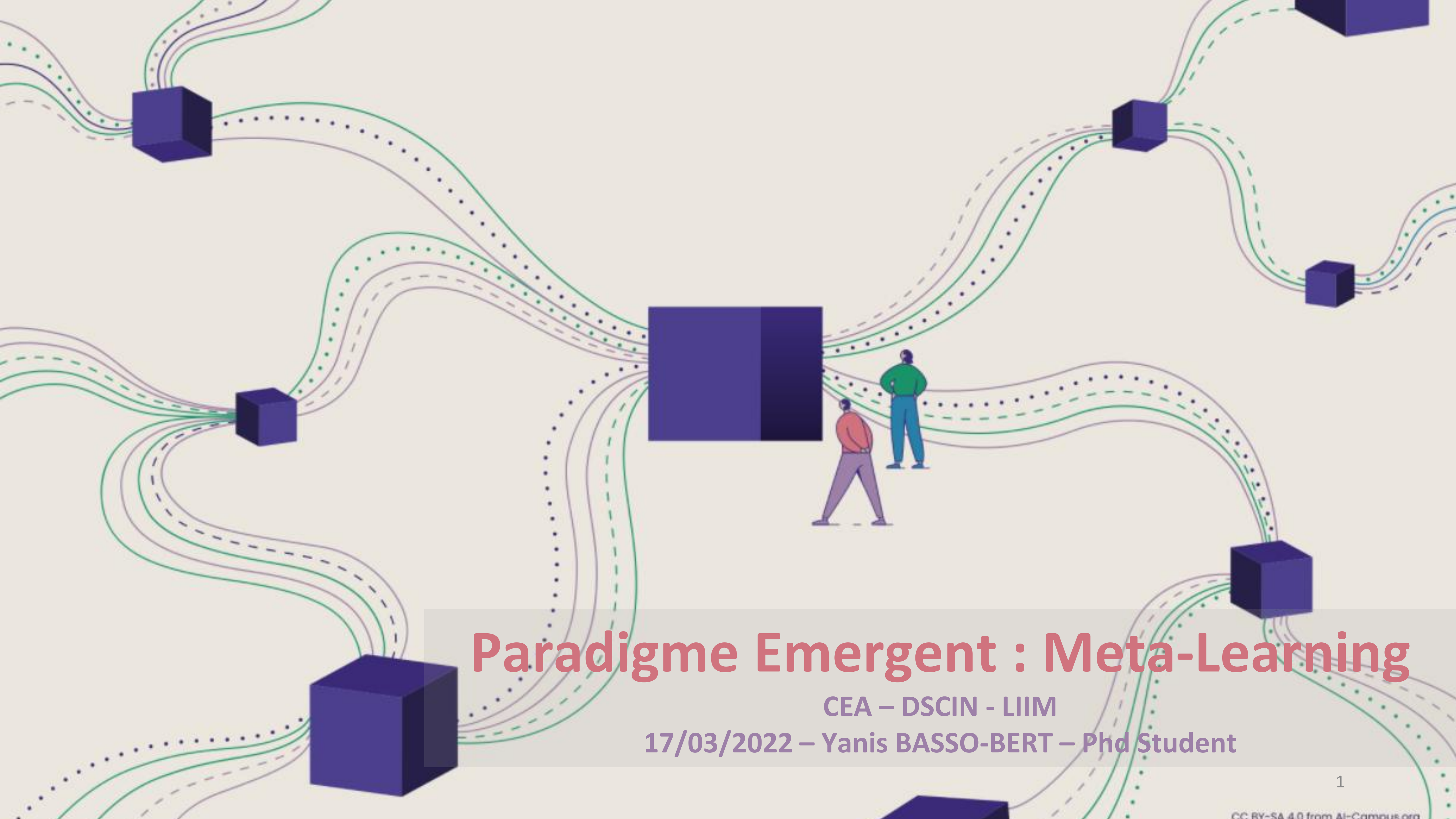


Paradigme Emergent : Meta-Learning

CEA – DSCIN - LIIM

17/03/2022 – Yanis BASSO-BERT – Phd Student

CC BY-SA 4.0 from AI-Campus.org

Trends in Cognitive Sciences

CellPress
REVIEWS

Series: Machine Behavior

Review

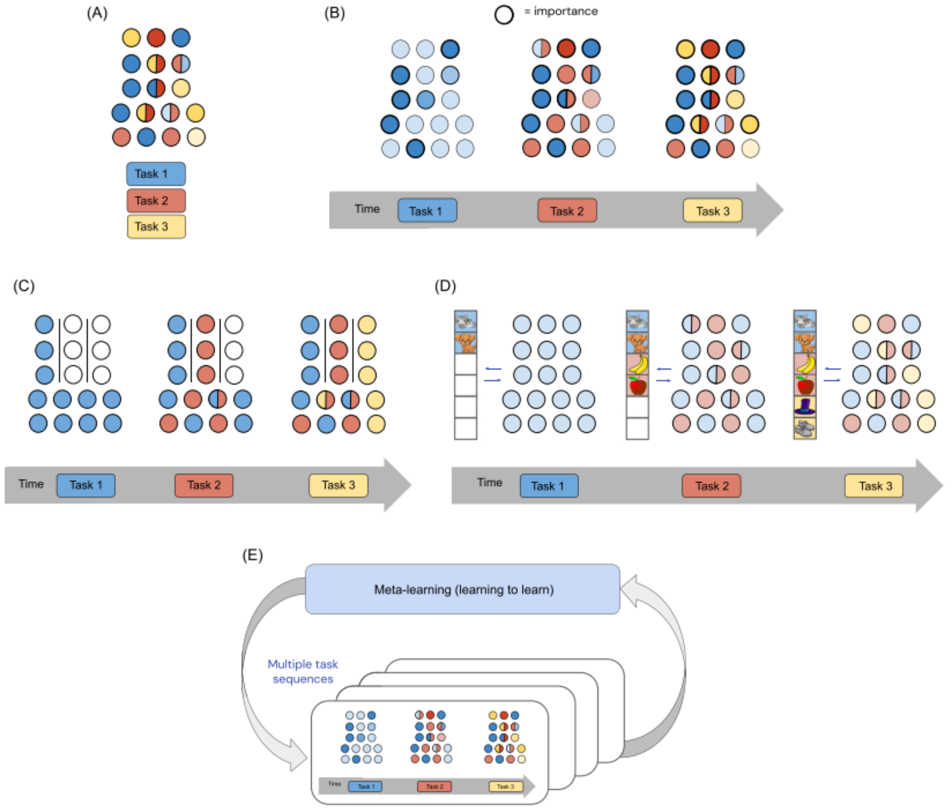
Embracing Change: Continual Learning in Deep Neural Networks

Raia Hadsell,[1,*,@] Dushyant Rao,[1,@] Andrei A. Rusu,[1,@] and Razvan Pascanu[1,@]

Artificial intelligence research has seen enormous progress over the past few decades, but it predominantly relies on fixed datasets and stationary environments. Continual learning is an increasingly relevant area of study that asks how artificial systems might learn sequentially, as biological systems do, from a continuous stream of correlated data. In the present review, we relate continual learning to the learning dynamics of neural networks, highlighting the potential it has to considerably improve data efficiency. We further consider the many new biologically inspired approaches that have emerged in recent years, focusing on those that utilize regularization, modularity, memory, and meta-learning, and highlight some of the most promising and impactful directions.

https://www.cell.com/action/showPdf?pii=S1364-6613%2820%2930219-9

Paradigms for Continual Learning



Meta Learning

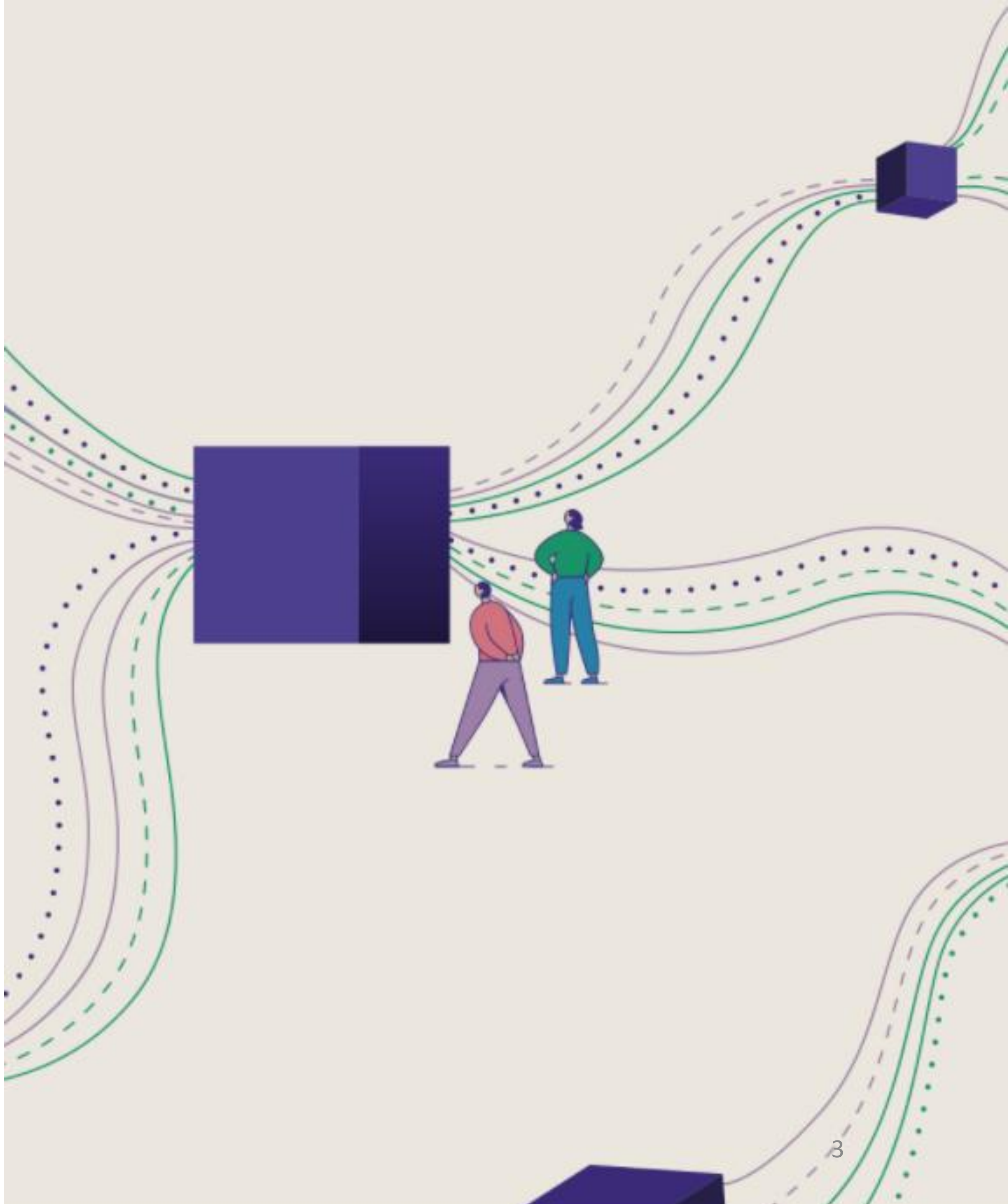
Un concept qui peut faire référence à beaucoup de chose :

- Conception Automatique d’architecture : NAS/AutoML
- Apprentissage d’hyperparamètres
- **Initialisation de réseaux**

Idée : Optimiser un modèle à deux niveaux :
- **Inner-loop :** Boucle d’optimisation au niveau de la tâche : Faire converger un modèle
- **Outer-loop :** Boucle d’optimisation à l’échelle du problème considéré : Améliorer les performance de la première boucle

→ Initialisation apprise
→ Améliorer la convergence
→ Prendre en compte les spécificité d’un scenario

Application :
Few-shot learning / Continual learning /Reinforcement learning

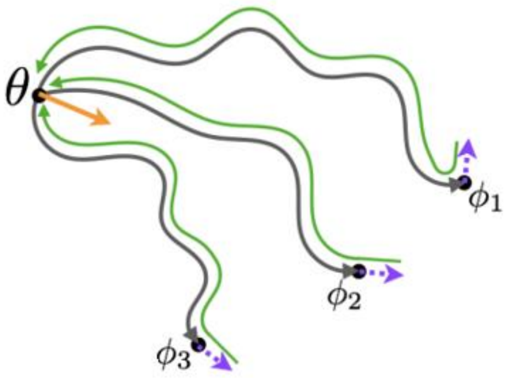


Travaux pionniers : Chelsea Finn

Idée : Remplacer le pré-entrainement classique sur ImageNet par un apprentissage des points initiaux

Transfer learning $\longrightarrow$ **Meta learning**

Tache 1 : cat vs dog
Tache 2 : MNIST classication
…



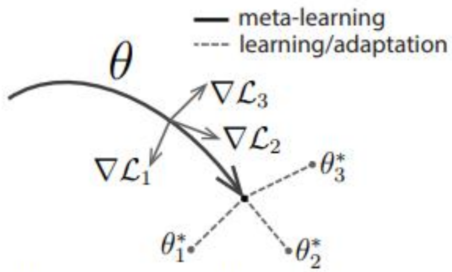


Figure 1. Diagram of our model-agnostic meta-learning algorithm (MAML), which optimizes for a representation θ that can quickly adapt to new tasks.

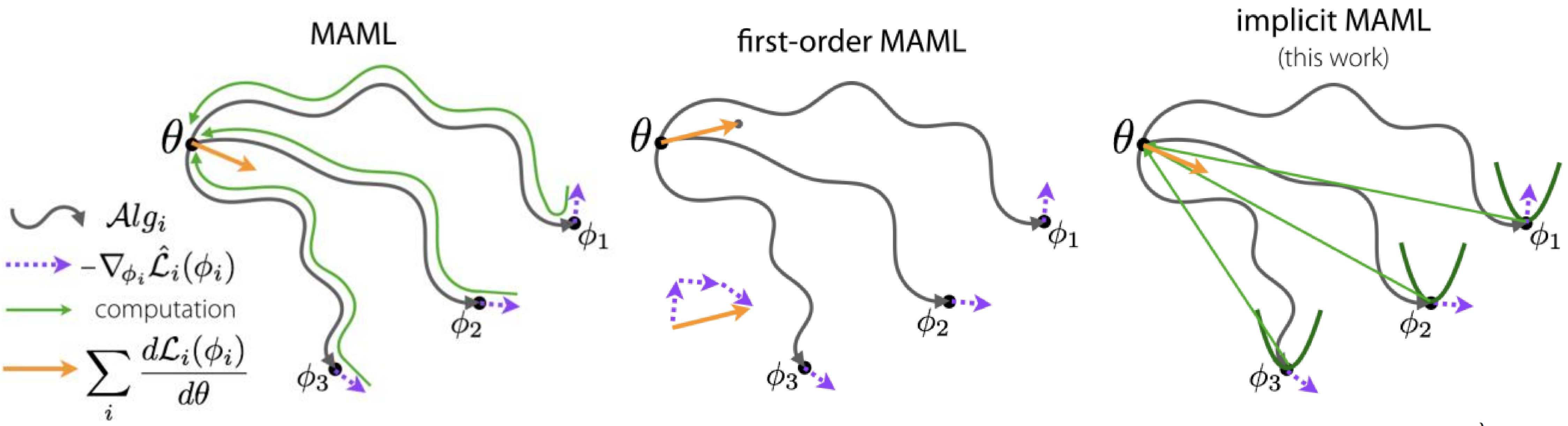
Algorithm 1 Model-Agnostic Meta-Learning

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters
1: randomly initialize θ
2: **while** not done **do**
3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
4: **for all** $\mathcal{T}_i$ **do**
5: Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ with respect to K examples
6: Compute adapted parameters with gradient descent: $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$
7: **end for**
8: Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$
9: **end while**

$$\overbrace{\boldsymbol{\theta}^*_{\mathrm{ML}} := \operatorname*{argmin}_{\boldsymbol{\theta} \in \Theta} F(\boldsymbol{\theta})}^{\text{outer}-\text{level}}, \text{ where } F(\boldsymbol{\theta}) = \frac{1}{M} \sum_{i=1}^{M} \mathcal{L}\Big(\overbrace{\mathcal{A}lg\big(\boldsymbol{\theta}, \mathcal{D}_i^{\mathrm{tr}}\big)}^{\text{inner}-\text{level}}, \mathcal{D}_i^{\mathrm{test}}\Big).$$

Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks, Finn et al, 2017 : http://proceedings.mlr.press/v70/finn17a/finn17a.pdf

Améliorations



$$\mathcal{A}lg^\star(\boldsymbol{\theta}, \mathcal{D}_i^{\mathrm{tr}}) = \underset{\phi' \in \Phi}{\operatorname{argmin}}\ \mathcal{L}(\phi', \mathcal{D}_i^{\mathrm{tr}}) + \frac{\lambda}{2}\,||\phi' - \boldsymbol{\theta}||^2.$$

Meta-Learning with Implicit Gradients, Finn et al 2019 : https://arxiv.org/pdf/1909.04630.pdf

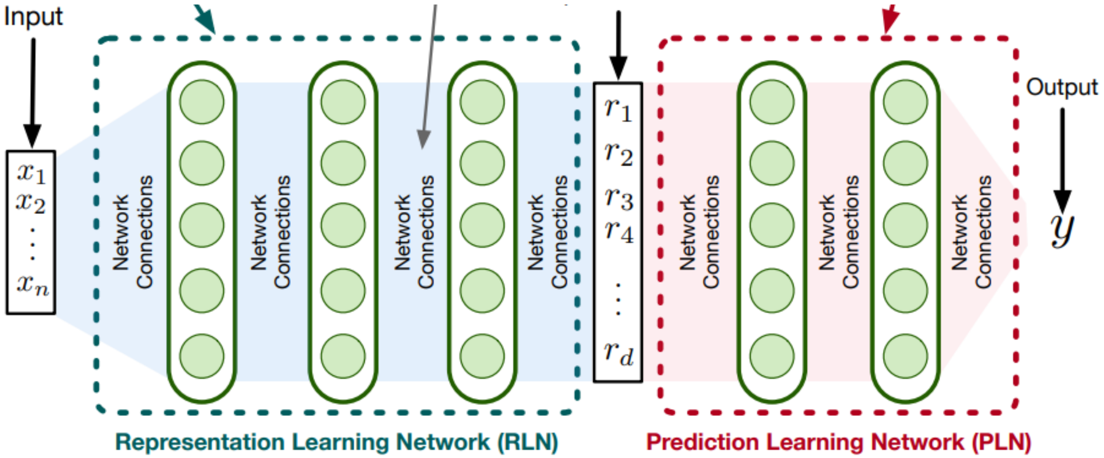
On First-Order Meta-Learning Algorithms, Nichol, 2018, https://arxiv.org/abs/1803.02999

Meta learning : Apprendre des représentations

$$\mathcal{T} = (X_1, Y_1), (X_2, Y_2), \ldots, (X_t, Y_t), \ldots$$

Outer-loop

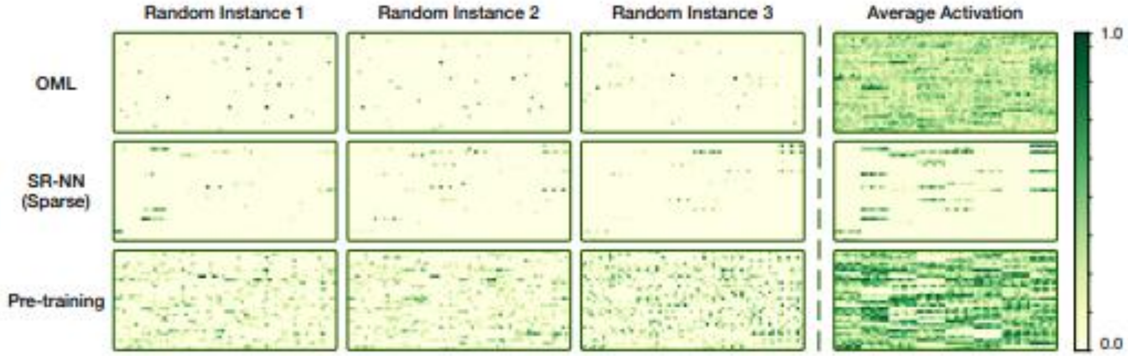
Inner-loop



Algorithm 2: Meta-Training : OML

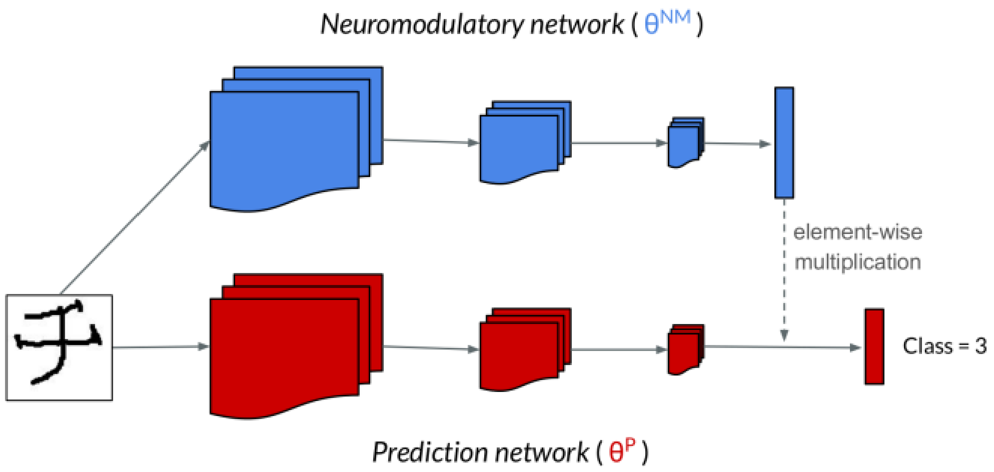
Require: $p(\mathcal{T})$: distribution over CLP problems
Require: α, β: step size hyperparameters

randomly initialize θ
while not done **do**
  randomly initialize W
  Sample CLP problem $\mathcal{T}_i \sim p(\mathcal{T})$
  Sample $\mathcal{S}_{train}$ from $p(\mathcal{S}_k|\mathcal{T}_i)$
  $W_0 = W$
  for $j = 1, 2, \ldots, k$ **do**
    $(X_j, Y_j) = \mathcal{S}_{train}[j]$
    $W_j = W_{j-1} - \alpha \nabla_{W_{j-1}} \ell_i(f_{\theta, W_{j-1}}(X_j), Y_j)$
  end for
  Sample $\mathcal{S}_{test}$ from $p(\mathcal{S}_k|\mathcal{T}_i)$
  Update $\theta \leftarrow \theta - \beta \nabla_\theta \ell_i(f_{\theta, W_k}(\mathcal{S}_{test}[:, 0]), \mathcal{S}_{test}[:, 1])$
end while



Meta-Learning Representations for Continual Learning, Javed et al, 2019 https://papers.nips.cc/paper/2019/file/f4dd765c12f2ef67f98f3558c282a9cd-Paper.pdf

Meta learning : Apprendre un réseau modulaire



Algorithm 1 A Neuromodulated Meta-Learning algorithm (ANML)

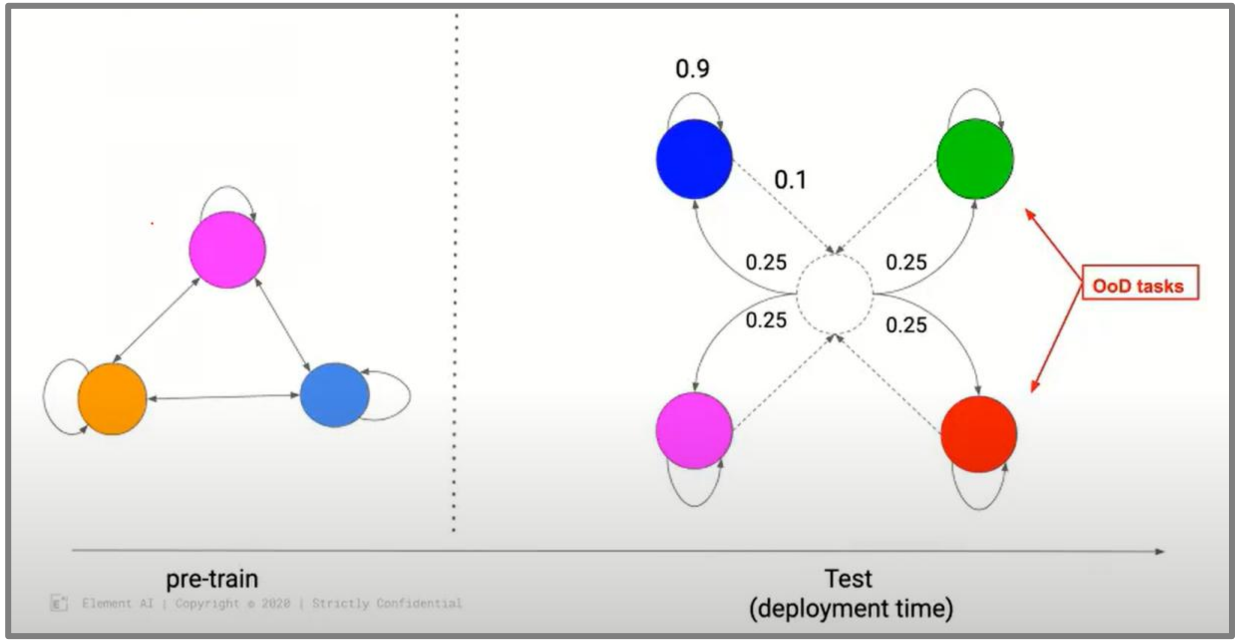
Require: $\mathcal{T} \leftarrow$ trajectory of N sequential meta-training tasks
Require: $\theta^{\mathrm{NM}} \leftarrow$ weights of the neuromodulatory network
Require: $\theta^{P} \leftarrow$ weights of the prediction network
Require: $\alpha, \beta \leftarrow$ learning-rate hyperparameters

```
initialize θ^NM, θ^P
for n = 1, 2, ... do                                        ▷ meta-learning outer-loop
    S_traj = T_n                                            ▷ trajectory for inner-loop training
    S_rem ~ T                                               ▷ sample instances from all tasks to remember
    θ_0^P = θ^P                                             ▷ create inner-loop copy of prediction net
    for i = 1, 2, ..., k do                                 ▷ task-learning inner-loop
        θ_i^P ← θ_{i-1}^P − β∇_{θ_{i-1}^P} L(θ^NM, θ_{i-1}^P, S_traj)   ▷ SGD on θ_{i-1}^P
    end for
    θ^{NM, P} ← θ^{NM, P} − α∇_{θ^{NM, P}} L(θ^NM, θ_k^P, S_traj, S_rem)
                                ▷ meta-update on θ^NM, θ^P w.r.t. final inner-loop θ_k^P
end for
```

“However, a perspective we advocate is that, when possible, we should not optimize for one thing and hope doing so leads to another thing : Instead, we should optimize directly for what we want (here, learning without forgetting)”

Learning to Continually Learn, Beaulieu et al, ECAI 2020 https://arxiv.org/pdf/2002.09571v1.pdf

Meta learning : fast adapation and fast recovery



[Online fast adapatation and online knowledge accumulation : a new approach to continual learning (Caccia et al,NIPS, 2020)]()

Meta Learning

Le concept présente pas mal d'enjeux encore à résoudre :

- **Cout de calcul conséquent**
- **Stabilité de convergences (MAML ++)**

Une attention particulière doit être portée à la formulation du problème

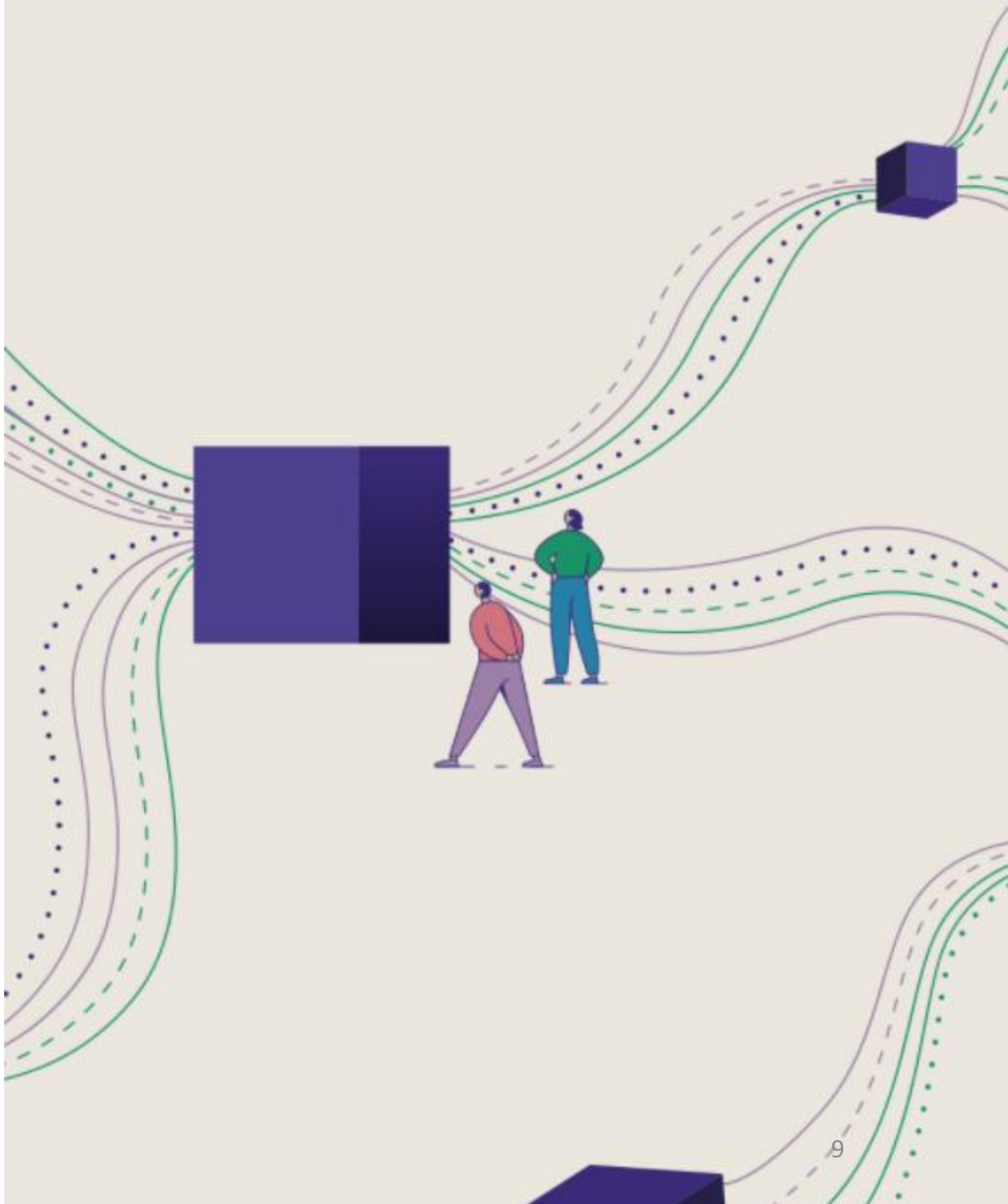
Le méta-learning a apporter un changement de paradigme en Continual learning :

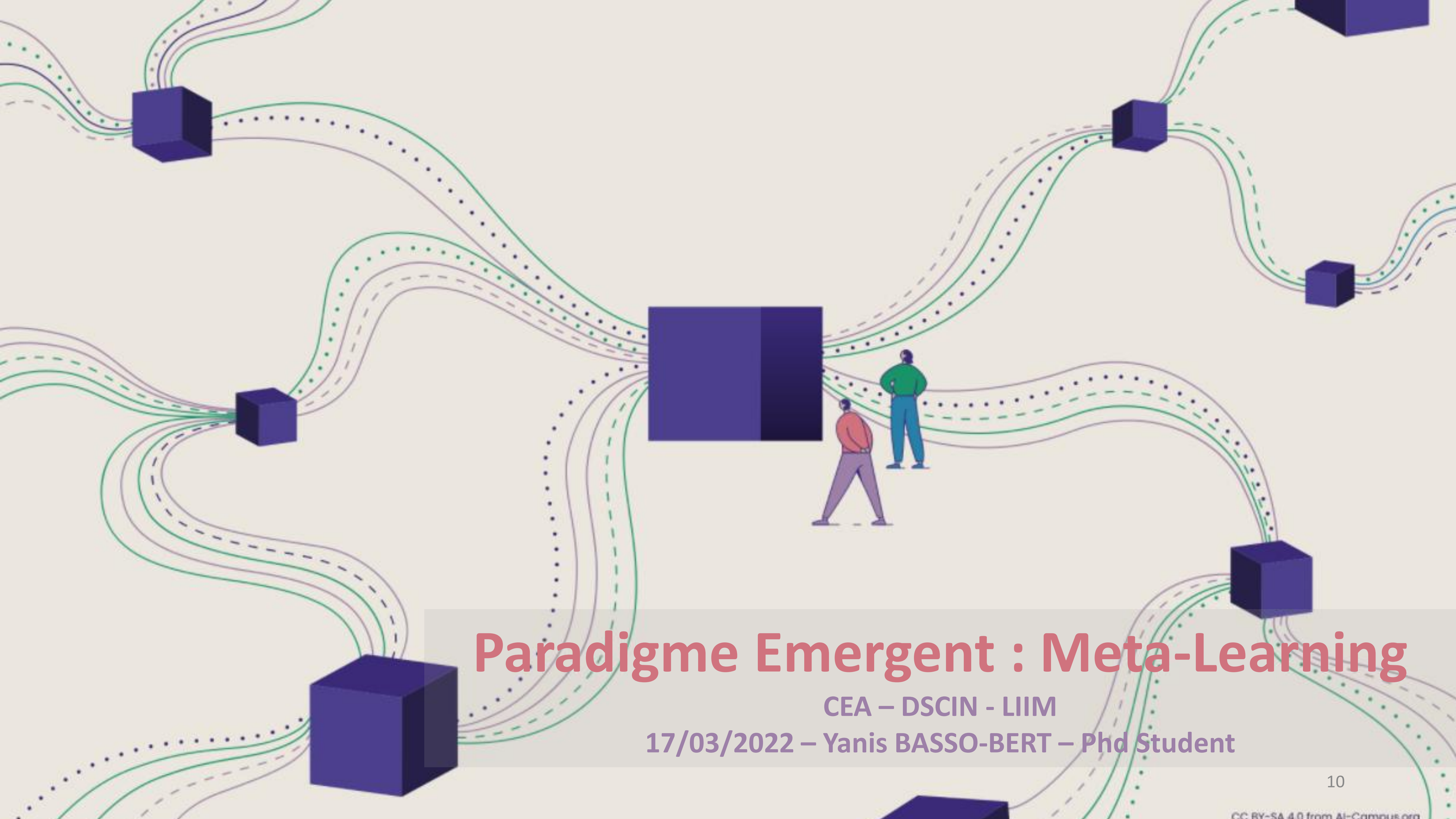
Desiderata :

- Accumulation de connaissance
- Rétention parfaite
- Forward Transfer
- Backward Transfer

Desiderata :

- Fast adaptation
- Fast recovery
- Dynamic representation
- Oublie possible





Paradigme Emergent : Meta-Learning

CEA – DSCIN - LIIM

17/03/2022 – Yanis BASSO-BERT – Phd Student

CC BY-SA 4.0 from AI-Campus.org

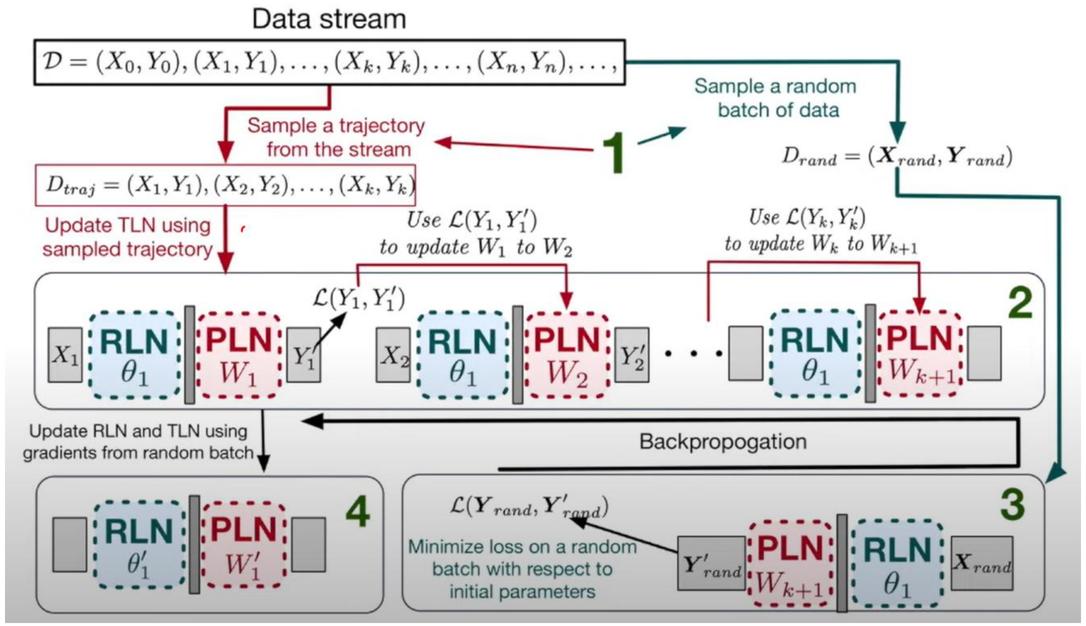
Meta learning Representation for continual learning

Javed et al – 2019 - NIPS

Tag : Continual learning/Streaming/Sparse Representation/Meta learning

Problem : Avoid Catastrophic forgetting in Deep Neural Network

Intuition : Catastrophic Forgetting is due to Dense/Distributed Represention



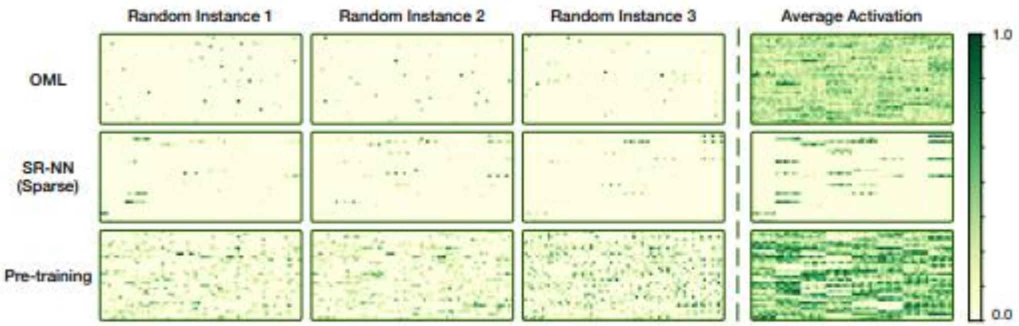
Solution : Explicitely learn an appropriate represenation via Meta-learning algorithm

Method :

- Meta-optimization on a CLP problem with naive solution
- MAML-Reptile like algorithm
- Compatible with other CLP solutions

Result :

- Natural semi-distributed Representation



Main related Article :

French 1991 : semi distributed representation
Riemer 2019 : weight regularisation for CL
Nichol & Schulman 2018 : MAML-Rept