Continual Learning: On Machines that can Learn Continually

Official Open-Access Course @ University of Pisa, ContinualAI, AIDA

Lecture 8: Frontiers in Continual Learning

Vincenzo Lomonaco
University of Pisa & ContinualAI
vincenzo.lomonaco@unipi.it

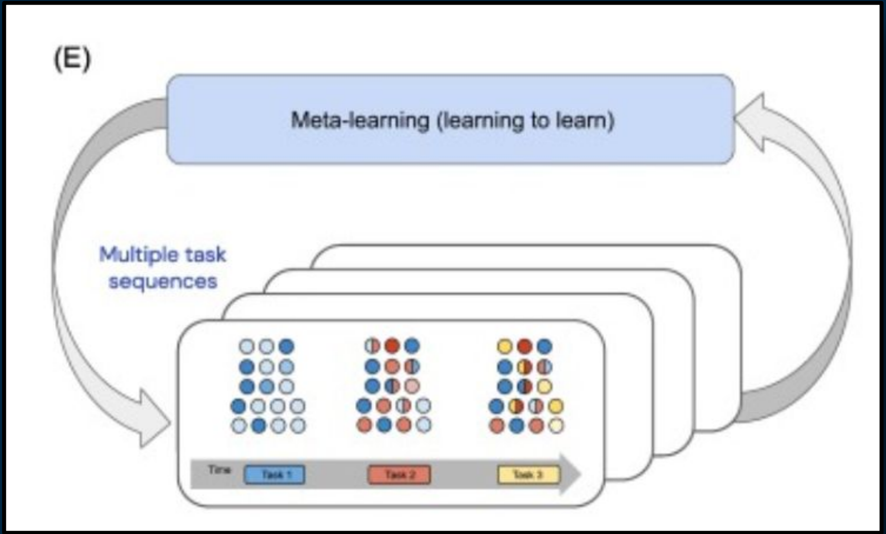
TABLE OF CONTENTS

Advanced Topics & Future Directions

Meta Learning & Continual Learning

Difference in Focus

- **Meta-Learning** is about "learning to learn"
- **Continual Learning** is about learning reusable representations from non-stationary data
- Two main categories:
  - **Meta Continual-Learning**
  - **Continual Meta-Learning**

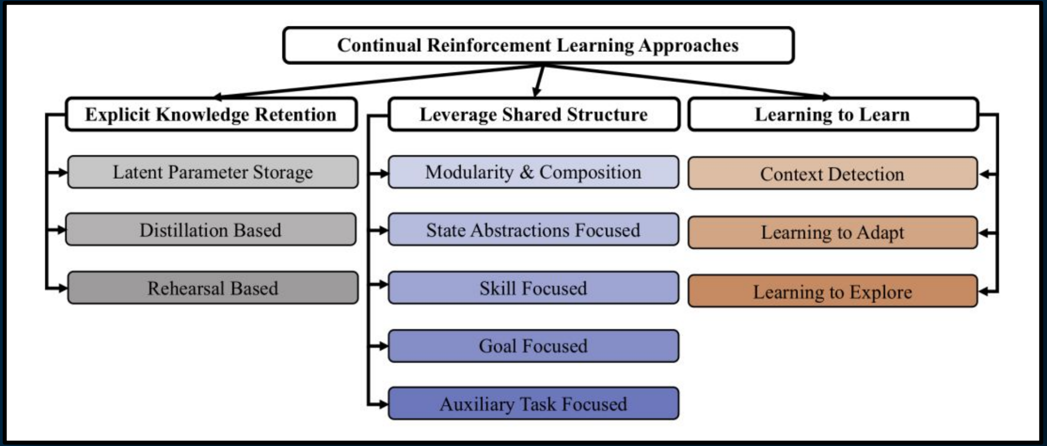


Embracing Change: Continual Learning in Deep Neural Networks, Trends in Cognitive Science, 2020

Continual Reinforcement Learning

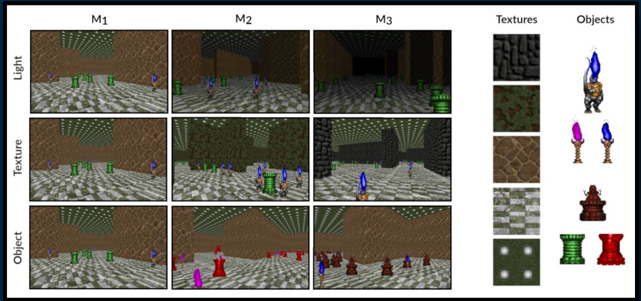
Difference in Focus

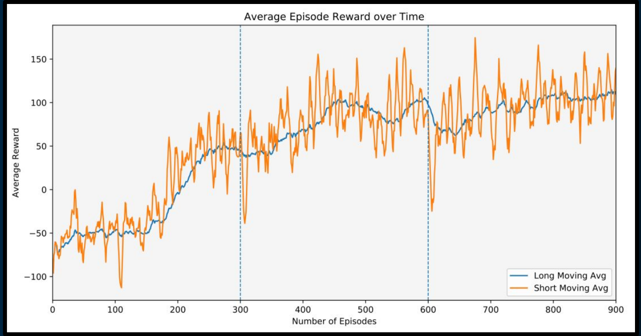
- **Reinforcement Learning** is about "learning from (sparse) rewards"
- **Continual Learning** is about learning reusable representations from non-stationary data
- Quite **orthogonal objectives** but some **shared constraints** (single agent view, non-stationary envs, sample bias, etc..)



Towards Continual Reinforcement Learning: A Review and Perspectives, Khetarpal et al, 2020.
Continual Reinforcement Learning in 3D Non-stationary Environments, Lomonaco et al, 2020.

Continual Reinforcement Learning in 3D Non-stationary Environments



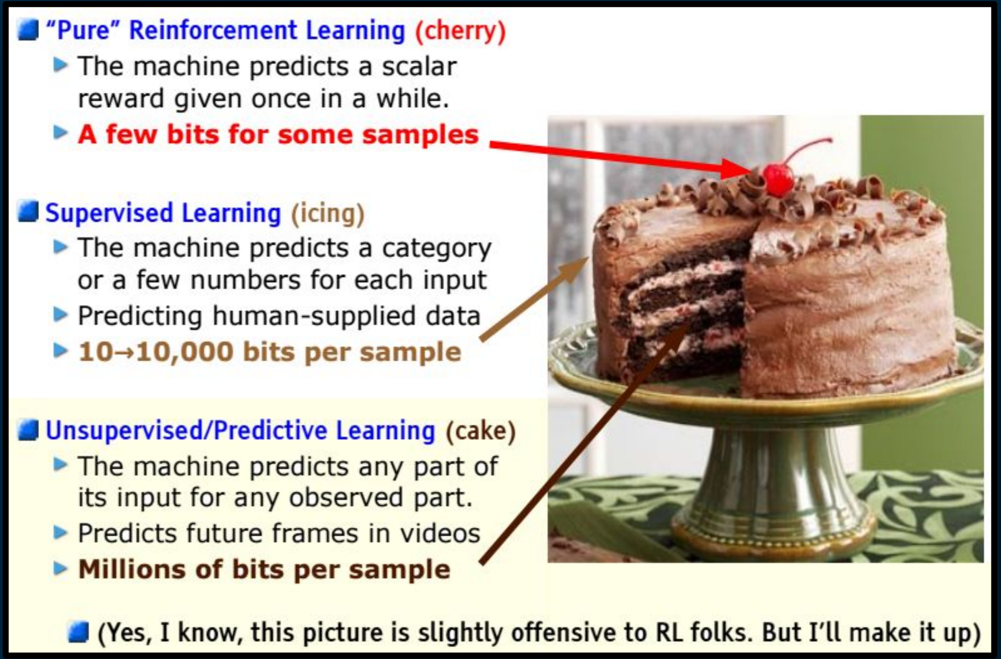


[Towards Continual Reinforcement Learning: A Review and Perspectives](), Khetarpal et al, 2020.
[Continual Reinforcement Learning in 3D Non-stationary Environments](), Lomonaco et al, 2020.

Continual Unsupervised Learning

Ideal Paradigm to Combine with CL

- **No Continual Labeling**
- **Less Bias**
- Why this is still not the case?
    - ***Changing the paradigm***: More Data, Less Supervision
    - **Less impactful applications** (for now)



Y. LeCun, NeurIPS 2016

Continual Unsupervised Representation Learning

Key Ideas

- **Fully Generative Approach**
- **y** can be interpreted as representing some **discrete clusters** in the data
- **Mixture of Gaussian with Dynamic Expansion**
- ***Difficult to scale***: tested only on MNIST and Omniglot

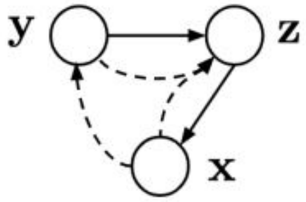


Figure 1: Graphical model for CURL. The categorical task variable **y** is used to instantiate a latent mixture-of-Gaussians **z**, which is then decoded to **x**.

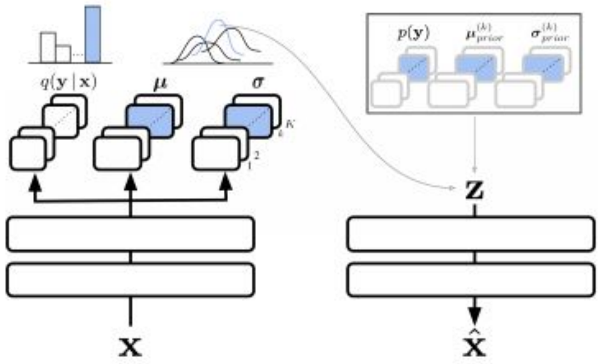


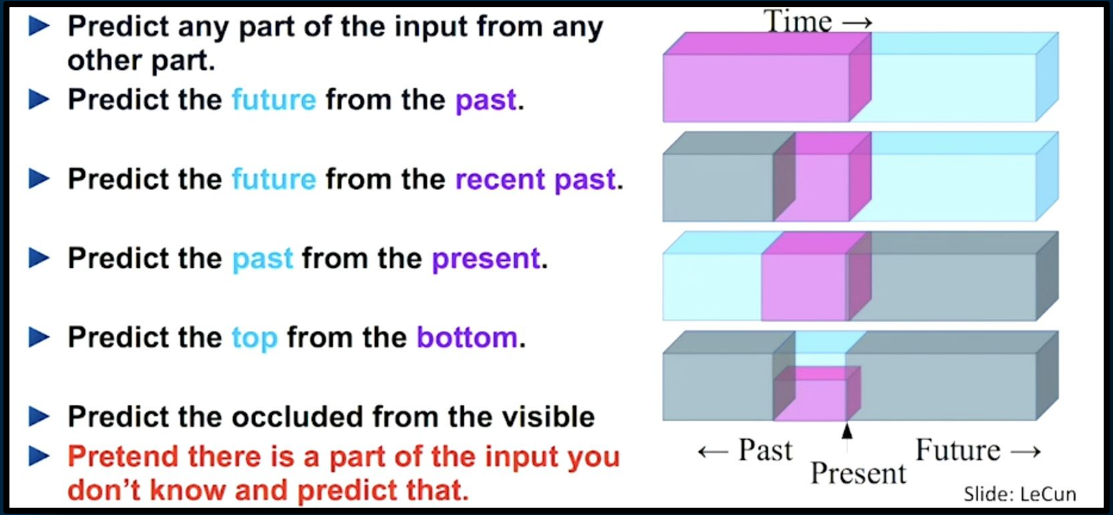
Figure 2: Diagram of the proposed approach, showing the inference procedure and architectural components used.

D. Rao et al. ***Continual Unsupervised Representation Learning***. NeurIPS 2019.

Continual Unsupervised Learning

Huge Exploration Opportunities

- **Self-Supervised Learning**
- Sequence Learning
- Contrastive Learning
- Hebbian-like Learning
- Active Learning
- Weakly/Semi-Supervised Learning
- Randomized Networks

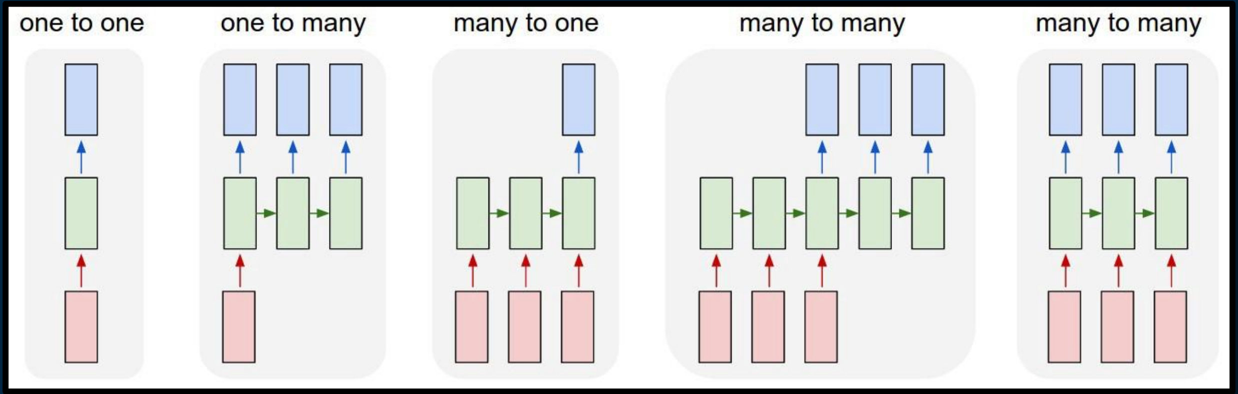


S. Zhang et Al. ***Self-Supervised Learning Aided Class-Incremental Lifelong Learning***. CLVision Workshop Findings at CVPR, 2021.
J. Gallardo et Al. ***Self-Supervised Training Enhances Online Continual Learning***. arXiv, 2021.

Continual Unsupervised Learning

Huge Exploration Opportunities

- Self-Supervised Learning
- **Sequence Learning**
- Contrastive Learning
- Hebbian-like Learning
- Active Learning
- Weakly/Semi-Supervised Learning
- Randomized Networks

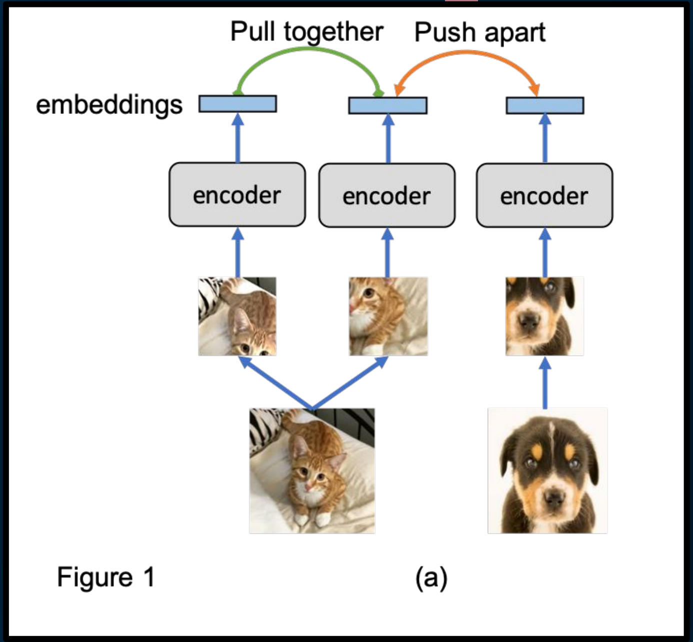


Y. Cui et al. ***Continuous online sequence learning with an unsupervised neural network model***. *Neural Computation, 2016.*
A. Cossu et Al. ***Continual Learning for Recurrent Neural Networks: an Empirical Evaluation***. Elsevier Neural Networks, 2021.
B. Ehret et Al. ***Continual learning in Recurrent Neural Networks.*** ICLR 2021.

Continual Unsupervised Learning

Huge Exploration Opportunities

- Self-Supervised Learning
- Sequence Learning
- **Contrastive Learning**
- Hebbian-like Learning
- Active Learning
- Weakly/Semi-Supervised Learning
- Randomized Networks



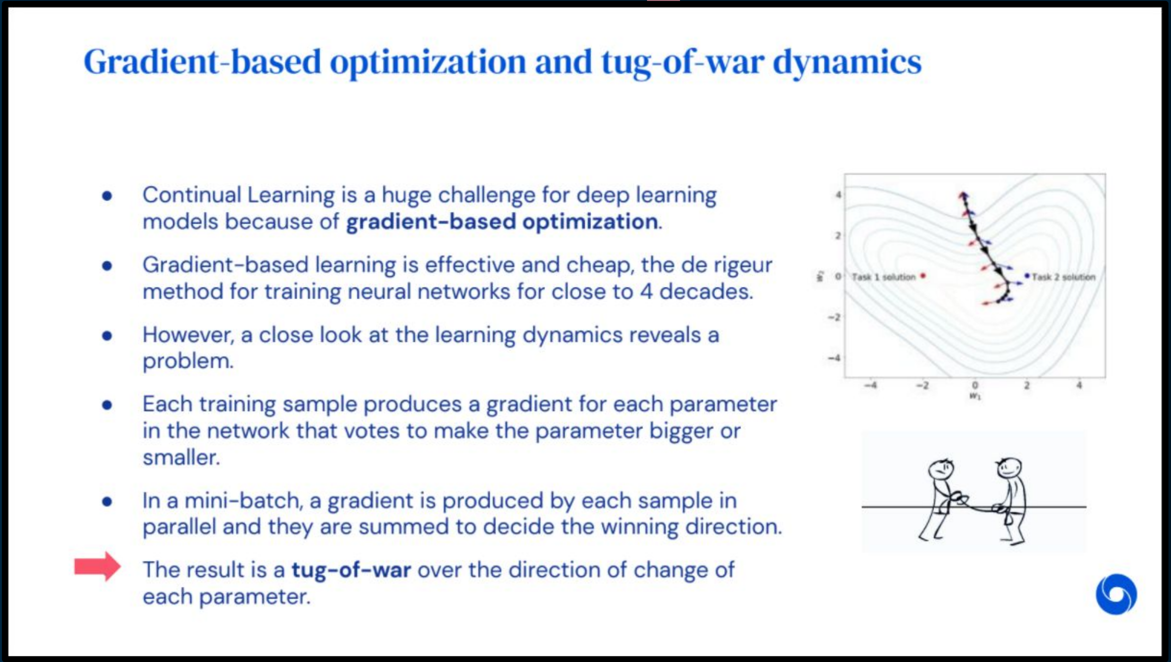
Junnan Li, 2020

M. Zheda et al. ***Supervised Contrastive Replay: Revisiting the Nearest Class Mean Classifier in Online Class-Incremental Continual Learning***. CLVision Workshop at CVPR 2021.
C. Hyuntak et al. ***CO2L: Contrastive Continual Learning***. arXiv, 2021.

Continual Unsupervised Learning

Huge Exploration Opportunities

- Self-Supervised Learning
- Sequence Learning
- Contrastive Learning
- **Hebbian-like Learning**
- Active Learning
- Weakly/Semi-Supervised Learning
- Randomized Networks



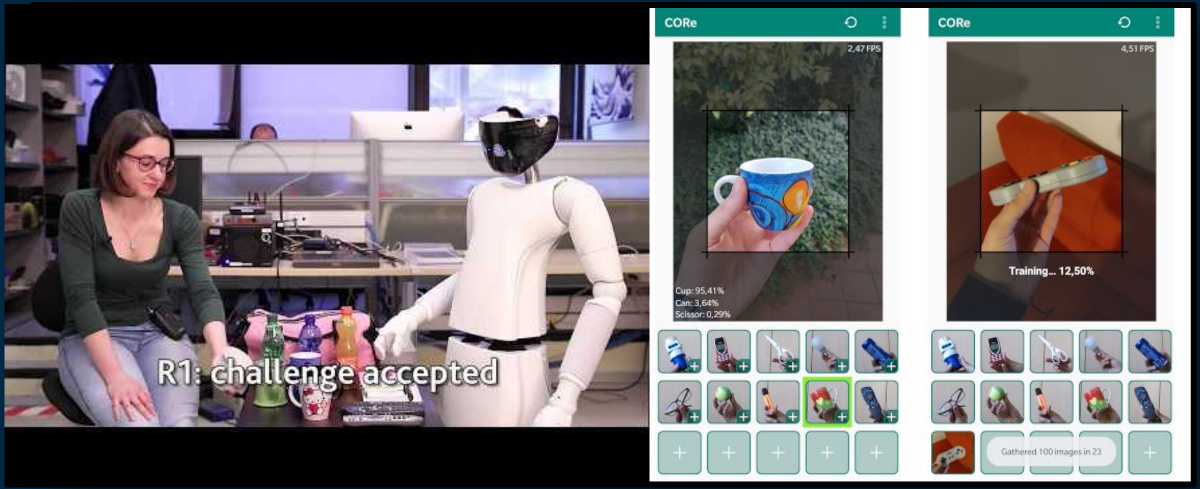
R. Pascanu, 2021

G. Parisi et al. ***Lifelong learning of spatiotemporal representations with dual-memory recurrent self-organization***. Frontiers in neurorobotics, 2018.
P. Bashivan et al. ***Continual learning with self-organizing maps***. CL Workshop at NeurIPS 2018

Continual Unsupervised Learning

Huge Exploration Opportunities

- Self-Supervised Learning
- Sequence Learning
- Contrastive Learning
- Hebbian-like Learning
- **Active Learning**
- Weakly/Semi-Supervised Learning
- Randomized Networks

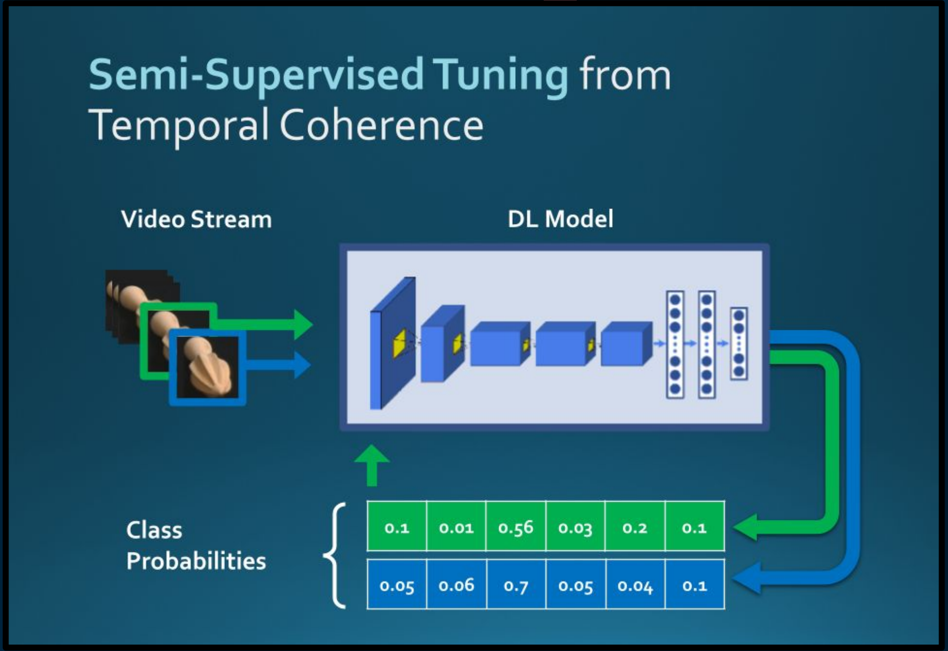


L. Pellegrini et al. ***Continual Learning at the Edge: Real-Time Training on Smartphone Devices***. ESANN, 2021.
R. Camoriano et al. ***Incremental robot learning of new objects with fixed update time***. ICRA, 2017.

Continual Unsupervised Learning

Huge Exploration Opportunities

- Self-Supervised Learning
- Sequence Learning
- Contrastive Learning
- Hebbian-like Learning
- Active Learning
- **Weakly/Semi-Supervised Learning**
- Randomized Networks



Lomonaco V. and Maltoni D. ***Semi-Supervised Tuning from Temporal Coherence***. ICPR 2016.
L. Wang et al. **Ordisco: Effective and efficient usage of incremental unlabeled data for semi-supervised continual learning**. CVPR 2021.

Continual Unsupervised Learning

Huge Exploration Opportunities

- Self-Supervised Learning
- Sequence Learning
- Contrastive Learning
- Hebbian-like Learning
- Active Learning
- Weakly/Semi-Supervised Learning
- **Randomized Networks**

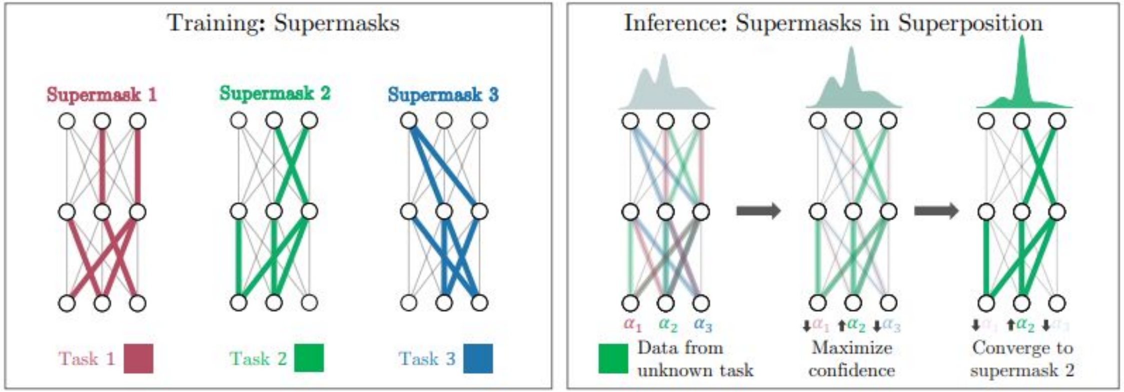


Figure 1: **(left)** During training SupSup learns a separate supermask (subnetwork) for each task. **(right)** At inference time, SupSup can infer task identity by superimposing all supermasks, each weighted by an α_i, and using gradients to maximize confidence.

M. Wortsman, 2020

A. Cossu et al. ***Continual Learning with Echo State Networks***. ESANN 2021.
M. Wortsman et al. ***Supermasks in superposition***. NeurIPS 2020.

Continual Unsupervised Learning

Other relevant works in this area

- A. Bertugli et al. ***Few-Shot Unsupervised Continual Learning through Meta-Examples***. Workshop on Meta-Learning at NeurIPS 2020.
- I. Muñoz-Martín et al. ***Unsupervised learning to overcome catastrophic forgetting in neural networks***. IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, 2019.
- L. Caccia et al. ***SPeCiaL: Self-Supervised Pretraining for Continual Learning***, arXiv 2021.
- W. Sun et al. ***ILCOC: An Incremental Learning Framework Based on Contrastive One-Class Classifiers.*** CLVision Workshop at CVPR 2021.
- J. He et al. ***Unsupervised Continual Learning Via Pseudo Labels.*** arXiv 2020.
- S. Khar et al. ***Unsupervised Class-Incremental Learning through Confusion***. arXiv 2021.

Impact on Sustainable AI

Sustainable AI Principles

General Principles

- **Accuracy** & **Robustness**
- **Explainability**, **Transparency** & **Accountability**
- **Bias** & **Fairness**
- **Privacy** & **Security**
- **Human**, **Social** and **Environmental Wellbeing**

L. Royakkers et al. ***Societal and ethical issues of digitization***. Ethics and Information Technology, 2018.
B.D. Mittelstadt et al. ***The ethics of algorithms: Mapping the debate***. Big Data & Society, 2016.
A. Jobin et al. ***The global landscape of AI ethics guidelines***. Nature Machine Intelligence, 2019.
Cossu et al. ***Sustainable Artificial Intelligence through Continual Learning***. CAIP 2021.
https://www.aiforpeople.org/ethical-ai/

Continual Learning Impact

...On each Principle:

- Accuracy & Robustness ➞ **Robustness & Autonomy, Continual & Fast Improvement**
- Bias & Fairness ➞ **CL as the new Agile: Bias & Fairness Patches**
- Privacy & Security ➞ **Security Patches**
- Human, Social and Environmental Wellbeing ➞ **improved efficiency & scalability: less energy consumption, CO2 emission; sustainable & "progressive" by design**
- Explainability, Transparency & Accountability ➞ **Neuroscience-grounded, Human-centered AI**

L. Royakkers et al. ***Societal and ethical issues of digitization***. Ethics and Information Technology, 2018.
B.D. Mittelstadt et al. ***The ethics of algorithms: Mapping the debate***. Big Data & Society, 2016.
A. Jobin et al. ***The global landscape of AI ethics guidelines***. Nature Machine Intelligence, 2019.
Cossu et al. ***Sustainable Artificial Intelligence through Continual Learning***. CAIP 2021.
https://www.aiforpeople.org/ethical-ai/

Open Questions

Open Questions (1/2)

1. Is it possible to learn **robust**, **deep representations continually**?
2. Are currently addressed **scenarios** and **eval metrics enough**?
3. What is the right **level of supervision**?
4. How to know **what to forget** and **what to remember**?
5. What's the relationship with **concept drift**?
6. Is **replay** a research direction worth pursuing?
7. Is **computation** more important than **memory**?
8. Is **gradient descent** the right algorithm to learn continually?
9. **Continual Meta-Learning** & **Meta-Continual Learning**: what's the right relationship?
10. What is the relationship with **Sequence** and **Continual Learning**?

N. Díaz-Rodríguez et al. **Don't Forget, There is More than Forgetting: new Metrics for Continual Learning. CL Workshop at NeurIPS 2018.**
A Prabhu. **Gdumb: A simple approach that questions our progress in continual learning. ECML, 2020.**

Open Questions (2/2)

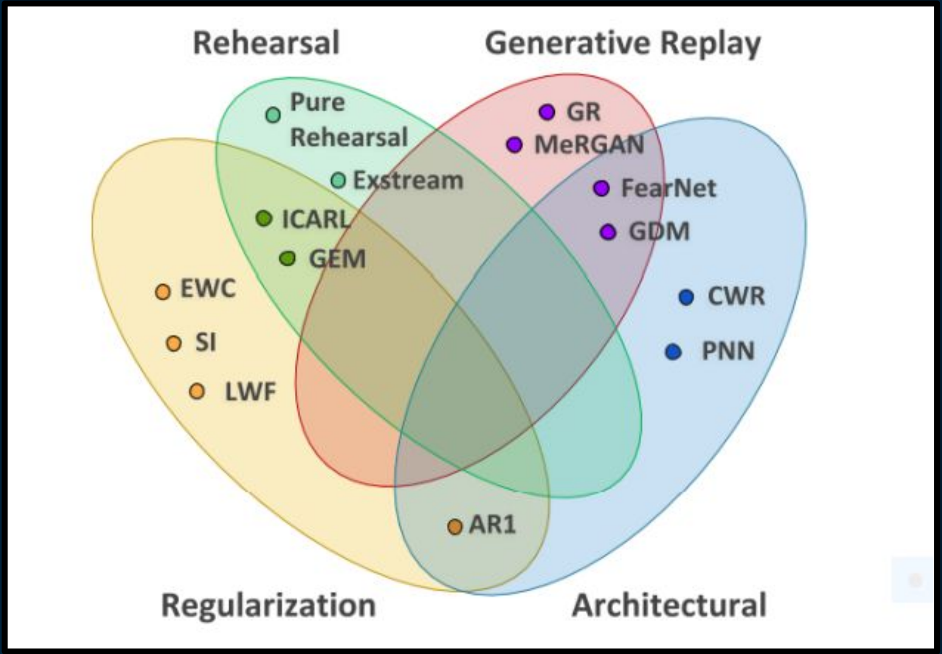
1. Is **curiosity** important for continual learning?
2. What about **Curriculum Learning**?
3. **Compositionality** is a key aspect of human intelligence: what to expect for CL Systems?
4. **Self-Reflection***: accuracy of learned functions, given only unlabeled data?
5. **Self-reflection** that can detect every possible shortcoming (called impasse) of the agent*
6. (External) **Knowledge and Reasoning***

...and much more!

*T. Mitchell and P. Talukdar. ***Never-Ending Learning***. Tutorial at ICML 2019.
J.A. Mendez et al. ***Lifelong learning of compositional structures***. ICLR 2021.

On the Future of CL (Short-Medium Term)

1. **More Natural Scenarios**
   - **Domain**, **Task** and **Class-Incremental** are not enough.
   - Longer streams of "*experiences*".
   - More metrics, focus on scalability.
2. **Move towards unsupervised training**
   - Mostly **Semi-Supervised**, **Self-Supervised** and **Sequence Learning**.
3. **Hybrid Continual Learning Strategies**
4. **Continual Learning Applications**

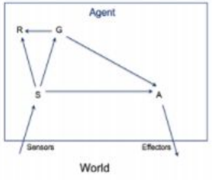


T. Lesort et al. ***Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges***. Information Fusion, 2020.
G.I. Parisi et al. ***Online Continual Learning on Sequences.*** Studies in Computational Intelligence, 2020.

On the Future of CL (Long-Term)

1. Fundamentally a question of **agent architecture***
2. **Two main paths for (deep) CL**
   a. Neuroscience-Inspired
   b. Distributed Continual Learning

What should a theory of Learning Agents answer?



might model learning agent A as tuple <S,E,M,F,G,L>

- S = sensors
- E = effectors
- F = set of functions
- M = set of memory units
- G = graph specifying data flow among F, M, S, E
- L = learning mechanism

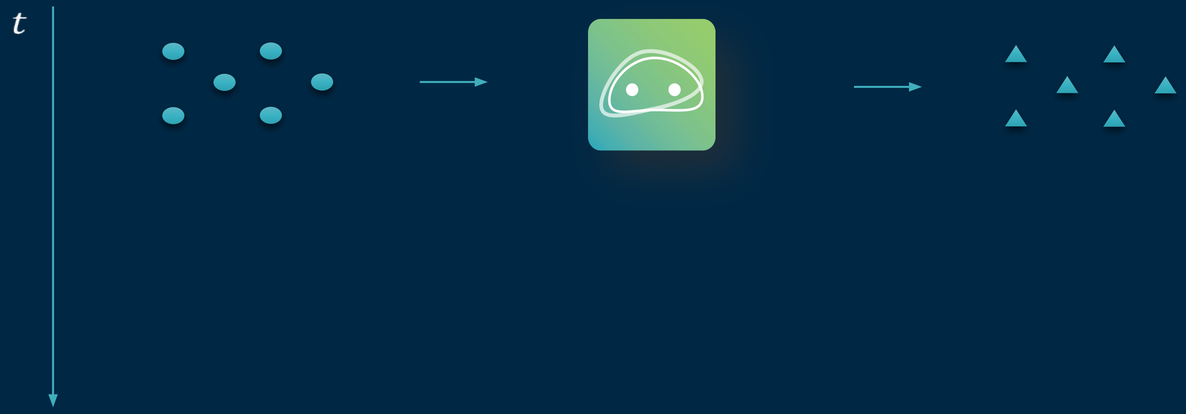
might model L as another agent $L = \langle S_L, E_L, M_L, F_L, G_L \rangle$

- where S_L, E_L sense and act on Agent, especially its F, M, G

*T. Mitchell and P. Talukdar. **Never-Ending Learning**. Tutorial at ICML 2019.

Distributed
Continual Learning

Continual Learning (CL)



X_1 $\quad$ $f_\theta : X_1 \to Y_1$ $\quad$ Y_1

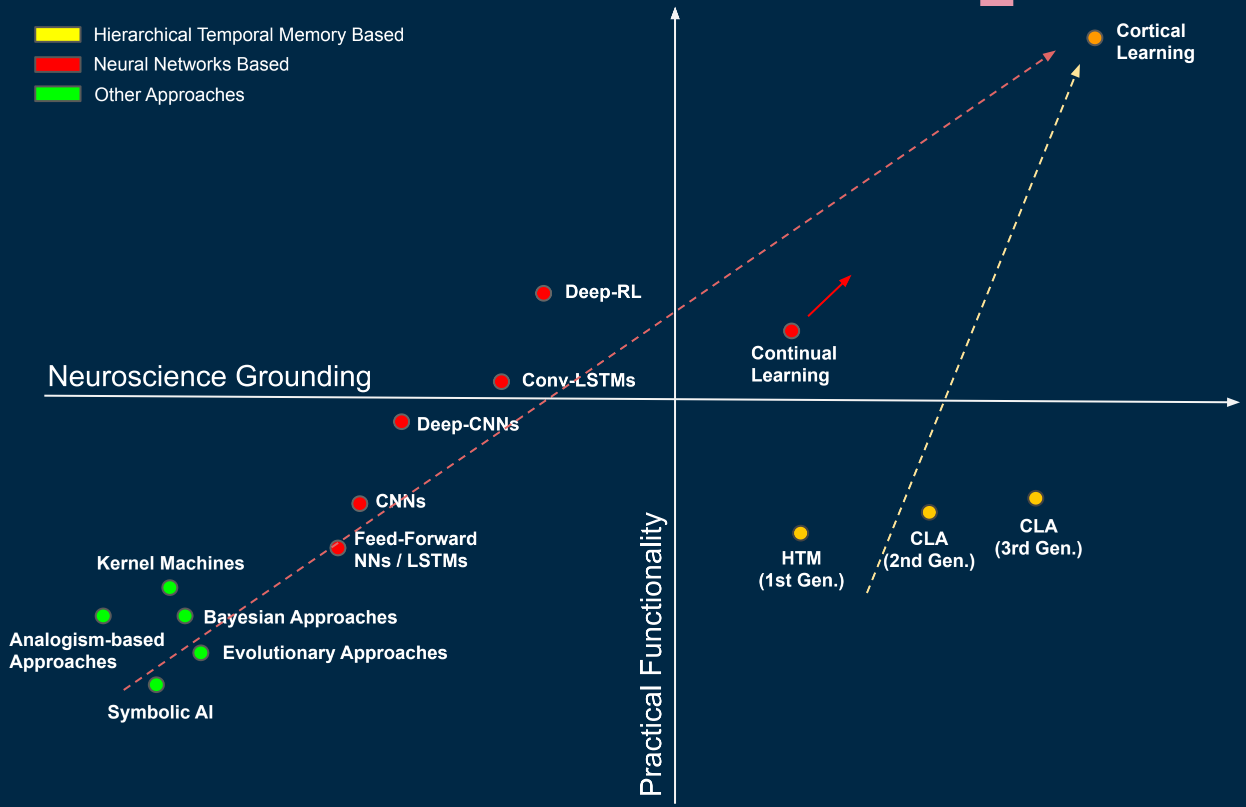
Continual Learning (CL)

X_2

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

Y_2

Biologically-Inspired Continual Learning



Agent-Centric Learning

A Continual Learning algorithm trains a **single agent**

Example: a robot that learns to grasp different objects over time.

Desiderata

- **Replay-free CL**
- **Limited computational** resources
- **Task-free CL**
- **Online CL**

A Fork in the CL Road

Distributed Continual Learning

A Continual Learning algorithm trains a **single agent** (as before).

Example: a fleet of independent robots that learn to grasp different objects over time.

Desiderata

- **Reuse of expert knowledge**
- **Efficient** and **distributed** learning
- **Independent agents** (unlike federated learning)
- **Privacy** (at will)

Ex-Model Continual Learning (ExML)

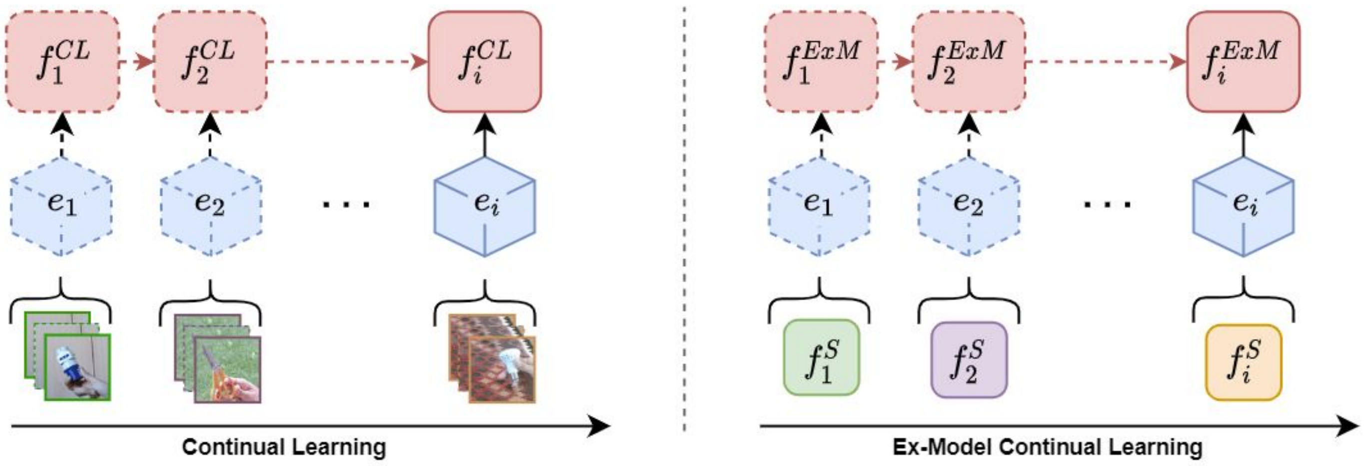


Figure 1. Ex-model Continual Learning. The stream is composed of expert models, without access to the original data.

Ex-Model: Continual Learning from a Stream of Pre-Trained Models. A. Carta, A. Cossu, V. Lomonaco, D. Bacciu. 2021.

Issues and Missed Opportunities

- **Expert models**: continual learning should reuse knowledge from expert agents (i.e. the model's parameters), such as local personalized models or large pretrained models
- **Distributed learning**: agents in a distributed environment should be able to learn independently and to share knowledge efficiently at the same time (sharing the models).
- **Sample efficiency**: learning from raw data may be in-efficient due to noise and redundancy inherent to high-dimensional perceptual data.
- **Privacy**: sharing knowledge between agents must be limited by privacy constraints, and each agent should be allowed to set its privacy constraints.

Ex-Model Continual Learning (ExML)

- The Continual Learning model **never gets access to the original data**, only the expert's model.
- We **cannot maintain all the experts in memory**.
- We are allowed to maintain **a memory of generated / out-of-distribution examples**.
- This approach open the doors to the **efficient on-demand integrations of "neural skills".**
- **Privacy by design**: we never share private data which can stay on the source device.

ExML Scenario The objective of the ExML scenario is to continuously update a model f_i^{ExM} whenever a new expert f_i^S becomes available. Notice that the loss $\mathcal{L}_{exp}(f_i^{ExM}, \mathcal{D}_{train}^i)$ cannot be evaluated since we do not have access to the original data. Since the stream of models may be unbounded, training strategies must be scalable up to a large number of experts. Therefore, ex-model algorithms cannot keep in memory all the previous experts. As a result, there are two constraints in an ExML scenario: *lack of access to the original data and limited computational resources.*

Overall, an ExML algorithm $\mathcal{A}^{ExM}$ is a function with the following signature:

$$\mathcal{A}^{ExM} : \langle f_{i-1}^{ExM}, f_i^S, \mathcal{M}_{i-1}^{ex}, t_i \rangle \rightarrow \langle f_i^{ExM}, \mathcal{M}_i^{ex} \rangle, \quad (4)$$

where f_i^{ExM} is the current model, f_i^S the current expert from the stream, $\mathcal{M}_{i-1}^{ex}$ is a set of samples from out-of-distribution data or synthetically generated and currently available to the model (Section 4), and t_i the task label information. Again, notice that task labels are optional and they may not be available in many scenarios. The objective of ex-model algorithms is to minimize Eq. 2, the loss over the original (and unavailable) data stream.

Ex-Model: Continual Learning from a Stream of Pre-Trained Models. A. Carta, A. Cossu, V. Lomonaco, D. Bacciu. 2021.

Ex-Model Distillation

- **Data-free Distillation** as an elegant way to merge two pre-trained models
- **We assume no access to the original data** as we should stay agnostic w.r.t. to the experts training constraints
    - **Synthetic data generation via optimization** or **auxiliary data**
- Each expert's model can be trained separately, completely **agnostic w.r.t. the continual learning scenario**

Algorithm 1 Ex-Model Distillation

Require: Stream of pretrained experts S and a continually learned model f^{ExM}.

$\mathcal{M}_0^{ex} \leftarrow \{\}$ $\qquad \triangleright$ empty buffer

for f_i^S in S **do**

$\quad \mathcal{D}_i^{ex} \leftarrow \mathcal{A}^{gen}(f_i^S, \frac{N}{i})$

$\quad \tilde{\mathcal{M}}_{i-1}^{ex} \leftarrow subsample(\mathcal{M}_{i-1}^{ex})$

$\quad \mathcal{M}_i^{ex} \leftarrow \tilde{\mathcal{M}}_{i-1}^{ex} \cup \mathcal{D}_i^{ex}$

$\quad$ **for** k in $1, \dots, n_{iter}$ **do** $\qquad \triangleright$ Knowledge Distillation

$\quad\quad \langle \boldsymbol{x}^k, \boldsymbol{y}^k \rangle \leftarrow sample(\mathcal{M}_i^{ex})$

$\quad\quad \boldsymbol{y}^{curr} \leftarrow f^{ExM}(\boldsymbol{x}^k)$

$\quad\quad \tilde{\boldsymbol{y}} \leftarrow get\_target(\boldsymbol{x}^k)$ $\qquad \triangleright$ Eq. 8

$\quad\quad L \leftarrow \mathcal{L}_{ED}(\boldsymbol{y}^k, \tilde{\boldsymbol{y}}^k, y^k)$

$\quad\quad$ do SGD step on L

$\quad$ **end for**

end for

<u>**Ex-Model: Continual Learning from a Stream of Pre-Trained Models**</u>. A. Carta, A. Cossu, V. Lomonaco, D. Bacciu. 2021.

Some Preliminary Results

Table 2. Stream accuracy computed on the test set for MNIST and CIFAR10 continual learning scenarios. Ensemble methods' results are not shown for joint scenarios because ensembling is not necessary when there is a single model.

	Ex-model scenario	MNIST		CIFAR10		
		Joint	NC	Joint	NC	MT
Oracle	✗	93.71±0.28	99.42±0.19	87.37±1.11	96.58±0.86	96.58±0.86
Ensemble Avg.	✗	–	33.40±4.74	–	51.85±2.37	–
Min Entropy	✗	–	39.41±5.27	–	52.03±2.67	–
Param. Avg.	✓	–	20.11±0.97	–	10.00±0.00	51.85±2.37
Model Inversion ED	✓	93.09±1.43	43.23±3.00	64.55±3.25	17.40±3.96	61.71±7.52
Data Impression ED	✓	92.12±0.88	36.05±6.74	52.64±5.82	24.70±6.85	61.15±3.92
Aux. Data ED	✓	89.35±0.18	35.48±6.35	76.94±2.68	41.35±5.83	60.72±3.70

Ex-Model: Continual Learning from a Stream of Pre-Trained Models. A. Carta, A. Cossu, V. Lomonaco, D. Bacciu. 2021.

Challenges

- Model distillation is quite a **complex task without the original data**
- **Generating training samples** without the original data **is very challenging** (especially in terms of quality and diversity)
- **What about efficiency?** Data-free knowledge distillation can be computationally intensive.
- Some efficient **selective pruning + ensembling** methods may be interesting to study
- More specific distributed scenarios may allow a **simplification of the problem's constraints**. For example, a shares training protocol for the experts, or access to a pretrained generator, which may significantly reduce the overall problem complexity

Opportunities

- **Federated Learning** requires **frequent sync** (large bandwidth) and a **shared single stakeholder training protocol**
- Federated learning assumes **homogeneity in the model architecture**
- Federated Learning is **not designed to handle non-stationarity**
- Federated Learning can be seen a **constrained version of Ex-Model Continual Learning**, where the learning agents are controlled by a centralized protocol and synchronized frequently.
- **ExML**: opening the path for a **Marketplace of Neural Skills** for AI systems
- **ExML**: a new exciting path for **Distributed Continual Learning** machines!

Continual Sequence Learning

SCUOLA NORMALE SUPERIORE

UNIVERSITÀ DI PISA

Continual Sequence Learning

learning from a dynamic stream of temporally-correlated patterns

$$\mathbf{x}^1 \xrightarrow{\sim} \mathbf{x}^2 \longrightarrow \mathbf{x}^3 \longrightarrow \mathbf{x}^4 \xrightarrow{\sim} \mathbf{x}^5 \longrightarrow \cdots\cdots$$

Andrea Cossu – Continual Learning course @ Unipi

PhD student @ SNS, Unipi
PAI Lab, ContinualAI, CIML
https://andreacossu.github.io/
andrea.cossu@sns.it

TL;DR

- Temporal correlation is a powerful source of information with many applications
- Continual Sequence Learning opens to new CL scenarios
- Applying CL strategies in recurrent models is not as straightforward as one may think
- NLP is currently the driving field for Continual Sequence Learning

Why should I care?

- *Sequential data processing tasks are widespread*

Stock prediction

Urban mobility

Natural Language Processing

Robot control

Video Processing

Human Activity Recognition

Sequences: challenges and opportunities

Temporal correlation is a (powerful) source of information!

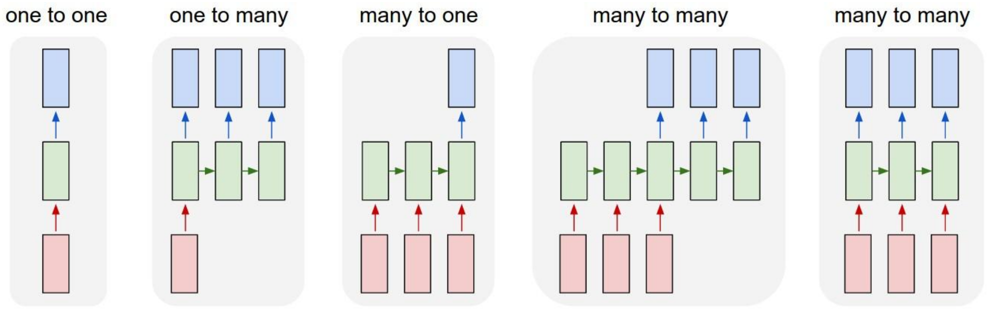
- What is **important** and what can be **discarded**?
- What do you **expect** to receive next?
  - Can you lower the amount of **supervision**?
- Replay!

Temporal correlation introduces new challenges

- Long/Short-term **memory**
- Computational **efficiency** (e.g. RNNs)

Scenarios for Continual Sequence Learning

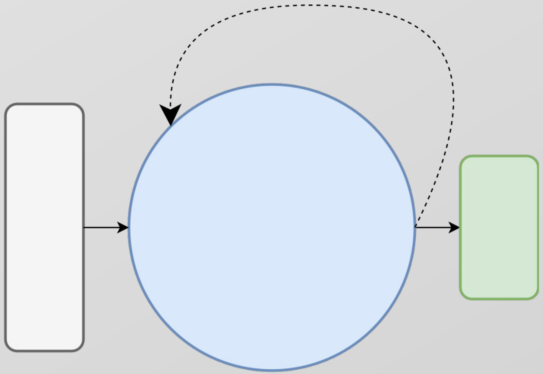
- X-incrementals are still there
- Online / streaming continual sequence learning
  - next-item prediction (unsupervised!)
  - item-to-item prediction (classification/regression at each timestep)
    - We will see another example in the following (NLP)



Andrej Karpathy, The Unreasonable Effectiveness of Recurrent Neural Networks

A brief tour of CL with RNNs

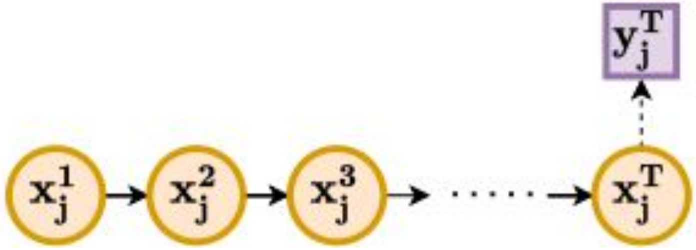
- Few contributions available
- RNNs ad-hoc models for temporal correlations
- Architectural strategies, regularization strategies, …



Continual Learning with Gated Incremental Memories for sequential data processing, A. Cossu, A. Carta, D. Bacciu, IJCNN, 2020.

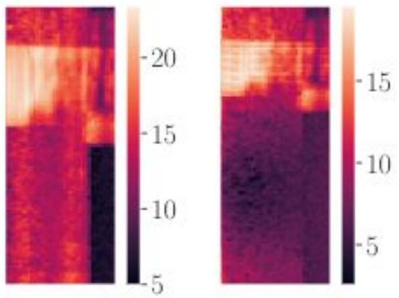
Benchmarks for Continual Sequence Learning

- *Sequence classification tasks*

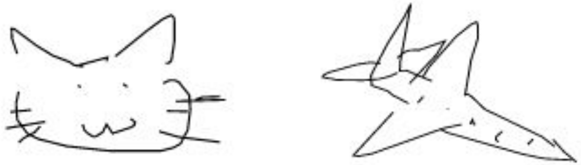


- Split / Permuted MNIST pixel-wise

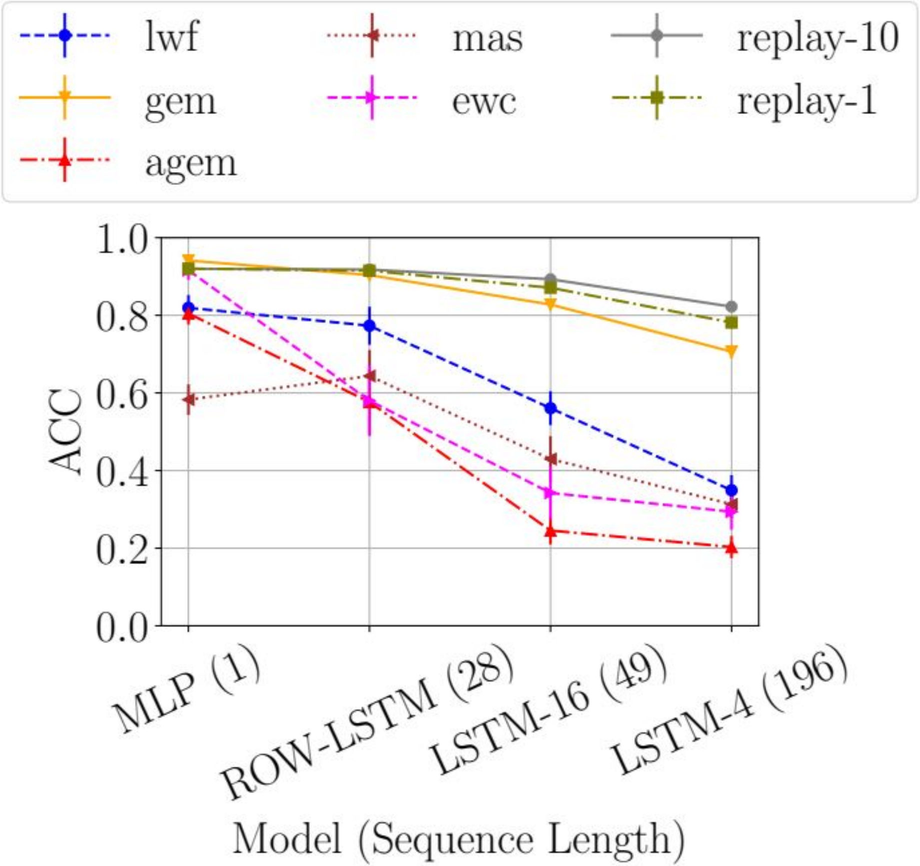
Synthetic Speech Commands



Quick, Draw!



[Continual learning for recurrent neural networks: An empirical evaluation](), A. Cossu, A. Carta, V. Lomonaco, D. Bacciu, *Neural Networks*, 2021.



Continual learning for recurrent neural networks: An empirical evaluation,
A. Cossu, A. Carta, V. Lomonaco, D. Bacciu, *Neural Networks,* 2021.

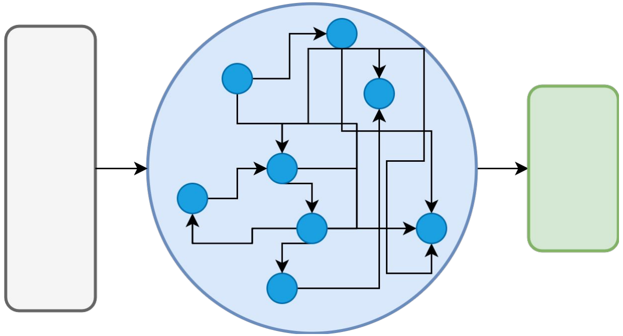
SCUOLA NORMALE SUPERIORE

UNIVERSITÀ DI PISA

Sequence Length effect on forgetting

Echo State Networks and CL

- Untrained recurrent connections
  - you cannot forget, if you are not changing
- Treat reservoir as pretrained model
- Apply CL strategies only on the trained output layer
  - which is often linear
  - allows for the design of simple and efficient strategies
  - Deep Streaming LDA
- Neuromorphic deployment



Continual Learning with Echo State Networks,
A. Cossu, D. Bacciu, A. Carta, C. Gallicchio, V. Lomonaco, ESANN, 2021.

SCUOLA NORMALE SUPERIORE

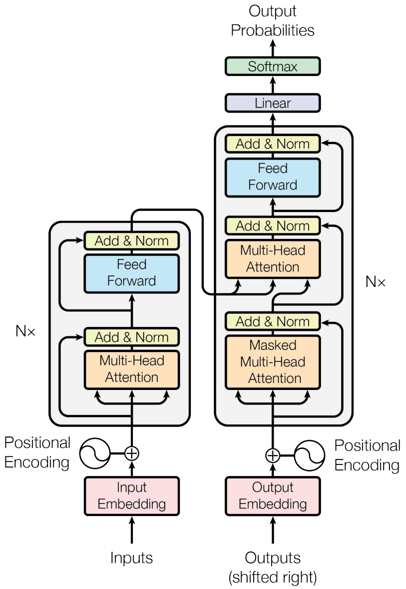
UNIVERSITÀ DI PISA

Trends of NLP in CL

Keep your knowledge updated

NLP is driving the Continual Sequence Learning topic

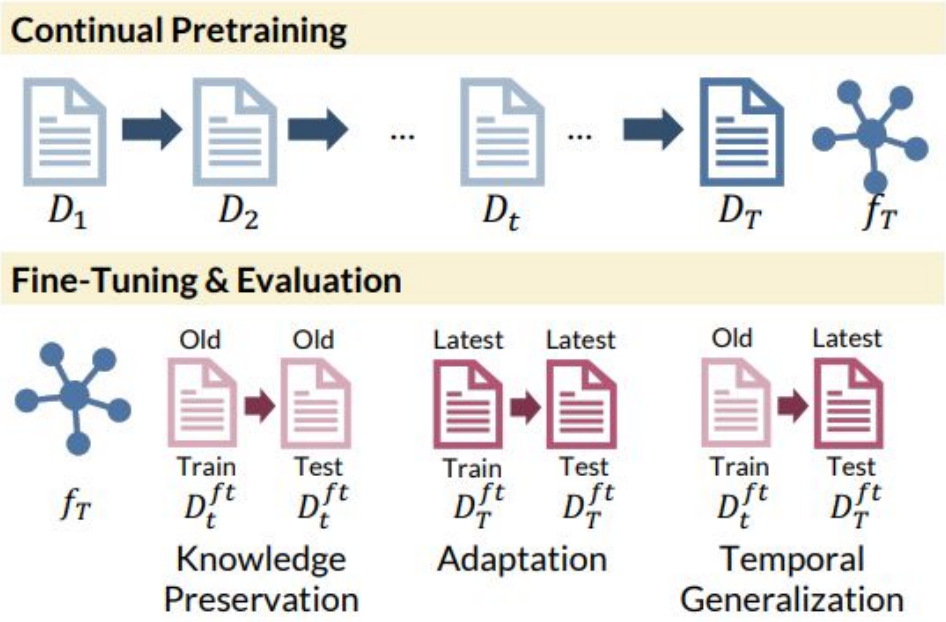
- Transformers: standard *de-facto* in NLP
- Promising scenario: Dynamic Language Modelling
  - Language models can be used to solve many downstream tasks
  - Keep your language model updated
  - Temporal generalization, adapting to new information

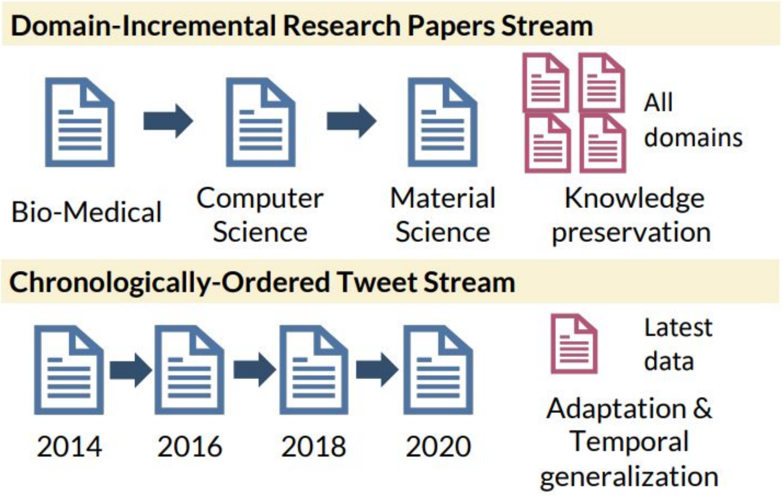


[Attention is All you Need](), A. Vaswani et al, NeurIPS, 2017.

[Mind the Gap: Assessing Temporal Generalization in Neural Language Models](), NeurIPS, 2021. A. Lazaridou et al.,

A Continual Sequence Learning scenario for NLP





[Lifelong Pretraining: Continually Adapting Language Models to Emerging Corpora](), arXiv, 2021
X. Jin et al.

The future ahead

- Move away from traditional CL scenarios
- Understand how to exploit temporal correlation
- Adapt CL objectives – we can forget, can’t we?
- Real-world applications are out there, waiting

https://andreacossu.github.io/
andrea.cossu@sns.it

Conclusions

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**

Next:

Avalanche Dev Day

Do you have any questions?

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

THANKS