Continual Learning: An Overview

Vincenzo Lomonaco
vincenzo.lomonaco@unipi.it

ContinualAI

A Non-profit Research Organization and Open Community on
Continual Learning for **AI**

Home | News & Events | Research | Lab | Forum

Supporters | About us | Join us!

ContinualAI.org

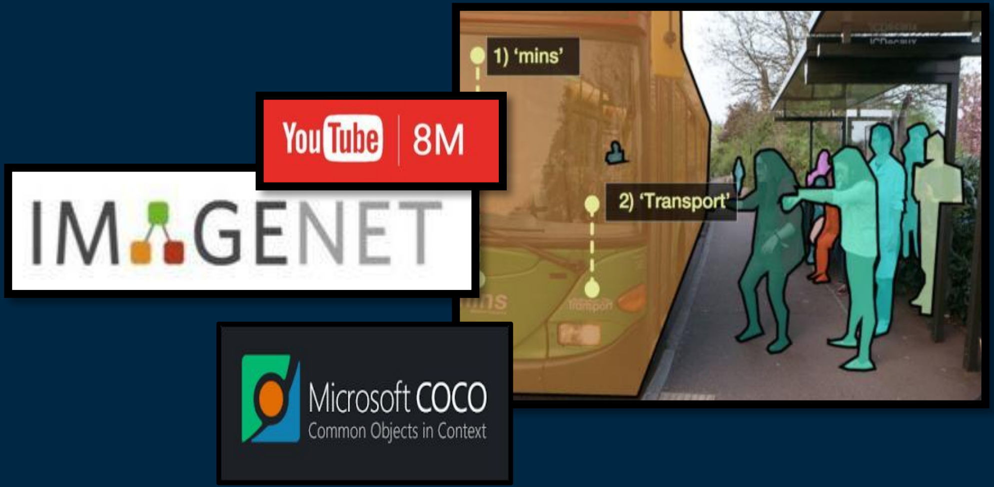
Focus on Deep Neural Networks

TABLE OF CONTENTS

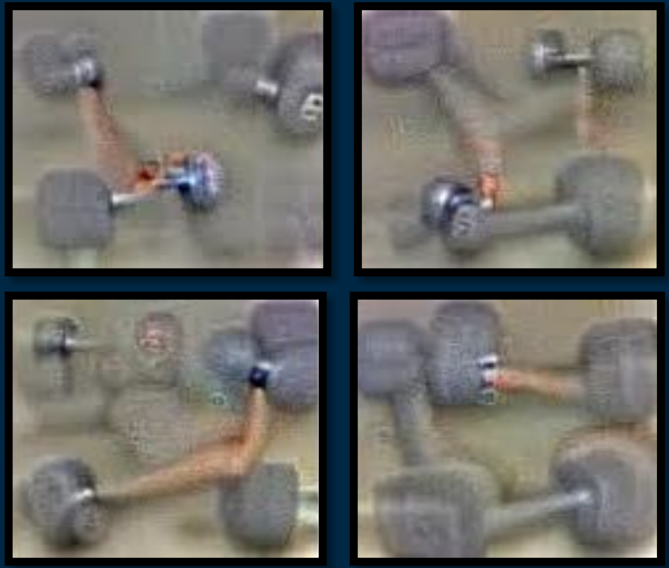
What’s Continual Learning?

Machine Learning: State-Of-The-Art

- **Deep Learning** holds state-of-the-art performances in many tasks.
- Mainly supervised training with **huge** and **fixed** datasets.



The Curse of Dimensionality

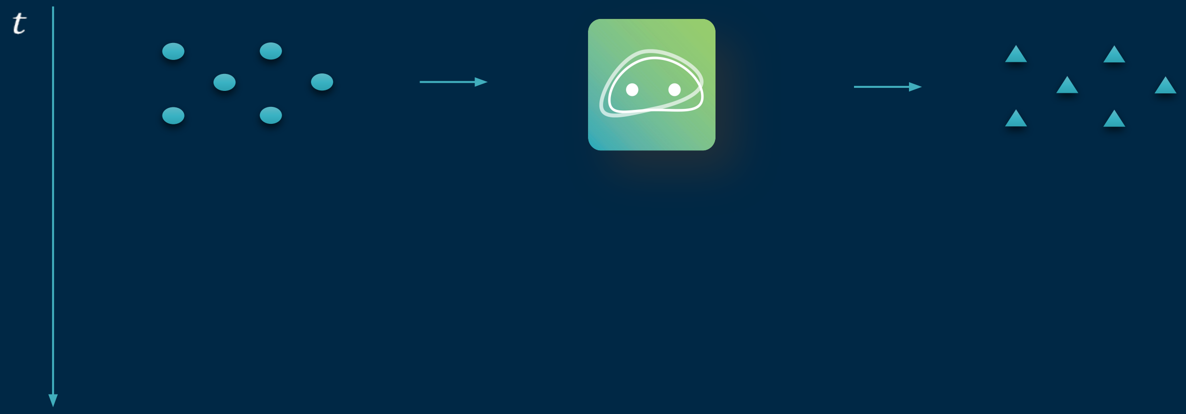


a group of people standing on top of a beach

How can we improve AI efficiency, scalability and adaptability?

*(Hence making them **sustainable in the long term**)*

Continual Learning (CL)



X_1 $\qquad$ $f_\theta : X_1 \rightarrow Y_1$ $\qquad$ Y_1

Continual Learning

X_2 $\quad f_\theta \colon X_1 \cup X_2 \to Y_1 \cup Y_2$ $\quad Y_2$

Continual Learning

- Higher and **realistic time-scale** where data (and tasks) become available only during time.
- **No access** to previously encountered data.
- **Constant** computational and memory resources (efficiency)
- **Incremental development** of ever more complex knowledge and skills (scalability)
- **Efficiency** + **Scalability** = **Sustainability**

The Stability-Plasticity Dilemma

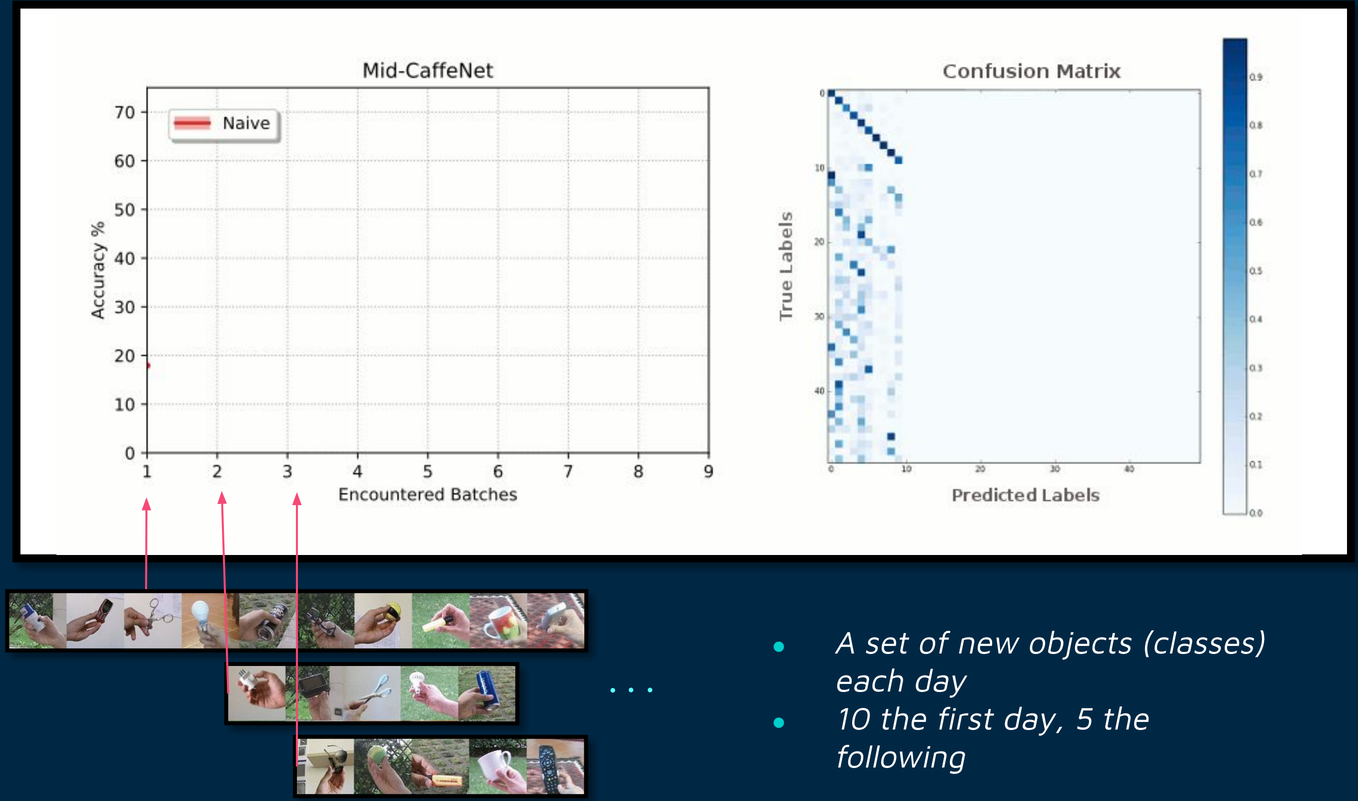
Stability-Plasticity Dilemma:

- **Remember past concepts**
- **Learn new concepts**
- **Generalize**

First Problem in Deep Learning:

- **Catastrophic Forgetting**

Catastrophic Forgetting



. . .

- *A set of new objects (classes) each day*
- *10 the first day, 5 the following*

CORe50: a new Dataset and Benchmark for Continuous Object Recognition, V. Lomonaco & D. Maltoni. Conference on Robot Learning (CoRL), 2017.

Nomenclature & Related Paradigms

Unconsolidated Nomenclature

- Continual Learning
- Incremental Learning
- Lifelong Learning
- Continuous Learning

Related Paradigm

- Multi-Task Learning
- Meta-Learning / Learning to Learn
- Transfer Learning & Domain Adaptation
- Online / Streaming Learning

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges, Lesort et al. Information Fusion, 2020.

History Timeline (Some Examples)

- Incremental learning with rule-based systems (**Diederich, 1987**)
- Forgetting in Neural Networks (**French, 1989**)
- Incremental learning with Kernel Machines (**Tat-Jun, 1999**)
- Continual Learning (**Ring, 1998**)
- Lifelong Learning (**Thrun, 1998**)
- Dataset Shift (**Quiñonero-Candela, 2008**)
- Never-Ending Learning (**Mitchell, 2009**)
- Concept Drift Adaptation (**Ditzler, 2015**)
- Deep Continual Learning (**Kirkpatrick, 2016**)
- Lifelong (Language) Learning (**Liu, 2018**)

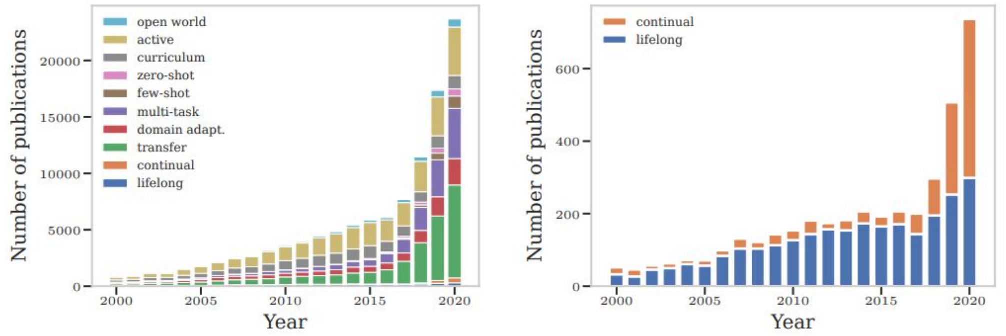


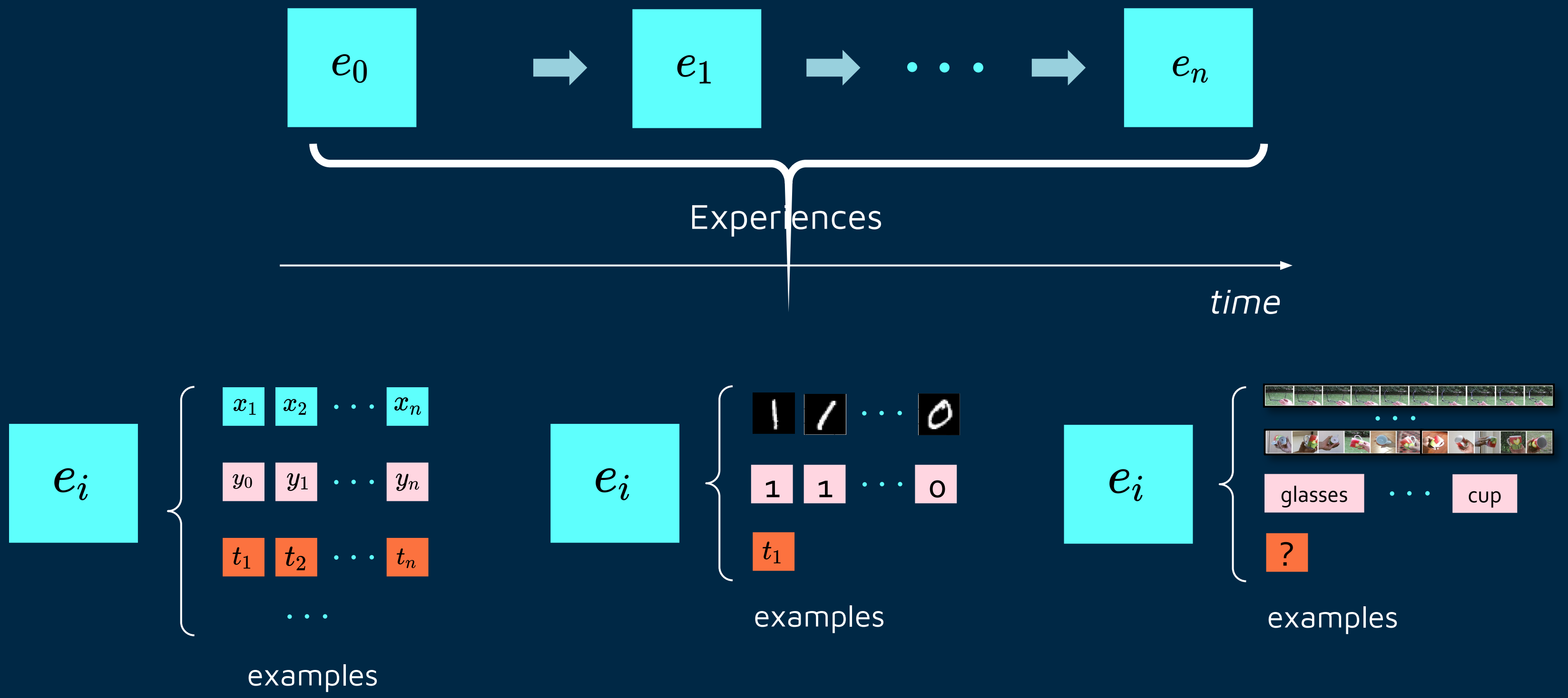
Figure 1: Per year machine learning publications. Left: cumulative amount of papers across keywords with continuous components that influence continual learning practice, see Section 2. Right: increasing use of "continual" machine learning, demonstrating a shift in use of terminology with respect to the preceding emphasis on the term "lifelong". Data queried using the Microsoft Academic Graph utilities (Sinha et al., 2015) based on keyword occurrence in the abstract.

CLEVA-Compass: A Continual Learning EValuation Assessment Compass to Promote Research Transparency and Comparability, Martin Mundt et al. preprint, 2021

“We are not looking for incremental improvements in state-of-the-art AI and neural networks, but rather paradigm-changing approaches to machine learning that will enable systems to continuously improve based on experience.”

— Hava Siegelmann, 2018

Continual Learning (more formally)



Continual Learning (more formally)

In continual learning (CL) data arrives in a streaming fashion as a (possibly infinite) sequence of learning experiences $S = e_1, \ldots, e_n$. For a supervised classification problem, each experience e_i consists of a batch of samples $\mathcal{D}^i$, where each sample is a tuple $\langle x_k^i, y_k^i \rangle$ of input and target, respectively, and the labels y_k^i are from the set $\mathcal{Y}^i$, which is a subset of the entire universe of classes $\mathcal{Y}$. Usually $\mathcal{D}^i$ is split into a separate train set $\mathcal{D}^i_{train}$ and test set $\mathcal{D}^i_{test}$.

A continual learning algorithm $\mathcal{A}^{CL}$ is a function with the following signature:

$$\mathcal{A}^{CL} : \langle f^{CL}_{i-1}, \mathcal{D}^i_{train}, \mathcal{M}_{i-1}, t_i \rangle \rightarrow \langle f^{CL}_i, \mathcal{M}_i \rangle \quad (1)$$

where f^{CL}_i is the model learned after training on experience e_i, $\mathcal{M}_i$ a buffer of past knowledge, such as previous samples or activations, stored from the previous experiences and usually of fixed size. The term t_i is a task label which may be used to identify the correct data distribution.

The objective of a CL algorithm is to minimize the loss $\mathcal{L}_S$ over the entire stream of data S:

$$\mathcal{L}_S(f^{CL}_n, n) = \frac{1}{\sum_{i=1}^{n} |\mathcal{D}^i_{test}|} \sum_{i=1}^{n} \mathcal{L}_{exp}(f^{CL}_n, \mathcal{D}^i_{test}) \quad (2)$$

$$\mathcal{L}_{exp}(f^{CL}_n, \mathcal{D}^i_{test}) = \sum_{j=1}^{|\mathcal{D}^i_{test}|} \mathcal{L}(f^{CL}_n(\boldsymbol{x}^i_j), y^i_j), \quad (3)$$

where the loss $\mathcal{L}(f^{CL}_n(\boldsymbol{x}), y)$ is computed on a single sample $\langle \boldsymbol{x}, y \rangle$, such as cross-entropy in classification problems.

Desiderata

- Replay-Free Continual Learning
- Memory and Computationally Bounded
- Task-free Continual Learning
- Online Continual Learning

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges, Lesort et al. Information Fusion, 2020.

Benchmarks and Algorithms

Benchmarks

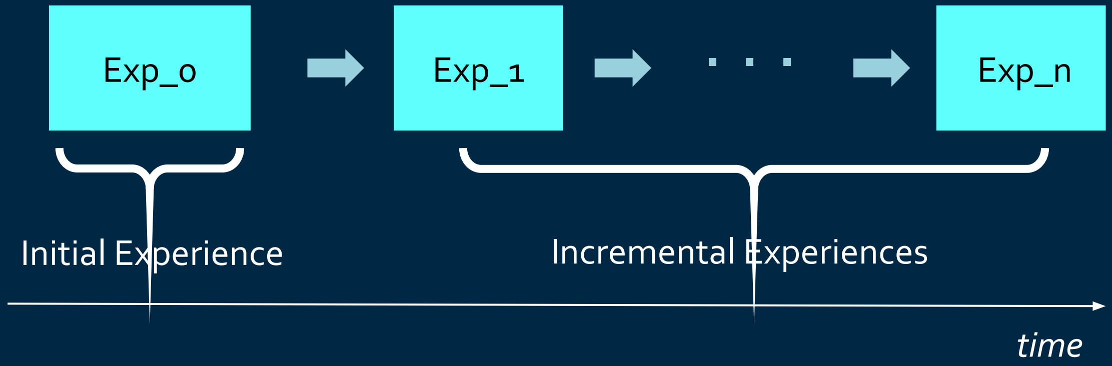
Current Focus

- Multi-Task (Often with Task Supervised Signals)
- I.I.D by Parts
- Few Big Tasks
- Unrealistic / Toy Datasets
- Mostly Supervised
- Accuracy

Better Focus

- Single-Incremental-Task
- High-Dimensional Data Streams (highly non-i.i.d.)
- Natural / Realistic Datasets
- Mostly Unsupervised
- Scalability and Efficiency

Scenarios



1. **Task-Incremental**: every experience is a different task.
2. **Class-Incremental**: every experience contains examples of different classes.
3. **Domain-Incremental**: every experience contains examples of the same classes.

...and many others!

Common CL benchmarks

Table 3: Benchmarks and environments for continual learning. For each resource, paper use cases in the NI, NC and NIC scenarios are reported.

Benchmark	NI	NC	NIC	Use Cases
Split MNIST/Fashion MNIST		✓		[83, 81, 57, 130]
Rotation MNIST	✓			[92, 83, 127]
Permutation MNIST	✓			[53, 73, 43, 150, 176, 83, 57, 127]
iCIFAR10/100		✓		[125, 97, 70]
SVHN		✓		[71, 145, 130]
CUB200	✓			[80]
CORe50	✓	✓	✓	[91, 115, 97]
iCubWorld28	✓			[116, 90]
iCubWorld-Transformation		✓		[117, 16]
LSUN		✓		[171]
ImageNet		✓		[125, 95]
Omniglot		✓		[77, 144]
Pascal VOC		✓		[104, 151]
Atari	✓			[136, 73, 144]
RNN CL benchmark			✓	[153]
CRLMaze (based on VizDoom)	✓			[89]
DeepMind Lab	✓			[99]

Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. Lesort et al, Information Fusion, 2020.

Not only Data Streams but Sequences!

Continual Learning needs the presence of multiple (temporal coherent and unconstrained) views of the same objects taken in different sessions.

Lomonaco V. and Maltoni D. ***CORe50: a New Dataset and Benchmark for Continuous Object Recognition***. CoRL2017.

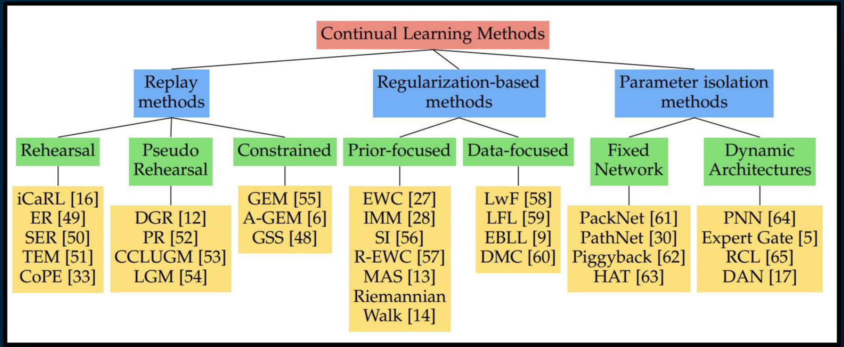
CORe50: a Video Benchmark for CL and Object Recognition/Detection/Segmentation

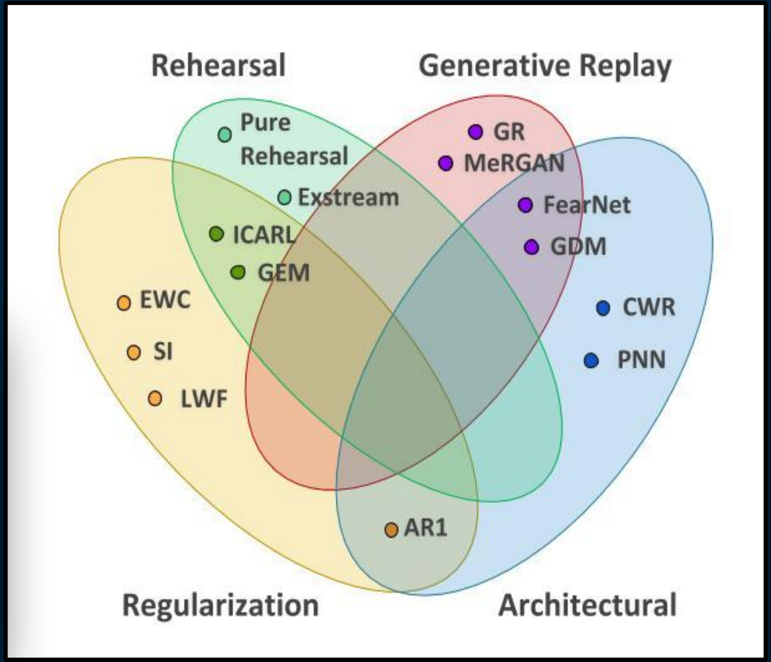
Lomonaco V. and Maltoni D. ***CORe50: a New Dataset and Benchmark for Continuous Object Recognition***. CoRL2017.

Possible 4-way Fuzzy Categorization

With some twists

- **No formal definition**
- **Alternative categorizations** are possible



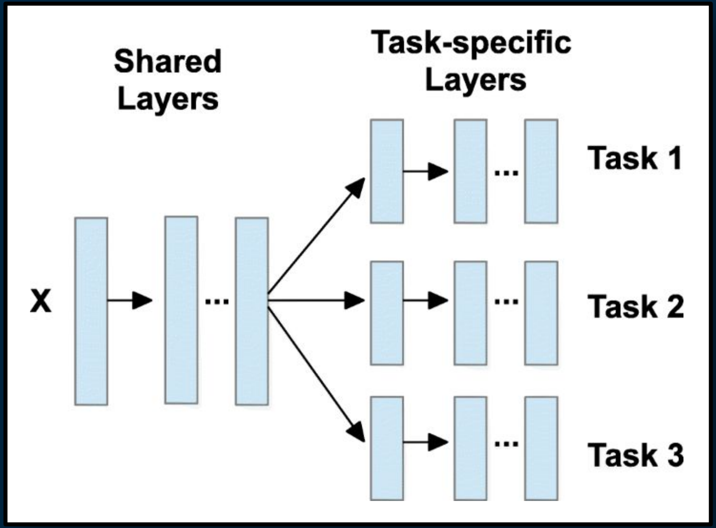


Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges, Lesort et al. Information Fusion, 2020.
A continual learning survey: Defying forgetting in classification tasks. De Lange et al, TPAMI 2021.

Continual Learning Baselines

Common Baselines / Control Algorithms

- **Naive** / **Finetuning** (just continuing backprop)
- **JointTraining** / **Offline** (pure Multi-task learning): The best you can do with all the data starting from scratch
- **Ensemble**: one model for each experience
- **Cumulative**: for every experience, accumulate all data and re-train from scratch.



A brief review on multi-task learning. Thung et al, 2018.

Random Replay

A basic approach

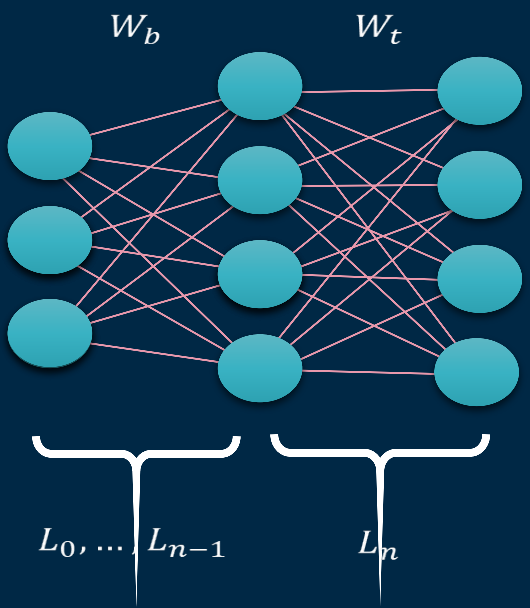
- **Sample randomly** from the current experience data
- **Fill your fixed Random Memory** (RM)
- **Replace examples randomly** to maintain an approximate equal number of examples for experience

Algorithm 1 Pseudocode explaining how the external memory RM is populated across the training batches. Note that the amount h of patterns to add progressively decreases to maintain a nearly balanced contribution from the different training batches, but no constraints are enforced to achieve a class-balancing.

$RM = \varnothing$
RM_{size} = number of patterns to be stored in RM
for each training batch B_i:
    train the model on shuffled $B_i \cup RM$
    $h = \frac{RM_{size}}{i}$
    R_{add} = random sampling h patterns from B_i
    $R_{replace} = \begin{cases} \varnothing & \text{if } i == 1 \\ \text{random sample } h \text{ patterns from } RM & \text{otherwise} \end{cases}$
    $RM = (RM - R_{replace}) \cup R_{add}$

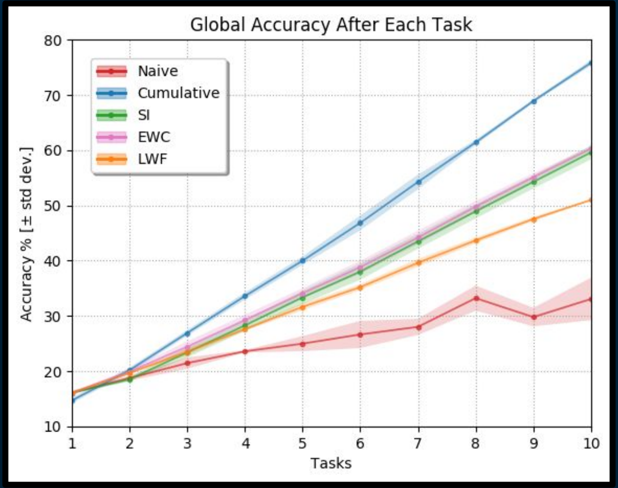
Latent Replay for Real-Time Continual Learning. Pellegrini et al. IROS, 2019.

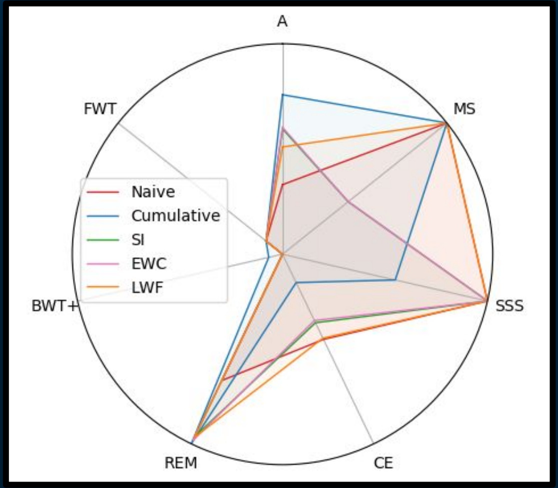
Elastic Weights Consolidation (EWC)



$$\underbrace{\begin{gathered}\tilde{L}_\mu = L_\mu + \lambda \sum_k \Omega_k^\mu \left(\bar{\theta}_k - \theta_k\right)^2 \\ w_k^v = \int \left(\frac{\partial}{\partial \theta_k} \log f(x; \theta_k)\right)^2 f(x; \theta_k)\, dx\end{gathered}}_{\text{Fisher Information}}$$

Evaluation & Metrics





N. Díaz-Rodríguez, V. Lomonaco et al. ***Don't forget, there is more than forgetting: new metrics for Continual Learning***. CL Workshop, NeurIPS 2018.

Real-world Applications

Continual Learning Applications

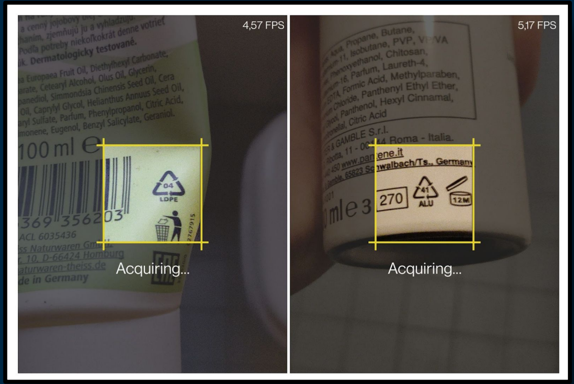
Main Possibilities (Grouped by "Where" Computation Happen)

- **Edge**
  - **Embedded systems and Robotics:** +privacy, +efficiency, +fast adaptation, +on the edge, -Internet connection (e.g. Autonomous Cars, Robotics Arms/Hands)
- **Cloud**
  - **AutoML and CI systems for AI models:** +scalability, +efficiency, +fast adaptation, -energy consumption, -$$$ (e.g. Recommendation Systems)
- **Continuum Edge-Cloud**
  - **Pervasive AI systems**: Efficient Communication, fluid & dynamic computation
  - **Neural Patches**: +security patches, +fairness patches, +fast update
  - **Continual Distributed Learning**: understudied relationship with parallel and federated learning

On-Device Personalization without Forgetting

L. Pellegrini et al. ***Latent Replay for Real-Time Continual Learning***, IROS 2020.
L. Pellegrini et al. ***Continual Learning at the Edge: Real-Time Training on Smartphone Devices***. ESANN, 2021.
G. Demosthenous et al. ***Continual Learning on the Edge with TensorFlow Lite***. *arXiv 2021.*
L. Ravaglia et al. ***Memory-Latency-Accuracy Trade-offs for Continual Learning on a RISC-V Extreme-Edge Node***. *SiPS 2020.*

Use-Case: Recycling Codes Recognition



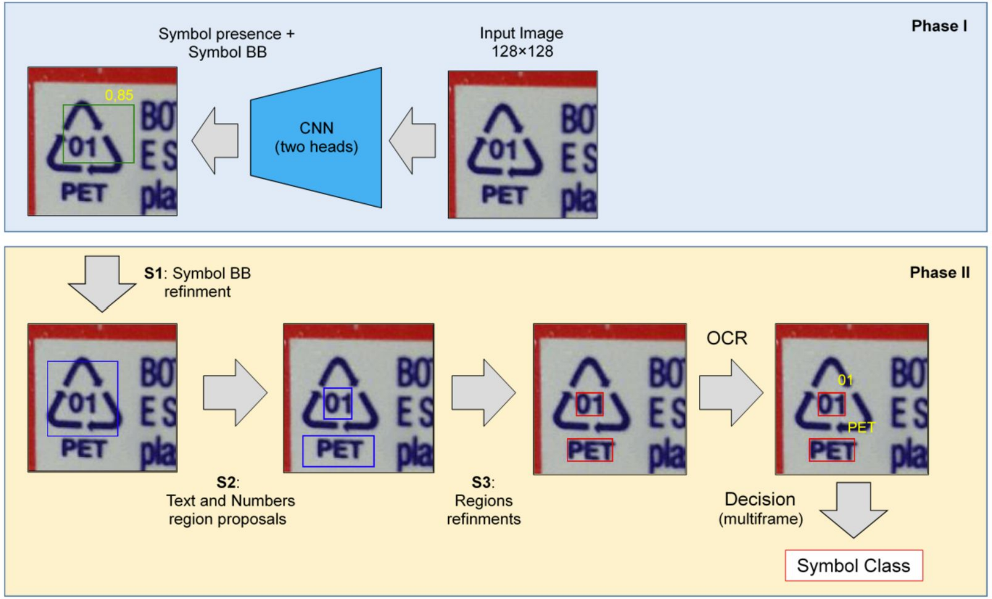


Figure 4. Overall schema of the proposed approach.

A Weakly Supervised Approach for Recycling Code Recognition, Pellegrini et al, under review, 2021.

Continual Learning for Particle Accelerators

278

petabytes of data

In the last decade, LHC experiments collected almost 280 petabytes of data, which scientists recorded on tape. You would need to stream Netflix 24/7 for more than 15,000 years to eventually use that much data! But from another perspective, platforms like Facebook (which has 2.5 billion users) collect that much data in 70 days!

Interesting Properties

- Massive amount of data
- Non-stationary stream
- Need for fast adaptation

Questions

- Supervision level
- Types of non-stationarity

10 years of LHC physics, in numbers, Sarah Charley, 2020. [link]

Useful Tools for Prototyping

Avalanche

- Focus on deep continual learning (representation learning & knowledge accumulation)
- Python & Pytorch based
- https://avalanche.continualai.org

River

- Focus on online/streaming learning on low-dimensional data
- Python-based
- https://riverml.xyz/0.11.1/

10 years of LHC physics, in numbers, Sarah Charley, 2020. [link]

Conclusions

What we have seen

- Significant and **growing Interest** in the last few years on Continual learning within Deep Learning
- Significant **improvements over standard benchmark** but **focus still mostly on simplified scenarios** and forgetting centered metrics
- **Huge space of possible and significant explorations**

Take-Home Messages

1. Continual Learning is a **paradigm-changing approach** trying to break the fundamental i.i.d. assumption in statistical learning
2. CL pushes for the **next step in Neuroscience-grounded approaches** to learning
3. CL pushes for the next generation of truly intelligent robust and autonomous AI systems: **efficient, effective, scalable, hence sustainable**

Continual Learning: On Machines that Can Learn Continually

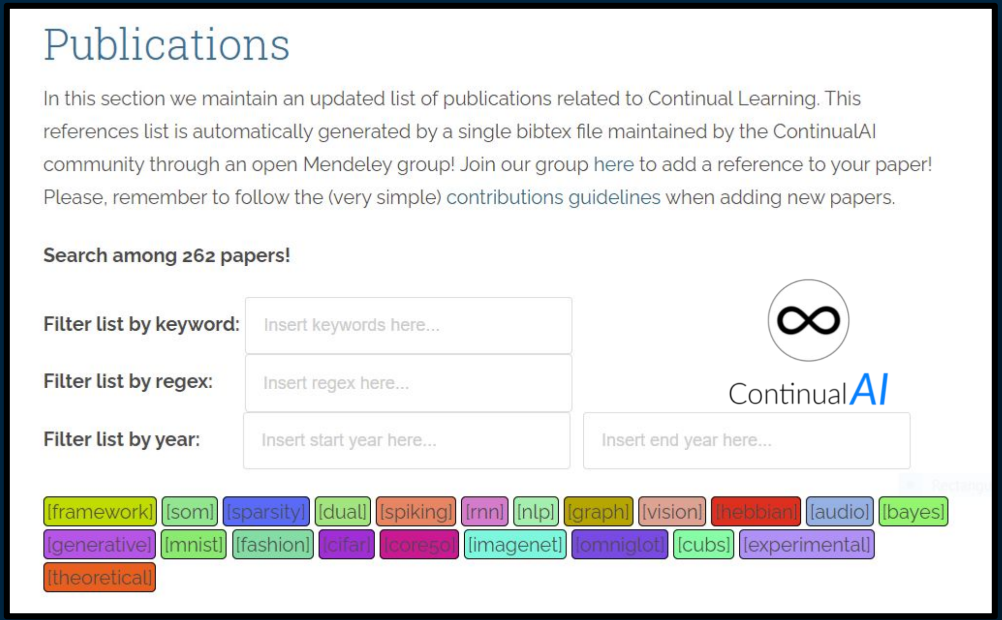
1st Open-Access Course on CL
Offered by Unipi & ContinualAI

course.continualai.org

Available on YouTube!

Additional Materials

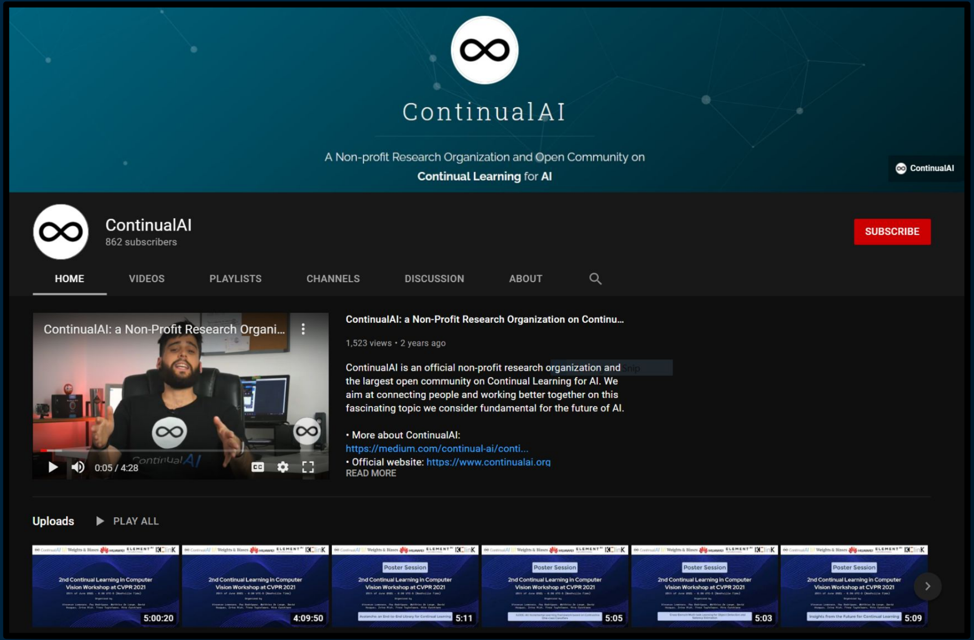
- **ContinualAI Wiki**: a shared and collaboratively maintained *knowledge base* for Continual Learning: tutorials, workshops, demos, tutorials, courses, etc.
- **Continual Learning Papers**: curated list of CL papers & books with meta-data by ContinualAI
- **ContinualAI Forum** + Slack: discussions / Q&As about Continual Learning
- **ContinualAI Research Consortium**: networks of Top CL Labs across the world.



You can find more at: *www.continualai.org*

Additional Materials

- **ContinualAI Publication**: a curated list of original blog posts on CL.
- **Continual Learning & AI Mailing List**+: open mailing-list moderated by the ContinualAI community.
- **ContinualAI Newsletter**:news from the ContinualAI community and the CL World in one place.
- **ContinualAI Seminars**: weekly invited talks on CL.
- **ContinualAI YouTube**: collection of videos about CL.



You can find more at: *www.continualai.org*

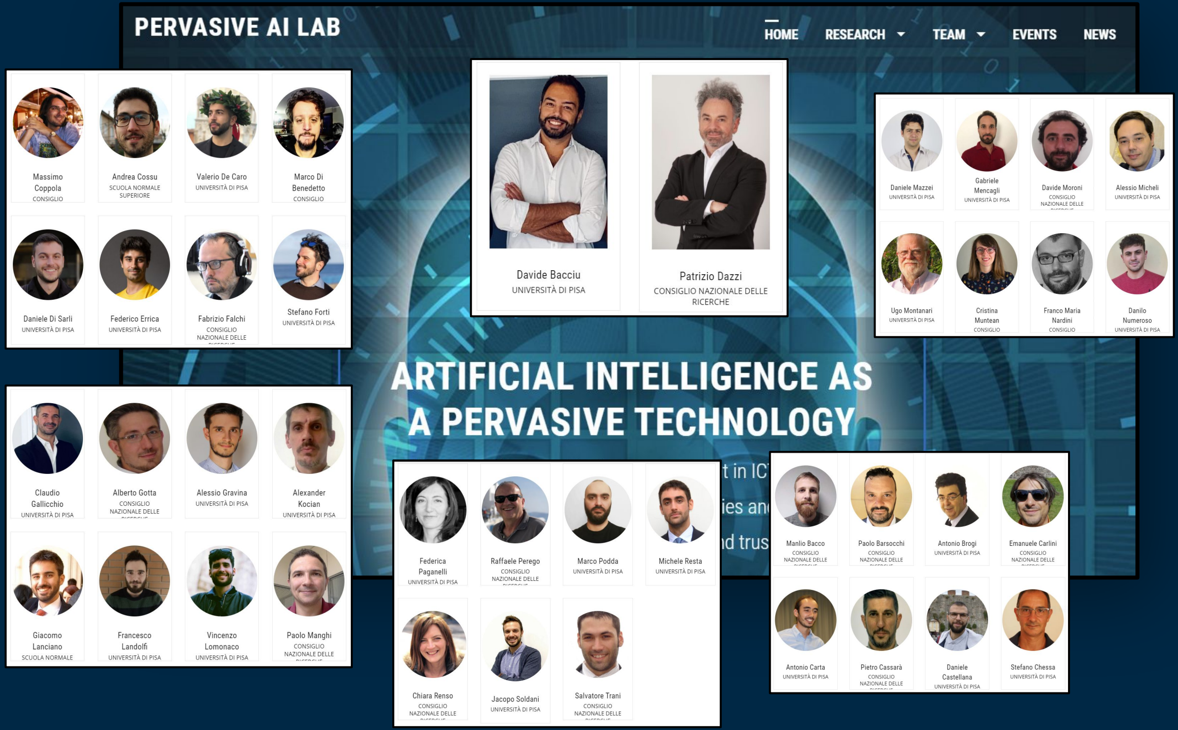
Some Must-Read References

Here a **few reviews** and **books** you can use as main references for CL:

1. **Lifelong Machine Learning**, Second Edition. by Zhiyuan Chen and Bing Liu. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2018.
2. **A Continual Learning Survey: Defying Forgetting in Classification Tasks** by Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Gregory Slabaugh and Tinne Tuytelaars. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
3. **Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges** by Timothée Lesort, Vincenzo Lomonaco, Andrei Stoian, Davide Maltoni, David Filliat and Natalia Díaz-Rodr\ǵuez. Information Fusion, 52--68, 2020.
4. **Continual Lifelong Learning with Neural Networks: A Review** by German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan and Stefan Wermter. Neural Networks, 54--71, 2019.
5. **Continual Learning for Recurrent Neural Networks: An Empirical Evaluation** by Andrea Cossu, Antonio Carta, Vincenzo Lomonaco and Davide Bacciu. Neural Networks, 607--627, 2021.
6. **Replay in Deep Learning: Current Approaches and Missing Biological Elements** by Tyler L. Hayes, Giri P. Krishnan, Maxim Bazhenov, Hava T. Siegelmann, Terrence J. Sejnowski and Christopher Kanan. arXiv, 2021.
7. **Embracing Change: Continual Learning in Deep Neural Networks** by Raia Hadsell, Dushyant Rao, Andrei A Rusu and Razvan Pascanu. Trends in Cognitive Sciences, 2020.
8. **Towards Continual Reinforcement Learning: A Review and Perspectives** by Khimya Khetarpal, Matthew Riemer, Irina Rish, Doina Precup. arXiv:2012.13490, 2020.
9. A **Wholistic View of Continual Learning with Deep Neural Networks**: Forgotten Lessons and the Bridge to Active and Open World Learning by Martin Mundt, Yong Won Hong, Iuliia Pliushch and Visvanathan Ramesh. arXiv, 32, 2020.

.... More here: https://github.com/ContinualAI/continual-learning-papers#review-papers-and-books

Join our Pervasive AI Lab!



Teaching & Supervision

Research

Spin-off

Consultancy

The lab is in Pisa, Italy! Feel free to visit and get in touch with us anytime! Official website: [Pervasive AI Lab (unipi.it)]

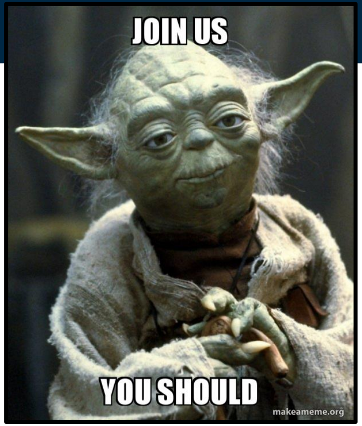
CL @ PAI Lab

At the **PAI Lab** we design and implement **deep continual learning algorithms** for enabling the next generation AI systems and study their applications to real-world problems. In particular, we are interested in:

- *Unsupervised / Self-Supervised/ Weekly/ Semi-Supervised Continual Learning*
- *Continual Sequence Learning*
- *Neuroscience-Inspired Continual Learning*
- *Continual Reinforcement Learning*
- *Continual Learning R&D Frameworks & Tools*
- *Continual Robot Learning*
- *Continual learning on the Edge*
- *Distributed Continual Learning*
- *Real-World Continual Learning Applications*
- *...and much more!*

Team

- **Davide Bacciu** – Associate Professor
- **Vincenzo Lomonaco** – Assistant Professor
- **Claudio Gallicchio** – Assistant Professor
- **Antonio Carta** – Post-Doc
- **Andrea Cossu** – PhD Student
- **Rudy Semola** – PhD Student
- **Michele Resta** – PhD Student
- **Valerio De Caro** – PhD Student
- **Hamed Hemati** – PhD Student (co-supervised with Damian Borth at University of St. Gallen)



The lab is in Pisa, Italy! Feel free to visit and get in touch with us anytime! Official website: Pervasive AI Lab (unipi.it)

Do you have any questions?

vincenzo.lomonaco@unipi.it
vincenzolomonaco.com
University of Pisa

THANKS