Apprentissage continu de représentations visuelles

ENSIMAG
2023-2024

Karteek Alahari & Diane Larlus

Apprentissage continu

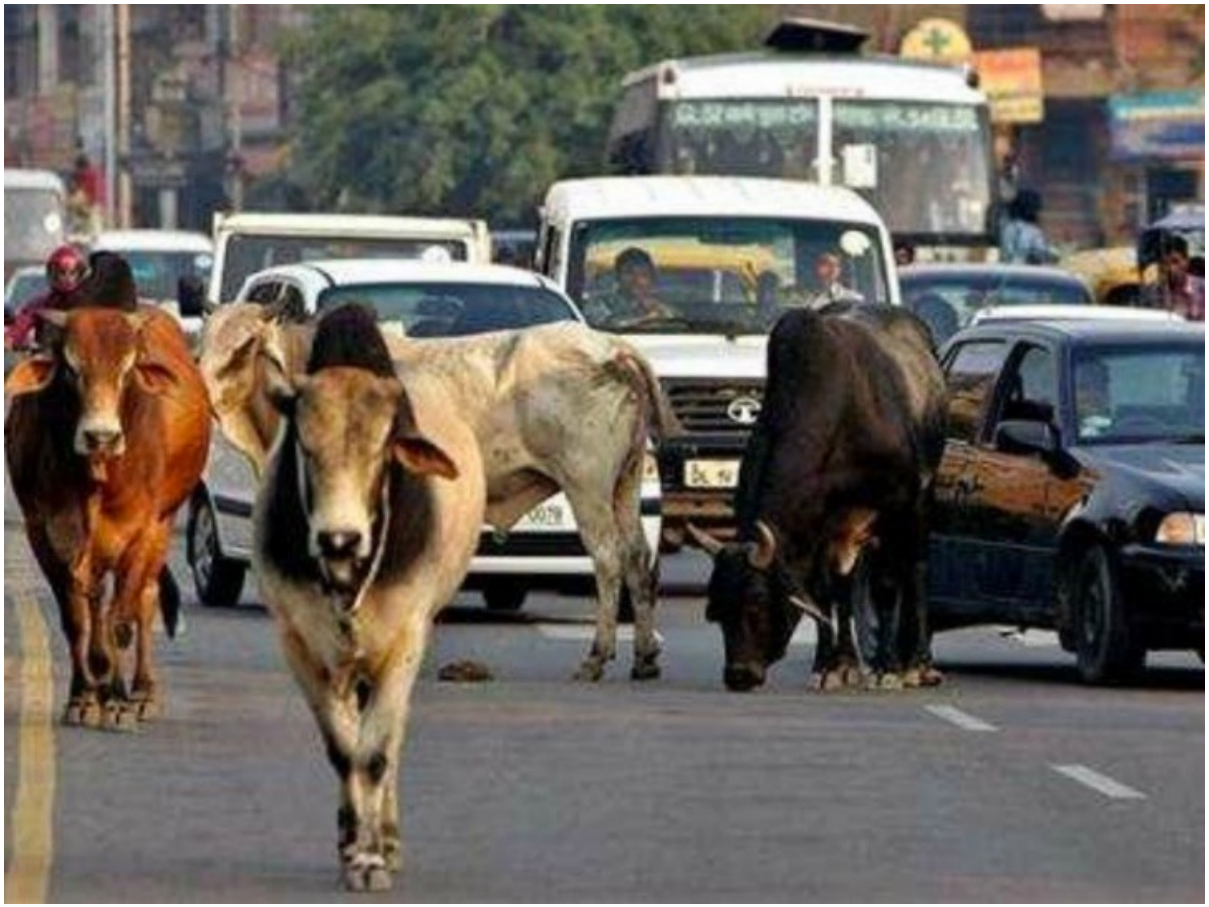
https://project.inria.fr/bigvisdata/

Continual Learning ?

- Incremental learning
- Lifelong learning
- Sequential learning
- Never-ending Learning

An Continual Learning Scenario

- Growing up in India

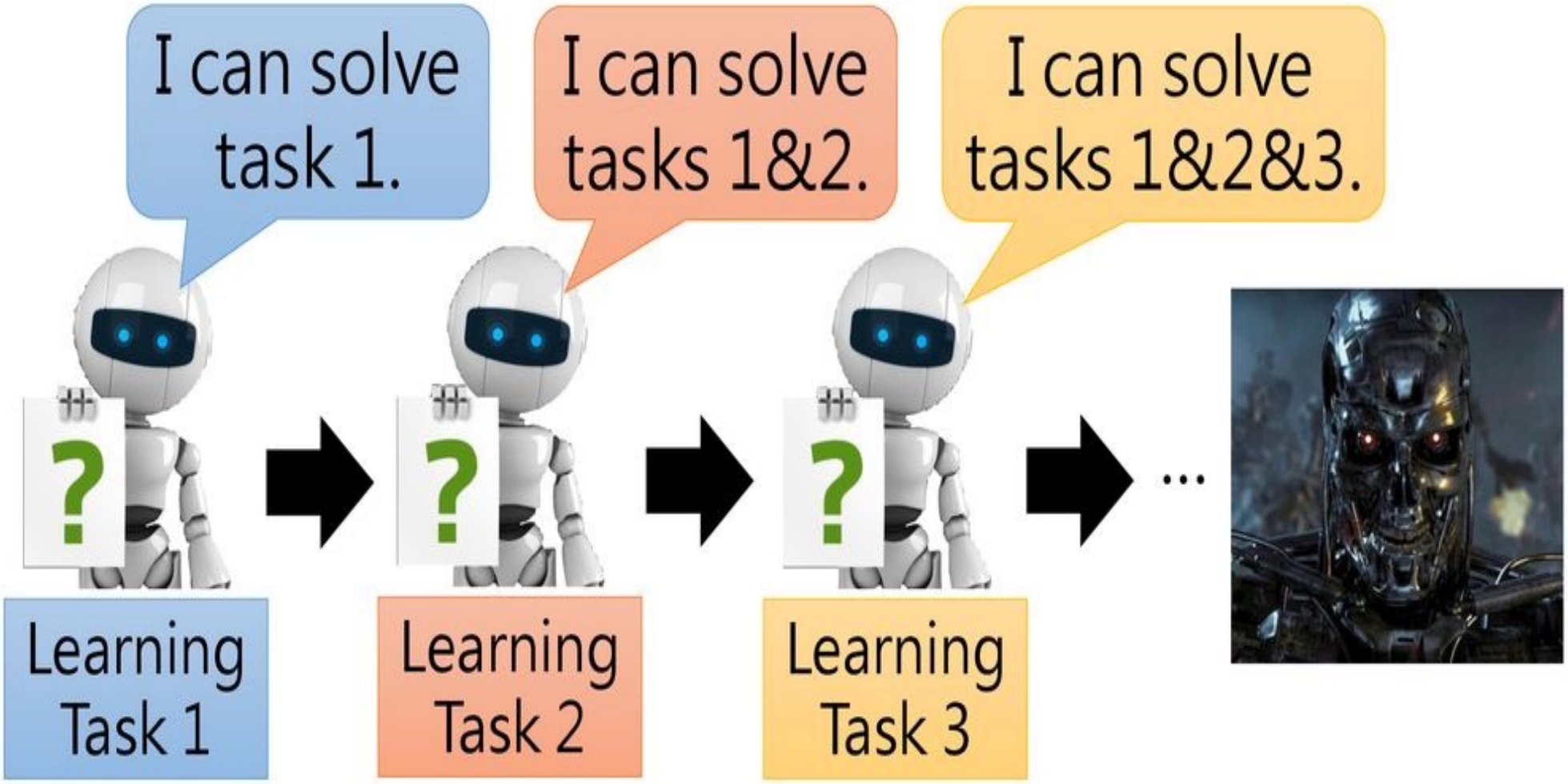


An Continual Learning Scenario

- And then during travels

An Continual Learning Scenario

- And then during travels



Slide credit: Hung-yi Lee

Standard Machine Learning

TRAIN – VALIDATION – TEST

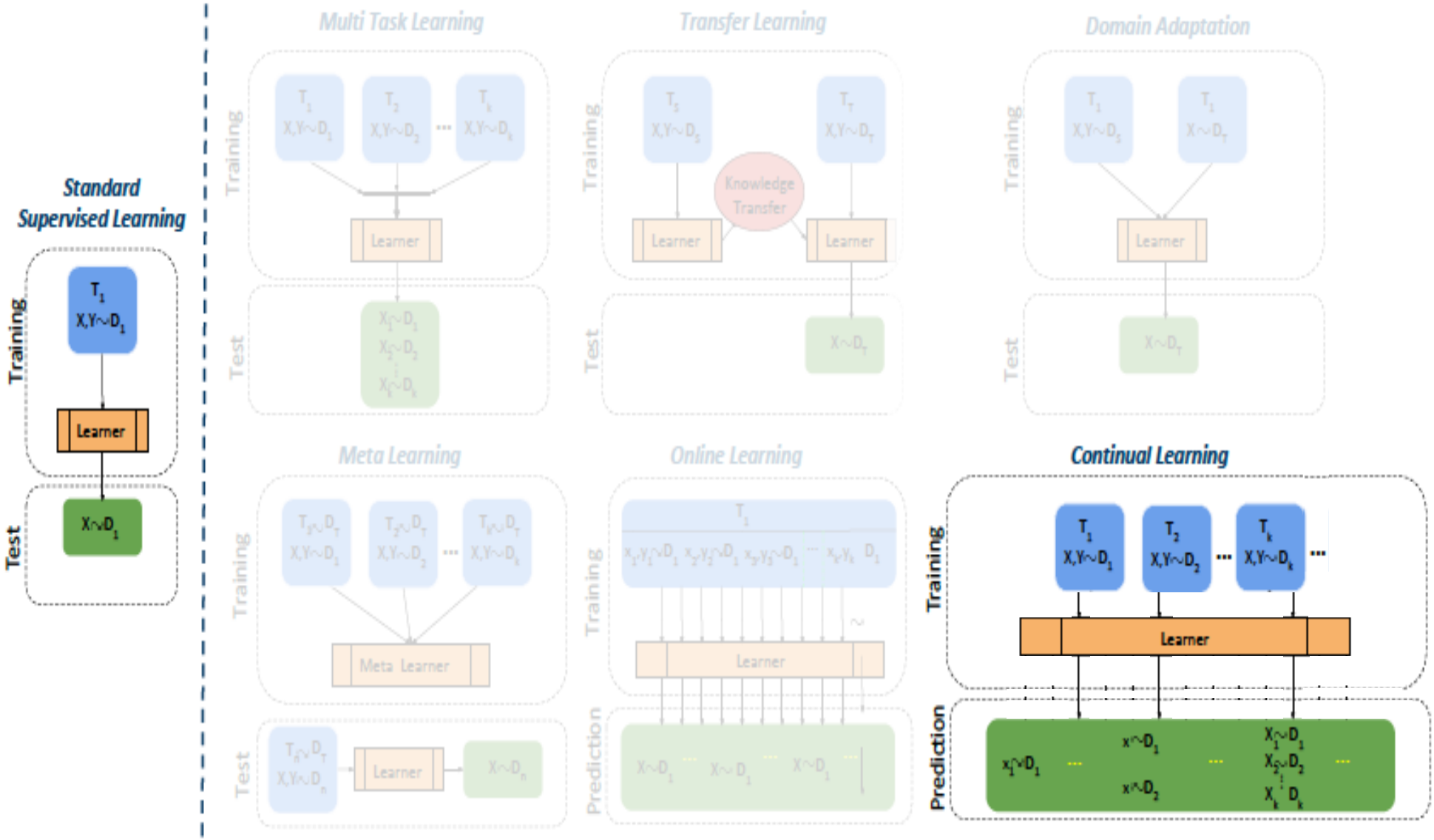
All sampled from the same distribution

-> benchmarks and academic datasets

-> real-world systems

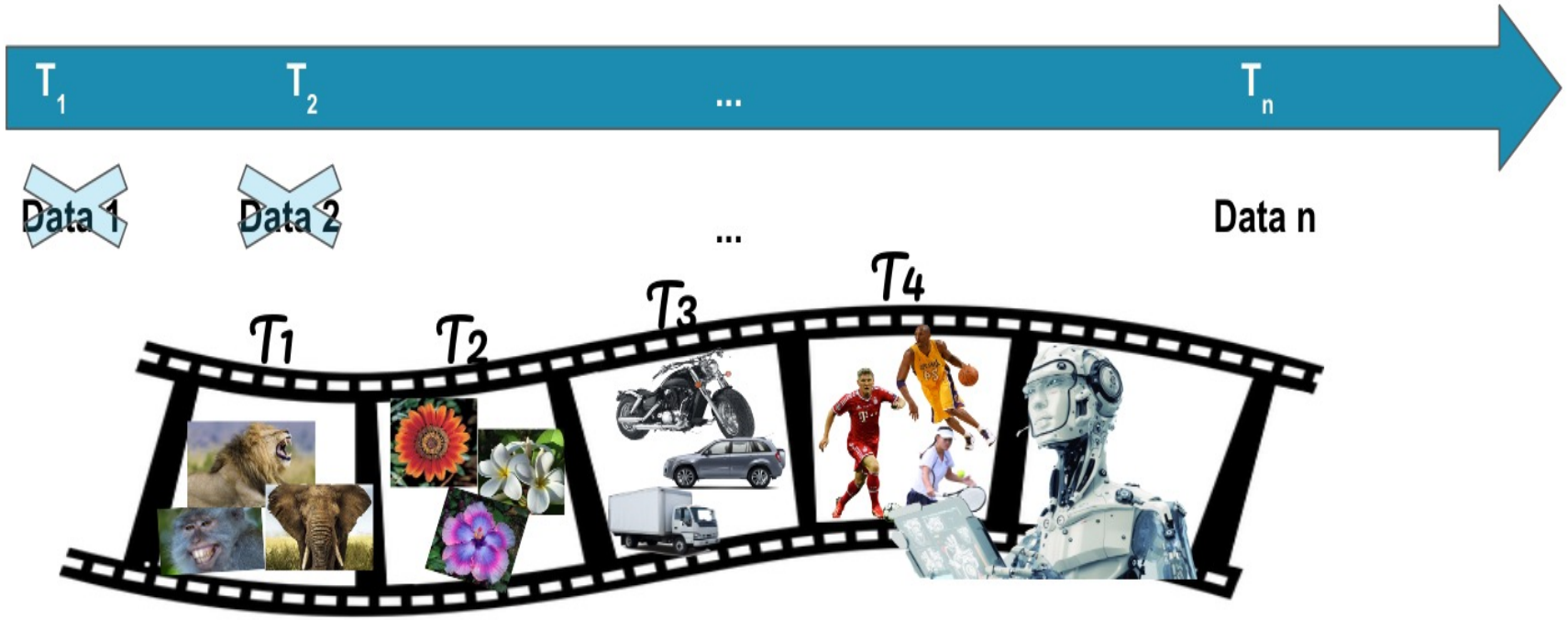
-> embodied learning

Slide credit: T. Tuytelaars



Slide credit: T. Tuytelaars

Incremental Learning Setup



- Task-incremental learning
- Class-incremental learning
- Domain-incremental learning

Slide credit: T. Tuytelaars

Incremental Learning

- A classical problem in machine learning, e.g.,

  [Carpenter et al. ’92, Cauwenberghs and Poggio ’00, Polikar et al. ’01, Schlimmer and Fisher ’86, Thrun ’96]

- Some methods
  - Zero-shot learning, e.g., [Lampert et al. ’13]

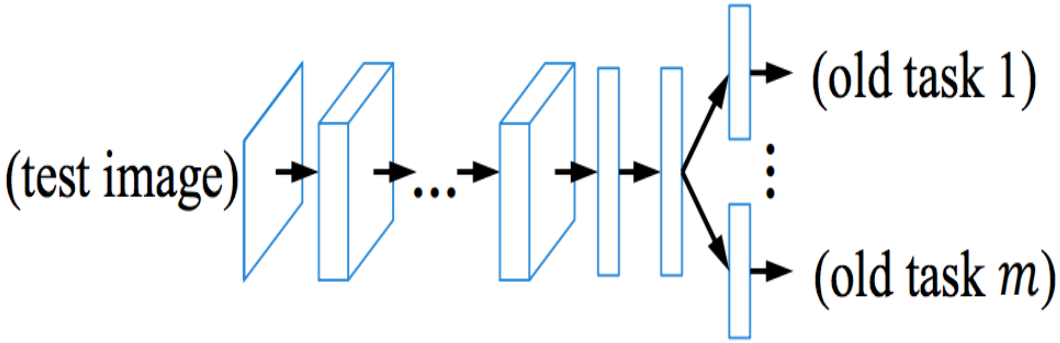
    No training step for unseen classes
  - Continuously update the training set, e.g., [Chen et al. ’13]

    Keep data and retrain
  - Use a fixed data representation, e.g., [Mensink et al. ’13]

    Simplify the learning problem

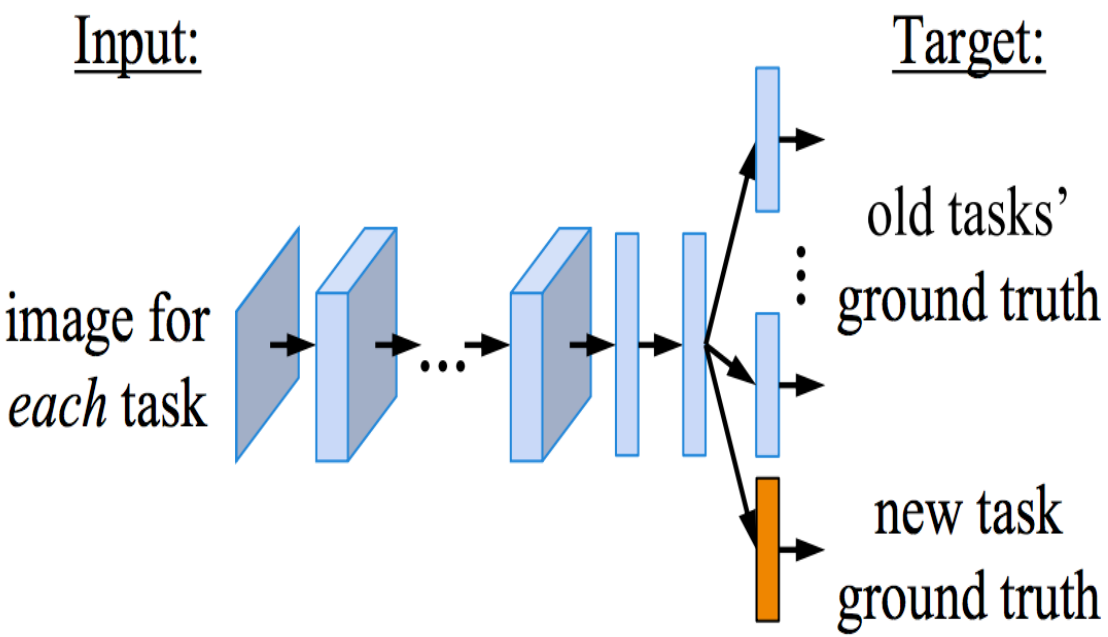
Brute Force Solution (non-incremental)

Original model



Joint training

“golden” baseline



random initialize + train
fine-tune
unchanged

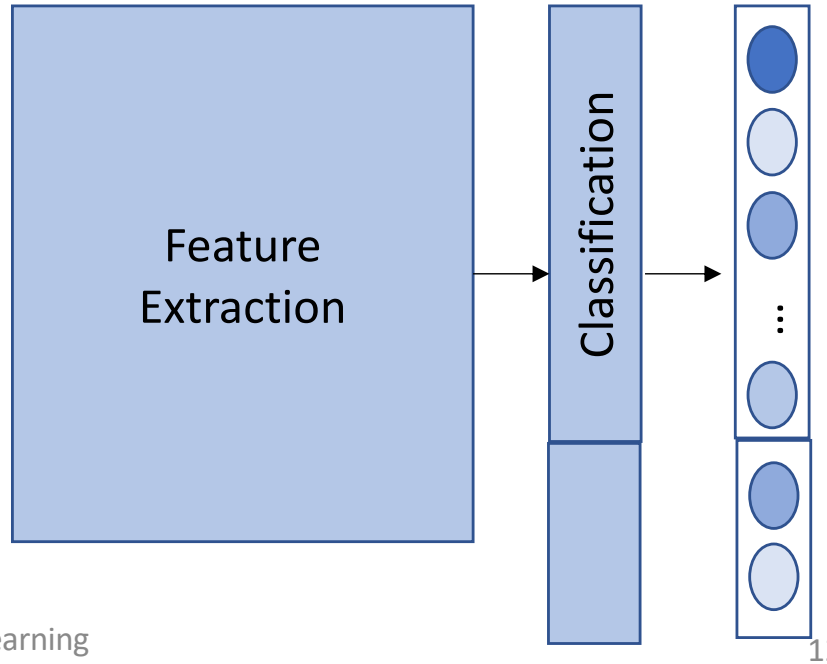
Figures from [Li and Hoiem 2016]

Brute Force Solution (non-incremental)

Retrain full model with both old and new data

- Computationally expensive

- Needs access to old data
  - Storage capacity limitations
  - Privacy issues
  - Scalability issues



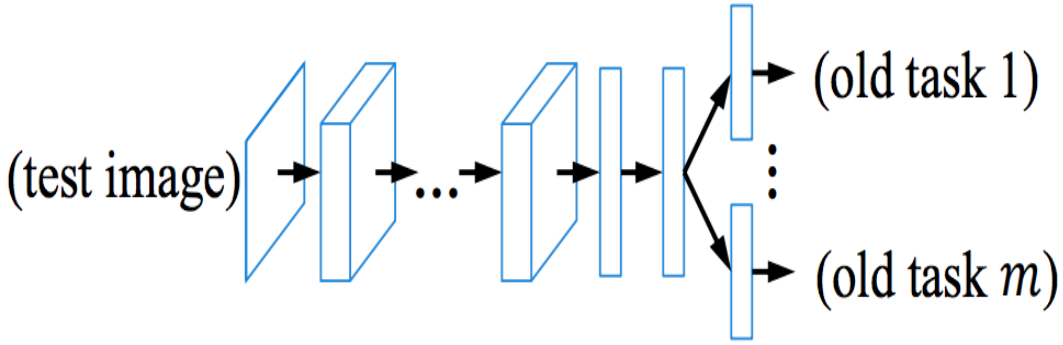
Slide credit: T. Tuytelaars

Why not brute force ?

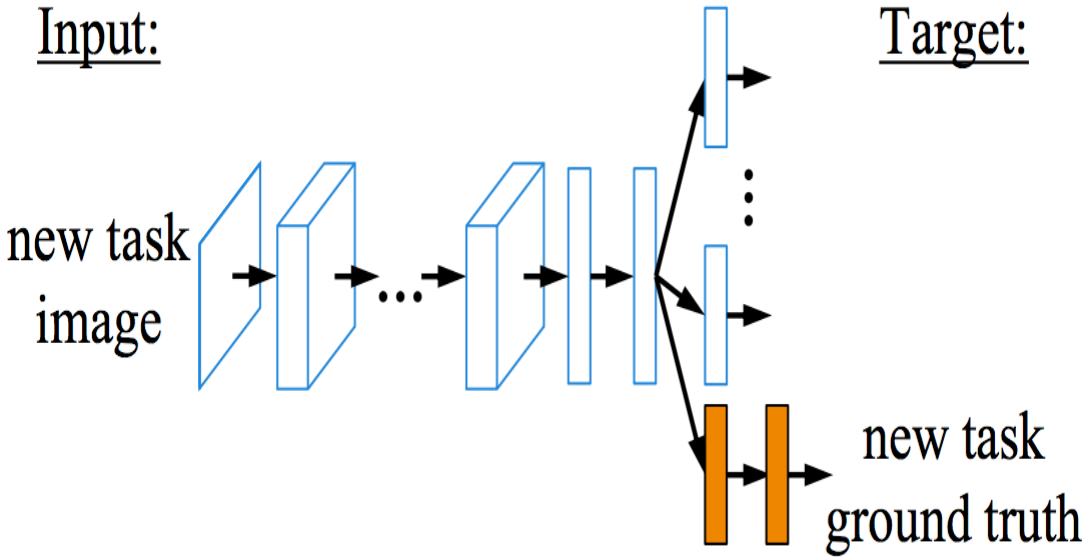
- No access to all the data
- Can not store all the data
- Access to only a previously learned model, e.g., trained by others

Naïve Solution 1

Original model



Feature extraction

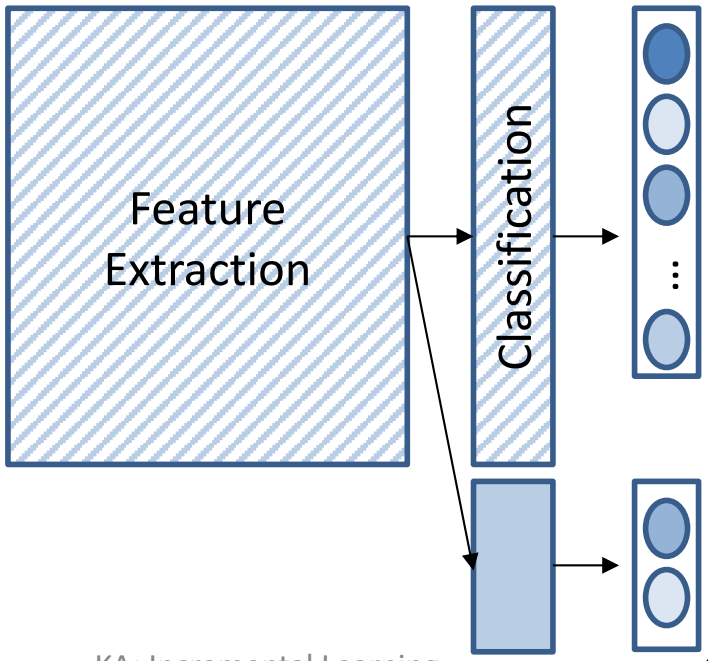


random initialize + train
fine-tune
unchanged

Figures from [Li and Hoiem 2016]

Naïve Solution 1

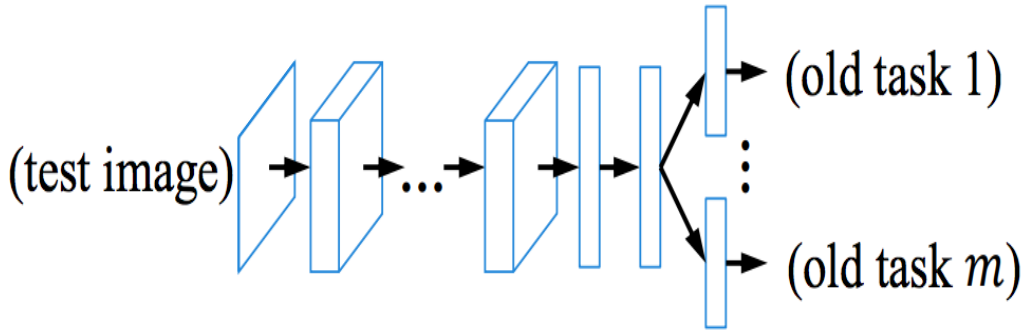
- Finetune only last layer using new data only
  - Leads to suboptimal results



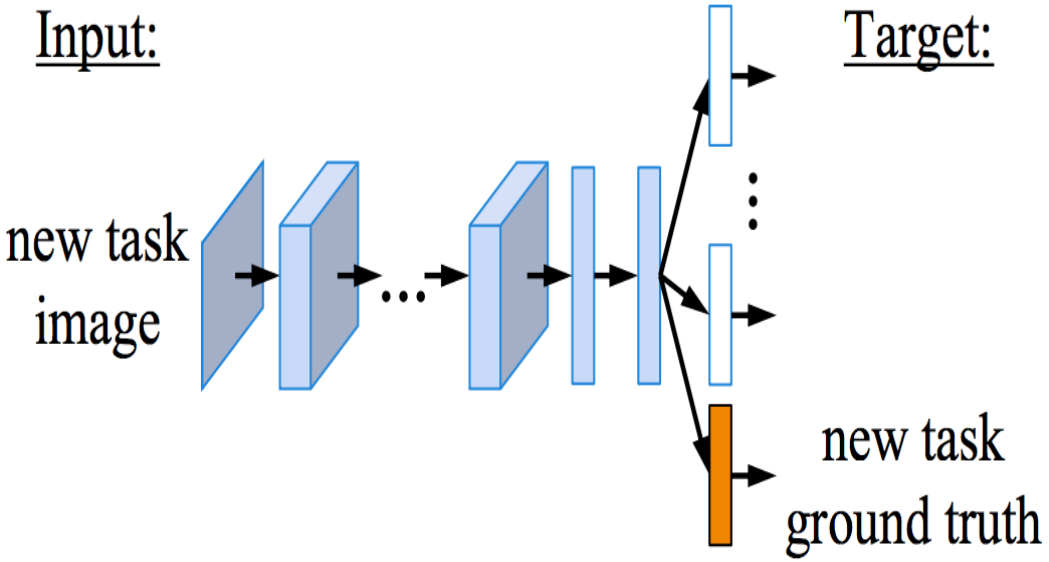
Slide credit: T. Tuytelaars

Naïve Solution 2

Original model



Fine tuning

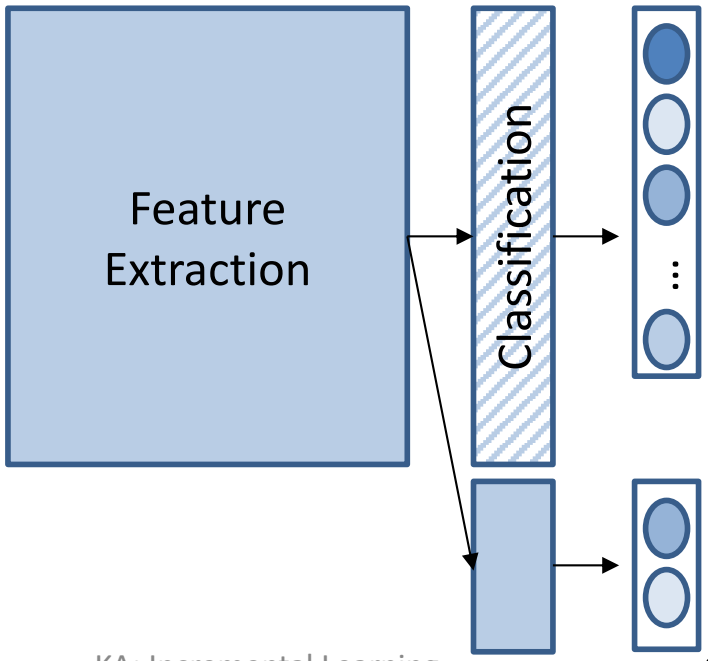


random initialize + train
fine-tune
unchanged

Figures from [Li and Hoiem 2016]

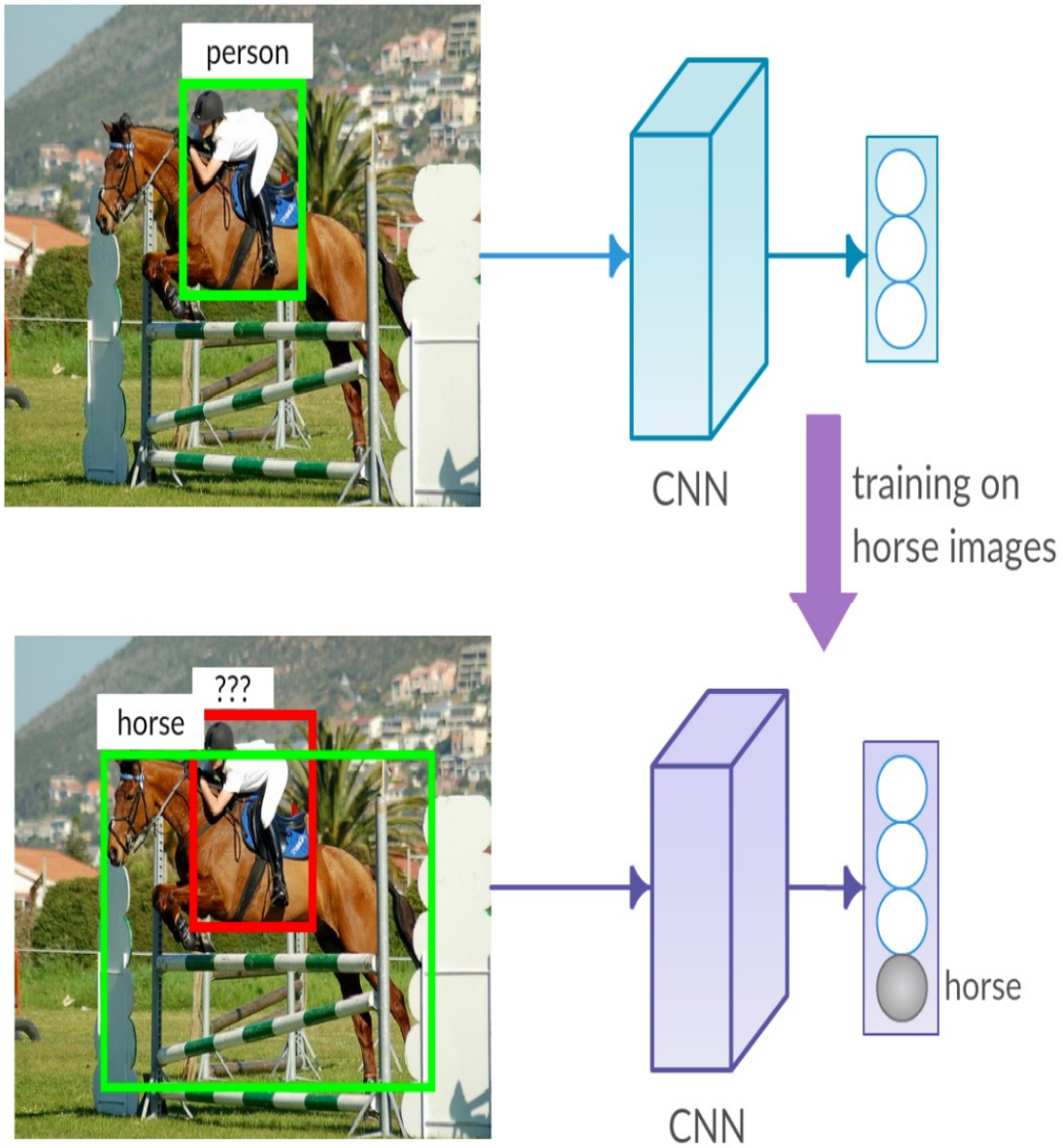
Naïve Solution 2

- Finetune the network using new data only
  - Leads to catastrophic forgetting



Slide credit: T. Tuytelaars

Incremental Learning: Computer Vision Task



How well does network B perform ?

Training with the **initial** set of classes

method	old	new	all
A(1-10)	65.8	-	-
+B(11-20)	**12.8**	**64.5**	**38.7**
A(1-20)	**68.4**	**71.3**	**69.8**

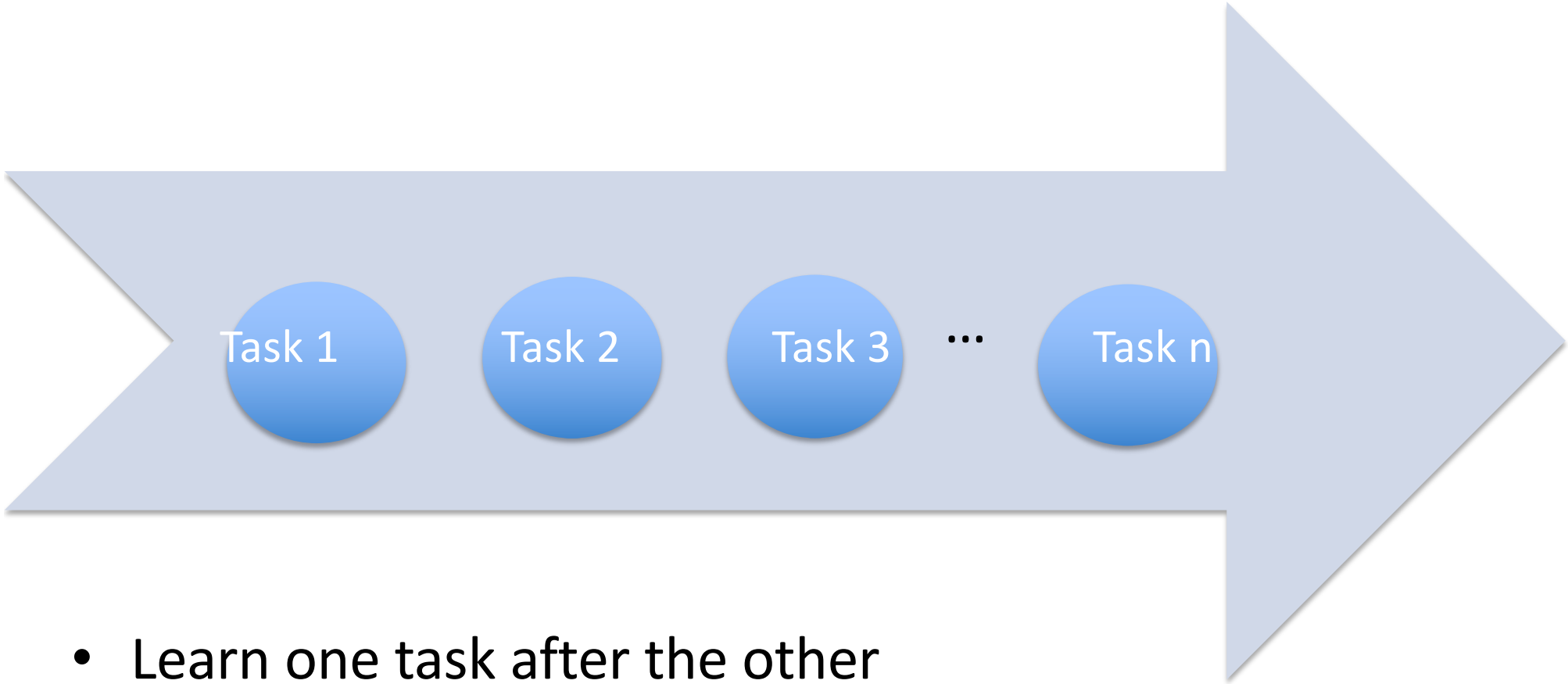
Training with the **new** set of classes

Baseline, i.e., training with all the classes

No guidance for retaining the old classes

[Catastrophic forgetting: McCloskey and Cohen 1989, Ratcliff 1990]

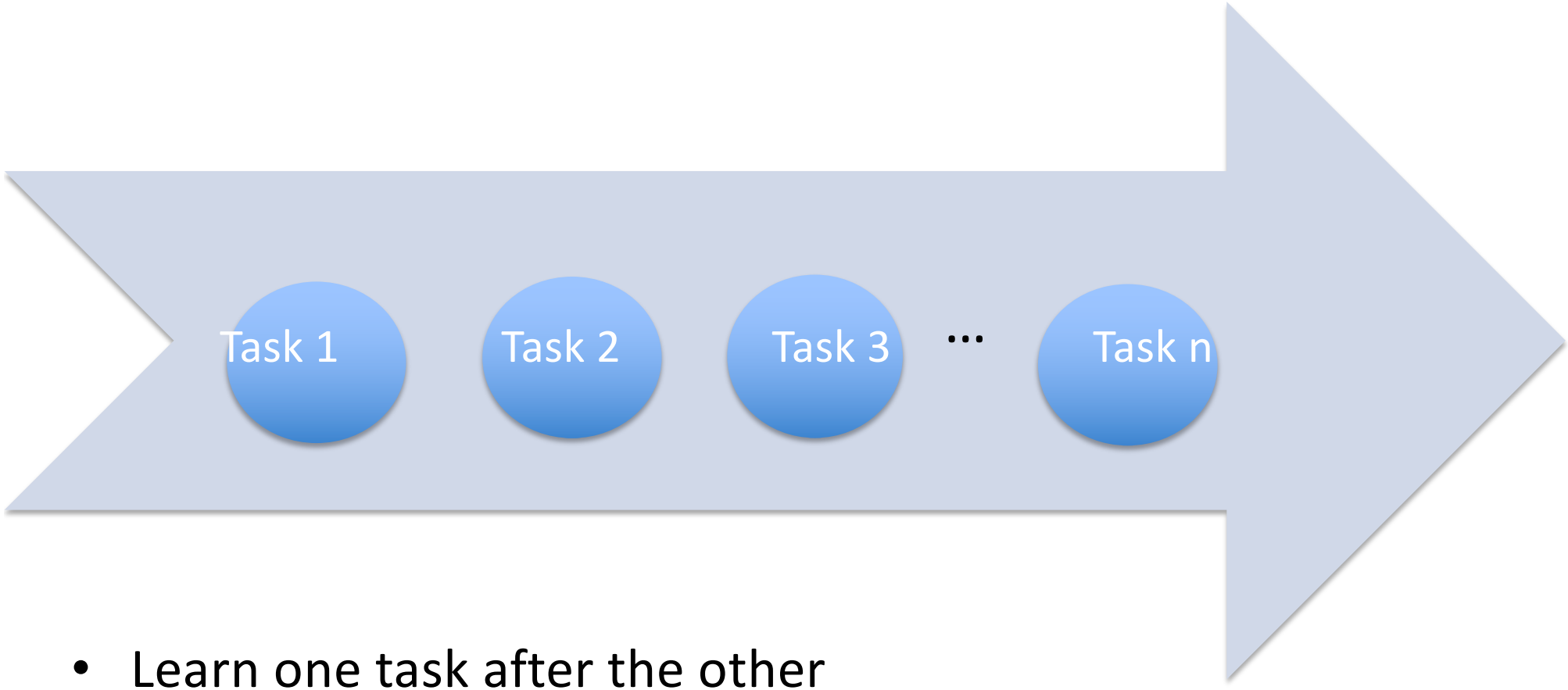
Incremental Learning: The Rules !



- Learn one task after the other
- Without storing data from previous tasks
- Without memory footprint growing over time
- Without forgetting old tasks

Slide credit: T. Tuytelaars

Incremental Learning: The Rules !



- Learn one task after the other
- Without storing (many) data from previous tasks
- Without memory footprint growing (significantly) over time
- Without (completely) forgetting old tasks

Slide credit: T. Tuytelaars

What else will we see today?

- Flavour of different approaches:
  1. Regularization based: LwF, EBLL, EWC, SI, MAS, IMM, …
  2. Rehearsal / Replay: iCaRL, DGR, GEM, …
  3. Architecture based: PackNet, progressive nets , HAT, …

- More than classification?

- Takeaways