Overview of continual learning and case experiment with Arabic Digits

Iliyas Bektas

Università degli Studi di Roma "La Sapienza", 2022

INTRODUCTION

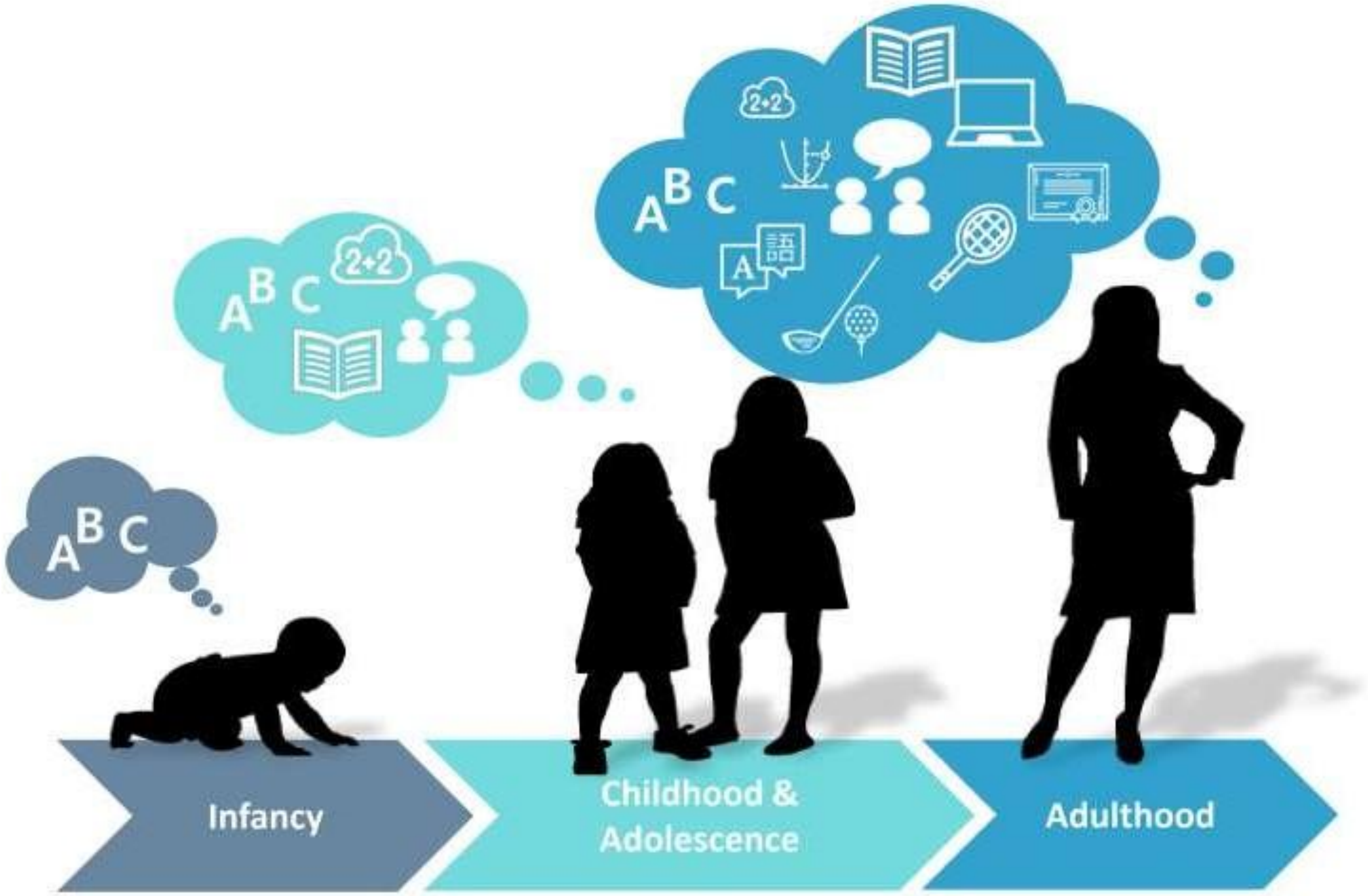
Continual learning

- Continual Learning (CL) is built on the idea of ***learning continuously and adaptively about the external world*** and enabling the *autonomous* ***incremental development*** of ever more complex *skills* and *knowledge*

Ex. From learning alphabet **to** learning other languages

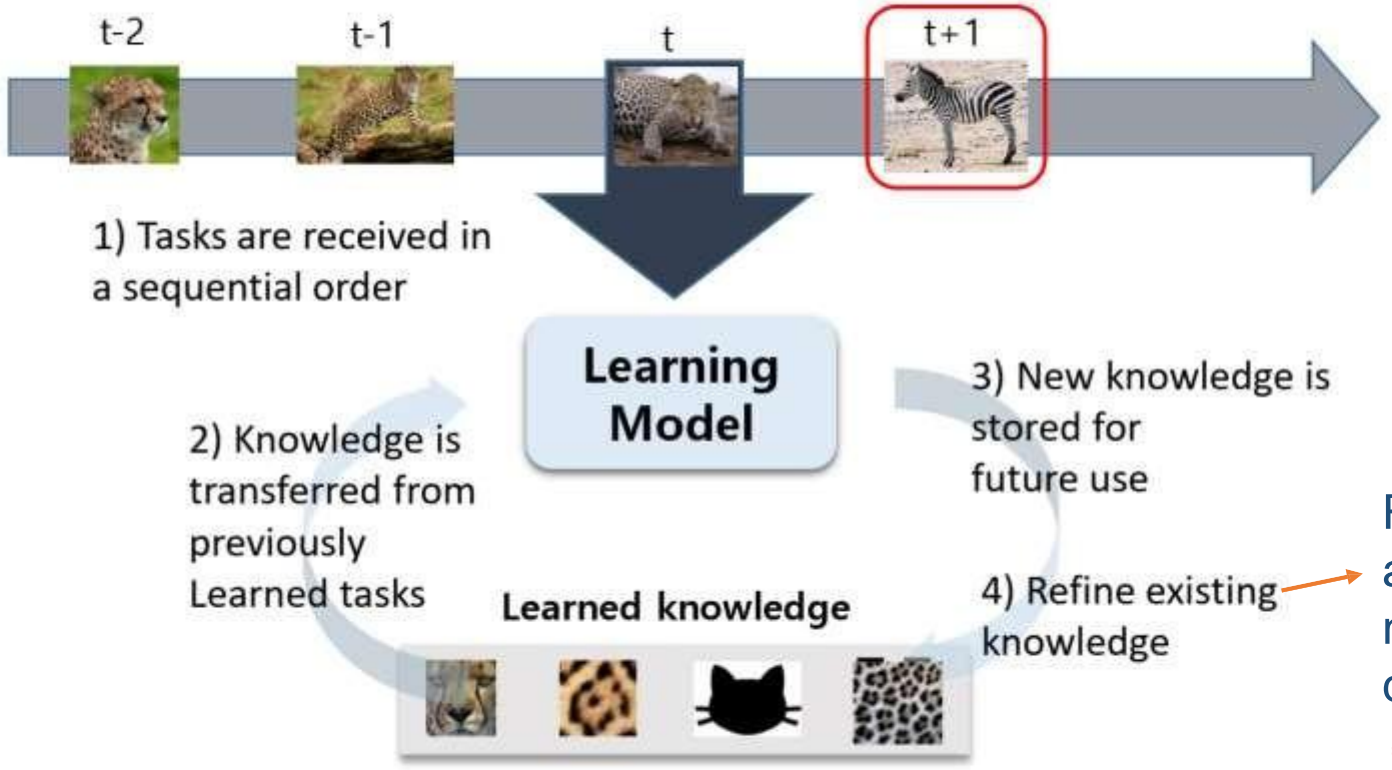
Continual learning

- Humans learn throughout their lives and retain/use the previously learned knowledge when learning for a new task



Continual learning

- Humans become increasingly smarter over time. Couldn't we build a similar system that basically learns forever?



Refine means we adapt the learned models with new data

⇒ weights of learn models could change to generalize the new data

CHALLENGES

Challenge: Incomplete, Growing Dataset

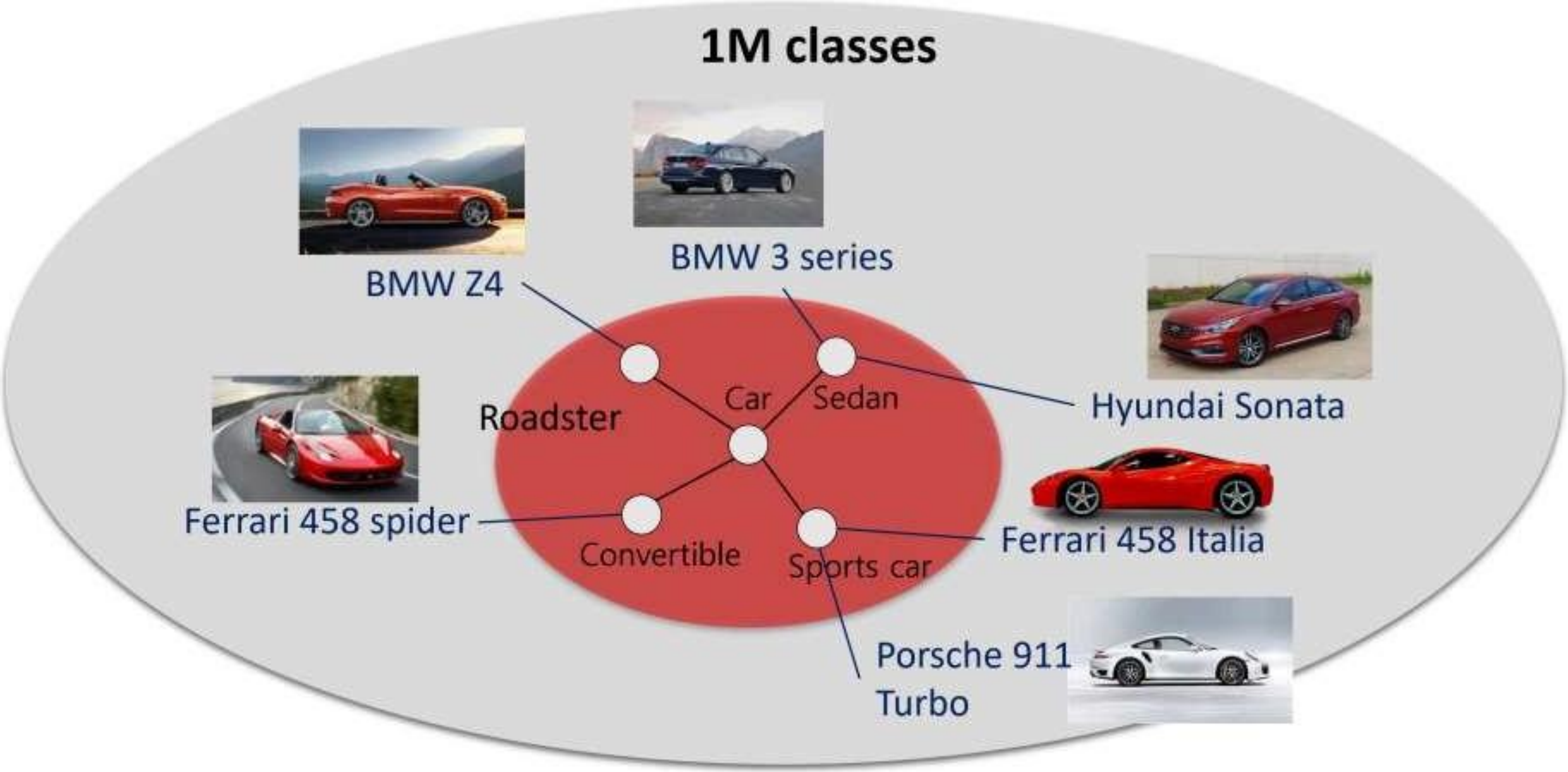
In many large-scale learning scenarios, not all training data might be available when we want to begin training the network.

ImageNet

22,000 classes

Challenge: Incomplete, Growing Dataset

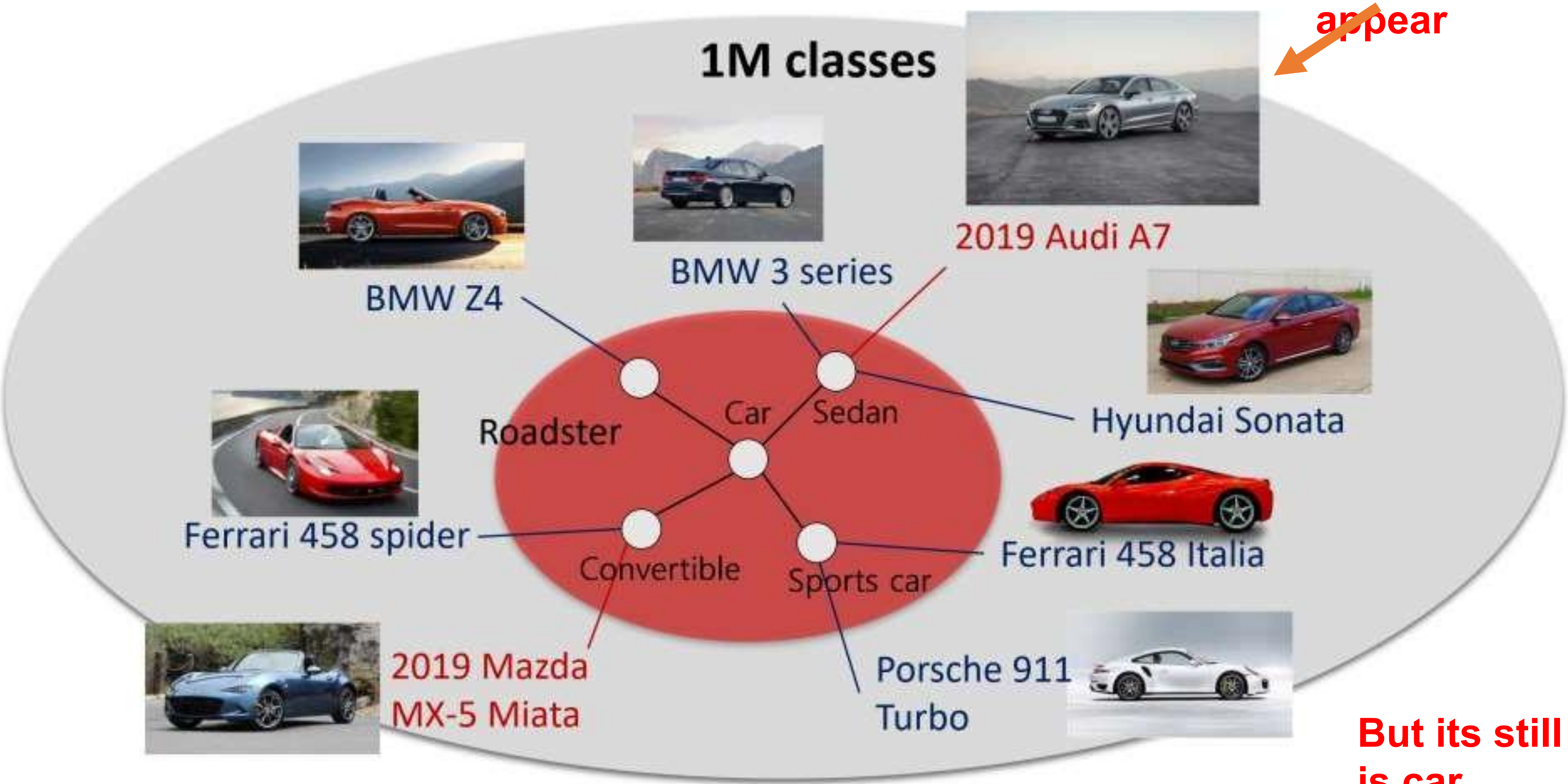
In many large-scale learning scenarios, not all training data might be available when we want to begin training the network.



Challenge: Incomplete, Growing Dataset

Even worse, the set of tasks may ***dynamically grow*** as new tasks are introduced.

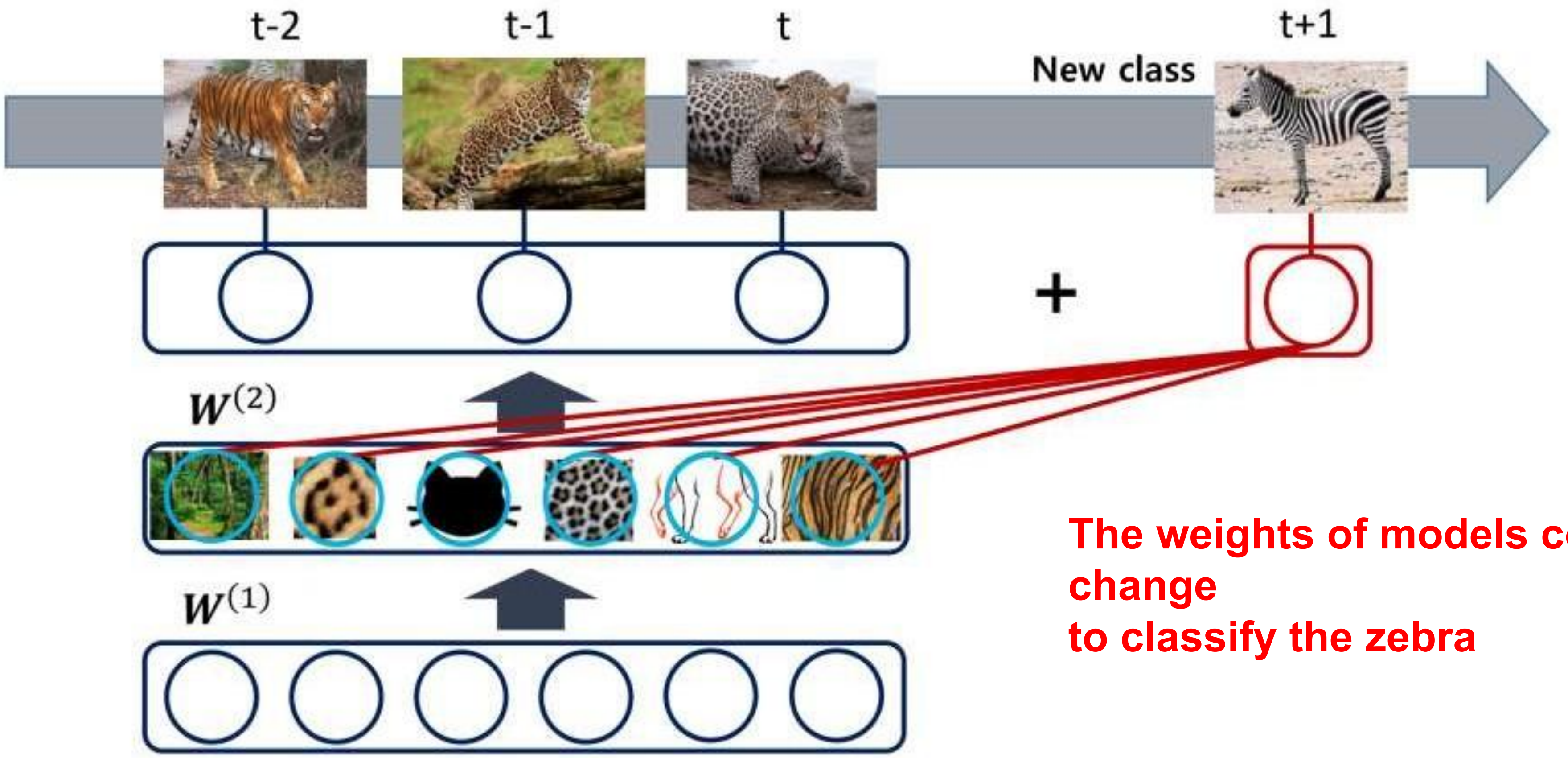
New class appear



But its still fine if the new data is car

Challenge

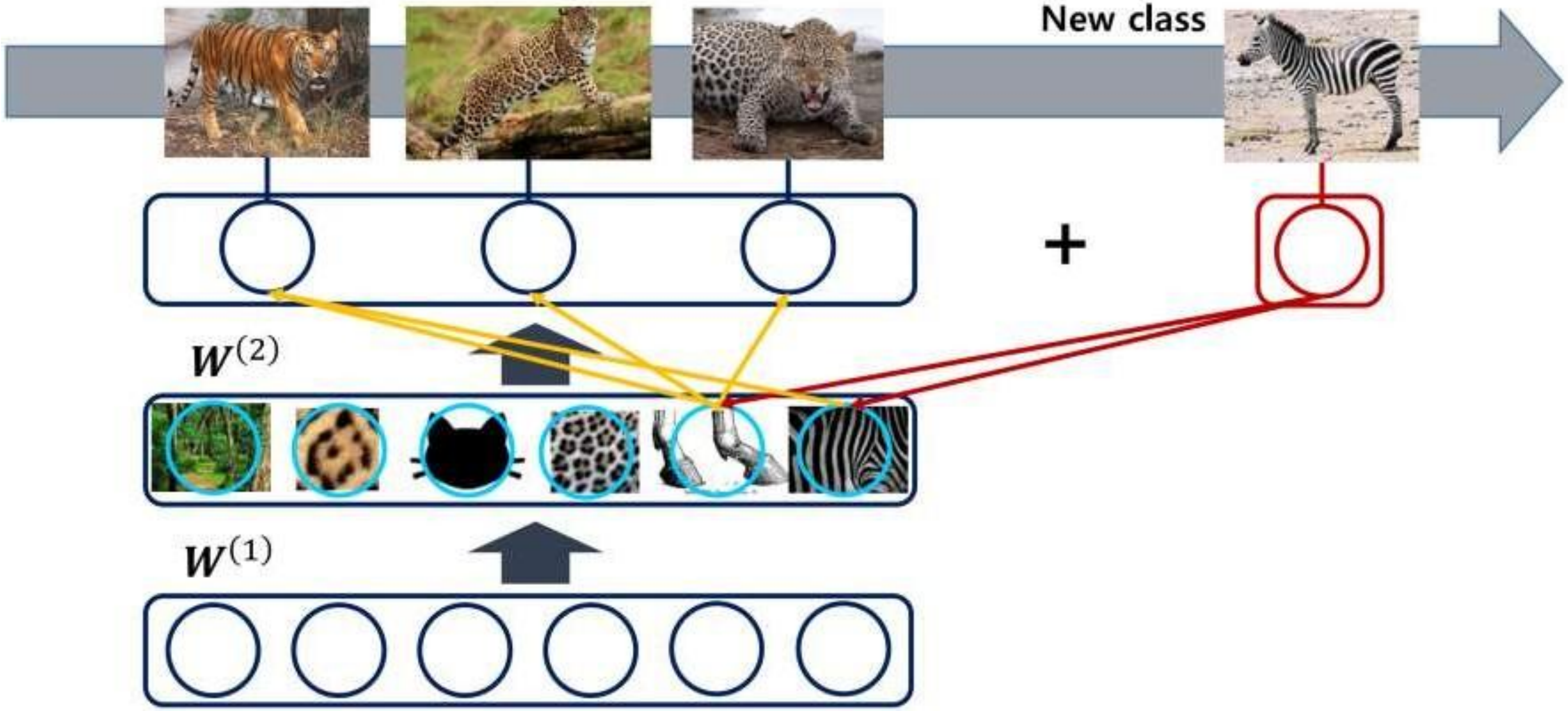
However, if the classes we had in the early stages of learning ***significantly differs from the new class***, utilization of prior knowledge may degenerate performance.



The weights of models could significantly change to classify the zebra

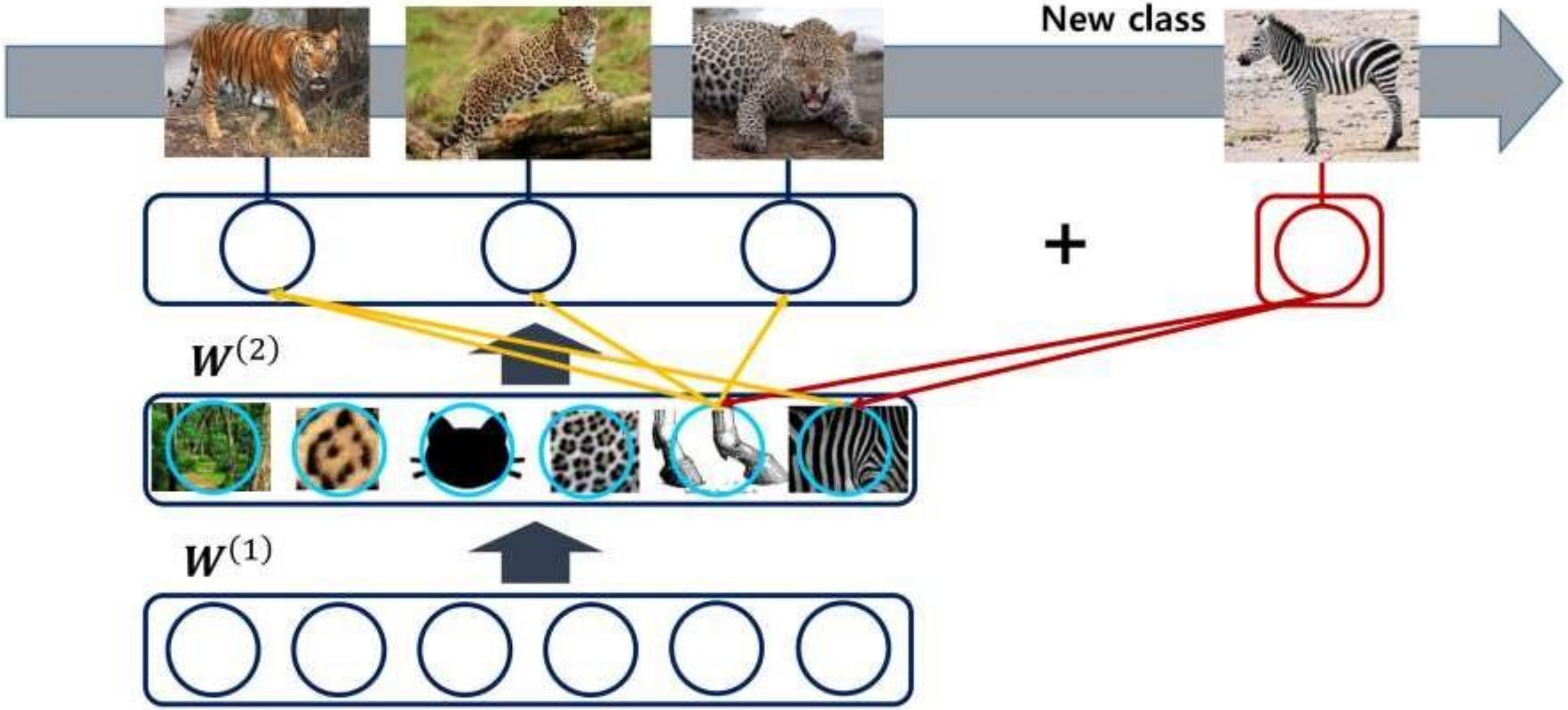
Catastrophic Forgetting

Introduction of new units can also result in ***semantic drift*** or ***catastrophic forgetting***, where original meaning of the features change as they fit to later tasks.



Catastrophic Forgetting

Introduction of new units can also result in ***semantic drift*** or ***catastrophic forgetting***, where original meaning of the features change as they fit to later tasks.



SOLUTIONS

Continual learning

- Expectations of CL
  - Online learning: learning occurs at every moment
  - Presence of transfer: able to transfer from previous tasks to new ones
  - Resistance to catastrophic forgetting
  - No direct access to previous experience

Challenge for Continual Learning

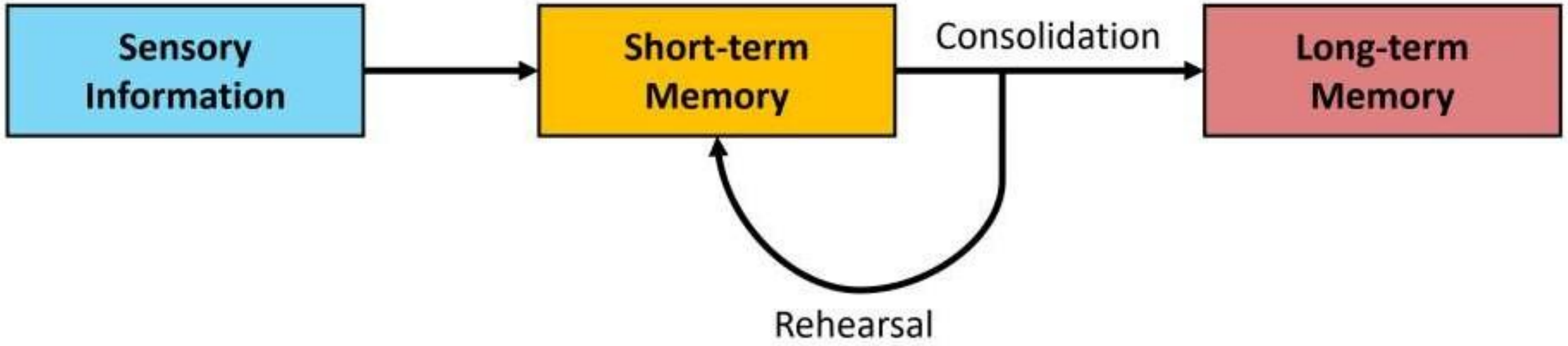
- We need a balance between adapting to recent data and retaining knowledge from old data because:
  - Too much plasticity leads to the catastrophic forgetting problem
  - Too much stability leads to an inability to adapt

Like our brain, sometimes we forget something invaluable to memorize valuable knowledge

Synaptic consolidation

Mammalian brain may avoid catastrophic forgetting by ***protecting previously-acquired knowledge*** in neocortical circuits.

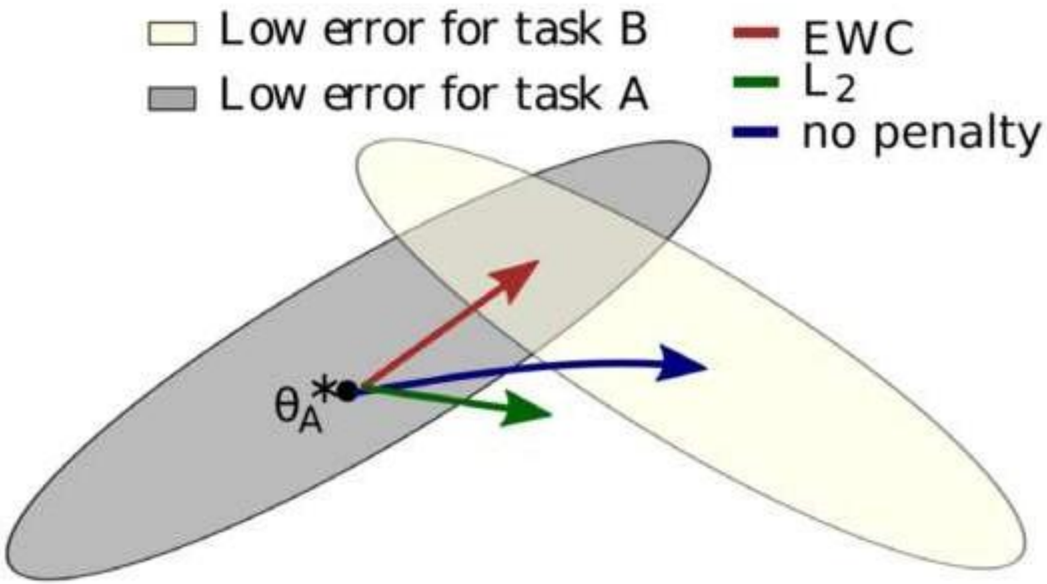
This how our CL system in our brain is working



Acquired knowledge is durably encoded in synapses that are rendered less plastic thus stable, called ***synaptic consolidation***.

Elastic Weight Consolidation

[Kirkpatrick17] tries to solve catastrophic forgetting by constraining important parameters to stay close to their old values.



This loss function not only care about loss of Task B but also performance of Task A

$$\mathcal{L}(\boldsymbol{\theta}) = \underbrace{\mathcal{L}_B(\boldsymbol{\theta})}_{\text{Task B loss}} + \underbrace{\sum_i \frac{\lambda}{2} \boldsymbol{F}_i \left(\boldsymbol{\theta}_i - \boldsymbol{\theta}^*_{A,i}\right)^2}_{\text{Transfer Constraints}}$$

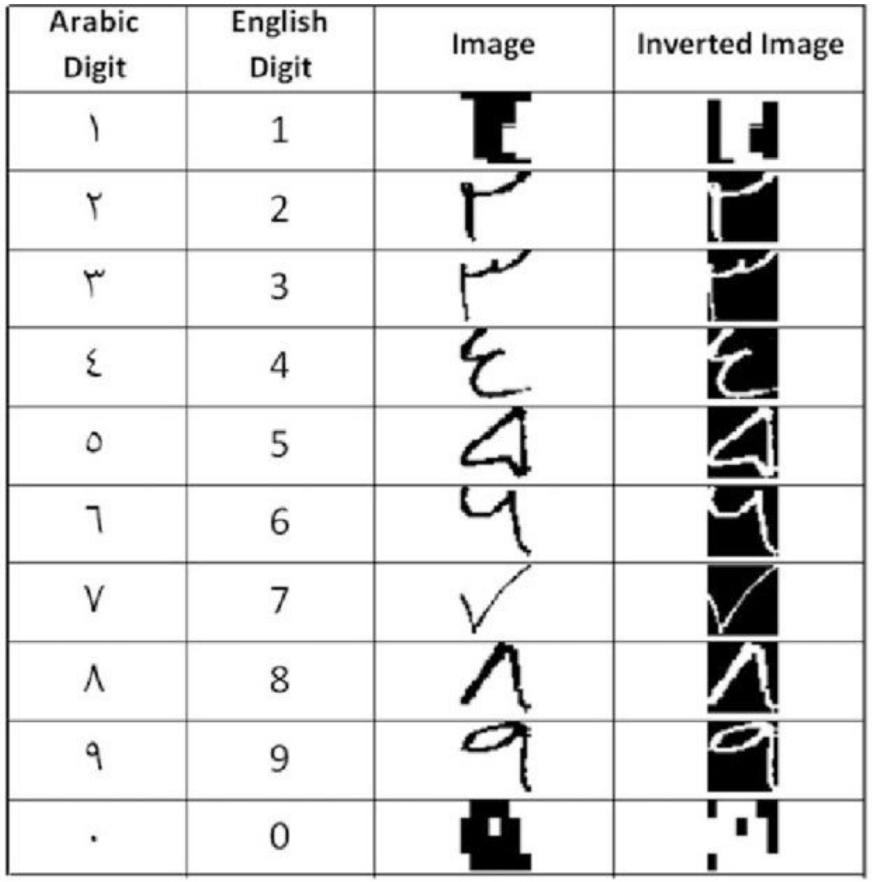
EWC protects the performance in task A by constraining the parameters to ***stay in a region of low error*** for task, where is around θ_A^*.

Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks." *Proceedings of the national academy of sciences* (2017): 201611835.

Application of the CL to Arabic Digits

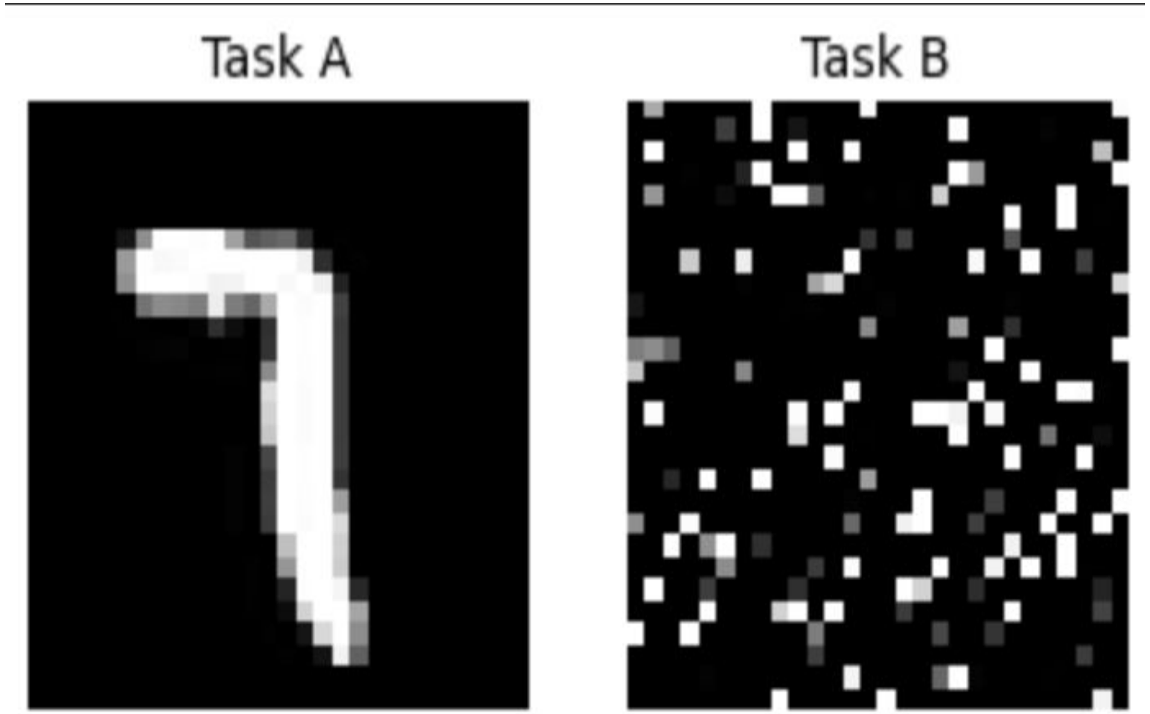
About the Dataset

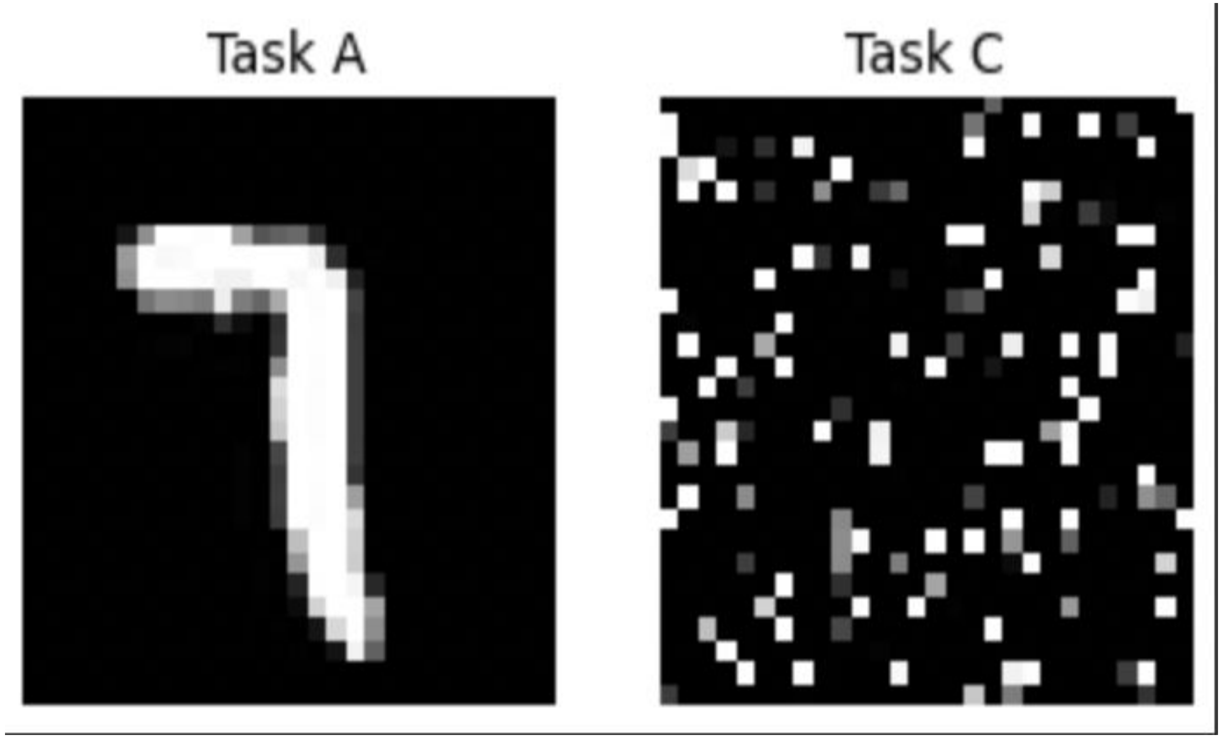
In recent years, handwritten digits recognition has been an important area due to its applications in several fields. This dataset is a collection of handwritten Arabic digits recognition that face several challenges, including the unlimited variation in human handwriting and the large public databases. It contains 60000 training and 10000 testing images. I downloaded the dataset from Kaggle.



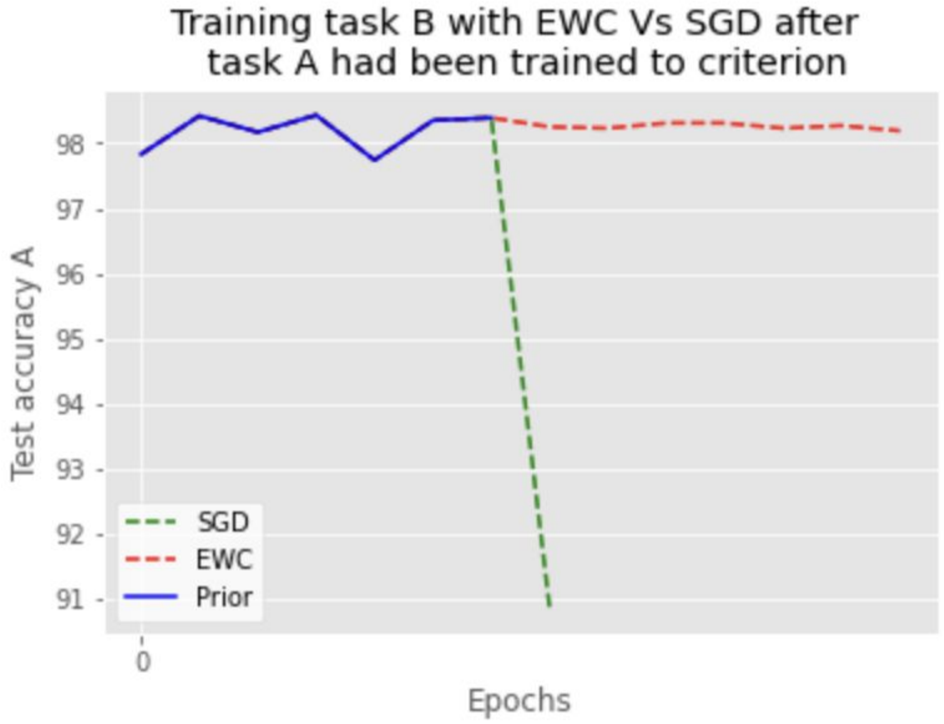
Permuting the Dataset

- Permuting the dataset as if it is coming from the other distribution
- The same idea as in "Overcoming catastrophic forgetting in neural networks" paper applied to mnist dataset

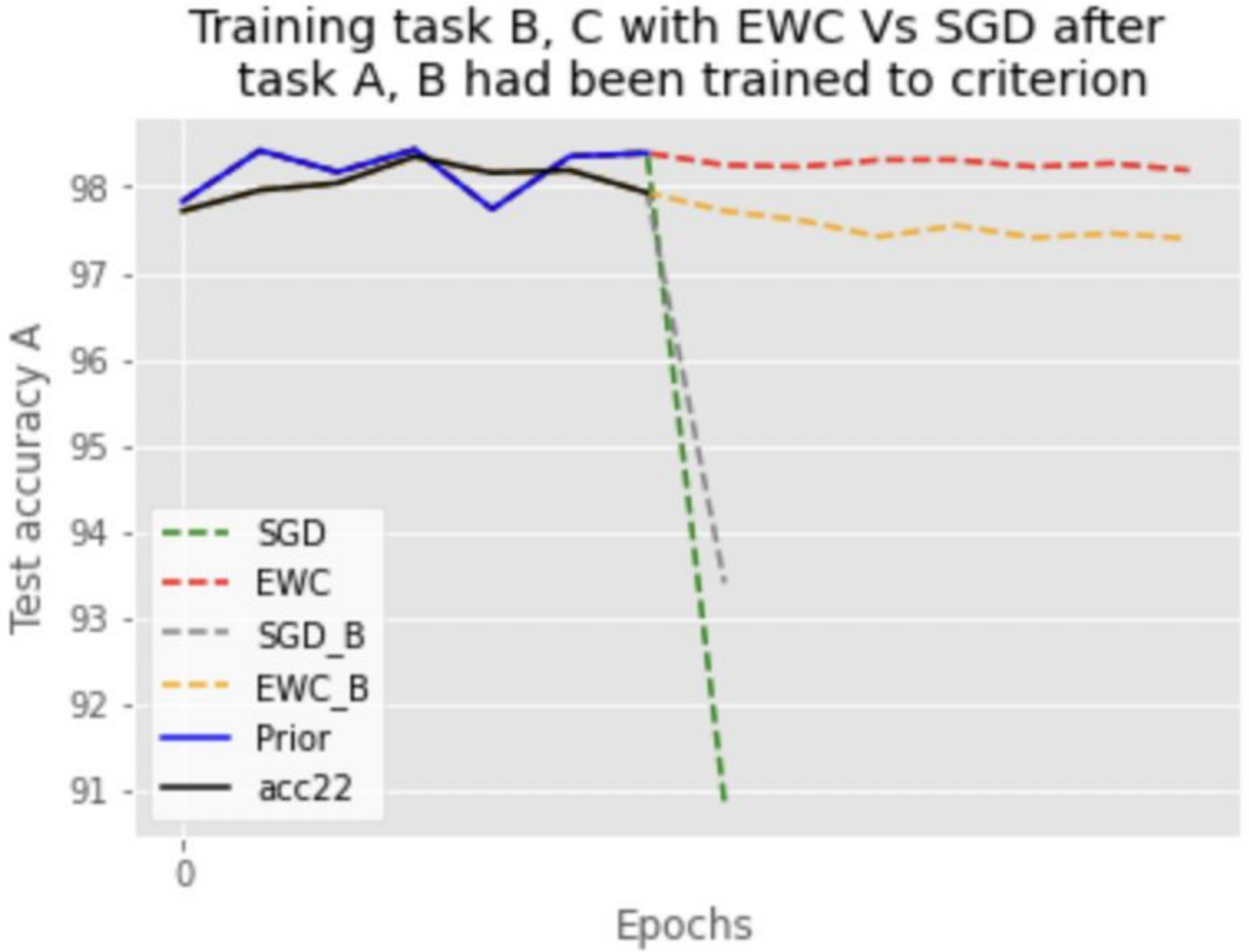




Task B results



Task C Results



Metrics For continual learning

$$\text{Accuracy} = \frac{\sum_{i\leq j} R_{i,j}}{\frac{1}{2}t(t+1)}.$$

$$\text{Backward transfer} = \frac{\sum_{j\geq 2, i<j} R_{i,j} - R_{i,i}}{\frac{t(t-1)}{2}}$$

After observing t different task, the performance of the model can be summarized in a R(t,t) matrix, where Ri,j is the performance on task i after training on task j. Several CL metrics can be obtained from this matrix. For example, a global accuracy.

Another interesting metric is backward transfer, i.e., how much training on a new task has improved or worsened the previous tasks.

Global Accuracy: 93.65
Backward Transfer: -9.09

Take-home messages

- CL is very new domain, and its potential is immerse, not already at its explosion
- The experiments now are on really simple task and not in large scale
- In Image Captioning, it’s possible to come with CL, but the way is so far and depends on state-of-the-art works on CL

Github repository of the code and the slide

https://github.com/iliyasbektas/Continual_learning_arabic_digits_dataset