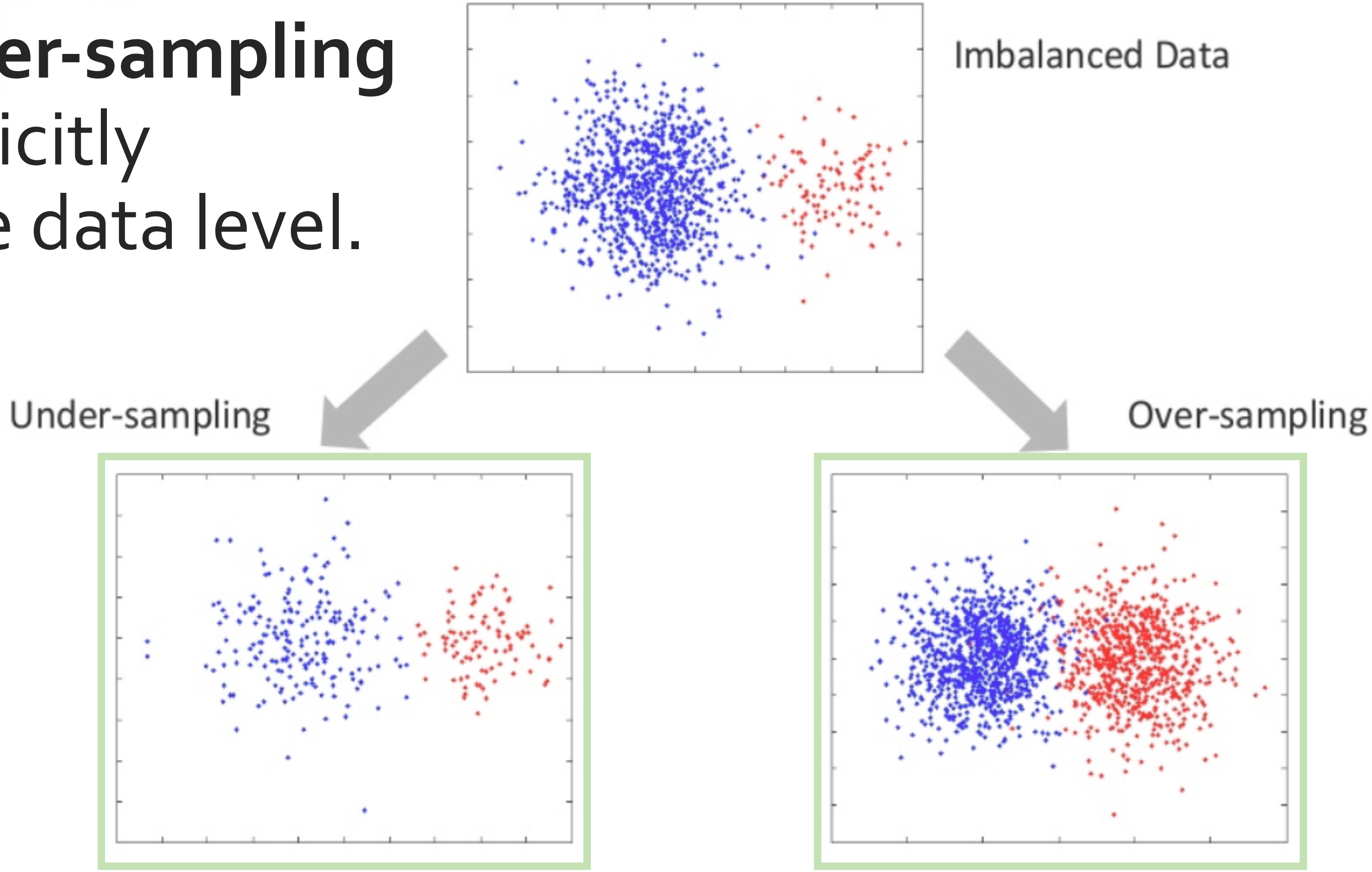
(a) replay-based continual learning (CRS)

(b) prior-focused continual learning (EWC)

Chaudhry et al., *Riemannian walk for incremental learning: Understanding forgetting and intransigence*, ECCV, 2018

Previous works on Imbalance

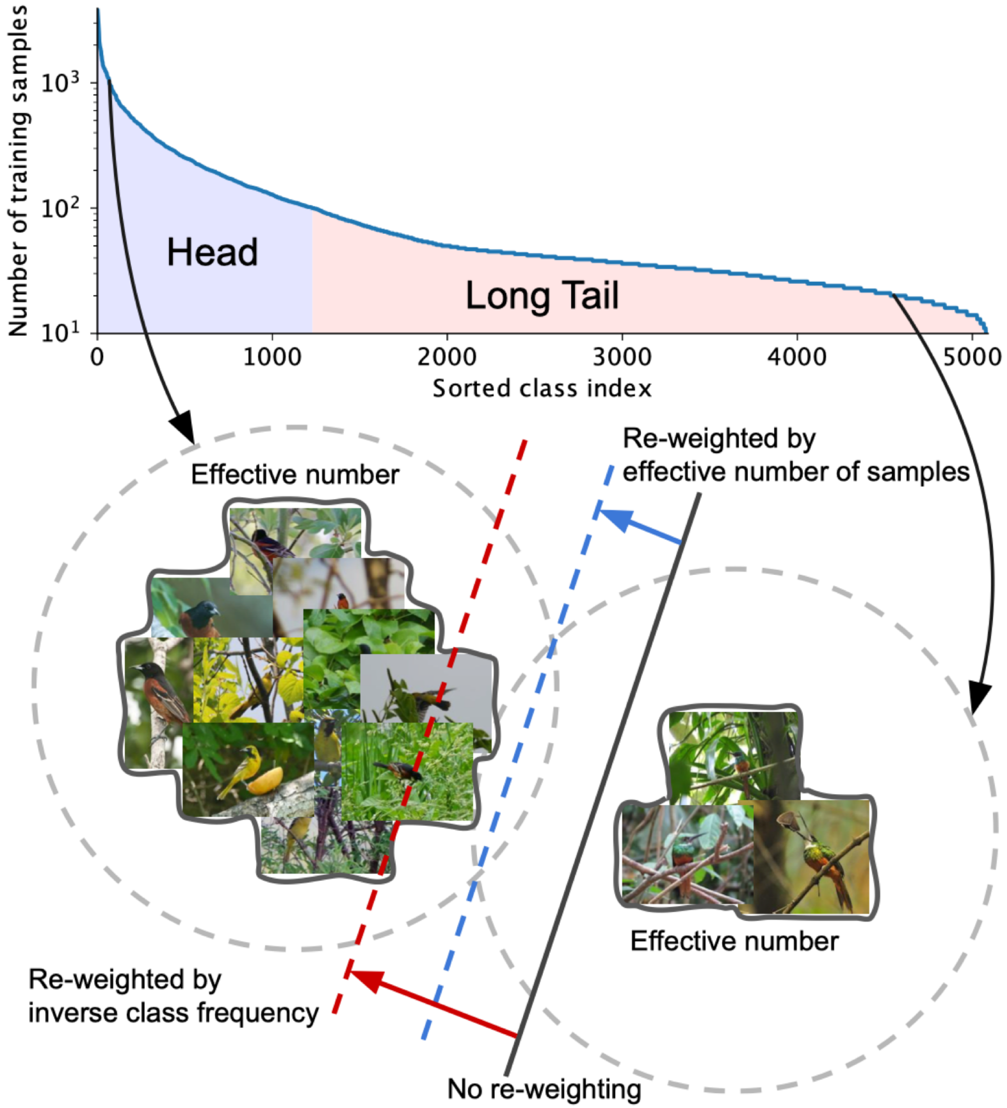
1. **Over-sampling, Under-sampling** and **Generating** to explicitly maintain **balance** at the data level.



Batista et al., *A study of the behavior of several methods for balancing machine learning training data*, SIGKDD, 2004
Buda et al., *A systematic study of the class imbalance problem in convolutional neural networks*, Neural Networks, 2018
Chawla et al., *Smote: synthetic minority over-sampling technique*, Journal of artificial intelligence research, 2002

Previous works on Imbalance

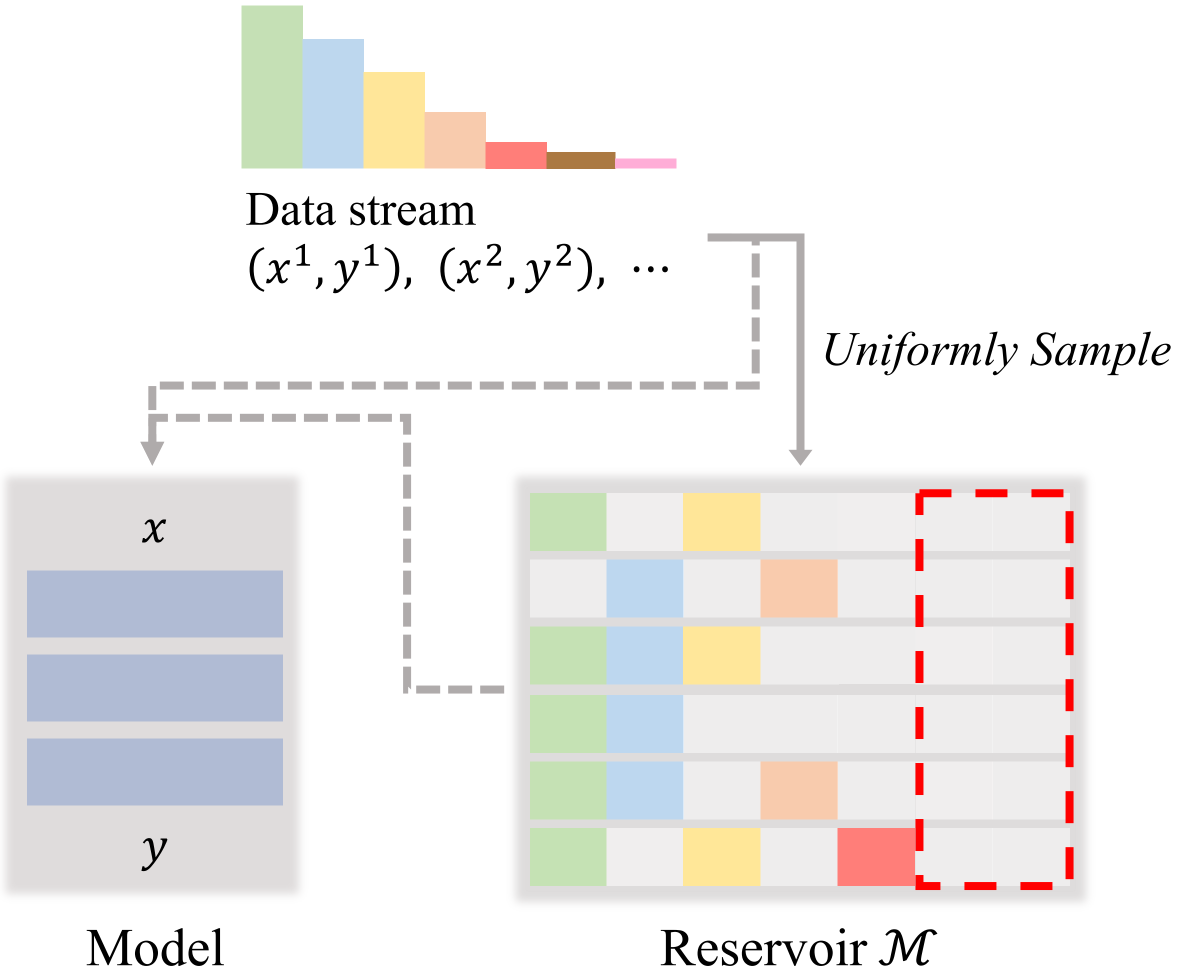
2. Use of **Cost-sensitive** loss approaches to **balance** the training dynamics.



Cui et al., *Class-balanced loss based on effective number of samples.*, CVPR, 2019
Huang et al., *Learning Deep Representation for Imbalanced Classification*, CVPR, 2016

Previous works on Continual Learning

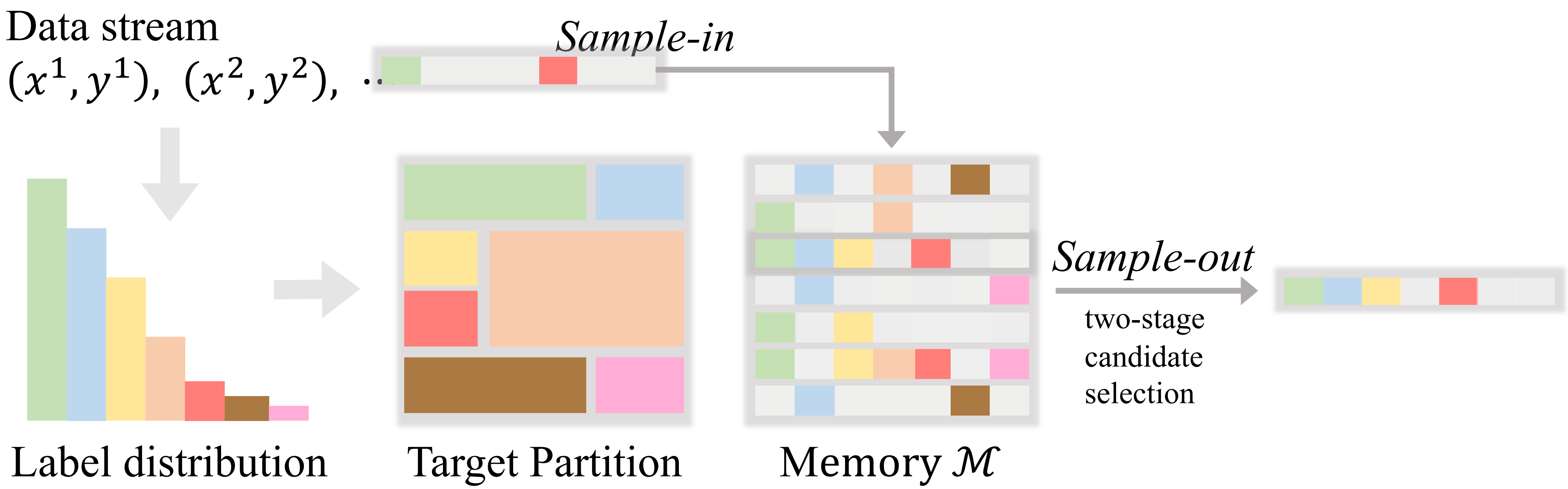
- **Reservoir Sampling** is still a strong baseline in CL.
- It aims to represent the input distribution via simple random sampling



Vitter et al., *Random sampling with a reservoir.*, ACM Transactions on Mathematical Software, 1985

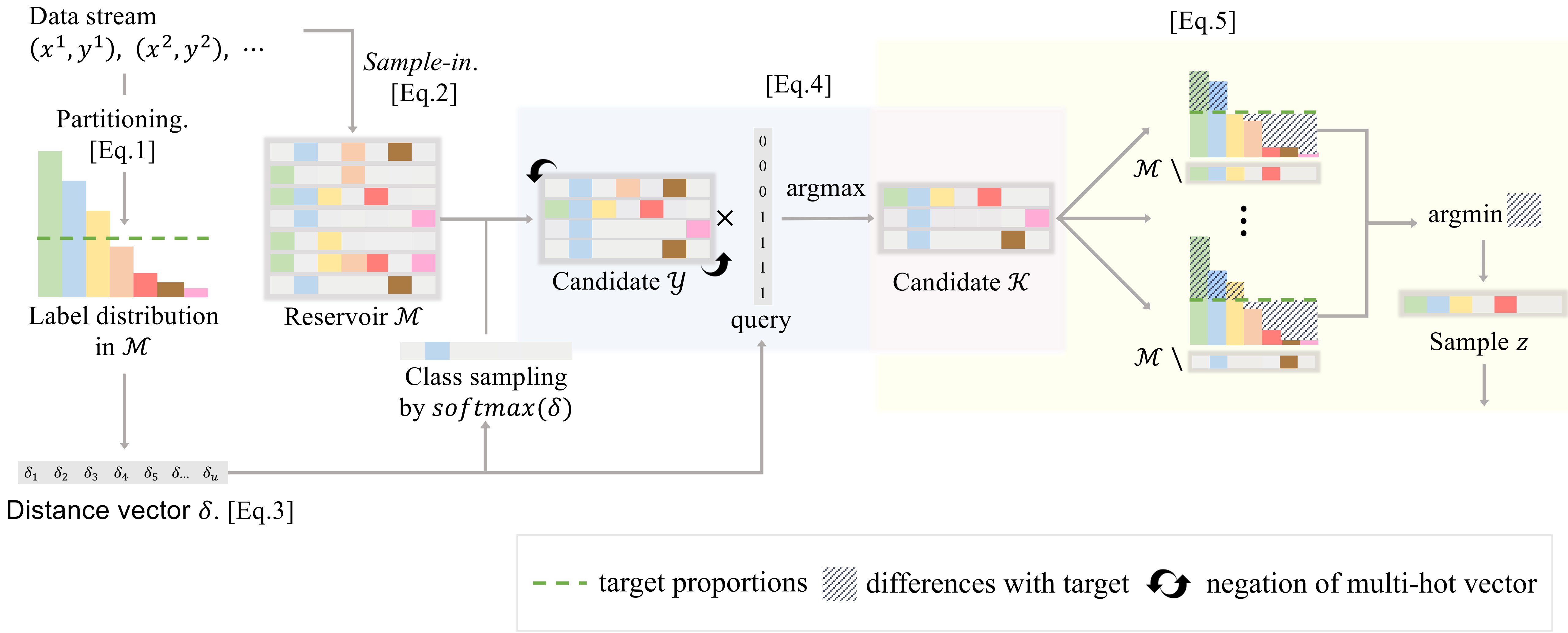
Overview: Partitioning Reservoir Sampling

✓ Easily integrable with other CL approaches

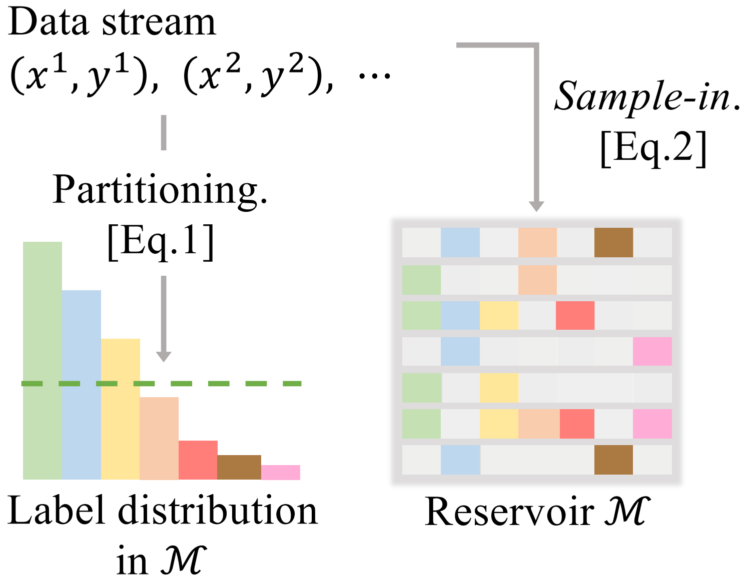


Code & Dataset: http://vision.snu.ac.kr/projects/PRS

Approach: Partitioning Reservoir Sampling



Approach: Partitioning Reservoir Sampling



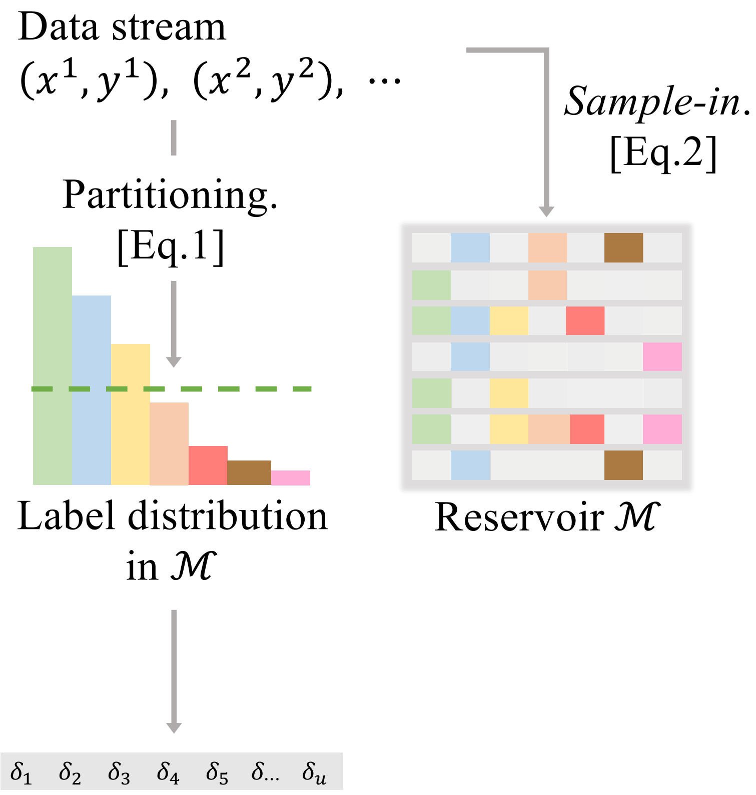
$$s = \sum \frac{m_i}{n_i} \cdot w_i$$

$$w_i = \frac{y_i e^{-n}}{\sum_{j=1} y_j e^{-n}}$$

$$p_i = \frac{n_i^{\rho}}{\sum_j n_j^{\rho}}$$

target proportions   differences with target   negation of multi-hot vector

Approach: Partitioning Reservoir Sampling



Distance vector δ. [Eq.3]

$$\delta_i = l_i - p_i \cdot \sum_j l_j$$

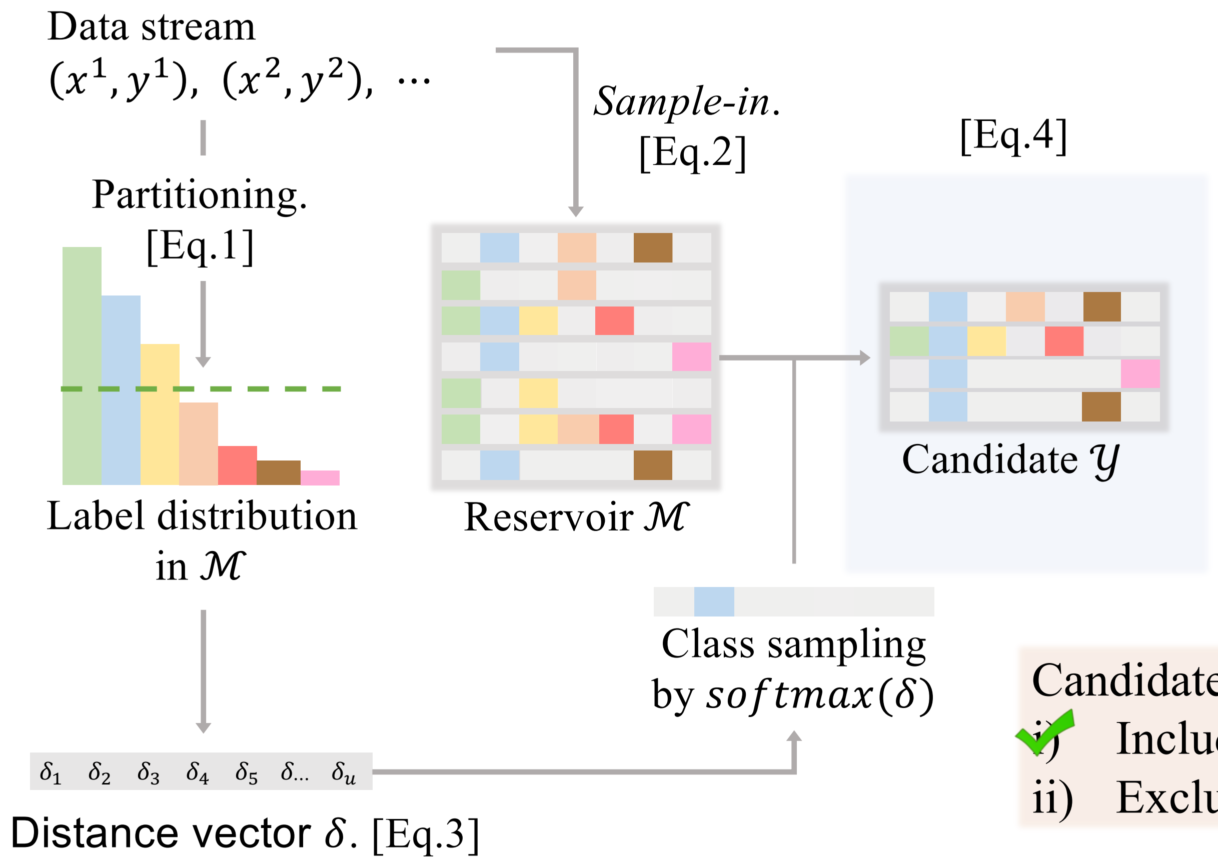
Candidates that satisfy the Desiderata:

i) Include classes that over-occupy the memory than the target. i.e. $\delta_i > 0$

ii) Exclude classes that under-occupy or satisfy the target. i.e. $\delta_i \leq 0$

target proportions | differences with target | negation of multi-hot vector

Approach: Partitioning Reservoir Sampling



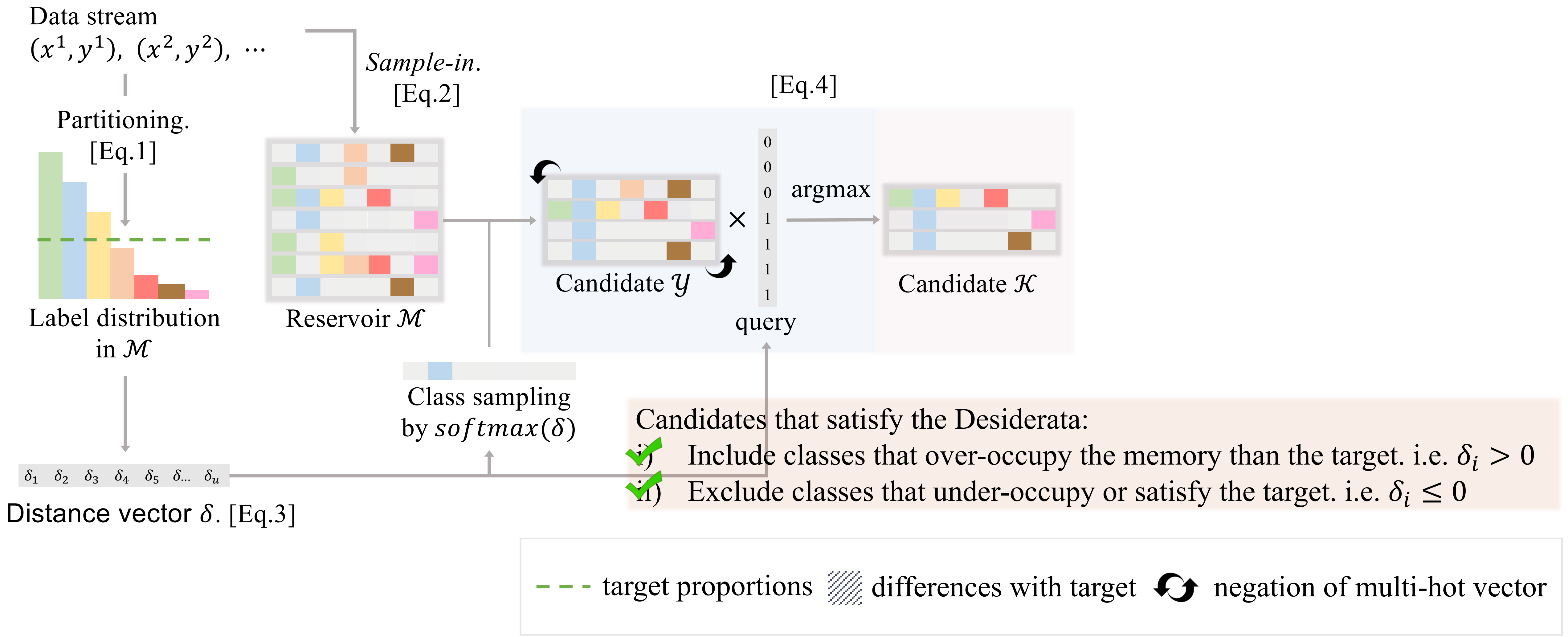
Candidates that satisfy the Desiderata:

i) Include classes that over-occupy the memory than the target. i.e. $\delta_i > 0$

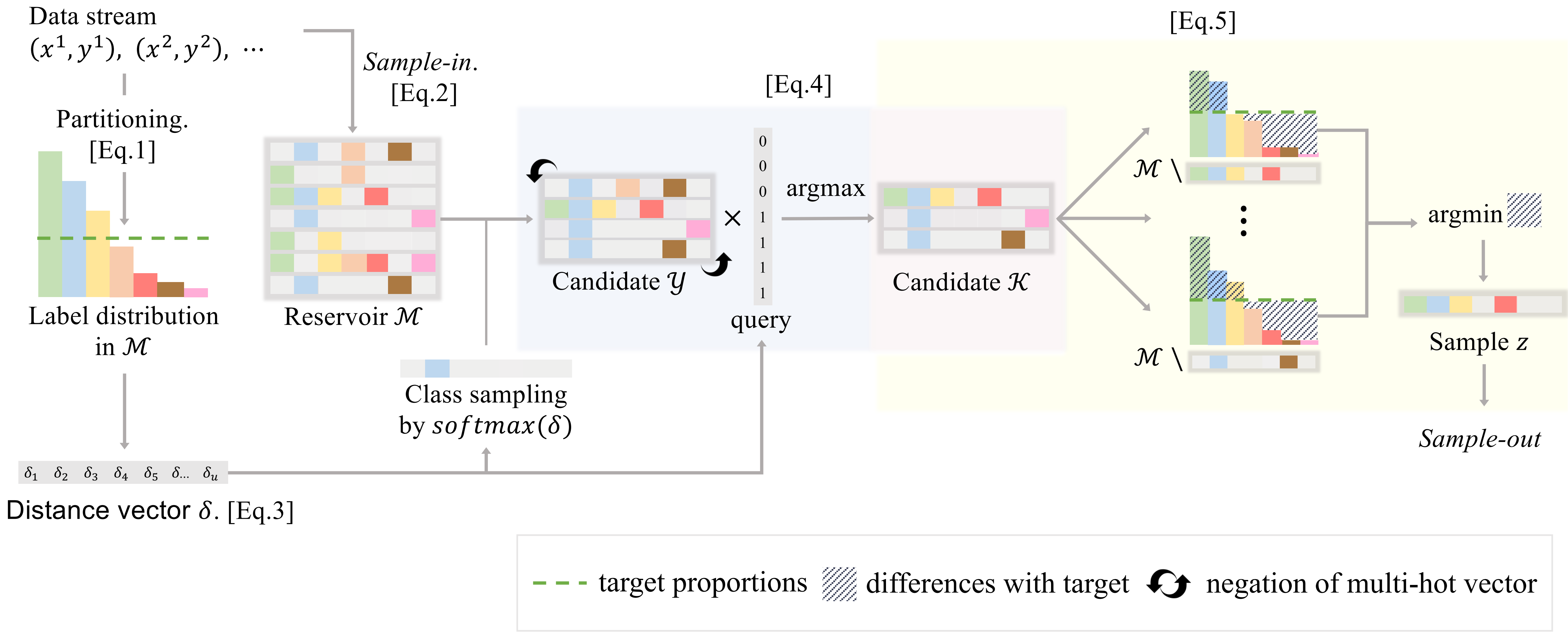
ii) Exclude classes that under-occupy or satisfy the target. i.e. $\delta_i \leq 0$

target proportions   differences with target   negation of multi-hot vector

Approach: Partitioning Reservoir Sampling



Approach: Partitioning Reservoir Sampling



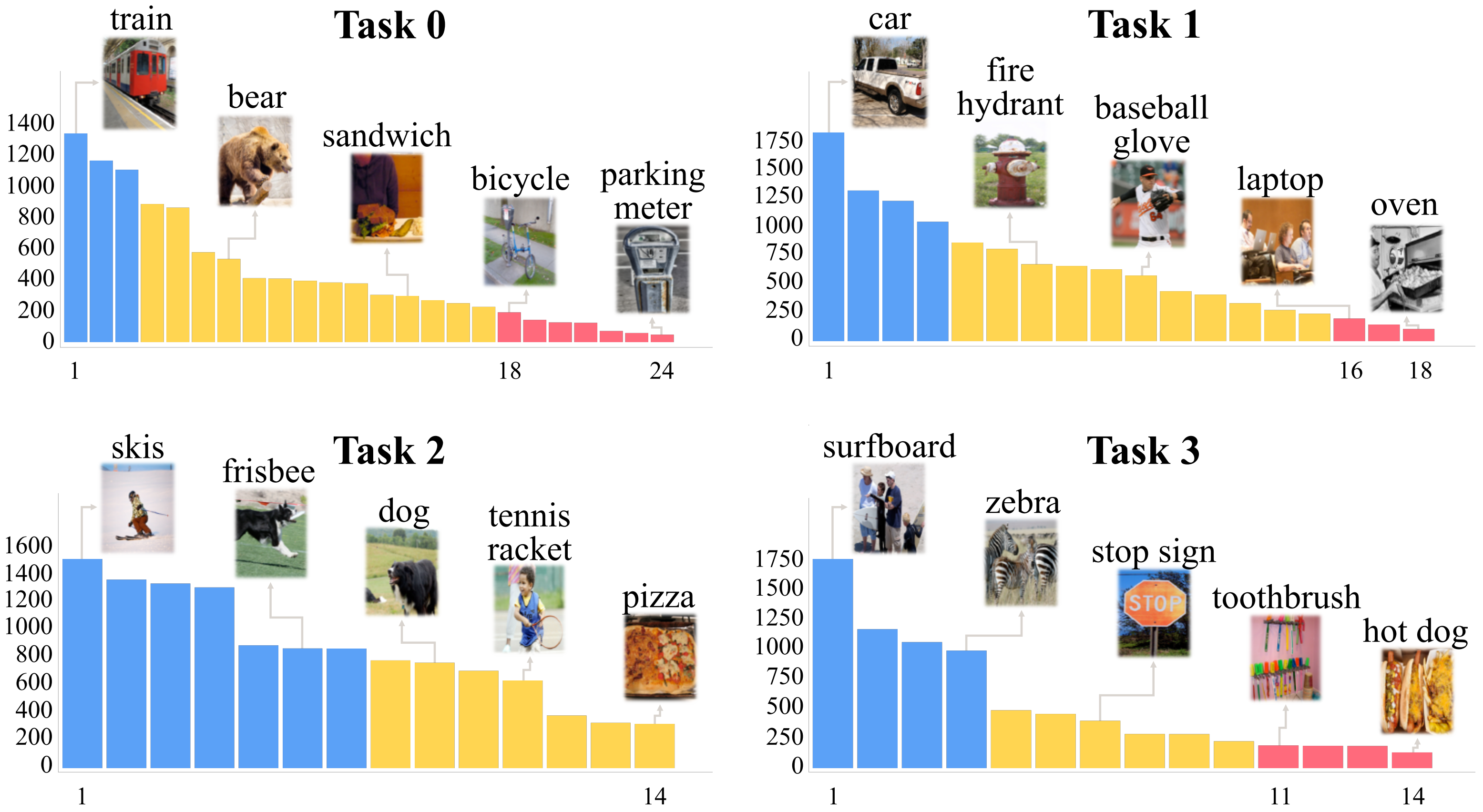
Experiments

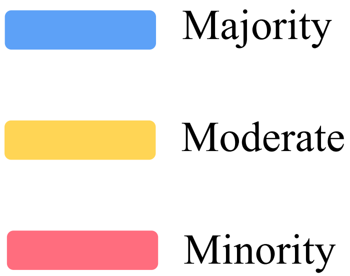
- **Task-free setting**: no task labels during training or test time.
- **Online Learning**: data stream is seen just once.

Aljundi et al., *Gradient based sample selection for online continual learning*, NIPS, 2020
Lee et al., *A neural dirichlet process mixture model for task-free continual learning*, ICLR, 2020

COCOseq

- Curate mutually exclusive tasks.
  - Bottom-up Hierarchical Clustering.
- 4 tasks

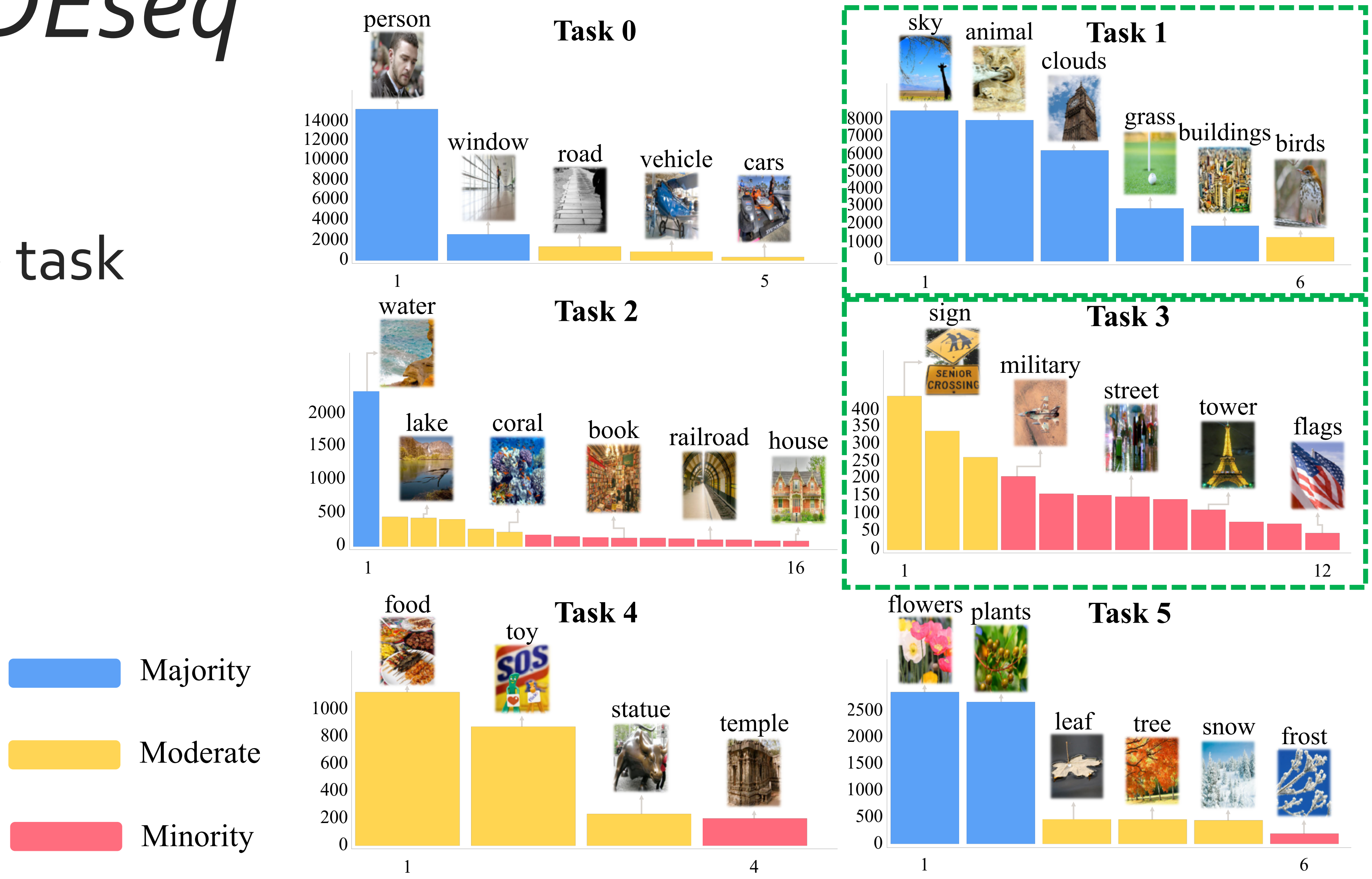




Lin et al., *Microsoft COCO: Common objects in Context*, ECCV, 2014

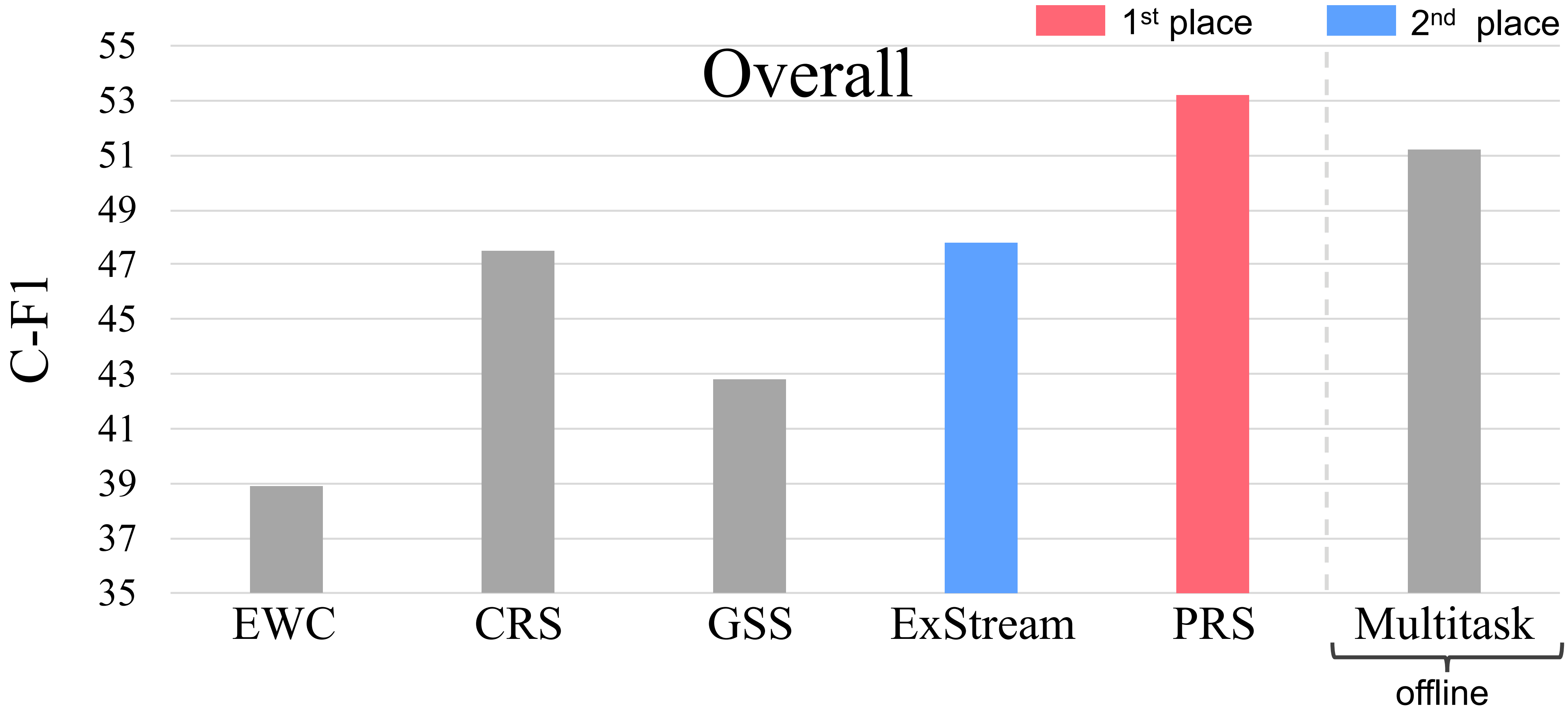
NUS-WIDEseq

- 6 tasks
- *intra-* & *inter-* task imbalances.



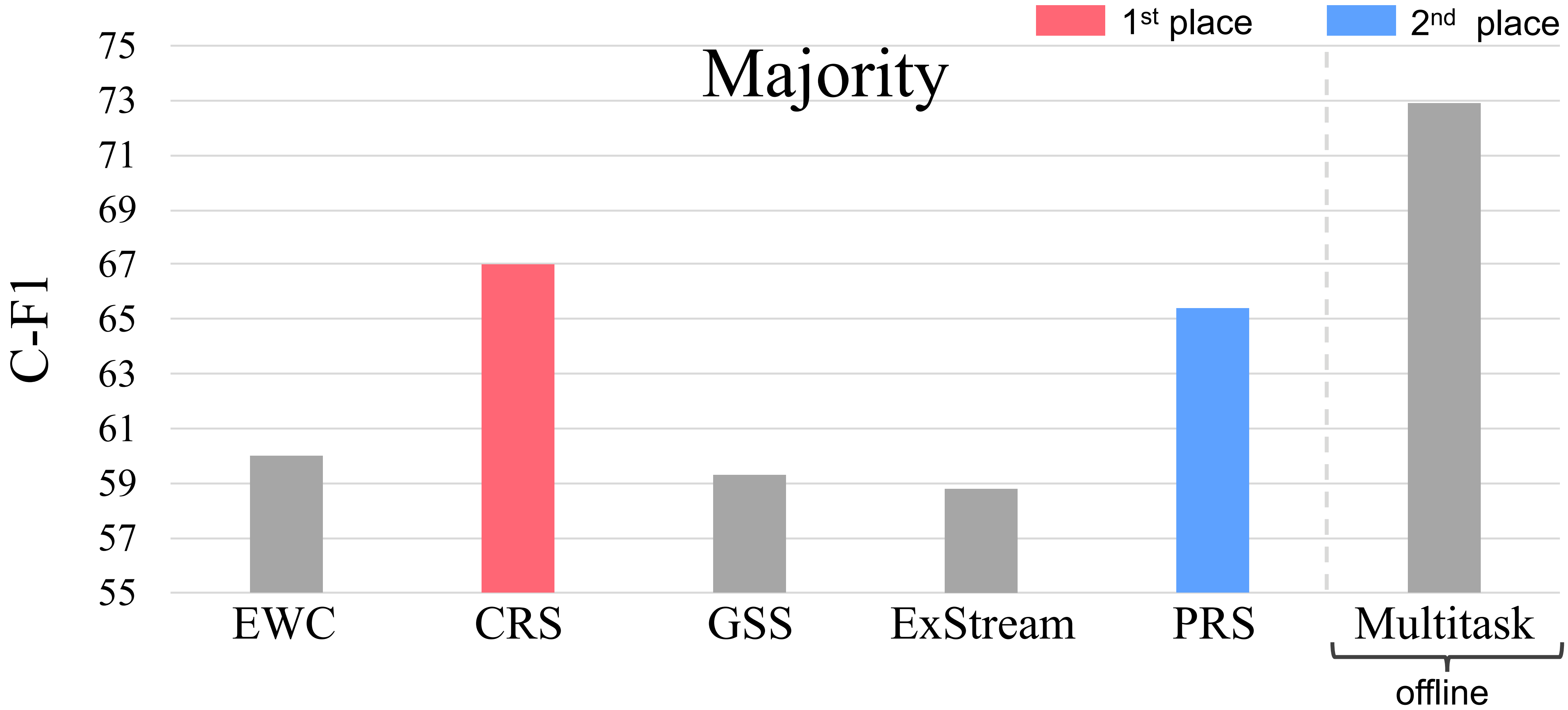
Chua et al., *Nus-wide: a real-world web image database from national university of singapore.*, ACM, 2009

Experiments on COCOseq



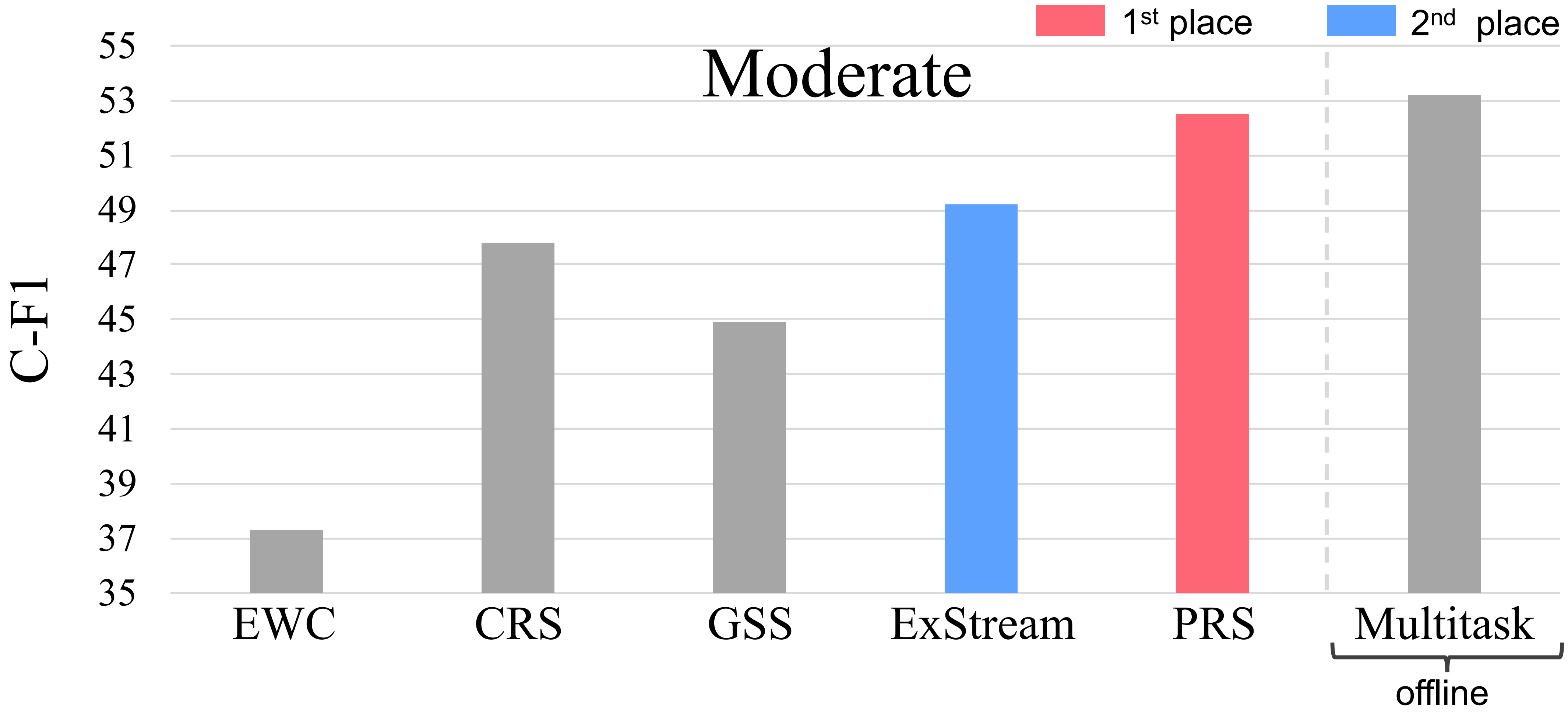
Kirkpatrick et al., *Overcoming catastrophic forgetting in neural networks*, NIPS, 2017
Chaudhry et al., *On Tiny Episodic Memories in Continual Learning*, arxiv, 2019
Aljundi et al., *Gradient based sample selection for online continual learning*, NIPS, 2019
Hayes et al., *Memory efficient experience replay for streaming learning*, ICRA, 2019
Caruaca et al., *Multitask learning*, Machine Learning, 1997

Experiments on COCOseq



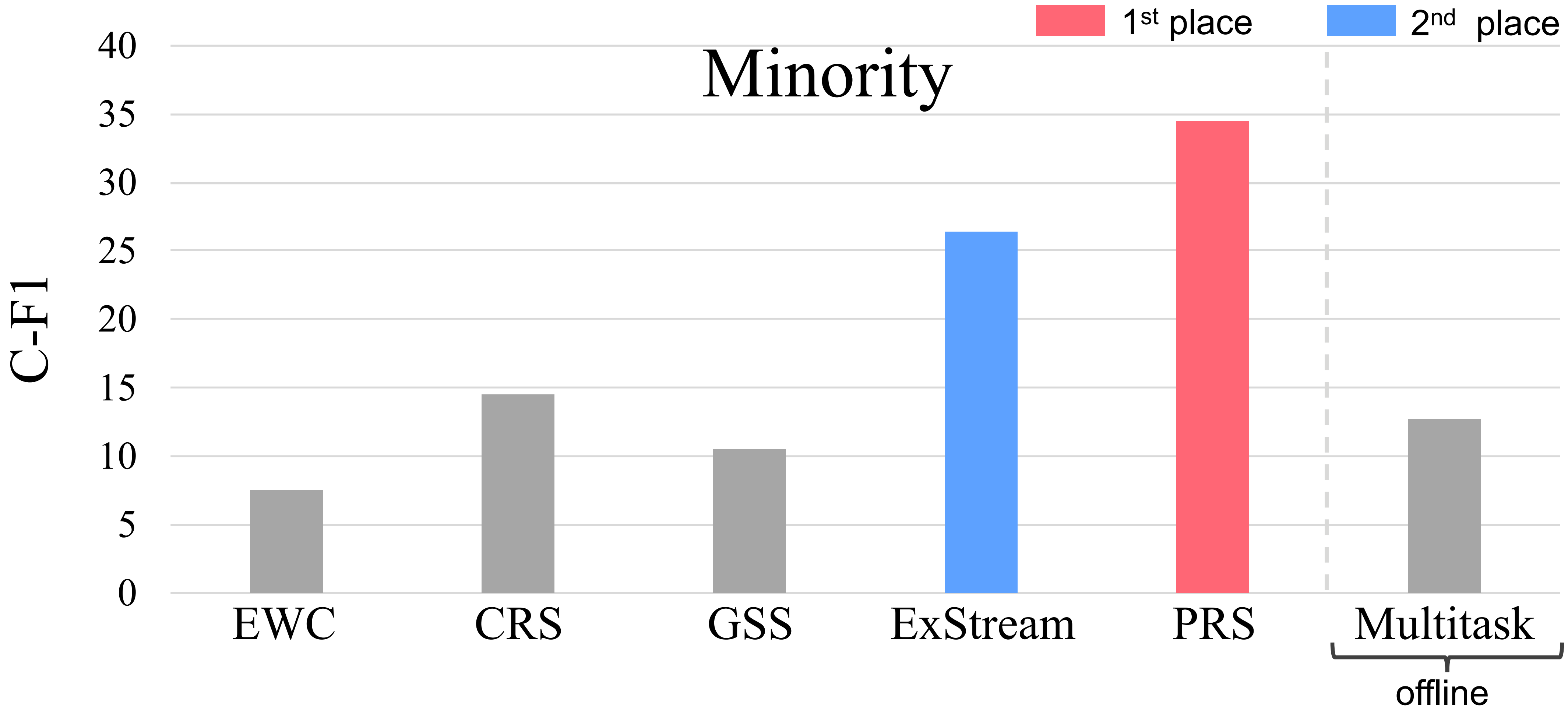
Kirkpatrick et al., *Overcoming catastrophic forgetting in neural networks*, NIPS, 2017
Chaudhry et al., *On Tiny Episodic Memories in Continual Learning*, arxiv, 2019
Aljundi et al., *Gradient based sample selection for online continual learning*, NIPS, 2019
Hayes et al., *Memory efficient experience replay for streaming learning*, ICRA, 2019
Caruaca et al., *Multitask learning*, Machine Learning, 1997

Experiments on COCOseq



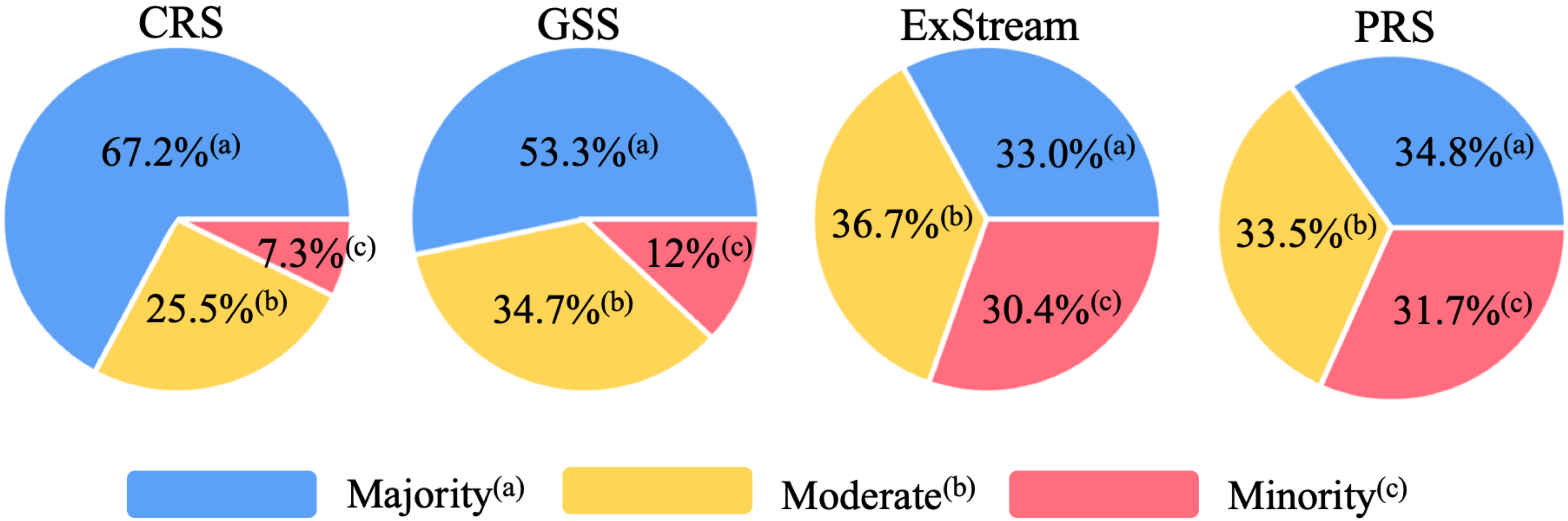
Kirkpatrick et al., *Overcoming catastrophic forgetting in neural networks*, NIPS, 2017
Chaudhry et al., *On Tiny Episodic Memories in Continual Learning*, arxiv, 2019
Aljundi et al., *Gradient based sample selection for online continual learning*, NIPS, 2019
Hayes et al., *Memory efficient experience replay for streaming learning*, ICRA, 2019
Caruaca et al., *Multitask learning*, Machine Learning, 1997

Experiments on COCOseq



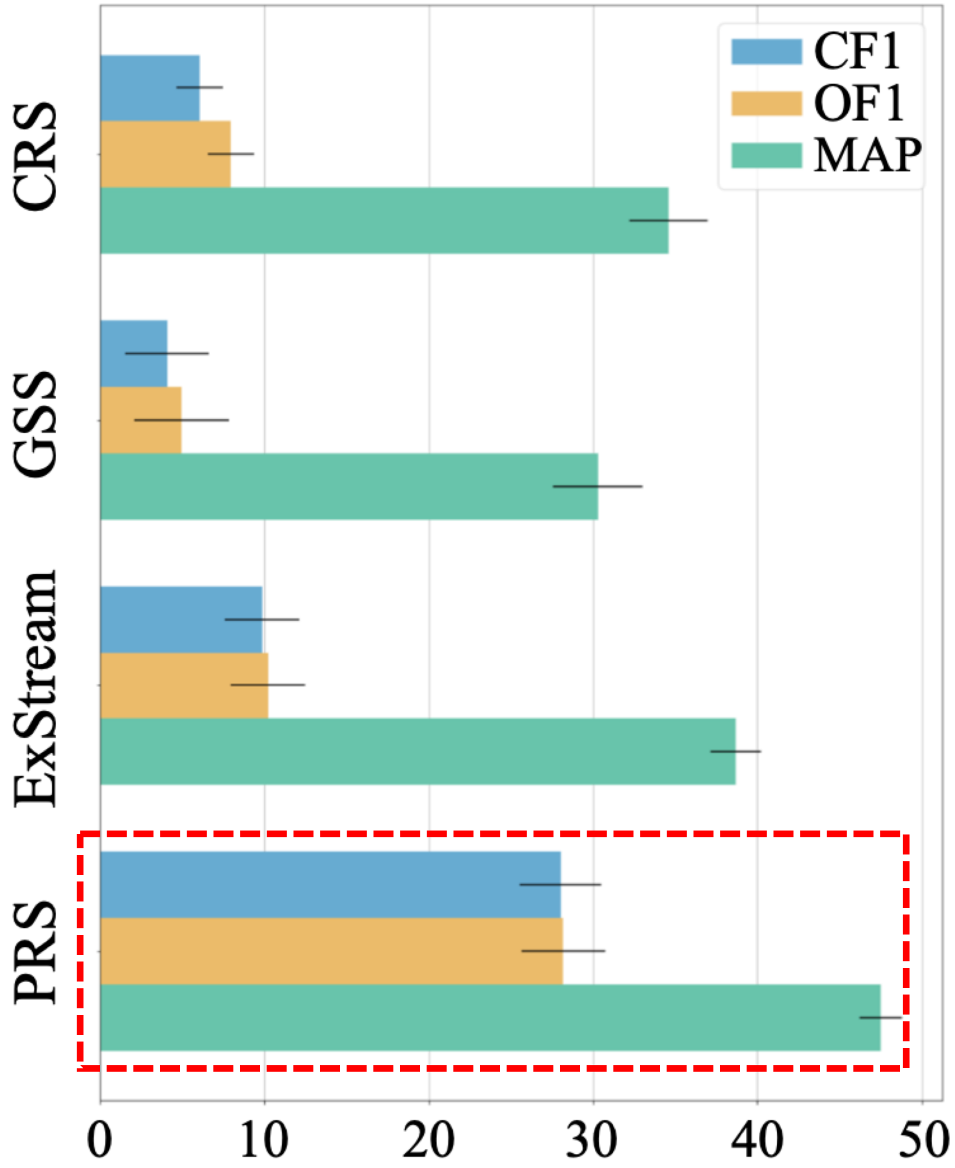
Kirkpatrick et al., *Overcoming catastrophic forgetting in neural networks*, NIPS, 2017
Chaudhry et al., *On Tiny Episodic Memories in Continual Learning*, arxiv, 2019
Aljundi et al., *Gradient based sample selection for online continual learning*, NIPS, 2019
Hayes et al., *Memory efficient experience replay for streaming learning*, ICRA, 2019
Caruaca et al., *Multitask learning*, Machine Learning, 1997

Post-training Memory Distribution



Inter-task Imbalance

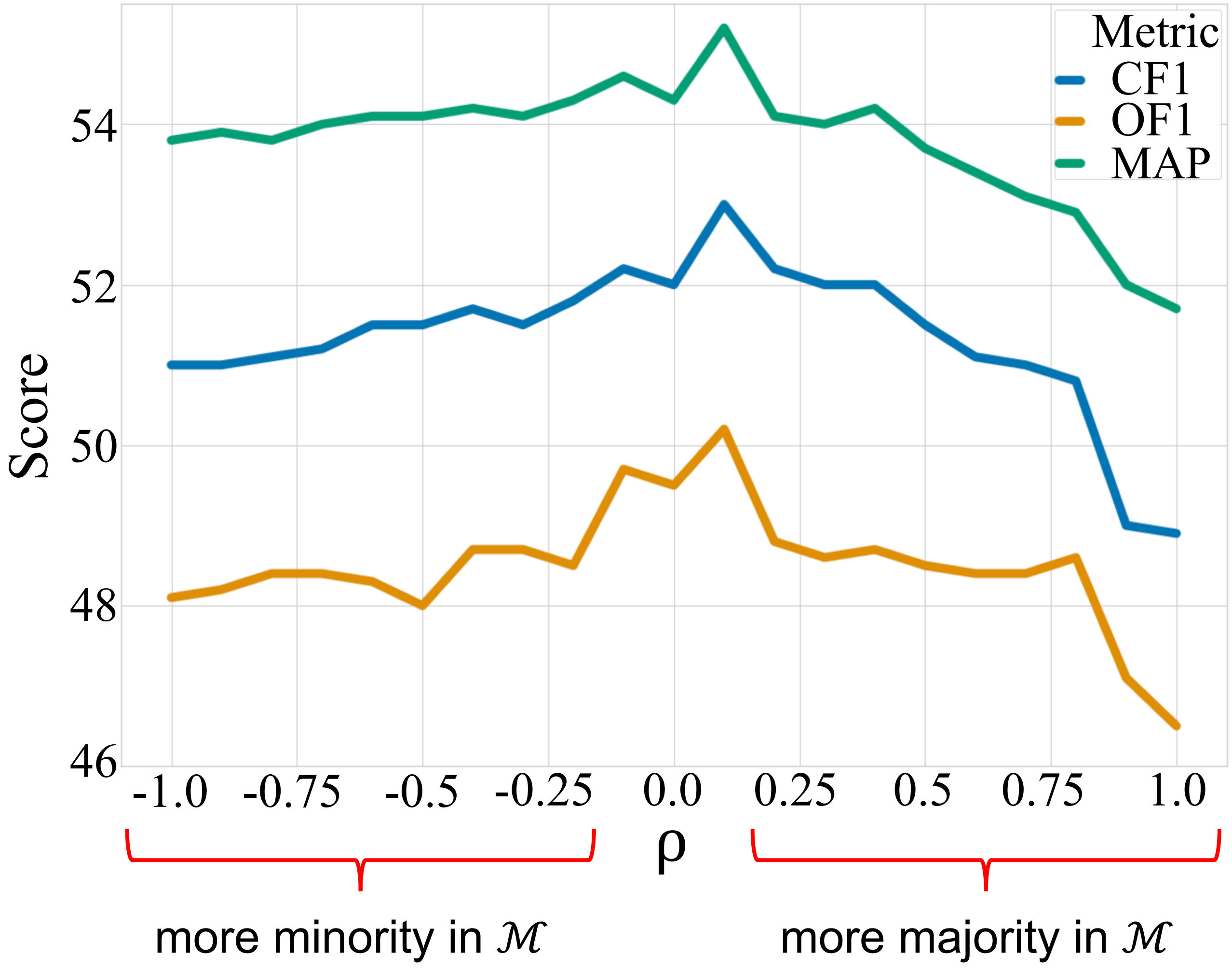
- Task 3 of NUS-WIDEseq is imbalanced relative to the other tasks.
- PRS remains very robust in inter-task imbalance environments.



Power of Allocation, ρ

- ρ = [-1, +1] are explored.
- Near balance (ρ =0.0) is best!
- Balance is indeed important!

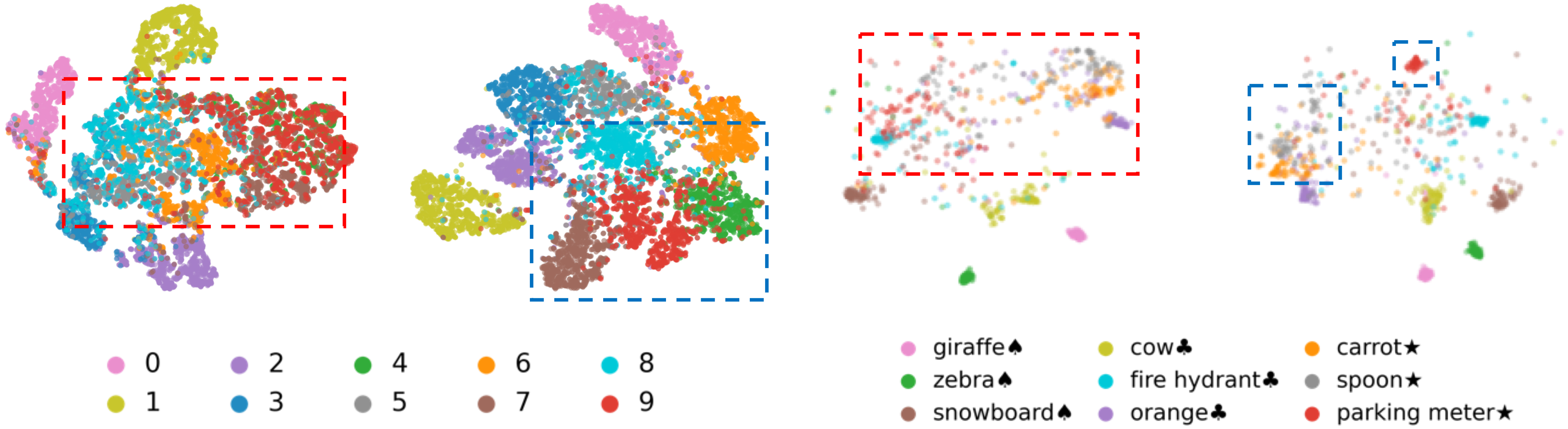
$$p_i = \frac{n_i^\rho}{\sum_j n_j^\rho}$$



Feature Analysis

- t-SNE feature projection of CRS & PRS.
- Results on MNIST & COCO*seq*.
- PRS much more discriminative on both!

(a) CRS-MNIST (b) PRS-MNIST (c) CRS-COCO*seq* (d) PRS-COCO*seq*



Concluding Remarks

- Explore a novel problem of multi-label continual learning (Clearly showing the problem arising due to data imbalances)
- Contribute two datasets (*COCOseq*, *NUS-WIDEseq*)to study the problem
- Devise a novel memory maintenance algorithm *Partitioning Reservoir Sampling*
- Extensive experiments and analysis to show the effectiveness of PRS.

Thank you

References

- Kirkpatrick et al., *Overcoming catastrophic forgetting in neural networks*, NIPS, 2017
- Chaudhry et al., *On Tiny Episodic Memories in Continual Learning*, arxiv, 2019
- Aljundi et al., *Gradient based sample selection for online continual learning*, NIPS, 2019
- Hayes et al., *Memory efficient experience replay for streaming learning*, ICRA, 2019
- Caruaca et al., *Multitask learning*, Machine Learning, 1997