# INSIGHTS FROM THE FUTURE FOR CONTINUAL LEARNING

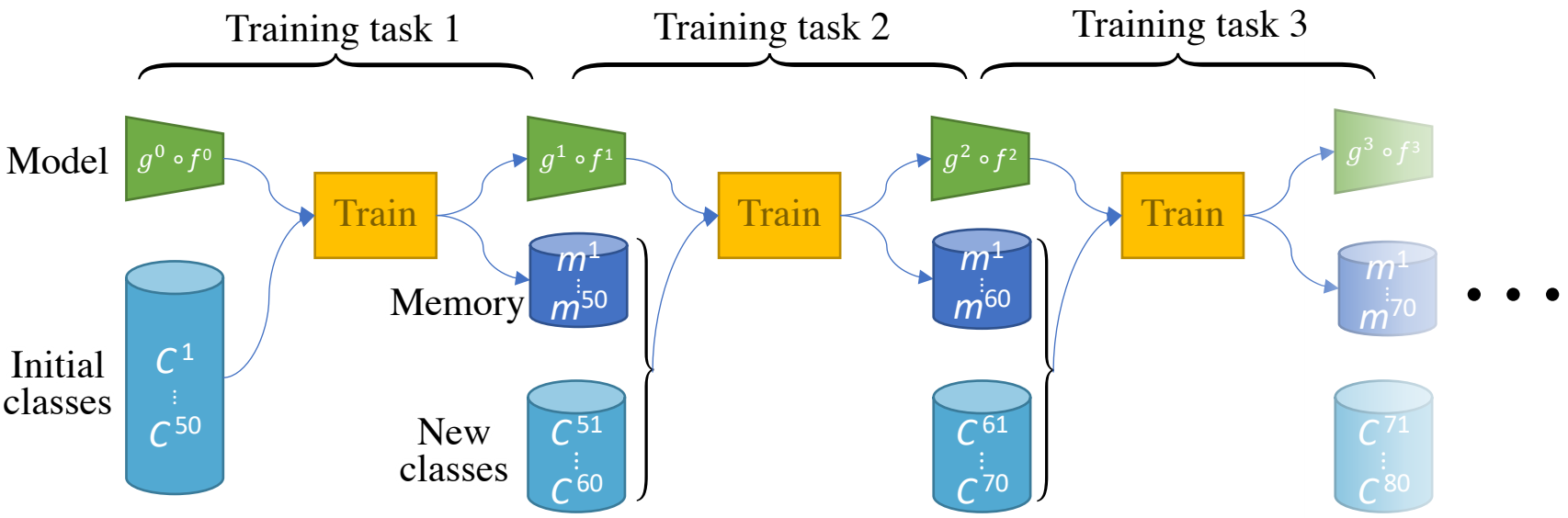
CVPR CLVISION WORKSHOP 2021

Arthur Douillard, *Sorbonne Université & Heuritech*
Eduardo Valle, *University of Campinas*
Charles Ollion, *CMAP École Polytechnique & Heuritech*
Thomas Robert, *Heuritech*
Matthieu Cord, *Sorbonne Université & Valeo.ai*

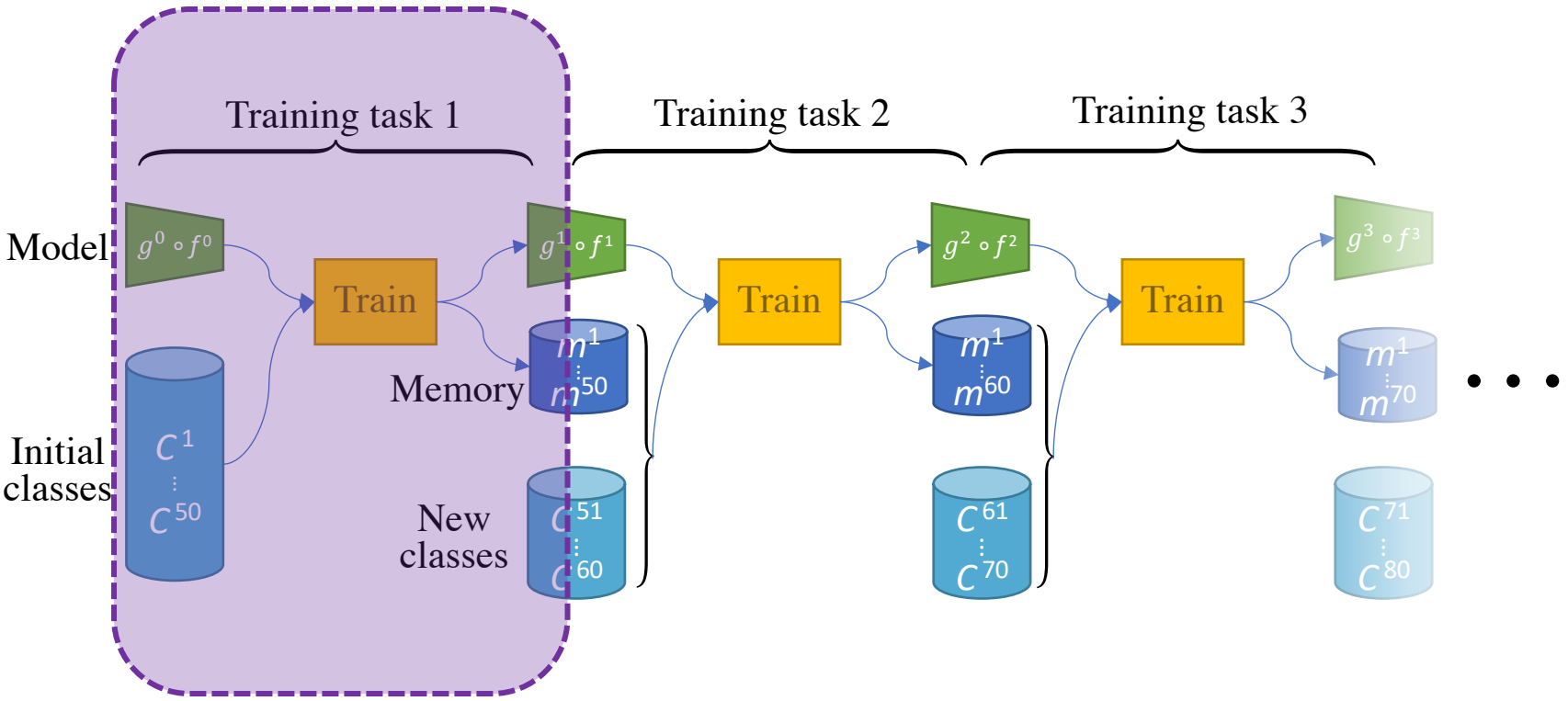
Machine Learning & Deep Learning for Information Access

# Prescient Continual Learning

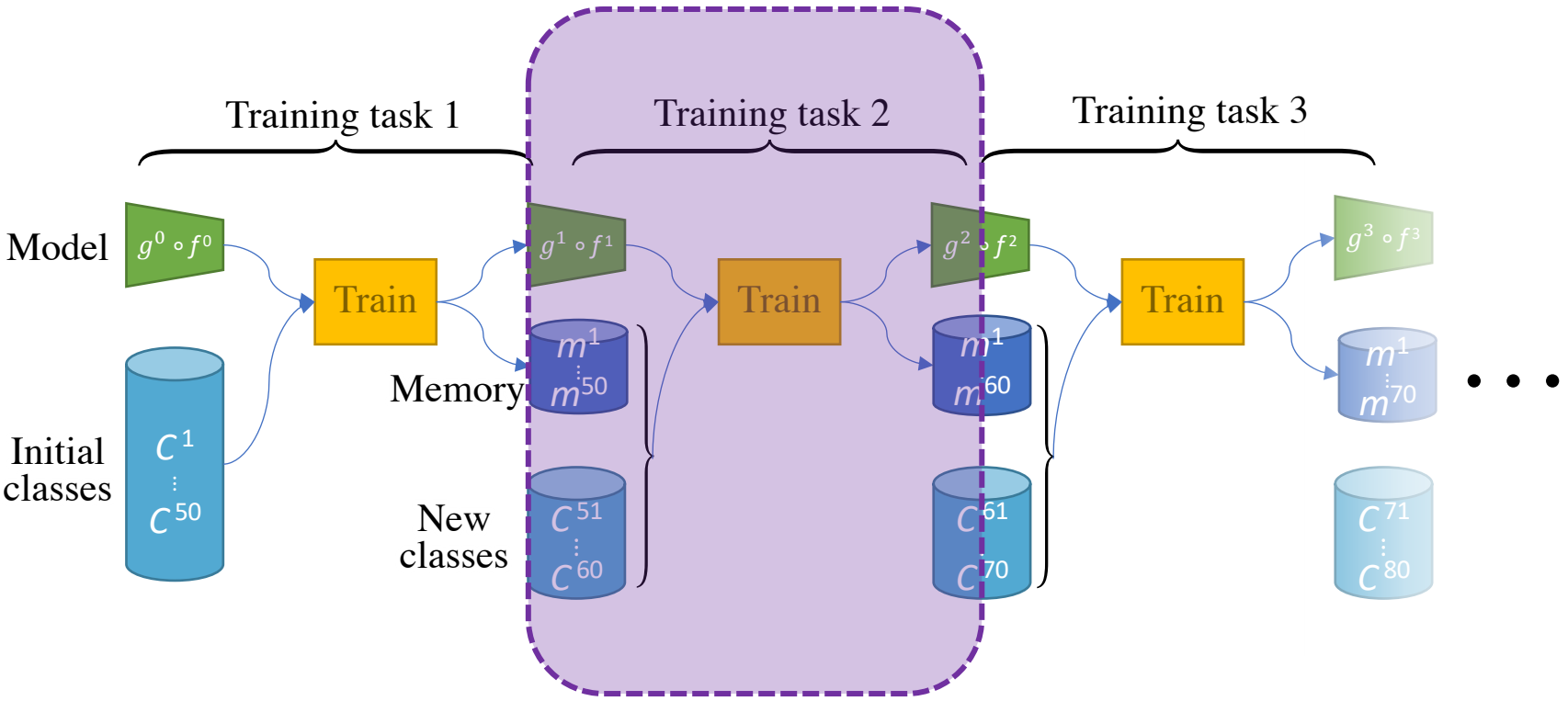
# Selflessness



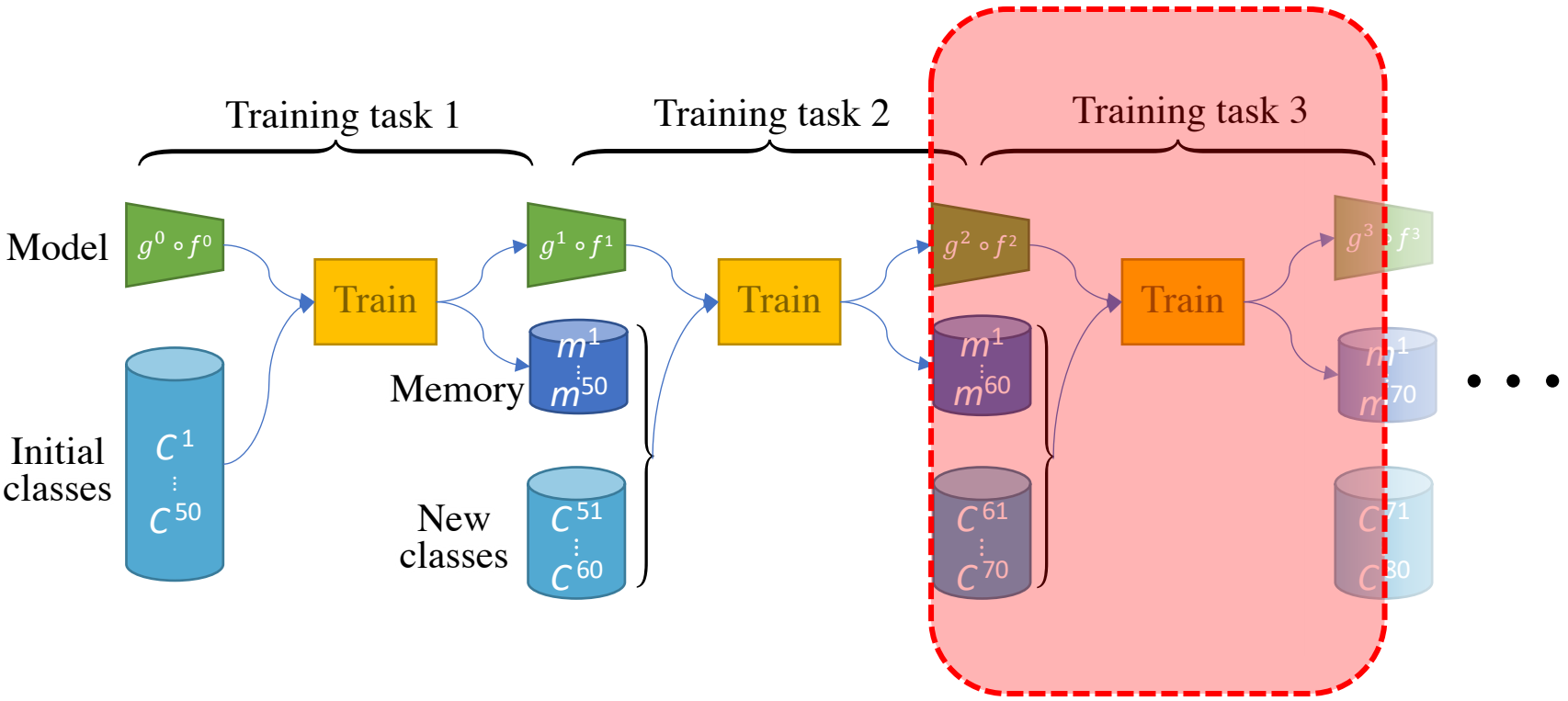
# Selflessness



# Selflessness



# Selflessness




Hard to adapt to new classes without forgetting

# Selfless Network

heuritech

SCIENCES SORBONNE UNIVERSITÉ

Selfless Network:

*A network leaving enough capacity to learn new classes.*

Selfless sequential learning, Aljundi and Tuytelaars, ICLR 2019

# Insights from the Future

heuritech

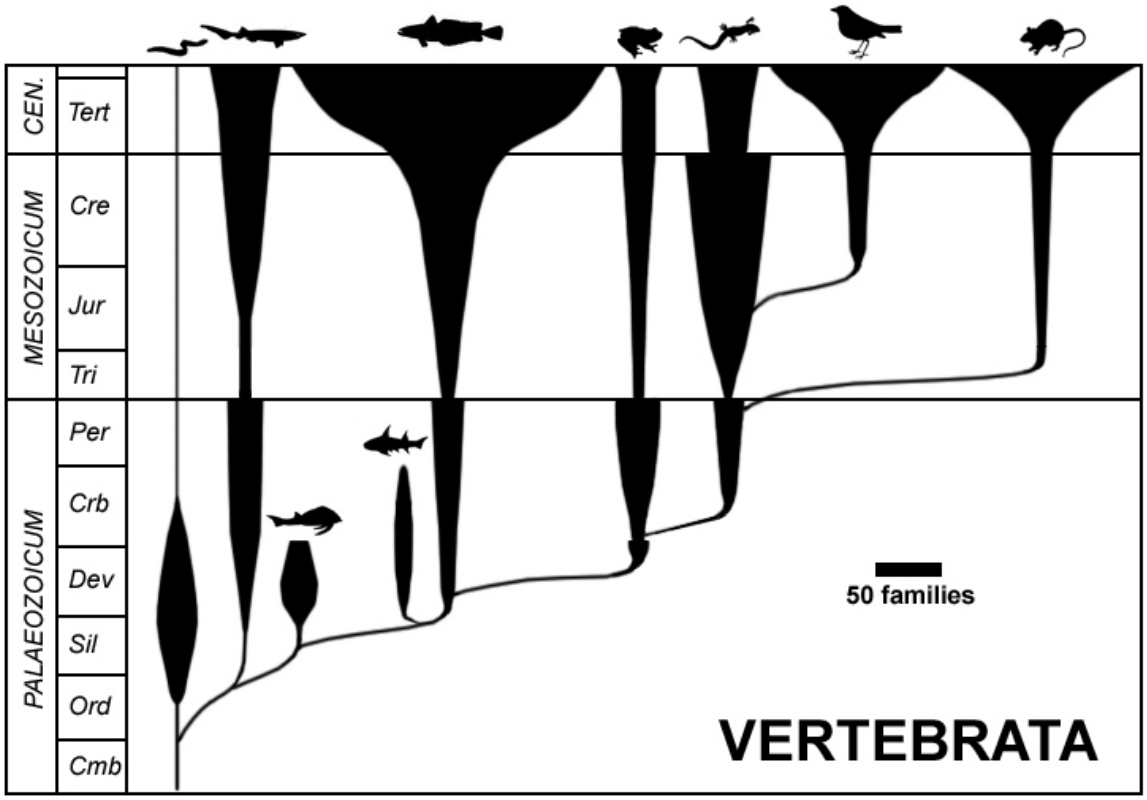
SCIENCES SORBONNE UNIVERSITÉ

What if we could have a prior on the future?

# Future's priors

**Phylogenetic-like tree** of the different classes to be learned

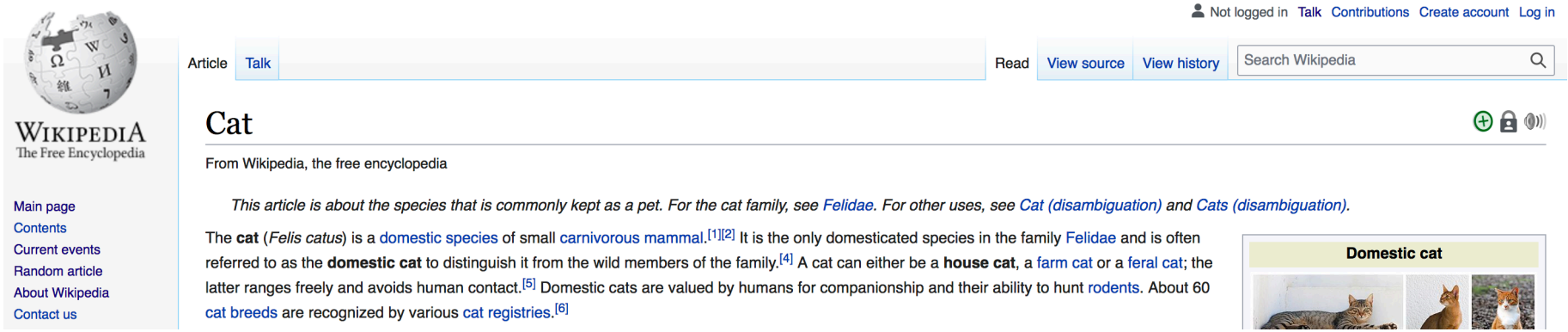
*e.g. ImageNet, iNaturalist…*



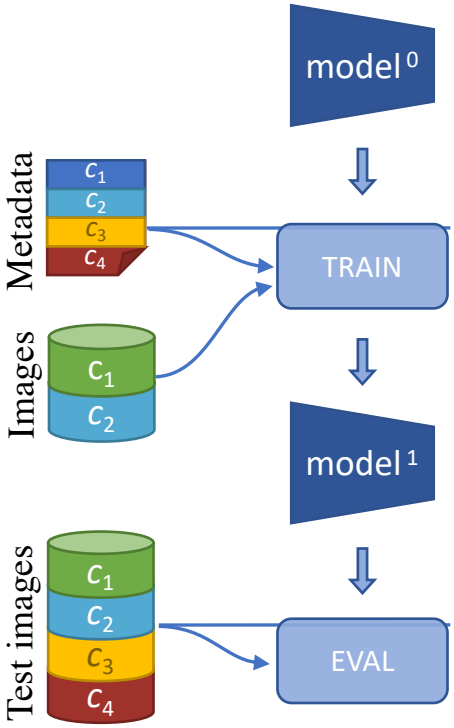
# Future's priors

World knowledge from **word embeddings** trained on encyclopedias

*e.g. word2vec from Wikipedia*

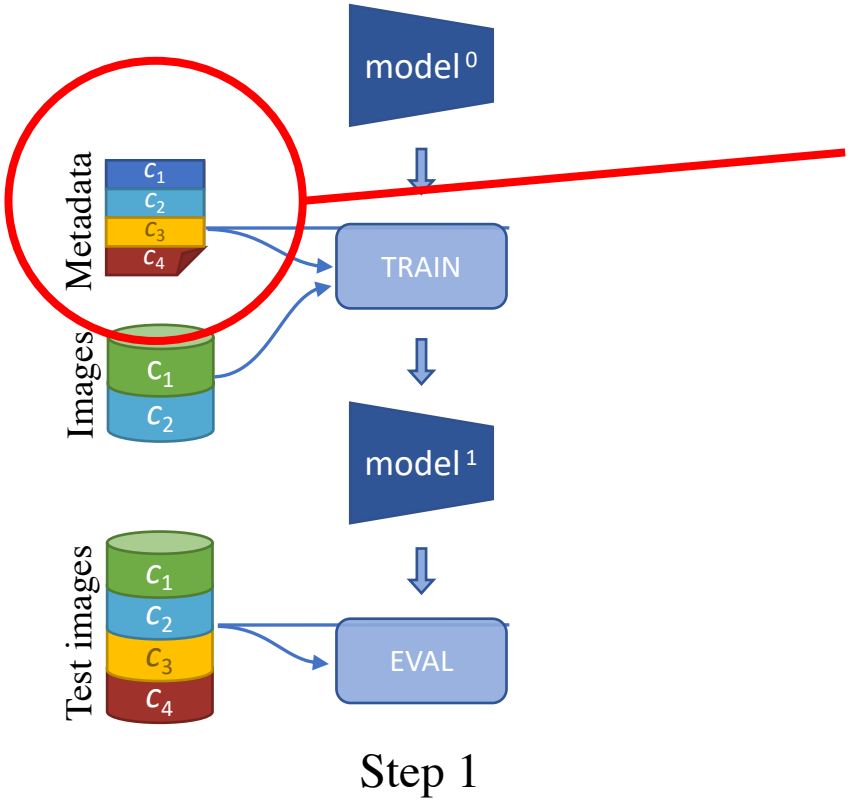


# Prescient Continual Learning



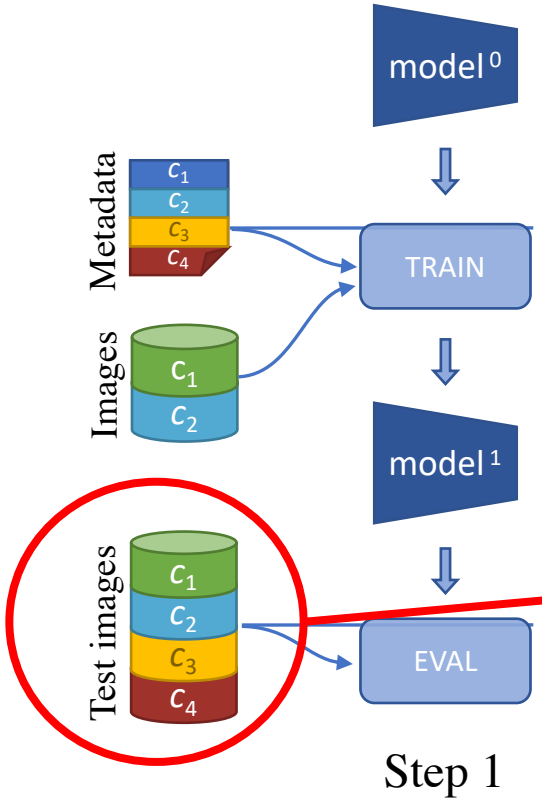

Step 1

# Prescient Continual Learning



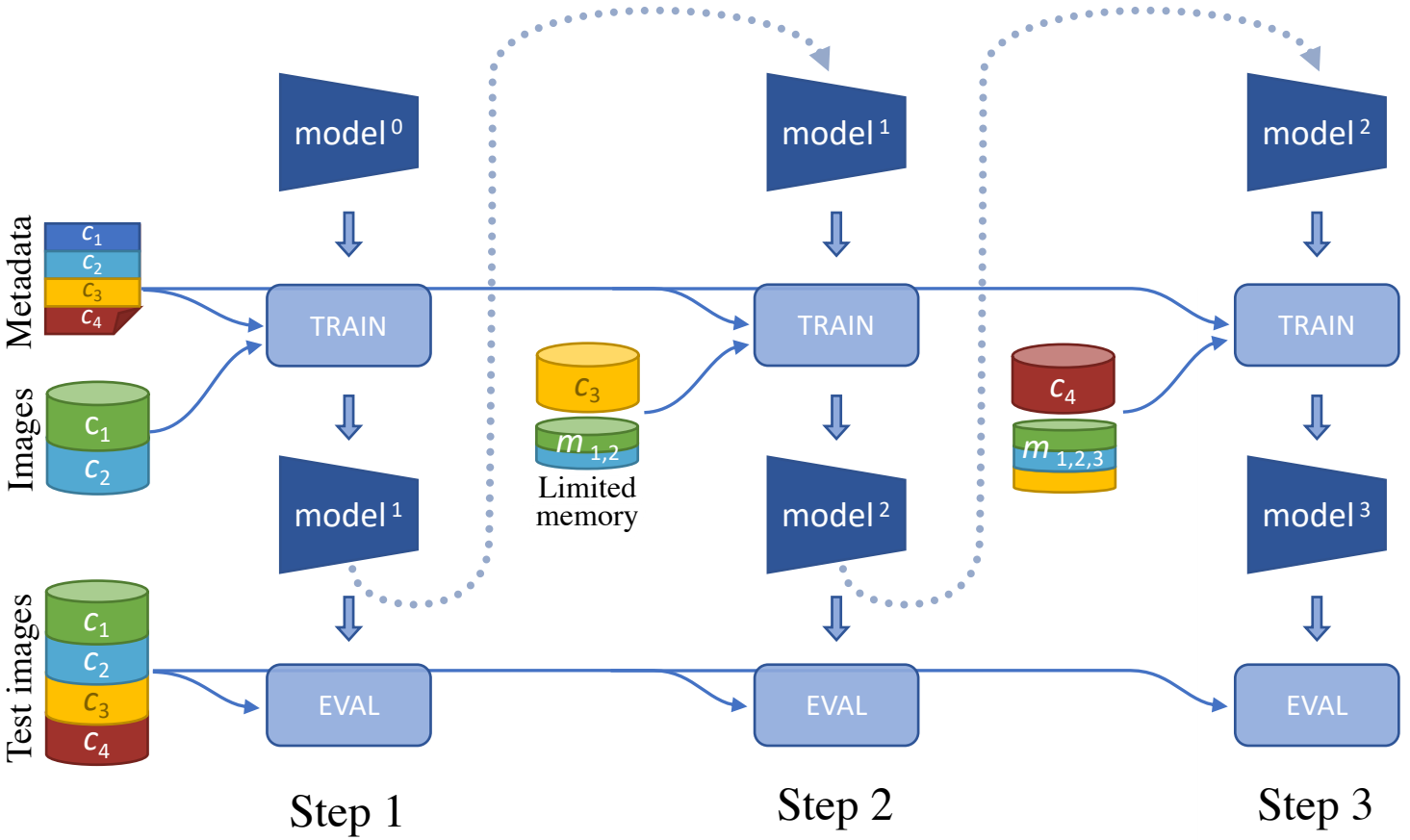

Prior on future classes

# Prescient Continual Learning

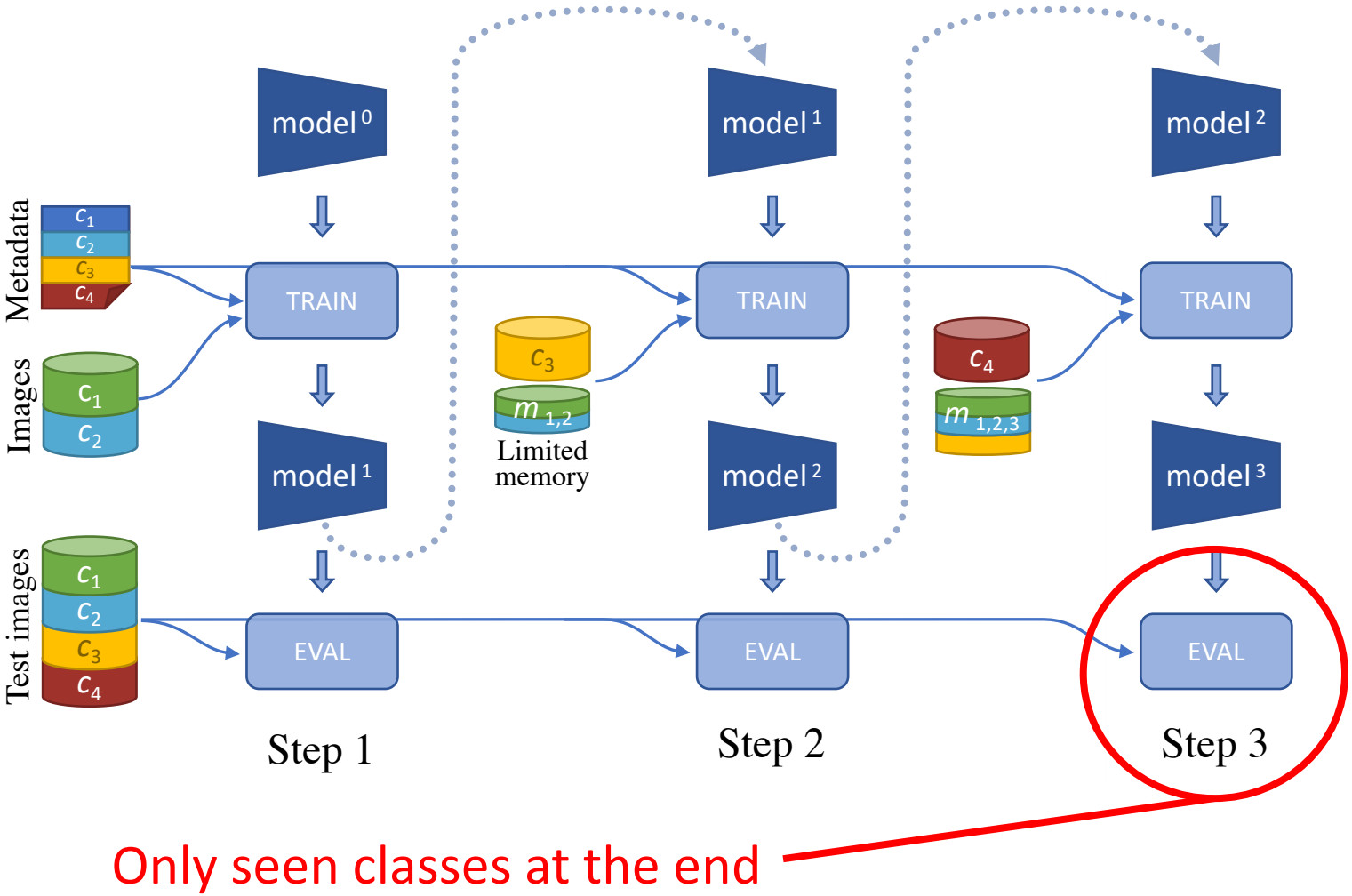



Evaluate on seen and unseen classes

# Prescient Continual Learning



# Prescient Continual Learning



# Ghost Strategy

# General Strategy

heuritech

SCIENCES SORBONNE UNIVERSITÉ

1. Train on task $T_1$
2. For all task $T_{i>1}$:
    1. Train features generator
    2. Generate fake unseen classes
    3. Train with real and fake features
    4. Eval on all classes

# General Strategy

heuritech

SCIENCES SORBONNE UNIVERSITÉ

1. **Train on task $T_1$**
2. For all task $T_{i>1}$:
    1. Train features generator
    2. Generate fake unseen classes
    3. Train with real and fake features
    4. Eval on all classes

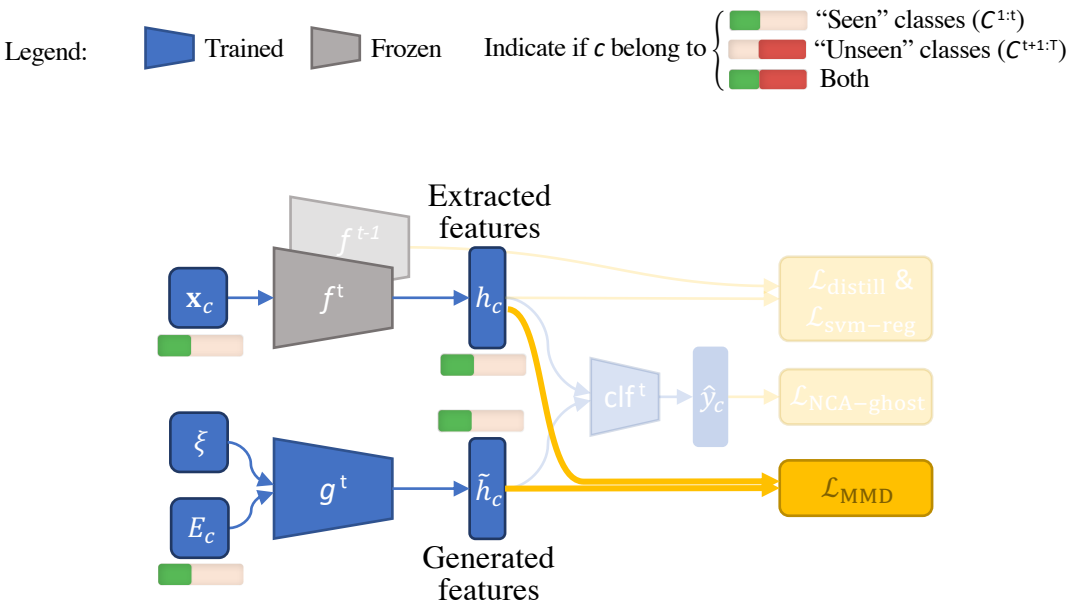
# General Strategy

heuritech

SCIENCES SORBONNE UNIVERSITÉ

1. Train on task $T_1$
2. For all task $T_{i>1}$:
    1. **Train features generator**
    2. Generate fake unseen classes
    3. Train with real and fake features
    4. Eval on all classes

# Feature Generation




Generative Moment Matching Network

generating **fake features** from metadata

to mimick real features of **seen classes**

# General Strategy

heuritech

SCIENCES SORBONNE UNIVERSITÉ

1. Train on task $T_1$
2. For all task $T_{i>1}$:
    1. Train features generator
    2. **Generate fake unseen classes**
    3. Train with real and fake features
    4. Eval on all classes

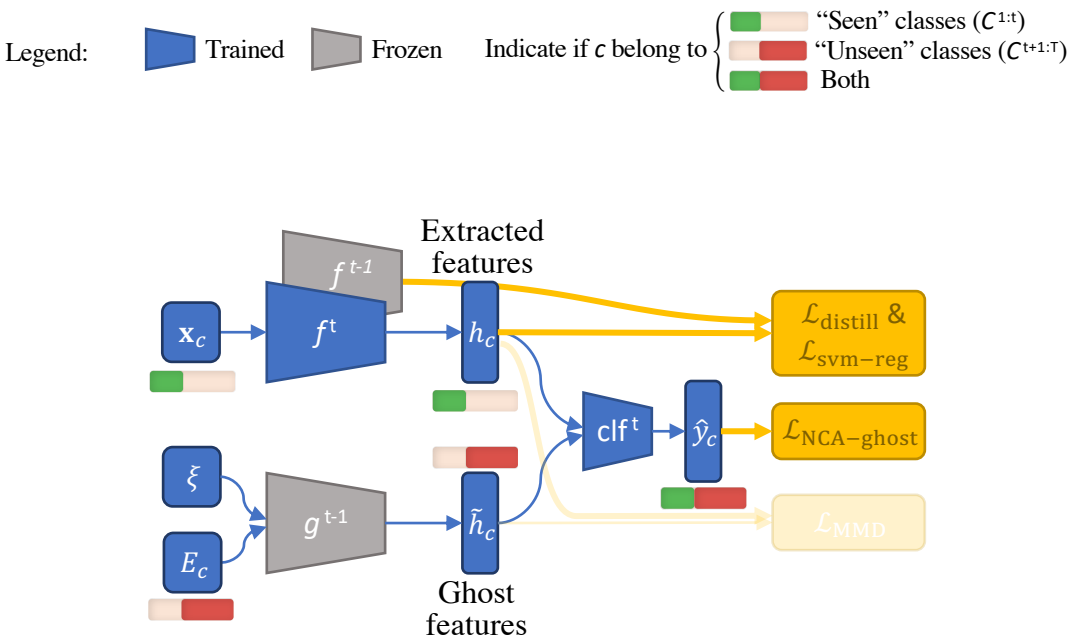
# General Strategy

heuritech

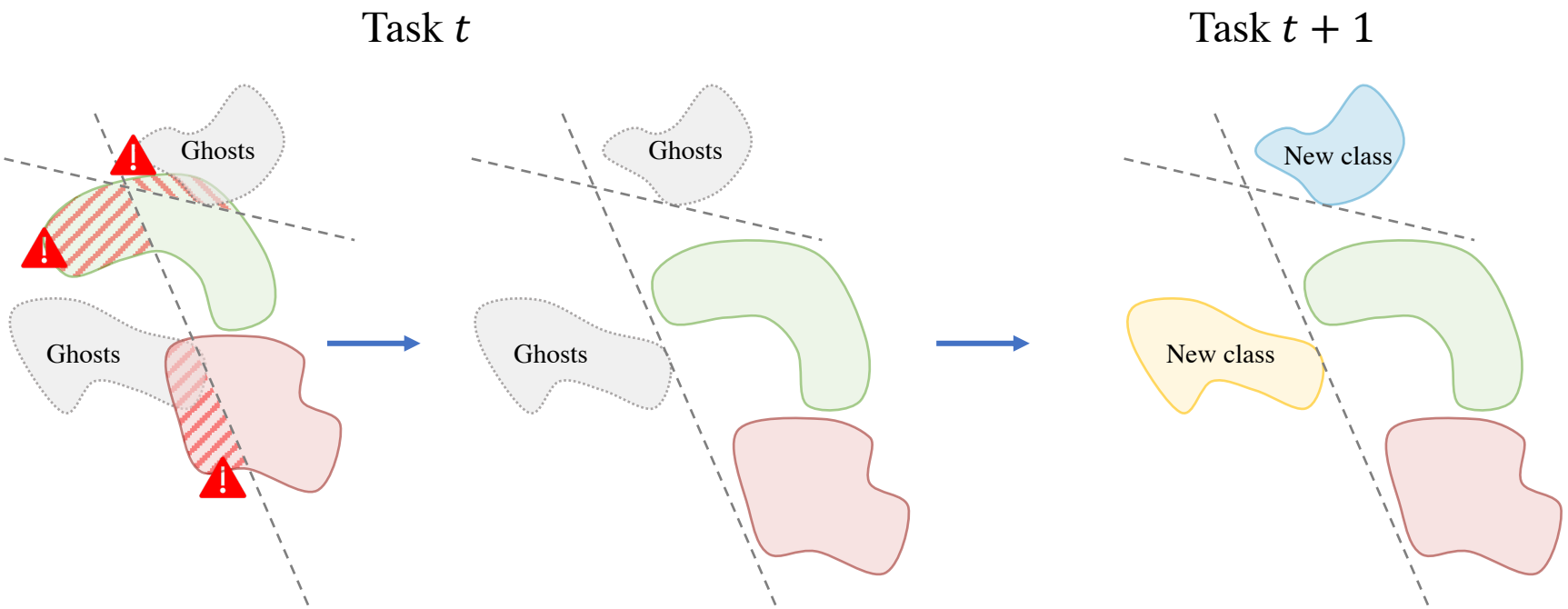
SCIENCES SORBONNE UNIVERSITÉ

1. Train on task $T_1$
2. For all task $T_{i>1}$:
    1. Train features generator
    2. Generate fake unseen classes
    3. **Train with real and fake features**
    4. Eval on all classes

# Train on real and *Ghost* features



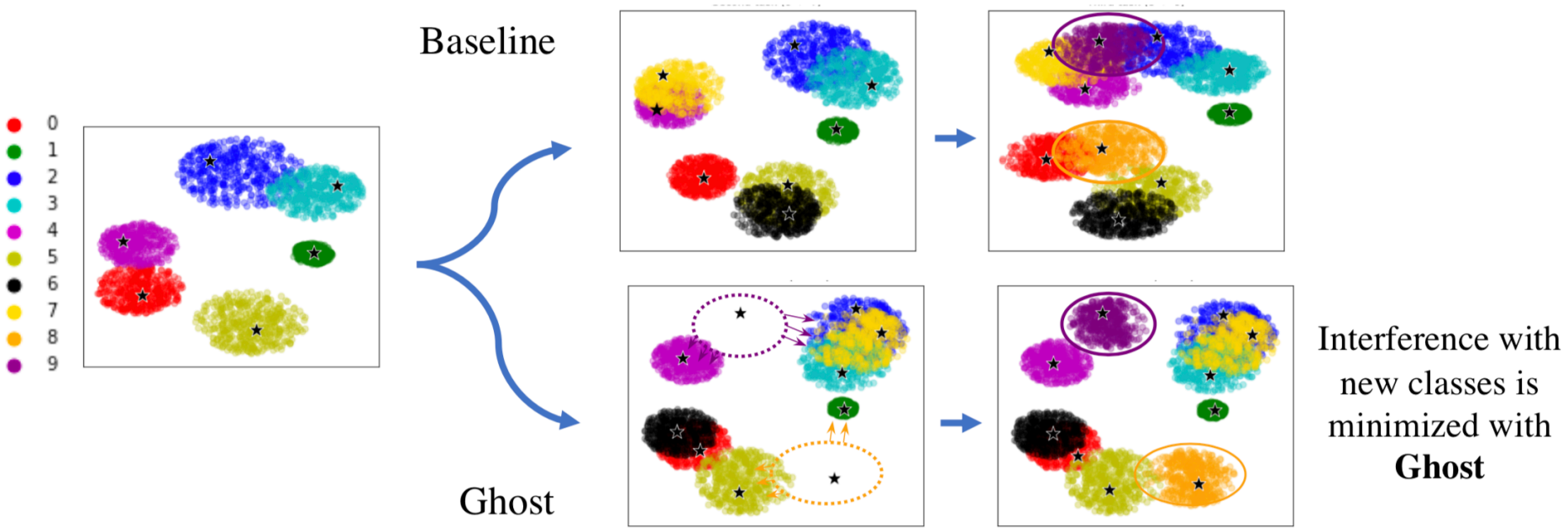
# Exclusion zones



# Results

# Qualitative Results

## Split-MNIST with a 2-dimensions features



# Quantitative Results

heuritech

SCIENCES SORBONNE UNIVERSITÉ

## Average incremental accuracy

| | AwA2<br>25 classes + 5 × 5 classes<br>PODNet | <br><br>UCIR |
|---|---|---|
| Base Continual Learner | $62.92 \pm 0.12$ | $62.66 \pm 0.17$ |
| + $\mathcal{L}^{\text{nca-ghost}}$ | $68.31 \pm 0.36$ | $64.96 \pm 0.12$ |
| + $\mathcal{L}^{\text{nca-ghost}}$ + $\mathcal{L}^{\text{svm-reg}}$ | $\mathbf{68.46 \pm 0.47}$ | $\mathbf{65.37 \pm 0.35}$ |
| | aP&Y<br>16 classes + 8 × 2 classes<br>PODNet | <br><br>UCIR |
| Base Continual Learner | $58.64 \pm 0.66$ | $45.96 \pm 0.22$ |
| + $\mathcal{L}^{\text{nca-ghost}}$ | $62.08 \pm 0.25$ | $59.67 \pm 0.84$ |
| + $\mathcal{L}^{\text{nca-ghost}}$ + $\mathcal{L}^{\text{svm-reg}}$ | $\mathbf{62.73 \pm 0.60}$ | $\mathbf{60.12 \pm 0.74}$ |

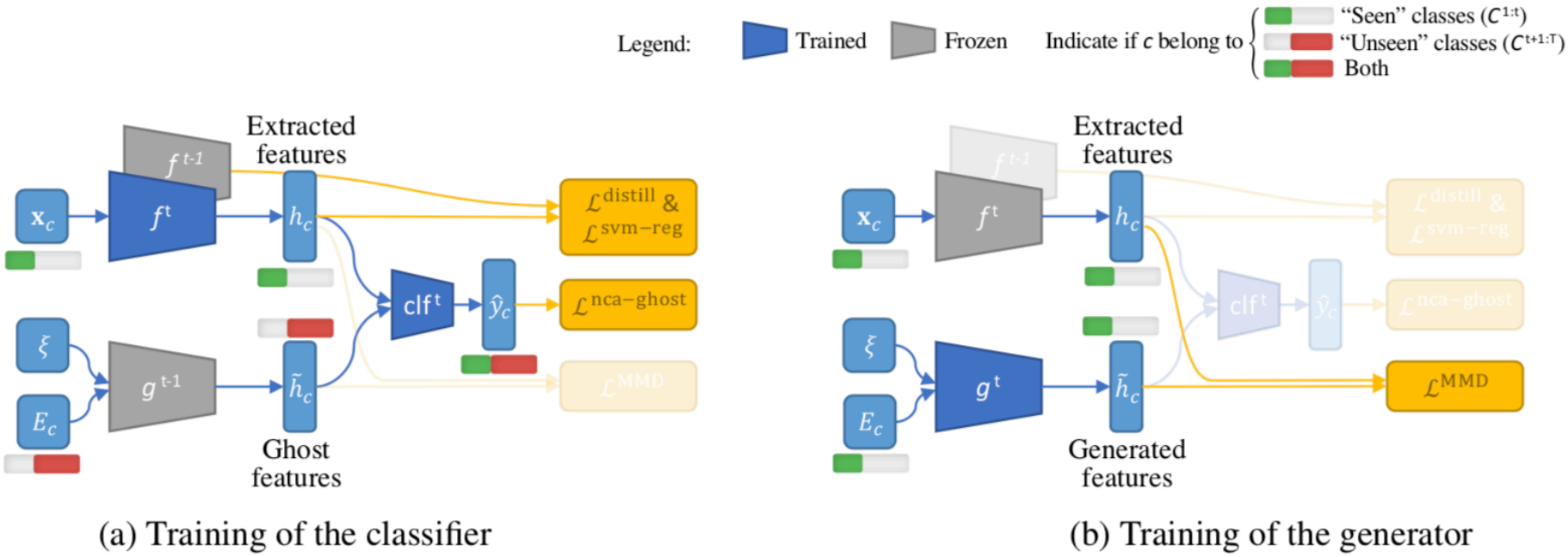
# Quantitative Results

heuritech SCIENCES SORBONNE UNIVERSITÉ

## Final accuracy

| | AwA2<br>25 classes + 5 × 5 classes<br>PODNet | <br><br>UCIR |
|---|---|---|
| Base Continual Learner | 77.63 ± 0.06 | 76.36 ± 0.28 |
| + $\mathcal{L}^{\text{nca-ghost}}$ | 78.70 ± 0.46 | **77.32 ± 0.40** |
| + $\mathcal{L}^{\text{nca-ghost}}$ + $\mathcal{L}^{\text{svm-reg}}$ | **79.08 ± 0.53** | 77.27 ± 0.35 |
| | aP&Y<br>16 classes + 8 × 2 classes<br>PODNet | <br><br>UCIR |
| Base Continual Learner | 57.80 ± 0.97 | 45.40 ± 0.17 |
| + $\mathcal{L}^{\text{nca-ghost}}$ | 62.47 ± 0.40 | 54.13 ± 0.90 |
| + $\mathcal{L}^{\text{nca-ghost}}$ + $\mathcal{L}^{\text{svm-reg}}$ | **63.30 ± 0.98** | **56.07 ± 1.29** |

# Thank you for listening




**Arthur Douillard**, *Sorbonne Université & Heuritech*
**Eduardo Valle**, *University of Campinas*
**Charles Ollion**, *CMAP École Polytechnique & Heuritech*
**Thomas Robert**, *Heuritech*
**Matthieu Cord**, *Sorbonne Université & Valeo.ai*