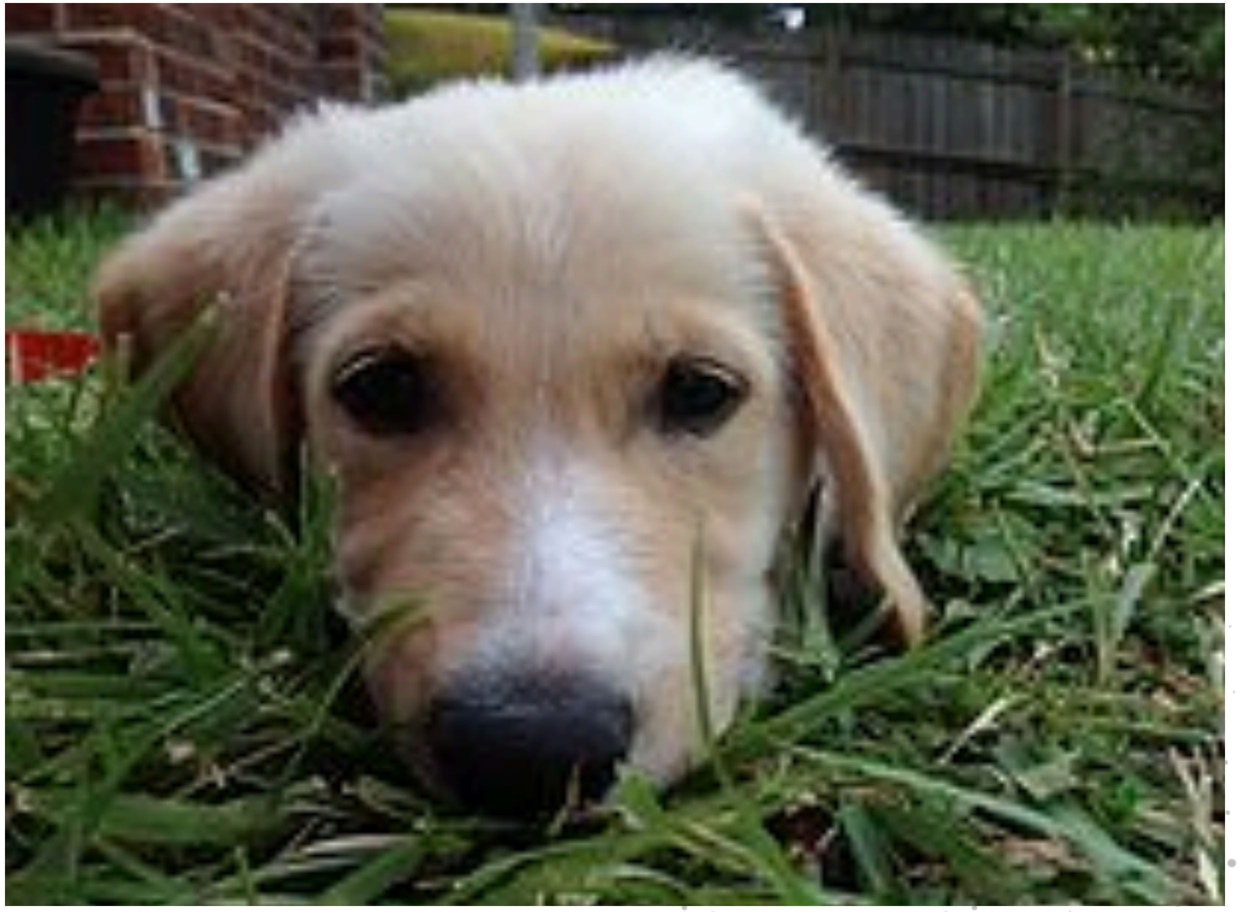
AIDA
ARTIFICIAL INTELLIGENCE
DOCTORAL ACADEMY

Keep on learning without forgetting

Tinne Tuytelaars, KU Leuven,
Tinne.Tuytelaars@esat.kuleuven.be

Data-driven AI

- Avoids the need for external domain knowledge



Data-driven AI

- Avoids the need for external domain knowledge

Data-driven AI

Underlying assumptions:

1. TRAINING DATA IS REPRESENTATIVE
2. ALL TRAINING DATA IS AVAILABLE
3. STATIC WORLD

Data-driven AI

Research

- Benchmark datasets
- Split in train/val/test

- Overfitting to dataset
- Exploiting dataset bias

Real-world applications

- Train/test data collected separately

- Domain shifts
- New tasks / classes
- Concept drift

- Generalization
- Adaptation

Data-driven AI

Real-world applications

- Train/test data collected separately

Research

- Benchmark datasets
- Split in train/val/

- Overfitting to
- Exploiting data

Continual Learning

- Learn from a stream of data
- Without storing old samples
- Without increasing memory footprint
- With ability to add new tasks/labels anytime
- Improving further over time

causality

meta-learning

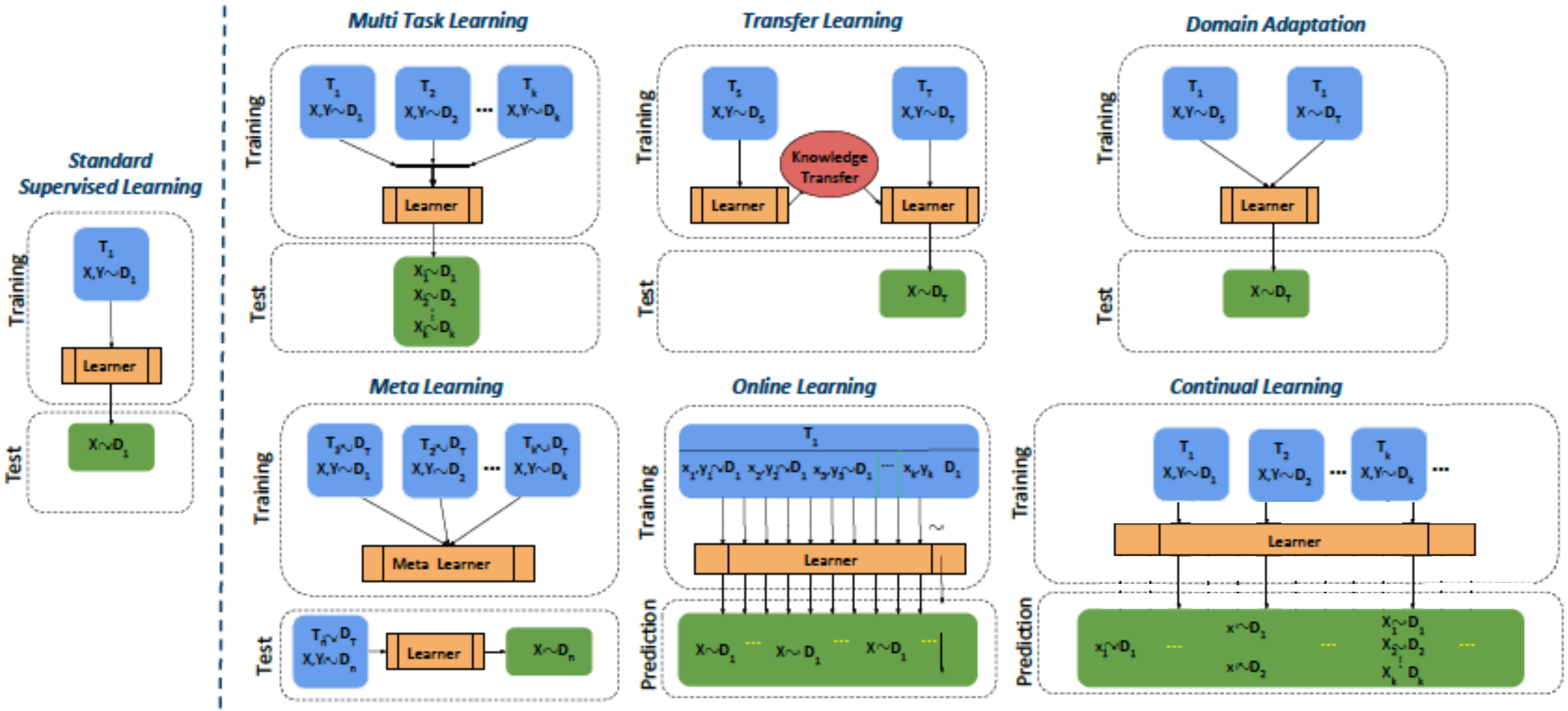
continual learning

domain adaptation

out-of-distribution detection

few-shot learning

zero-shot learning



Standard Supervised Learning
Training
Test
Learner
Multi Task Learning
Transfer Learning
Knowledge Transfer
Domain Adaptation
Meta Learning
Meta Learner
Online Learning
Continual Learning
Prediction

Continual learning

Catastrophic forgetting

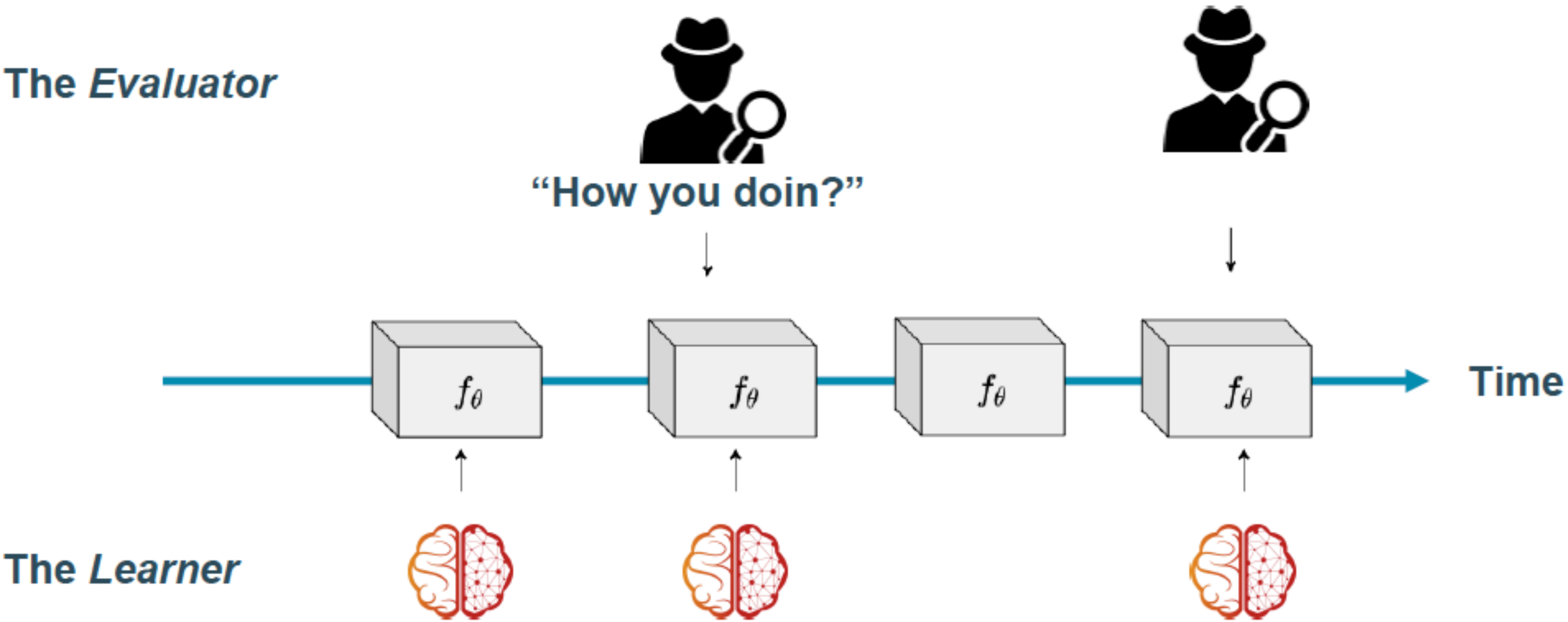
Continual learning: taxonomy

Scenario	*Required at test time*
Task-IL	Solve tasks so far, task-ID provided
Domain-IL	Solve tasks so far, task-ID not provided
Class-IL	Solve tasks so far *and* infer task-ID

[1]

[1] G. Van de Ven and A. S. Tolias, Three scenarios for continual learning, arxiv preprint 1904.07734 (2019)

Continual learning: taxonomy

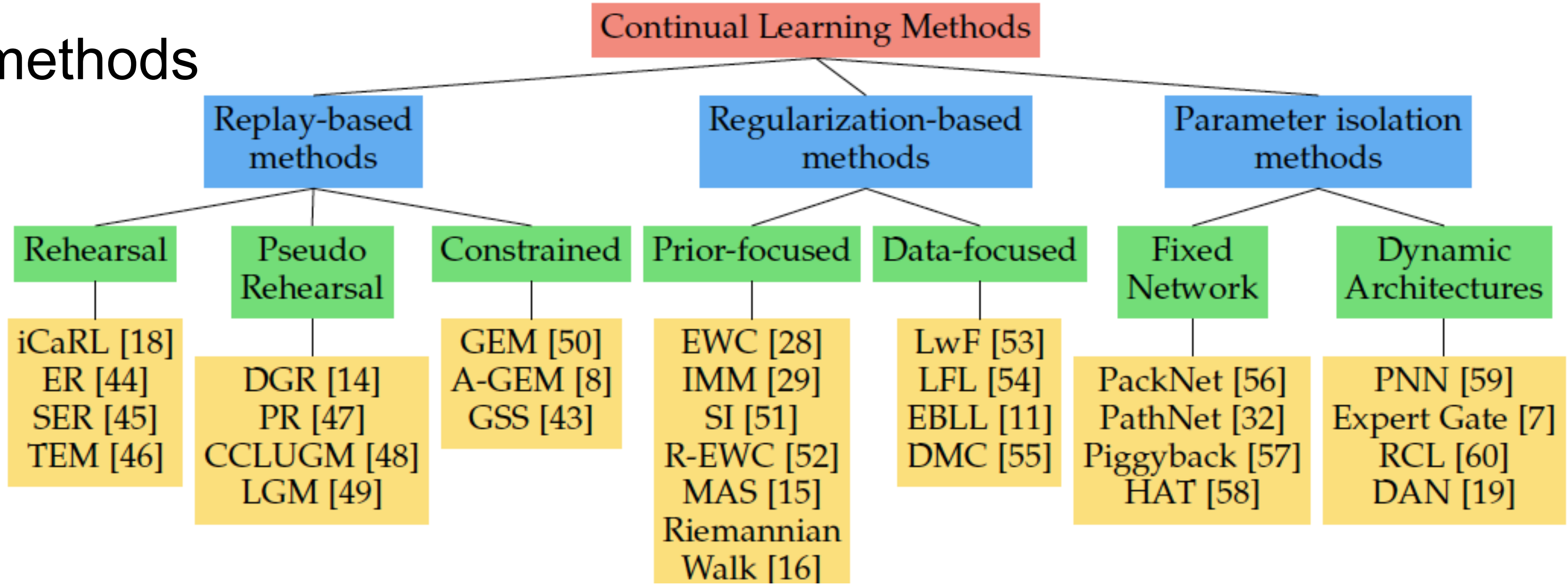


Continual learning: taxonomy

	information presented to	
	learner	**evaluator**
task incremental	$(\mathbf{x}_i, \mathbf{y}_i, t_i)$	$(\mathbf{x}_i, \mathbf{y}_i, t_i)$
class incremental	$(\mathbf{x}_i, \mathbf{y}_i, t_i)$	$(\mathbf{x}_i, \mathbf{y}_i)$
domain incremental	$(\mathbf{x}_i, \mathbf{y}_i, t_i)$	$(\mathbf{x}_i, \mathbf{y}_i)$

Continual learning: taxonomy

- Regularization-based methods
- Parameter isolation methods
- Replay-based methods



M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, T. Tuytelaars, A continual learning survey: Defying forgetting in classification tasks, arxiv 1909.08383 (2020).

COPE:
Continual Prototype Evolution

Matthias De Lange, Tinne Tuytelaars, Continual Prototype Evolution:
Learning Online from Non-Stationary Data Streams, https://arxiv.org/pdf/2009.00919

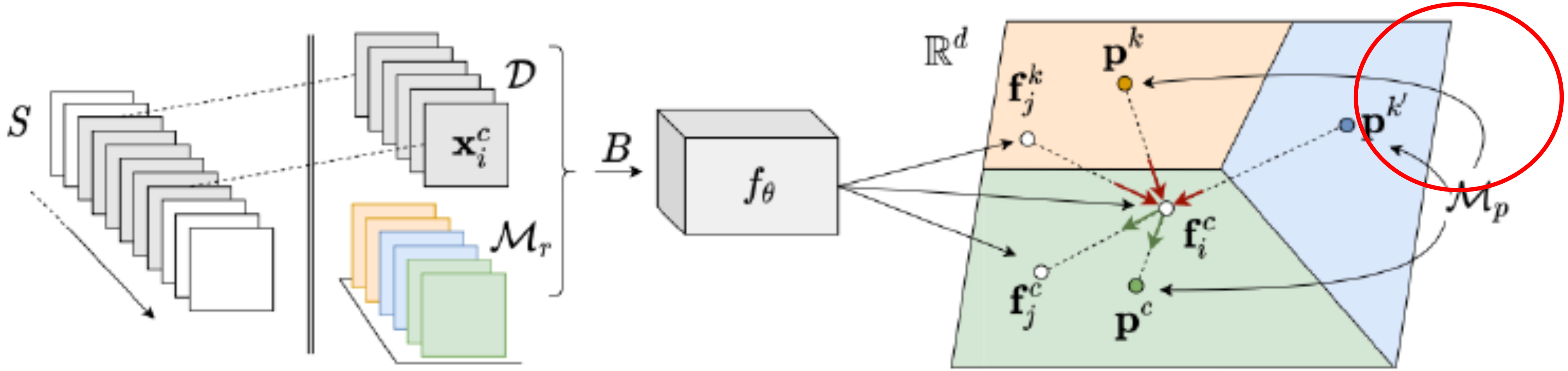
Learning online from non-stationary data streams

- Online data incremental learning
- Focus on imbalanced data
- Using experience replay

- Related work:
    - online incremental learning:
        - Replay: Reservoir, GSS, MIR
        - Parameter isolation: CURL, CN-DPM
    - Class-incremental learning:
        - iCARL, GEM

Learning online from non-stationary data streams

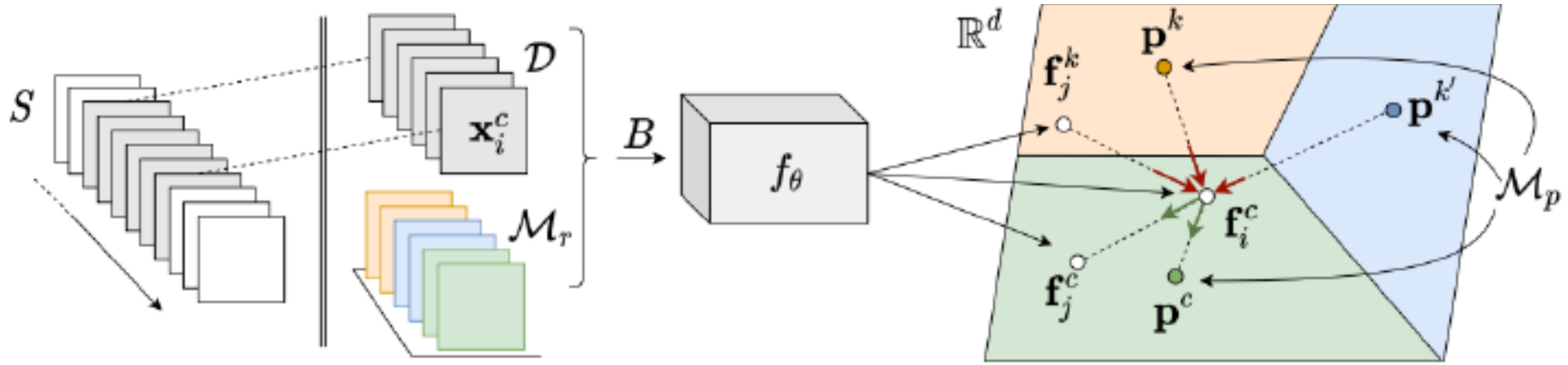
- **Use of prototypes**
- Evolving prototypes
- Balanced replay
- Pseudo-prototypical Proxy Loss



Learning online from non-stationary data streams

- Use of prototypes
- **Evolving prototypes**
- Balanced replay
- Pseudo-prototypical Proxy Loss

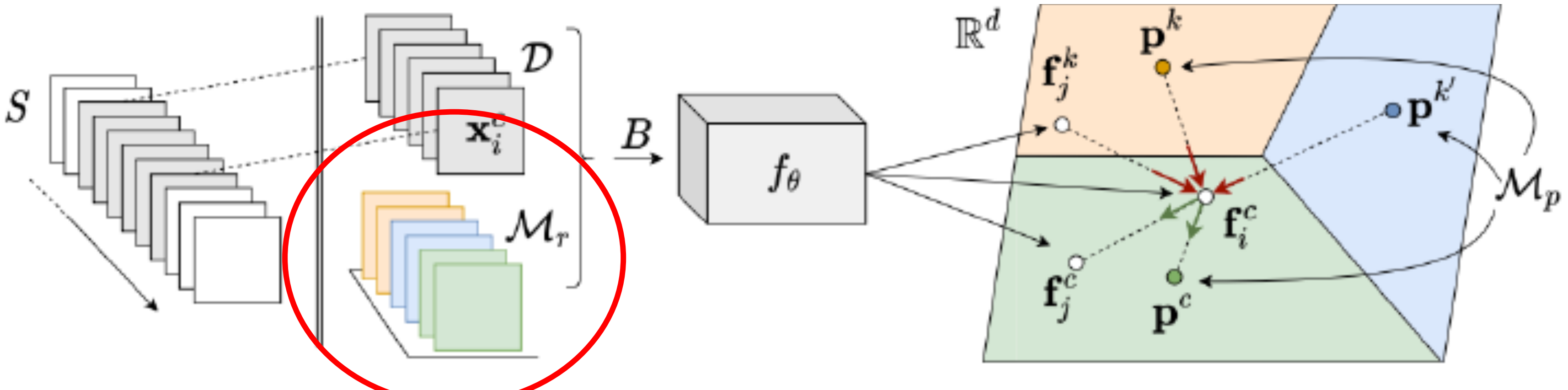
$$\mathbf{p}^c \leftarrow \alpha \mathbf{p}^c + (1-\alpha)\bar{\mathbf{p}}^c, \text{ s.t. } \quad \bar{\mathbf{p}}^c = \frac{1}{|B^c|} \sum_{\mathbf{x}^c \in B^c} f_\theta(\mathbf{x}^c)$$



Learning online from non-stationary data streams

- Use of prototypes
- Evolving prototypes
- **Balanced replay**
- Pseudo-prototypical Proxy Loss

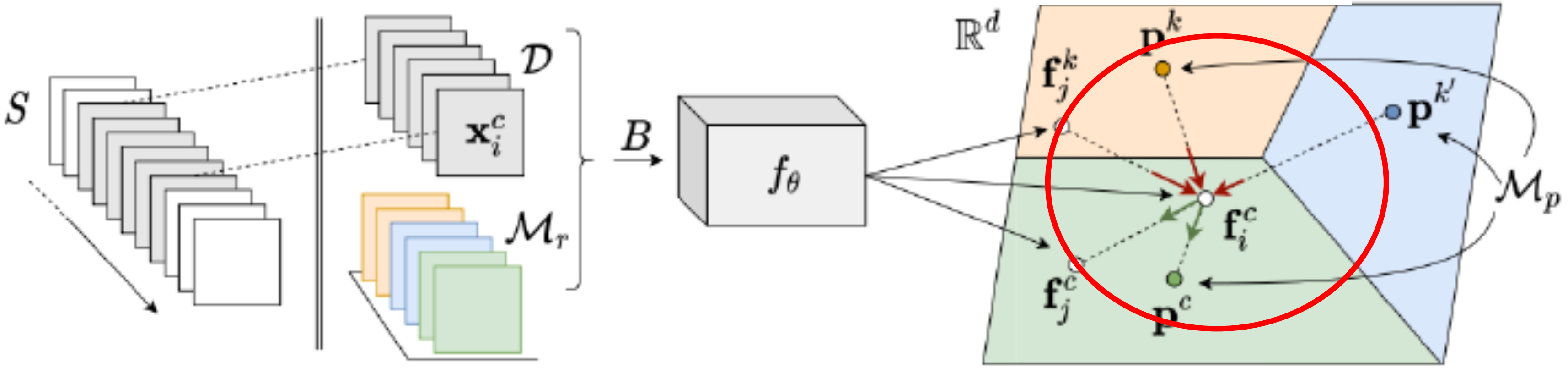
$$\mathbf{p}^c \leftarrow \alpha \mathbf{p}^c + (1-\alpha)\bar{\mathbf{p}}^c, \text{ s.t. } \quad \bar{\mathbf{p}}^c = \frac{1}{|B^c|} \sum_{\mathbf{x}^c \in B^c} f_\theta(\mathbf{x}^c)$$



Learning online from non-stationary data streams

- Use of prototypes
- Evolving prototypes
- Balanced replay
- **Pseudo-prototypical Proxy Loss**

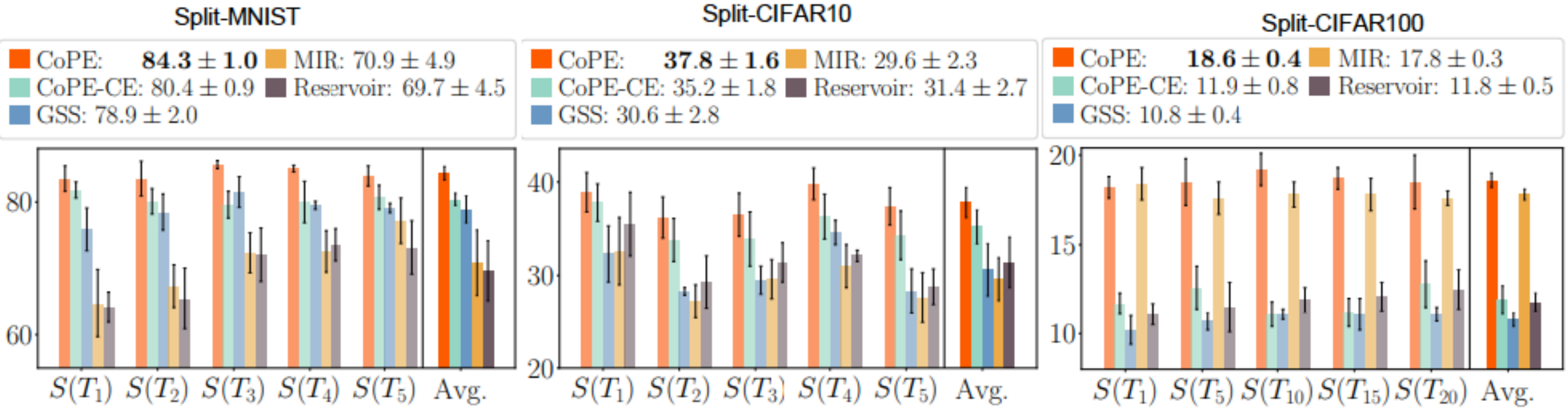
$$\mathcal{L} = -\frac{1}{|B|}\left[\sum_i \log P(c|\mathbf{x}_i^c) + \sum_i \sum_{\mathbf{x}_j^k} \log(1 - P_i(c|\mathbf{x}_j^k))\right]$$



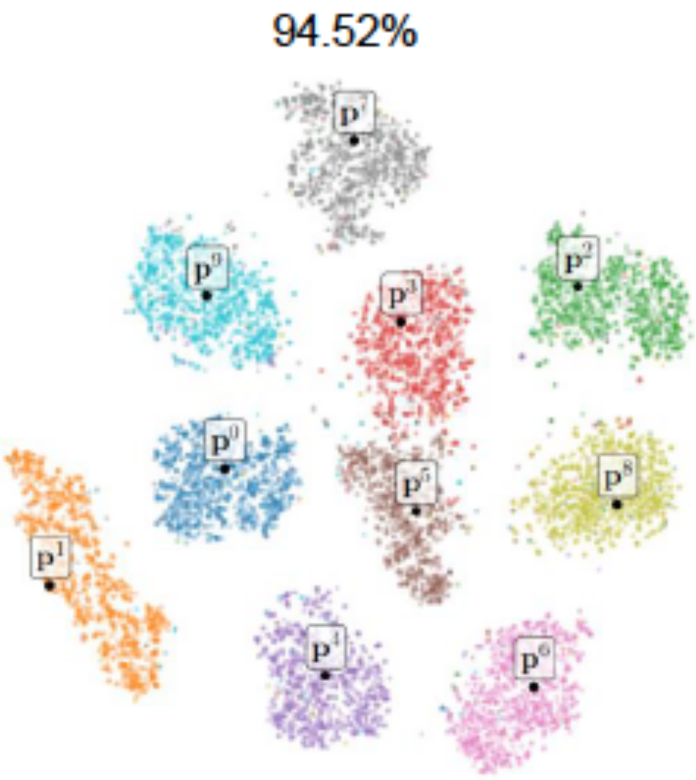
Balanced datastreams

	Split-MNIST	Split-CIFAR10	Split-CIFAR100
iid-offline	98.44 ± 0.02	83.02 ± 0.60	50.28 ± 0.66
iid-online	96.57 ± 0.14	62.31 ± 1.67	20.10 ± 0.90
finetune	19.75 ± 0.05	18.55 ± 0.34	3.53 ± 0.04
GEM	93.25 ± 0.36	24.13 ± 2.46	11.12 ± 2.48
iCARL	83.95 ± 0.21	37.32 ± 2.66	10.80 ± 0.37
CURL (Rao et al., 2019)	92.59 ± 0.66	–	–
DN-CPM (Lee et al., 2020)	93.23 ± 0.09	45.21 ± 0.18	20.10 ± 0.12
reservoir	92.16 ± 0.75	42.48 ± 3.04	19.57 ± 1.79
MIR	93.20 ± 0.36	42.80 ± 2.22	20.00 ± 0.57
GSS	92.47 ± 0.92	38.45 ± 1.41	13.10 ± 0.94
CoPE-CE	91.77 ± 0.87	39.73 ± 2.26	18.33 ± 1.52
CoPE (ours)	$\mathbf{93.94 \pm 0.20}$	$\mathbf{48.92 \pm 1.32}$	$\mathbf{21.62 \pm 0.69}$

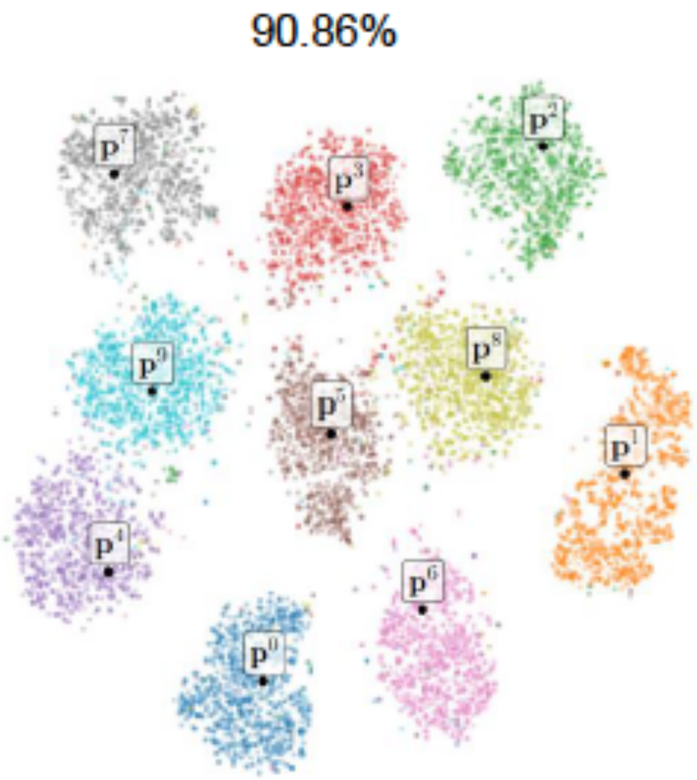
Unbalanced datastreams



Effect of PPP-loss



(a) PPP-loss – *incl.* $\hat{\mathbf{p}}$



(b) PPP-loss – *excl.* $\hat{\mathbf{p}}$

Take home message

- Rethink standard ML paradigm: beyond static train/val/test

- Open challenges:
  - Anything better than experience replay ?
  - Deeper understanding of learning process
  - Real-world use cases / benchmarks

ARTIFICIAL INTELLIGENCE DOCTORAL ACADEMY

THE INTERNATIONAL AI DOCTORAL ACADEMY (AIDA) IS A JOINT INITIATIVE OF THE EUROPEAN R&D PROJECTS AI4MEDIA AND VISION

Artificial Intelligence for the Media and Society

www.ai4media.eu

www.vision4eu.eu

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement No 951911 – AI4Media and Grant Agreement No. 952070