Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Bonn-Aachen
International Center for
Information Technology

Institute for AI and
Autonomous Systems

# Lifelong Learning

## An Overview for Cognitive Robotics

Dr. Alex Mitrevski
Master of Autonomous Systems

# Structure

- ▶ Overview of robot lifelong learning
- ▶ Class incremental learning
- ▶ Class incremental learning algorithm example: iCaRL
- ▶ Use case: Lifelong learning for action recognition

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

**Class-Incremental Learning: Survey and Performance Evaluation on Image Classification**

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

**Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges**

Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

**iCaRL: Incremental Classifier and Representation Learning**

Sylvestre-Alvise Rebuffi, University of Oxford/IST Austria
Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert, IST Austria

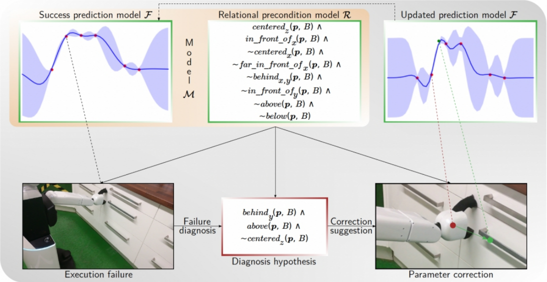
Master's Thesis

**Lifelong Action Learning for Socially Assistive Robots**

*Hasnainali Walli*

# Motivating Problem 1: Lifelong Skill Learning

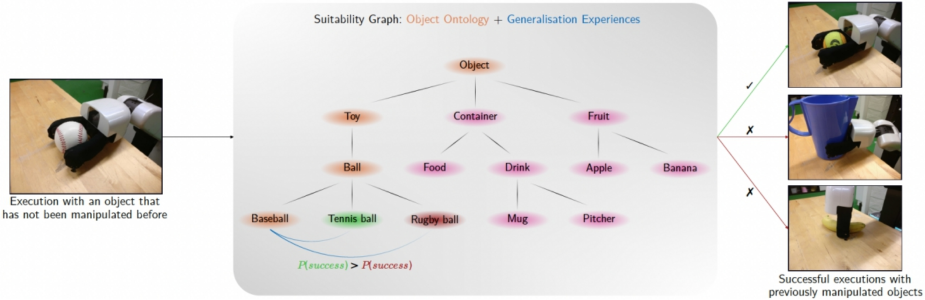
- An essential ability of a cognitive robot is to improve its skills based on new experiences as well as to acquire new skills
- Particularly in the case of execution failures, **a robot should be able to learn from the failure experience** so that it is less likely that it will repeat it again
- **Skill representations should thus have an inbuilt mechanism that enables lifelong skill learning**




A. Mitrevski, "Skill Generalisation and Experience Acquisition for Predicting and Avoiding Execution Failures," *Ph.D. dissertation*, Department of Computer Science, RWTH Aachen University, 2023. Available: https://publications.rwth-aachen.de/record/943042

# Motivating Problem 2: Lifelong Object Learning

- When learning object-centric manipulation skills, a robot will encounter a certain number of objects at training time, but **may need to interact with other — previously unseen — objects during its operation**
- Lifelong learning should thus enable a robot to:
  - **recognise new objects**
  - **learn how to handle those objects correctly** (e.g. by learning object interaction models)




A. Mitrevski, P. G. Plöger, and G. Lakemeyer, "A Hybrid Skill Parameterisation Model Combining Symbolic and Subsymbolic Elements for Introspective Robots," *Robotics and Autonomous Systems*, vol. 161, p. 104350:1–22, Mar. 2023. Available: https://doi.org/10.1016/j.robot.2022.104350

# Motivating Problem 3: Lifelong Learning for Action Recognition

- In robot-assisted therapy and education, the ability to recognise the actions performed by the person under therapy can help a robot adapt its behaviour or suggest corrective exercises
- Lifelong learning for action recognition is particularly useful **if therapists want to include new actions that the robot should recognise** — in this case, the system should be able to:
  - **Incorporate new actions over time** (not all actions are known at the initial training time)
  - **Learn from a few examples** (e.g. from therapist demonstrations)

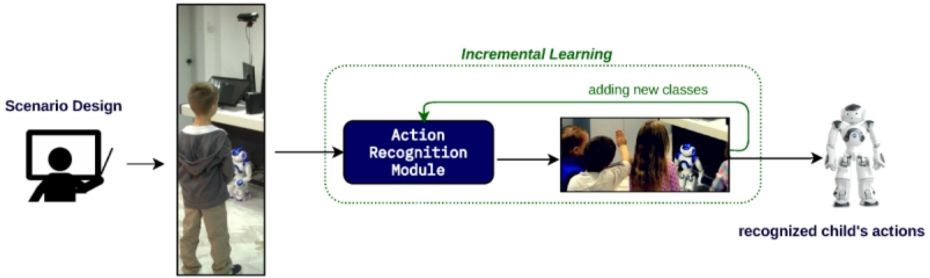
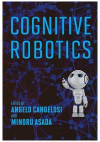
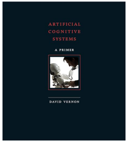


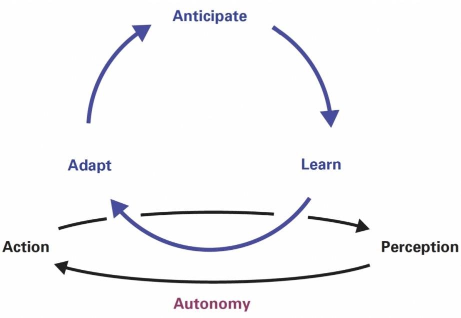

Diagram adapted from N. Efthymiou, P. P. Filntisis, G. Potamianos, and P. Maragos, "Visual Robotic Perception System with Incremental Learning for Child–Robot Interaction Scenarios," *Technologies*, vol. 9, no. 86, Nov. 2021. Available: https://doi.org/10.3390/technologies9040086

# Overview of Robot Lifelong Learning

# Lifelong Learning for Cognitive Robots

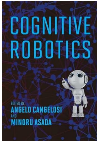


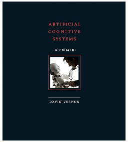



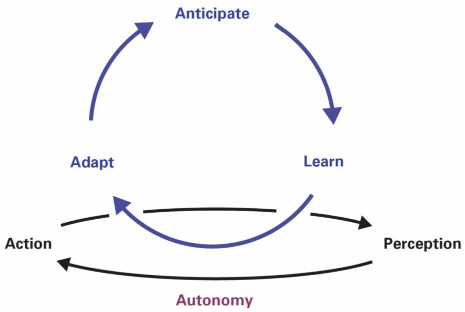




- The ability to learn continuously is one of the **central element of cognitive robots**

# Lifelong Learning for Cognitive Robots

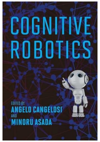


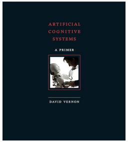



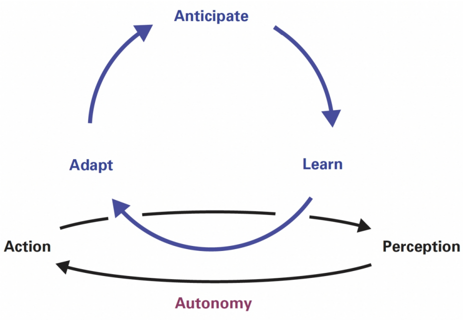




- The ability to learn continuously is one of the **central element of cognitive robots**
- Lifelong learning should be incorporated in different parts of a robotic system, such as:

# Lifelong Learning for Cognitive Robots

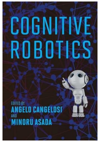


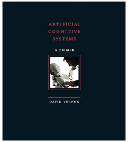



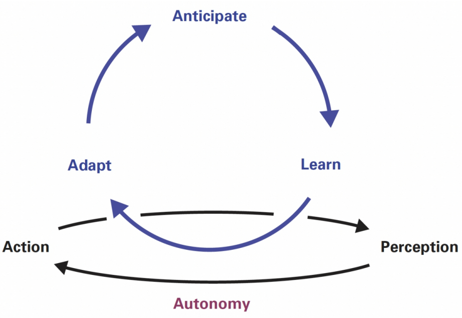




- The ability to learn continuously is one of the **central element of cognitive robots**
- Lifelong learning should be incorporated in different parts of a robotic system, such as:
  - **the perceptual subsystem** (e.g. so the robot can focus its attention better)

# Lifelong Learning for Cognitive Robots

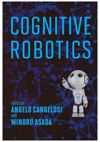


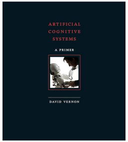



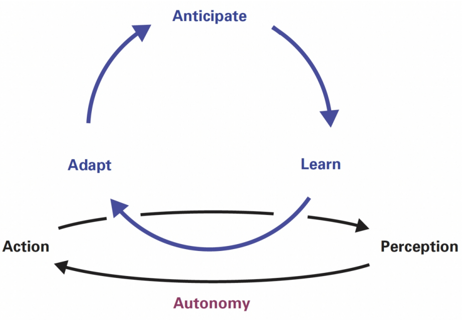




- The ability to learn continuously is one of the **central element of cognitive robots**
- Lifelong learning should be incorporated in different parts of a robotic system, such as:
  - **the perceptual subsystem** (e.g. so the robot can focus its attention better)
  - **the action subsystem** (so that a robot can improve its existing skills and learn new skills)

# Lifelong Learning for Cognitive Robots

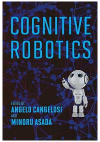


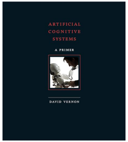



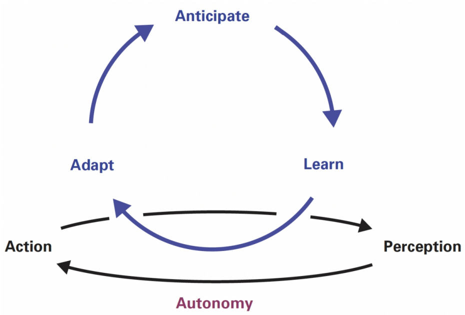




- The ability to learn continuously is one of the **central element of cognitive robots**
- Lifelong learning should be incorporated in different parts of a robotic system, such as:
  - **the perceptual subsystem** (e.g. so the robot can focus its attention better)
  - **the action subsystem** (so that a robot can improve its existing skills and learn new skills)
  - **the anticipation subsystem** (so that the robot can learn more accurate predictive models)

# Lifelong Learning for Cognitive Robots




- The ability to learn continuously is one of the **central element of cognitive robots**
- Lifelong learning should be incorporated in different parts of a robotic system, such as:
  - **the perceptual subsystem** (e.g. so the robot can focus its attention better)
  - **the action subsystem** (so that a robot can improve its existing skills and learn new skills)
  - **the anticipation subsystem** (so that the robot can learn more accurate predictive models)
- Note: **Lifelong learning is often synonymously called continual learning or incremental learning**

# Lifelong Learning in Cognitive Architectures

- **The cognitive architectures that we looked at a few lectures ago all have lifelong learning as part of the computational model**

# Lifelong Learning in Cognitive Architectures

- **The cognitive architectures that we looked at a few lectures ago all have lifelong learning as part of the computational model**
  - **Soar** uses learning in all of its memory types as well as for learning predictive models

# Lifelong Learning in Cognitive Architectures

- **The cognitive architectures that we looked at a few lectures ago all have lifelong learning as part of the computational model**
  - **Soar** uses learning in all of its memory types as well as for learning predictive models
  - **LIDA** uses learning to update its different memory types, to improve its perceptual models and the system's attention, and to update the system's behaviours

# Lifelong Learning in Cognitive Architectures

- **The cognitive architectures that we looked at a few lectures ago all have lifelong learning as part of the computational model**
  - **Soar** uses learning in all of its memory types as well as for learning predictive models
  - **LIDA** uses learning to update its different memory types, to improve its perceptual models and the system's attention, and to update the system's behaviours
  - **CLARION** performs both top-down and bottom-up learning for its explicit and implicit components, respectively

# Lifelong Learning in Cognitive Architectures

- **The cognitive architectures that we looked at a few lectures ago all have lifelong learning as part of the computational model**
  - **Soar** uses learning in all of its memory types as well as for learning predictive models
  - **LIDA** uses learning to update its different memory types, to improve its perceptual models and the system's attention, and to update the system's behaviours
  - **CLARION** performs both top-down and bottom-up learning for its explicit and implicit components, respectively
- Caveat: In the context of the architectures, learning is often only considered at a conceptual level, so it is not always clear how to implement specific types of learning practically

# Why is Lifelong Learning Difficult?

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges

Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

**Data forgetting**

It would be unreasonable to store all data that a robot collects over its operation, but not having access to old data can make it difficult to retain old knowledge

# Why is Lifelong Learning Difficult?



Timothée Lesort*[,1,2], Vincenzo Lomonaco*[,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez*[,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

### Data forgetting

It would be unreasonable to store all data that a robot collects over its operation, but not having access to old data can make it difficult to retain old knowledge

### New data demands

Lifelong learning requires new data, but updated models may require a significant amount of new data to achieve acceptable practical performance (few-shot or one-shot learning aim to address this problem)

# Why is Lifelong Learning Difficult?

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges

Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

**Data forgetting**

It would be unreasonable to store all data that a robot collects over its operation, but not having access to old data can make it difficult to retain old knowledge

**New data demands**

Lifelong learning requires new data, but updated models may require a significant amount of new data to achieve acceptable practical performance (few-shot or one-shot learning aim to address this problem)

**Representation updates**

Updating a learned model requires the model's representation to enable such updates easily — for some types of representations, dedicated updating procedures need to be defined

# Why is Lifelong Learning Difficult?

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges

Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

**Data forgetting**

It would be unreasonable to store all data that a robot collects over its operation, but not having access to old data can make it difficult to retain old knowledge

**New data demands**

Lifelong learning requires new data, but updated models may require a significant amount of new data to achieve acceptable practical performance (few-shot or one-shot learning aim to address this problem)

**Representation updates**

Updating a learned model requires the model's representation to enable such updates easily — for some types of representations, dedicated updating procedures need to be defined

**Data distribution drift**

In many cases, the distribution of the data that the robot receives can change over time; detecting such changes is another challenge

# Why is Lifelong Learning Difficult?

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges

Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

### Data forgetting

It would be unreasonable to store all data that a robot collects over its operation, but not having access to old data can make it difficult to retain old knowledge

### New data demands

Lifelong learning requires new data, but updated models may require a significant amount of new data to achieve acceptable practical performance (few-shot or one-shot learning aim to address this problem)

### Representation updates

Updating a learned model requires the model's representation to enable such updates easily — for some types of representations, dedicated updating procedures need to be defined

### Data distribution drift

In many cases, the distribution of the data that the robot receives can change over time; detecting such changes is another challenge

### Objective changes

In some cases, the objective that is optimised during learning may also need to be updated (e.g. in a social robotics context), but identifying that this is needed is challenging and, even if it can be done, obtaining the new objective is generally difficult

# Why is Lifelong Learning Difficult?

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges

Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

**Data forgetting**

It would be unreasonable to store all data that a robot collects over its operation, but not having access to old data can make it difficult to retain old knowledge

**New data demands**

Lifelong learning requires new data, but updated models may require a significant amount of new data to achieve acceptable practical performance (few-shot or one-shot learning aim to address this problem)

**Representation updates**

Updating a learned model requires the model's representation to enable such updates easily — for some types of representations, dedicated updating procedures need to be defined

**Data distribution drift**

In many cases, the distribution of the data that the robot receives can change over time; detecting such changes is another challenge

**Objective changes**

In some cases, the objective that is optimised during learning may also need to be updated (e.g. in a social robotics context), but identifying that this is needed is challenging and, even if it can be done, obtaining the new objective is generally difficult

**Data privacy**

Lifelong learning requires continuous data collection, which is a tricky aspect from a privacy point of view; in addition, users should be able to request data about themselves to be removed from the system (in Europe, we have the GDPR), but it is unclear how to do this in emergent systems (knowledge is distributed along the system)

# Related Learning Paradigms



Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

Lifelong learning has close relations with various other learning paradigms:

**Online / one-shot learning**

Leaning from a single example (can be done in the context of lifelong learning)

# Related Learning Paradigms



Timothée Lesort*[,1,2], Vincenzo Lomonaco*[,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez*[,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

Lifelong learning has close relations with various other learning paradigms:

**Online / one-shot learning**

Leaning from a single example (can be done in the context of lifelong learning)

**Few-shot learning**

Learning from a few examples (can also be done in lifelong learning)

# Related Learning Paradigms

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges

Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

Lifelong learning has close relations with various other learning paradigms:

**Online / one-shot learning**

Leaning from a single example (can be done in the context of lifelong learning)

**Transfer learning**

A learning paradigm in which previously learned knowledge is used when learning a new task

- In lifelong learning, knowledge reuse can be important, but so is preserving the original knowledge (unlike in transfer learning)

**Few-shot learning**

Learning from a few examples (can also be done in lifelong learning)

# Related Learning Paradigms



Lifelong learning has close relations with various other learning paradigms:

**Online / one-shot learning**

Leaning from a single example (can be done in the context of lifelong learning)

**Transfer learning**

A learning paradigm in which previously learned knowledge is used when learning a new task

- In lifelong learning, knowledge reuse can be important, but so is preserving the original knowledge (unlike in transfer learning)

**Few-shot learning**

Learning from a few examples (can also be done in lifelong learning)

**Curriculum learning**

A training strategy in which tasks to be learned are arranged into a sequence of increasing difficulty, with the task of interest being learned at the end

- The difference with lifelong learning is that curriculum learning really only cares about the last task, while lifelong learning cares about all of them

# Related Learning Paradigms



Timothée Lesort*[,1,2], Vincenzo Lomonaco*[,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez*[,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

Lifelong learning has close relations with various other learning paradigms:

**Online / one-shot learning**

Leaning from a single example (can be done in the context of lifelong learning)

**Few-shot learning**

Learning from a few examples (can also be done in lifelong learning)

**Transfer learning**

A learning paradigm in which previously learned knowledge is used when learning a new task

- In lifelong learning, knowledge reuse can be important, but so is preserving the original knowledge (unlike in transfer learning)

**Curriculum learning**

A training strategy in which tasks to be learned are arranged into a sequence of increasing difficulty, with the task of interest being learned at the end

- The difference with lifelong learning is that curriculum learning really only cares about the last task, while lifelong learning cares about all of them

**Meta-learning**

A learning strategy in which the objective is to learn some meta-information (hyperparameters) that can be used to speed up the learning of later downstream tasks

- Just as in the case of transfer learning, this can be a useful strategy to use for continual learning, but it also doesn't aim at preserving old knowledge

# Related Learning Paradigms

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges

Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

Lifelong learning has close relations with various other learning paradigms:

**Online / one-shot learning**

Leaning from a single example (can be done in the context of lifelong learning)

**Few-shot learning**

Learning from a few examples (can also be done in lifelong learning)

**Transfer learning**

A learning paradigm in which previously learned knowledge is used when learning a new task

- In lifelong learning, knowledge reuse can be important, but so is preserving the original knowledge (unlike in transfer learning)

**Curriculum learning**

A training strategy in which tasks to be learned are arranged into a sequence of increasing difficulty, with the task of interest being learned at the end

- The difference with lifelong learning is that curriculum learning really only cares about the last task, while lifelong learning cares about all of them

**Meta-learning**

A learning strategy in which the objective is to learn some meta-information (hyperparameters) that can be used to speed up the learning of later downstream tasks

- Just as in the case of transfer learning, this can be a useful strategy to use for continual learning, but it also doesn't aim at preserving old knowledge

**Active learning**

See the previous lecture (can be used in lifelong learning)

# Lifelong Learning Strategies



Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

There are different strategies to perform lifelong learning (methods often combine the strategies):

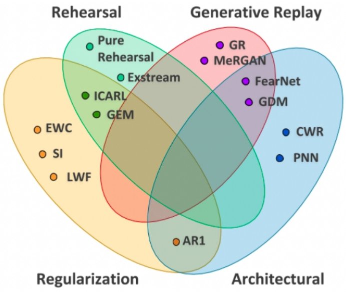
# Lifelong Learning Strategies

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges

Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

There are different strategies to perform lifelong learning (methods often combine the strategies):



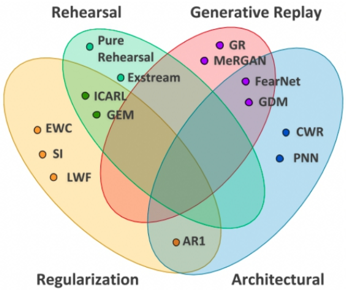
# Lifelong Learning Strategies



Timothée Lesort*[,1,2], Vincenzo Lomonaco*[,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez*[,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

There are different strategies to perform lifelong learning (methods often combine the strategies):

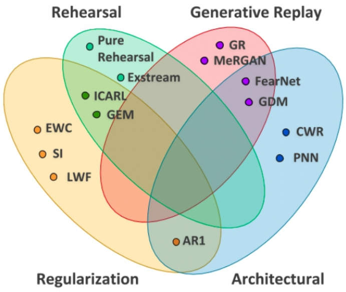



- **Architectural**: The underlying architectural model is changed dynamically when a new task needs to be learned (e.g. new neurons and connections are added in a neural network, or components are added to a GMM)

# Lifelong Learning Strategies



There are different strategies to perform lifelong learning (methods often combine the strategies):

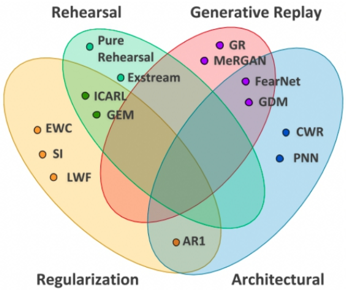



- **Architectural**: The underlying architectural model is changed dynamically when a new task needs to be learned (e.g. new neurons and connections are added in a neural network, or components are added to a GMM)
- **Regularisation**: A regularisation training term is used to mitigate the forgetting of previously learned knowledge

# Lifelong Learning Strategies


Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]
[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

There are different strategies to perform lifelong learning (methods often combine the strategies):




- **Architectural**: The underlying architectural model is changed dynamically when a new task needs to be learned (e.g. new neurons and connections are added in a neural network, or components are added to a GMM)
- **Regularisation**: A regularisation training term is used to mitigate the forgetting of previously learned knowledge
- **Rehearsal**: Data examples from old tasks are kept in memory and added to the dataset when a new task needs to be learned

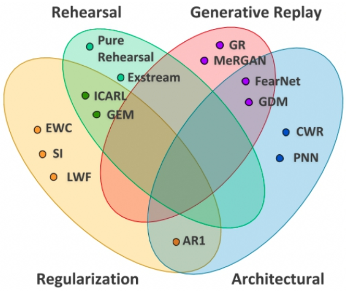
# Lifelong Learning Strategies

Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges

Timothée Lesort[*,1,2], Vincenzo Lomonaco[*,3], Andrei Stoian[2], Davide Maltoni[3], David Filliat[1], and Natalia Díaz-Rodríguez[*,1]

[1]Flowers Team (ENSTA Paris, Institute Polytechnique de Paris & INRIA).
[2]Thales, Theresis Laboratory.
[3]Department of Computer Science and Engineering - University of Bologna

There are different strategies to perform lifelong learning (methods often combine the strategies):




- **Architectural**: The underlying architectural model is changed dynamically when a new task needs to be learned (e.g. new neurons and connections are added in a neural network, or components are added to a GMM)
- **Regularisation**: A regularisation training term is used to mitigate the forgetting of previously learned knowledge
- **Rehearsal**: Data examples from old tasks are kept in memory and added to the dataset when a new task needs to be learned
- **Generative replay**: Related to rehearsal, but instead of keeping old data explicitly, a generative data model is trained instead so that experiences can be sampled from the data distribution

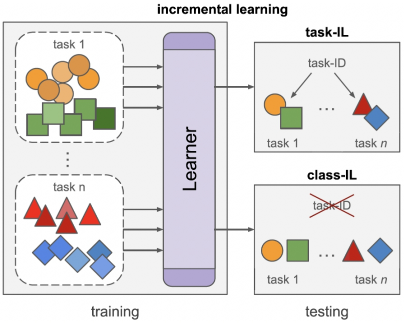
# Class Incremental Learning — Particularly for Neural Networks

# Class Incremental Learning

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer



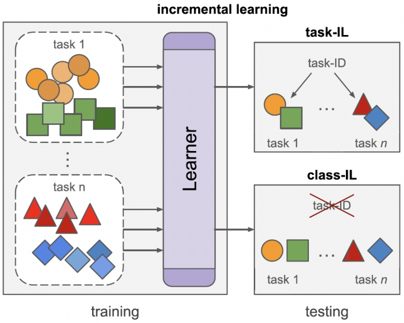

- Class incremental learning is a lifelong learning technique based on which **new classes are included in a recognition system over time**

# Class Incremental Learning

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer



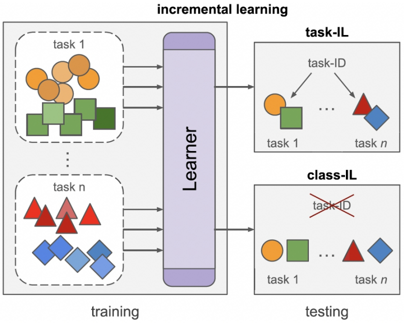

- Class incremental learning is a lifelong learning technique based on which **new classes are included in a recognition system over time**
- In the context of neural networks, this would mean retraining the network with the complete (growing) dataset — not computationally feasible

# Class Incremental Learning

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer



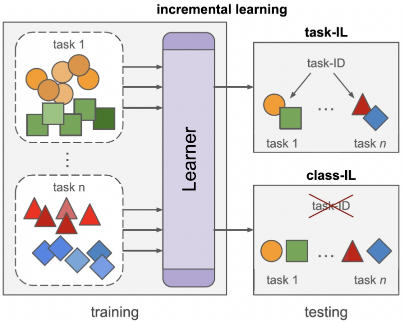

- Class incremental learning is a lifelong learning technique based on which **new classes are included in a recognition system over time**
- In the context of neural networks, this would mean retraining the network with the complete (growing) dataset — not computationally feasible
  - Typical techniques use **exemplars** (representative examples of old classes) instead of keeping the full dataset

# Class Incremental Learning

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer



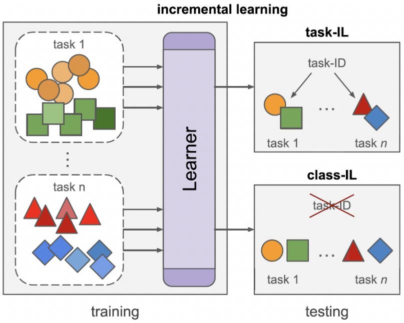

- Class incremental learning is a lifelong learning technique based on which **new classes are included in a recognition system over time**
- In the context of neural networks, this would mean retraining the network with the complete (growing) dataset — not computationally feasible
  - Typical techniques use **exemplars** (representative examples of old classes) instead of keeping the full dataset
- In the case of classification, we distinguish between two problems in the incremental learning context:

# Class Incremental Learning

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer



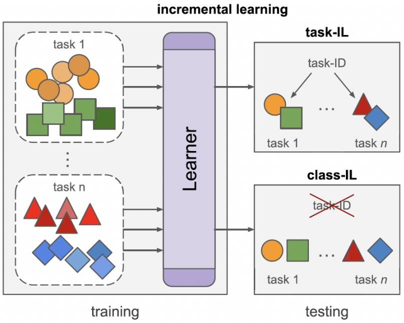

- Class incremental learning is a lifelong learning technique based on which **new classes are included in a recognition system over time**
- In the context of neural networks, this would mean retraining the network with the complete (growing) dataset — not computationally feasible
  - Typical techniques use **exemplars** (representative examples of old classes) instead of keeping the full dataset
- In the case of classification, we distinguish between two problems in the incremental learning context:
  - **Task-IL (task-aware)**: The ID of the task is known, so "only" the class within the task needs to be determined

# Class Incremental Learning

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer




- Class incremental learning is a lifelong learning technique based on which **new classes are included in a recognition system over time**
- In the context of neural networks, this would mean retraining the network with the complete (growing) dataset — not computationally feasible
  - Typical techniques use **exemplars** (representative examples of old classes) instead of keeping the full dataset
- In the case of classification, we distinguish between two problems in the incremental learning context:
  - **Task-IL (task-aware)**: The ID of the task is known, so "only" the class within the task needs to be determined
  - **Class-IL (task-agnostic)**: The ID of the task is unknown (practically a more realistic case)

# Class Incremental Learning Formulation


IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

- In class incremental learning, we assume that we have a sequence of tasks

$$\mathcal{T} = \left[\left(C^1, D^1\right), ..., \left(C^n, D^n\right)\right]$$

# Class Incremental Learning Formulation


IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

- In class incremental learning, we assume that we have a sequence of tasks

$$\mathcal{T} = \left[\left(C^1, D^1\right), ..., \left(C^n, D^n\right)\right]$$

where each task $t$ includes a number of classes

$$C^t = \left\{c_1^t, ..., c_{n^t}^t\right\}$$

# Class Incremental Learning Formulation


IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

- In class incremental learning, we assume that we have a sequence of tasks

$$\mathcal{T} = \left[\left(C^1, D^1\right), ..., \left(C^n, D^n\right)\right]$$

where each task $t$ includes a number of classes

$$C^t = \left\{c_1^t, ..., c_{n^t}^t\right\}$$

and a dataset

$$D^t = \left\{\left(\boldsymbol{x}_1, \boldsymbol{y}_1\right), ..., \left(\boldsymbol{x}_{m^t}, \boldsymbol{y}_{m^t}\right)\right\}$$

# Class Incremental Learning Formulation


IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta,
Andrew D. Bagdanov, and Joost van de Weijer

- In class incremental learning, we assume that we have a sequence of tasks

  $$\mathcal{T} = \left[\left(C^1, D^1\right), ..., \left(C^n, D^n\right)\right]$$

  where each task $t$ includes a number of classes

  $$C^t = \left\{c_1^t, ..., c_{n^t}^t\right\}$$

  and a dataset

  $$D^t = \left\{\left(\boldsymbol{x}_1, \boldsymbol{y}_1\right), ..., \left(\boldsymbol{x}_{m^t}, \boldsymbol{y}_{m^t}\right)\right\}$$

  with features $\boldsymbol{x}_i$ and labels $\boldsymbol{y}_i$ (one-hot encoded). When training on task $t$, only $D^t$ is available to the learning system; the datasets of earlier classes are not

# Class Incremental Learning Formulation


IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

- In class incremental learning, we assume that we have a sequence of tasks

$$\mathcal{T} = \left[\left(C^1, D^1\right), ..., \left(C^n, D^n\right)\right]$$

where each task $t$ includes a number of classes

$$C^t = \left\{c_1^t, ..., c_{n^t}^t\right\}$$

and a dataset

$$D^t = \left\{\left(\boldsymbol{x}_1, \boldsymbol{y}_1\right), ..., \left(\boldsymbol{x}_{m^t}, \boldsymbol{y}_{m^t}\right)\right\}$$

with features $\boldsymbol{x}_i$ and labels $\boldsymbol{y}_i$ (one-hot encoded). When training on task $t$, only $D^t$ is available to the learning system; the datasets of earlier classes are not

- Problem: **The tasks are not all known at the initial training time**

# Incremental Learning Metrics

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

- To measure the incremental learning performance, one common measure is the **average accuracy over the tasks**

$$A_t = \frac{1}{t}\sum_{i=1}^{t} a_{t,i}$$

where $a_{t,k}$ is the accuracy of task $k$ once task $t$ has been learned

# Incremental Learning Metrics

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

- To measure the incremental learning performance, one common measure is the **average accuracy over the tasks**

$$A_t = \frac{1}{t} \sum_{i=1}^{t} a_{t,i}$$

where $a_{t,k}$ is the accuracy of task $k$ once task $t$ has been learned

- Another useful metric is the **average forgetting**, which measures how much the accuracy of a task reduces after learning task $t$:

$$F_t = \frac{1}{t-1} \sum_{i=1}^{t-1} f_{t,i}$$

where the forgetting for task $k$ is defined as:

$$f_{t,k} = \max_{j \in \{1,\ldots,t-1\}} a_{j,k} - a_{t,k}$$

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

- In the context of neural networks (where incremental learning is gaining popularity), we consider a network parameterised by parameters $\boldsymbol{\theta}$ that calculates the logit output as

$$\boldsymbol{o}(\boldsymbol{x}) = h(\boldsymbol{x}, \boldsymbol{\theta})$$

# Incremental Learning for Neural Networks

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

- In the context of neural networks (where incremental learning is gaining popularity), we consider a network parameterised by parameters $\boldsymbol{\theta}$ that calculates the logit output as

$$\boldsymbol{o}(\boldsymbol{x}) = h(\boldsymbol{x}, \boldsymbol{\theta})$$

- We can decompose the network $h$ into a feature extractor $f$ and a linear classifier $g$:

$$\boldsymbol{o}(\boldsymbol{x}) = g\left(f(\boldsymbol{x}, \boldsymbol{\phi}), V\right)$$

  where $V$ are the parameters of the linear classifier

# Incremental Learning for Neural Networks


IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

- In the context of neural networks (where incremental learning is gaining popularity), we consider a network parameterised by parameters $\boldsymbol{\theta}$ that calculates the logit output as

$$\boldsymbol{o}(\boldsymbol{x}) = h(\boldsymbol{x}, \boldsymbol{\theta})$$

- We can decompose the network $h$ into a feature extractor $f$ and a linear classifier $g$:

$$\boldsymbol{o}(\boldsymbol{x}) = g\left(f(\boldsymbol{x}, \boldsymbol{\phi}), V\right)$$

  where $V$ are the parameters of the linear classifier

- As new tasks arrive, the network can be retrained with the cross-entropy loss function

$$\mathcal{L}_c(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}^t) = \sum_{k=1}^{N} y_k \log \frac{\exp\left(\boldsymbol{o}_k\right)}{\sum_{i=1}^{N^t} \exp\left(\boldsymbol{o}_i\right)}$$

  where $N^t$ is the number of all classes up to and including task $t$

# Incremental Learning Challenges

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

There are various challenges in incremental learning that can lead to catastrophic forgetting, (i.e. reduced performance on previously learned tasks), particularly for neural networks:

## Weight drift

The weights need to be updated every time a new task is added to the system; this can lead to **weights that are unfavourable to old tasks**

# Incremental Learning Challenges

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

There are various challenges in incremental learning that can lead to catastrophic forgetting, (i.e. reduced performance on previously learned tasks), particularly for neural networks:

## Weight drift

The weights need to be updated every time a new task is added to the system; this can lead to **weights that are unfavourable to old tasks**

## Activation drift

Weight changes will usually also **affect the outputs of activations functions** in different network layers, which can also have a detrimental effect on the overall performance

# Incremental Learning Challenges

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

There are various challenges in incremental learning that can lead to catastrophic forgetting, (i.e. reduced performance on previously learned tasks), particularly for neural networks:

## Weight drift

The weights need to be updated every time a new task is added to the system; this can lead to **weights that are unfavourable to old tasks**

## Activation drift

Weight changes will usually also **affect the outputs of activations functions** in different network layers, which can also have a detrimental effect on the overall performance

## Inter-task confusion

A problem during task-agnostic classification, which can arise if training is performed only on data from a new task, as **the classifier may not learn how to distinguish between all classes up to that point**

# Incremental Learning Challenges

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

There are various challenges in incremental learning that can lead to catastrophic forgetting, (i.e. reduced performance on previously learned tasks), particularly for neural networks:

## Weight drift

The weights need to be updated every time a new task is added to the system; this can lead to **weights that are unfavourable to old tasks**

## Activation drift

Weight changes will usually also **affect the outputs of activations functions** in different network layers, which can also have a detrimental effect on the overall performance

## Inter-task confusion

A problem during task-agnostic classification, which can arise if training is performed only on data from a new task, as **the classifier may not learn how to distinguish between all classes up to that point**

## Task-recency bias

Due to intensive training on the most recent task, **a network may be biased towards newly added classes**

# Strategies to Overcome the Challenges

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

There are different approaches that aim to address / mitigate the incremental learning challenges:

## Rehearsal

When learning a new task, **exemplar samples from previous tasks** are used in addition to data from the current task

# Strategies to Overcome the Challenges

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

There are different approaches that aim to address / mitigate the incremental learning challenges:

## Rehearsal

When learning a new task, **exemplar samples from previous tasks** are used in addition to data from the current task

## Regularisation

The network is updated not only by minimising a cross-entropy loss over the outputs, but also by including a **loss regularisation term** to control the weight updates

- Regularisation can be performed directly over the weights (weight regularisation) or over the data by using a distillation loss (data regularisation)
- Regularisation and rehearsal are sometimes combined as well

# Strategies to Overcome the Challenges


IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

There are different approaches that aim to address / mitigate the incremental learning challenges:

## Rehearsal

When learning a new task, **exemplar samples from previous tasks** are used in addition to data from the current task

## Regularisation

The network is updated not only by minimising a cross-entropy loss over the outputs, but also by including a **loss regularisation term** to control the weight updates

- Regularisation can be performed directly over the weights (weight regularisation) or over the data by using a distillation loss (data regularisation)
- Regularisation and rehearsal are sometimes combined as well

## Bias correction

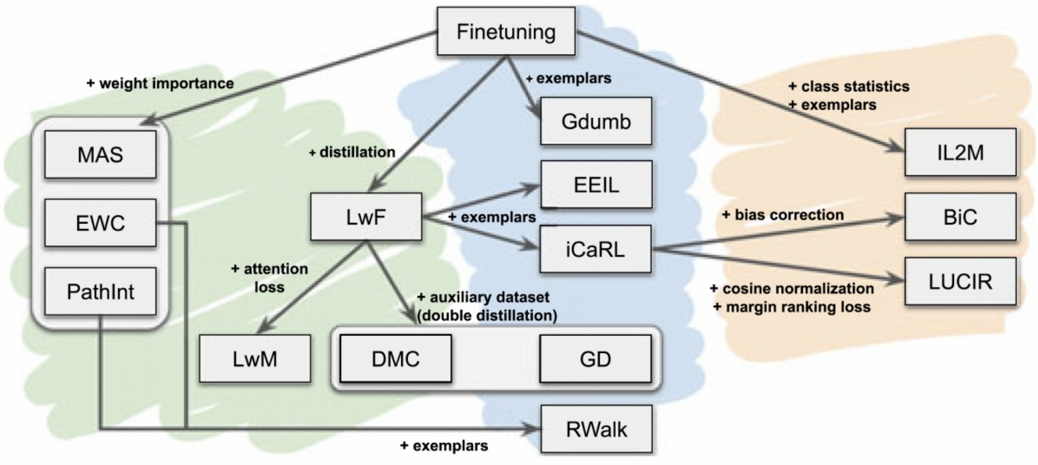
Aims to prevent bias towards new tasks, for instance by **using a different model for classification (not the network itself)** or by **introducing an extra training step to remedy the bias of the network predictions**

# Incremental Learning Algorithms Overview

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 5, MAY 2023 5513

Class-Incremental Learning: Survey and Performance Evaluation on Image Classification

Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer

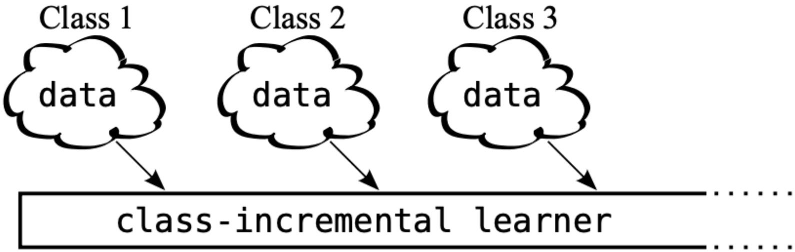


# Class Incremental Learning Algorithm Example: iCaRL

# iCaRL



- iCaRL (incremental classifier and representation learning) is a class incremental learning method for neural network-based classifiers, published at CVPR in 2017 (in machine learning time, that is a whole eternity)
- The method illustrates the techniques for overcoming the challenges of incremental learning — rehearsal, regularisation, and bias correction are all incorporated in iCaRL
- We will look at iCaRL in some detail on the next few slides



# Exemplar Set in iCaRL

- iCaRL does not preserve the raw data from old classes, but keeps **exemplars** (representative examples of each class)
  - The exemplars are extracted from the latent representation of the learned neural network

# Exemplar Set in iCaRL



- iCaRL does not preserve the raw data from old classes, but keeps **exemplars** (representative examples of each class)
  - The exemplars are extracted from the latent representation of the learned neural network
- The exemplar set is used both during learning (this is an example of **rehearsal** to facilitate remembering old classes) and for classification

# Exemplar Set in iCaRL



Sylvestre-Alvise Rebuffi
University of Oxford/IST Austria

Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert
IST Austria

- iCaRL does not preserve the raw data from old classes, but keeps **exemplars** (representative examples of each class)
  - The exemplars are extracted from the latent representation of the learned neural network
- The exemplar set is used both during learning (this is an example of **rehearsal** to facilitate remembering old classes) and for classification
- **The total number of exemplars in iCaRL is fixed**; thus, the exemplar set for old classes is reduced each time a new task is learned
  - Can be detrimental when many classes are learned, as there will only be a few exemplars per class in this case; growing memory can perform better

# Exemplar Set in iCaRL

iCaRL: Incremental Classifier and Representation Learning

Sylvestre-Alvise Rebuffi
University of Oxford/IST Austria

Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert
IST Austria

- iCaRL does not preserve the raw data from old classes, but keeps **exemplars** (representative examples of each class)
  - The exemplars are extracted from the latent representation of the learned neural network
- The exemplar set is used both during learning (this is an example of **rehearsal** to facilitate remembering old classes) and for classification
- **The total number of exemplars in iCaRL is fixed**; thus, the exemplar set for old classes is reduced each time a new task is learned
  - Can be detrimental when many classes are learned, as there will only be a few exemplars per class in this case; growing memory can perform better

**Algorithm 4** iCaRL ConstructExemplarSet

**input** image set $X = \{x_1, \dots, x_n\}$ of class $y$
**input** $m$ target number of exemplars
**require** current feature function $\varphi : \mathcal{X} \to \mathbb{R}^d$
  $\mu \leftarrow \frac{1}{n} \sum_{x \in X} \varphi(x)$ // current class mean
  **for** $k = 1, \dots, m$ **do**
    $p_k \leftarrow \operatorname*{argmin}_{x \in X} \left\| \mu - \frac{1}{k}[\varphi(x) + \sum_{j=1}^{k-1} \varphi(p_j)] \right\|$
  **end for**
  $P \leftarrow (p_1, \dots, p_m)$
**output** exemplar set $P$

**Algorithm 5** iCaRL ReduceExemplarSet

**input** $m$ // target number of exemplars
**input** $P = (p_1, \dots, p_{|P|})$ // current exemplar set
  $P \leftarrow (p_1, \dots, p_m)$ // *i.e.* keep only first $m$
**output** exemplar set $P$

Sylvestre-Alvise Rebuffi
University of Oxford/IST Austria
Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert
IST Austria

# iCaRL Training Process

The complete training process of iCaRL involves three main steps:

# iCaRL Training Process

The complete training process of iCaRL involves three main steps:

1. **The network representation is updated** given data about new tasks

# iCaRL Training Process

The complete training process of iCaRL involves three main steps:

1. **The network representation is updated** given data about new tasks
2. **The least important exemplars from the old classes are removed** (so that the fixed memory can be preserved)

# iCaRL Training Process



Sylvestre-Alvise Rebuffi
University of Oxford/IST Austria
Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert
IST Austria

The complete training process of iCaRL involves three main steps:

1. **The network representation is updated** given data about new tasks
2. **The least important exemplars from the old classes are removed** (so that the fixed memory can be preserved)
3. **An exemplar set for the newly added classes is created** and added to the memory

# iCaRL Training Process



Sylvestre-Alvise Rebuffi
University of Oxford/IST Austria

Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert
IST Austria

The complete training process of iCaRL involves three main steps:

1. **The network representation is updated** given data about new tasks
2. **The least important exemplars from the old classes are removed** (so that the fixed memory can be preserved)
3. **An exemplar set for the newly added classes is created** and added to the memory

**Algorithm 2** iCaRL INCREMENTALTRAIN

**input** $X^s, \dots, X^t$ // training examples in per-class sets
**input** $K$ // memory size
**require** $\Theta$ // current model parameters
**require** $\mathcal{P} = (P_1, \dots, P_{s-1})$ // current exemplar sets
  $\Theta \leftarrow$ UPDATEREPRESENTATION($X^s, \dots, X^t; \mathcal{P}, \Theta$)
  $m \leftarrow K/t$ // number of exemplars per class
  **for** $y = 1, \dots, s-1$ **do**
    $P_y \leftarrow$ REDUCEEXEMPLARSET($P_y, m$)
  **end for**
  **for** $y = s, \dots, t$ **do**
    $P_y \leftarrow$ CONSTRUCTEXEMPLARSET($X_y, m, \Theta$)
  **end for**
  $\mathcal{P} \leftarrow (P_1, \dots, P_t)$ // new exemplar sets

Hochschule Bonn-Rhein-Sieg University of Applied Sciences
b-it Bonn-Aachen International Center for Information Technology
Institute for AI and Autonomous Systems

# Learning in iCaRL



- When a new task needs to be learned, iCaRL updates the network using a dataset consisting of the **data from the new task** and **the exemplars of old classes**

# Learning in iCaRL

- When a new task needs to be learned, iCaRL updates the network using a dataset consisting of the **data from the new task** and **the exemplars of old classes**
- The loss function that is minimised during learning has two components:

# Learning in iCaRL



- When a new task needs to be learned, iCaRL updates the network using a dataset consisting of the **data from the new task** and **the exemplars of old classes**
- The loss function that is minimised during learning has two components:
  - **Classification loss**: The cross-entropy loss is used, where $g_y(x_i)$ is the network's output (obtained with the logistic function)

# Learning in iCaRL



- When a new task needs to be learned, iCaRL updates the network using a dataset consisting of the **data from the new task** and **the exemplars of old classes**
- The loss function that is minimised during learning has two components:
  - **Classification loss**: The cross-entropy loss is used, where $g_y(x_i)$ is the network's output (obtained with the logistic function)
  - **Distillation loss**: Serves as a regularisation term that should keep the network outputs close to the ones learned previously

# Learning in iCaRL



Sylvestre-Alvise Rebuffi
University of Oxford/IST Austria

Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert
IST Austria

- When a new task needs to be learned, iCaRL updates the network using a dataset consisting of the **data from the new task** and **the exemplars of old classes**
- The loss function that is minimised during learning has two components:
  - **Classification loss**: The cross-entropy loss is used, where $g_y(x_i)$ is the network's output (obtained with the logistic function)
  - **Distillation loss**: Serves as a regularisation term that should keep the network outputs close to the ones learned previously

**Algorithm 3** iCaRL UPDATEREPRESENTATION

**input** $X^s, \dots, X^t$ // training images of classes $s, \dots, t$
**require** $\mathcal{P} = (P_1, \dots, P_{s-1})$ // exemplar sets
**require** $\Theta$ // current model parameters
// form combined training set:

$$\mathcal{D} \leftarrow \bigcup_{y=s,\dots,t} \{(x,y) : x \in X^y\} \cup \bigcup_{y=1,\dots,s-1} \{(x,y) : x \in P^y\}$$

// store network outputs with pre-update parameters:
**for** $y = 1, \dots, s-1$ **do**
  $q_i^y \leftarrow g_y(x_i)$ for all $(x_i, \cdot) \in \mathcal{D}$
**end for**
run network training (*e.g.* BackProp) with loss function

$$\ell(\Theta) = -\sum_{(x_i,y_i)\in\mathcal{D}} \Big[ \sum_{y=s}^{t} \delta_{y=y_i} \log g_y(x_i) + \delta_{y\neq y_i} \log(1-g_y(x_i)) + \sum_{y=1}^{s-1} q_i^y \log g_y(x_i) + (1-q_i^y)\log(1-g_y(x_i)) \Big]$$

that consists of *classification* and *distillation* terms.

# iCaRL Classification



Sylvestre-Alvise Rebuffi
University of Oxford/IST Austria

Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert
IST Austria

- Classification in iCaRL is not performed by the neural network, but instead by a **nearest-mean-of-exemplars** classifier

# iCaRL Classification



- Classification in iCaRL is not performed by the neural network, but instead by a **nearest-mean-of-exemplars** classifier

- The neural network is used to extract $\varphi(\boldsymbol{x})$, the latent representation of the example $\boldsymbol{x}$ to be classified; the class of the example is then determined to be **the one whose mean of the exemplars is closest**

# iCaRL Classification



- Classification in iCaRL is not performed by the neural network, but instead by a **nearest-mean-of-exemplars** classifier
- The neural network is used to extract $\varphi(\boldsymbol{x})$, the latent representation of the example $\boldsymbol{x}$ to be classified; the class of the example is then determined to be **the one whose mean of the exemplars is closest**
- This is a strategy to reduce bias in the model (as discussed before), but also to make the classifier robust to large changes in the latent network representation (due to new classes)

# iCaRL Classification



Sylvestre-Alvise Rebuffi
University of Oxford/IST Austria

Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert
IST Austria

- Classification in iCaRL is not performed by the neural network, but instead by a **nearest-mean-of-exemplars** classifier
- The neural network is used to extract $\varphi(\boldsymbol{x})$, the latent representation of the example $\boldsymbol{x}$ to be classified; the class of the example is then determined to be **the one whose mean of the exemplars is closest**
- This is a strategy to reduce bias in the model (as discussed before), but also to make the classifier robust to large changes in the latent network representation (due to new classes)

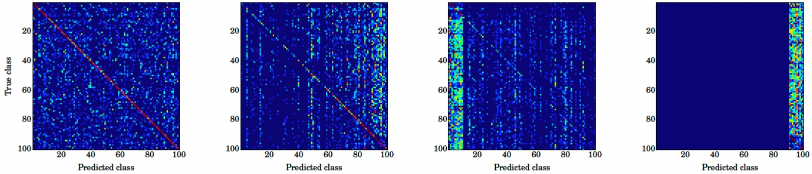
**Algorithm 1** iCaRL CLASSIFY

**input** $x$ // image to be classified
**require** $\mathcal{P} = (P_1, \ldots, P_t)$ // class exemplar sets
**require** $\varphi : \mathcal{X} \to \mathbb{R}^d$ // feature map
  **for** $y = 1, \ldots, t$ **do**
    $\mu_y \leftarrow \frac{1}{|P_y|} \sum_{p \in P_y} \varphi(p)$ // mean-of-exemplars
  **end for**
  $y^* \leftarrow \operatorname*{argmin}_{y=1,\ldots,t} \|\varphi(x) - \mu_y\|$ // nearest prototype
**output** class label $y^*$

# iCaRL Performance



Sylvestre-Alvise Rebuffi
University of Oxford/IST Austria
Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert
IST Austria




(a) iCaRL (b) LwF.MC (c) fixed representation (d) finetuning

- Here, we can see confusion matrices of iCaRL and a few other methods (Learning without Forgetting — another incremental learning algorithm — a model that uses a fixed latent representation, and one that simply performs finetuning on the new classes)
  - The evaluation is on the CIFAR-100 dataset with each task consisting of 10 classes
- iCaRL's confusion matrix is distributed among the classes, which means that the algorithm did not exhibit concrete biases (towards old or new classes)
- It should be noted that various newer algorithms, which are improved versions of iCaRL, have been published in the last few years

# Use Case: Lifelong Learning for Action Recognition

# Lifelong Learning for Action Recognition

- **In our MigrAVE project, one objective has been to recognise actions, but it should also be possible to incorporate new actions into the recognition system** (without significant data requirements, as actions may be demonstrated by therapists)

# Lifelong Learning for Action Recognition



- **In our MigrAVE project, one objective has been to recognise actions, but it should also be possible to incorporate new actions into the recognition system** (without significant data requirements, as actions may be demonstrated by therapists)
- **Action recognition is performed on the basis of skeleton data** in our case

# Lifelong Learning for Action Recognition



- **In our MigrAVE project, one objective has been to recognise actions, but it should also be possible to incorporate new actions into the recognition system** (without significant data requirements, as actions may be demonstrated by therapists)
- **Action recognition is performed on the basis of skeleton data** in our case
- In a completed master's thesis in this context, we investigated various action recognition models on publicly available datasets and then evaluated lifelong learning algorithms for action recognition

# Lifelong Learning for Action Recognition

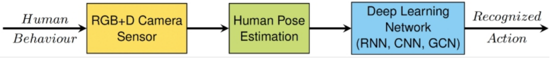
- **In our MigrAVE project, one objective has been to recognise actions, but it should also be possible to incorporate new actions into the recognition system** (without significant data requirements, as actions may be demonstrated by therapists)
- **Action recognition is performed on the basis of skeleton data** in our case
- In a completed master's thesis in this context, we investigated various action recognition models on publicly available datasets and then evaluated lifelong learning algorithms for action recognition
- We then performed a real-user evaluation in which multiple people demonstrated different actions to our QTrobot; these were then incorporated into an action recognition model

# Lifelong Learning for Action Recognition

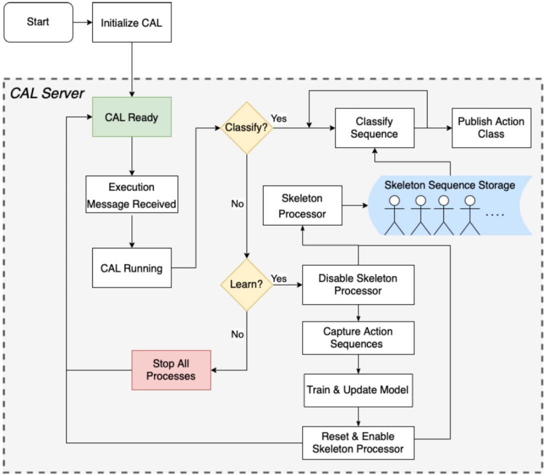


- **In our MigrAVE project, one objective has been to recognise actions, but it should also be possible to incorporate new actions into the recognition system** (without significant data requirements, as actions may be demonstrated by therapists)
- **Action recognition is performed on the basis of skeleton data** in our case
- In a completed master's thesis in this context, we investigated various action recognition models on publicly available datasets and then evaluated lifelong learning algorithms for action recognition
- We then performed a real-user evaluation in which multiple people demonstrated different actions to our QTrobot; these were then incorporated into an action recognition model
- A graph convolutional neural network was used for recognition (CTR-GCN) and BiC as a lifelong learning method (BiC is based on iCaRL, but performs bias correction by introducing a second training step for this purpose)

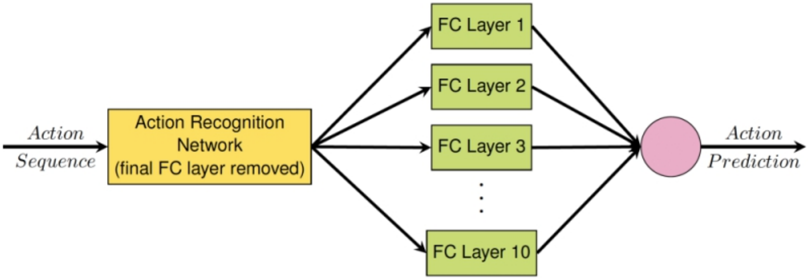
# System Overview




- **Skeleton data** are collected with QTrobot
- The action recognition model processes a **skeleton sequence** to recognise actions
- If desired, new actions can be **demonstrated and integrated into the recognition system**



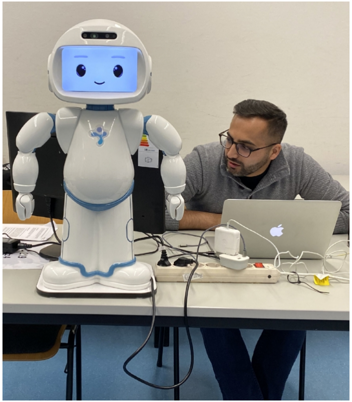
# Sample Model and Tasks




| Task | Actions in each task |
|---|---|
| 1 | Wipe<br>Eat |
| 2 | Cross hands<br>Clap |
| 3 | Step on something<br>Shake head |
| 4 | Sit down<br>Play on phone |
| 5 | Lift something<br>Brush teeth |
| 6 | Throw<br>Point at something |
| 7 | Wave<br>Stand up |
| 8 | Nod<br>Hop |
| 9 | Drop something<br>Drink water |
| 10 | Rub hands<br>Jump |

# Experiment Setup

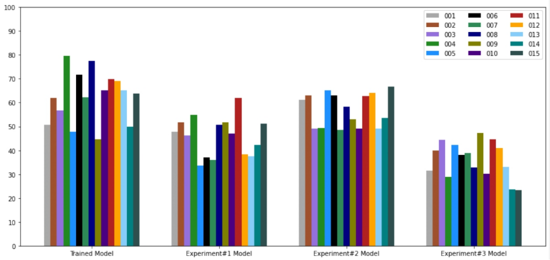
- In our evaluation with QTrobot, we had 15 participants
- The evaluation study involved two steps:
  1. **Actions were recognised using a pretrained baseline model**
  2. **Data for lifelong learning were collected** (two new actions) **and then actions were recognised using the new model** (three different models were learned)
- **Actions to recognise**: Play with phone, hop, rub hands, wave, shake head, drink water
- **Actions to learn**: Talk on the phone, cut food, wave, hop (the last two were finetuned with the new data)



# Evaluation Results

The diagram here shows the results of our evaluation (recognition accuracy per participant)

1. **Trained model**: The pretrained model
2. **Experiment#1 model**: New actions (cut food and talk on the phone) were integrated as a new task
3. **Experiment#2 model**: Hop and wave were retrained with the new data
4. **Experiment#3 model**: Cut food was added to an old task, while talk on the phone was added as a new task




Evaluation results per test user

The results clearly show a decrease in accuracy when adding new knowledge — particularly in the case where an old task is modified (indicating lack of robustness of the methods)

# Summary: Lifelong Learning

- Lifelong learning (aka incremental or continual learning) is a paradigm in which new knowledge is integrated into a learning-based model over time rather than all at once
- There are various ways in which lifelong learning can be achieved, such as by performing architectural changes, using regularisation, by performing rehearsal, or by generative replay
- Class incremental learning is an instance of the lifelong learning problem in which new classes are integrated into a classifier, such that there are two instances of the problem: task-aware and task-agnostic incremental learning
- iCaRL is one of the older class incremental learning algorithms that uses a combination of techniques to solve the problem, namely rehearsal (using exemplars), regularisation (through the learning loss), and bias correction (by performing classification using the nearest-mean-of-exemplars)
- Lifelong learning is associated with various challenges: it requires a suitable representation that is prone to being updated, but there are also problems related to data forgetting and privacy, data distribution shifts and new data demands, and changes in the optimisation objective