Continual Learning with deep neural networks
(and links to the brain)

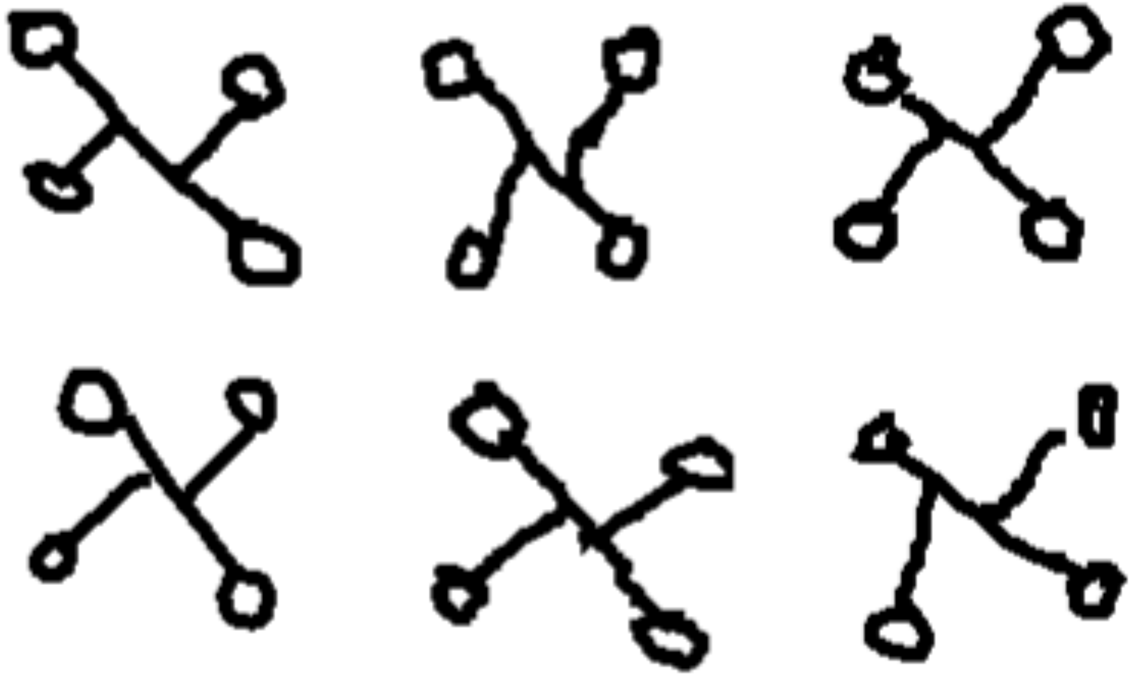
Gido van de Ven

Guest lecture, UMass

7 March 2023

Illustrative Problem

Class 1

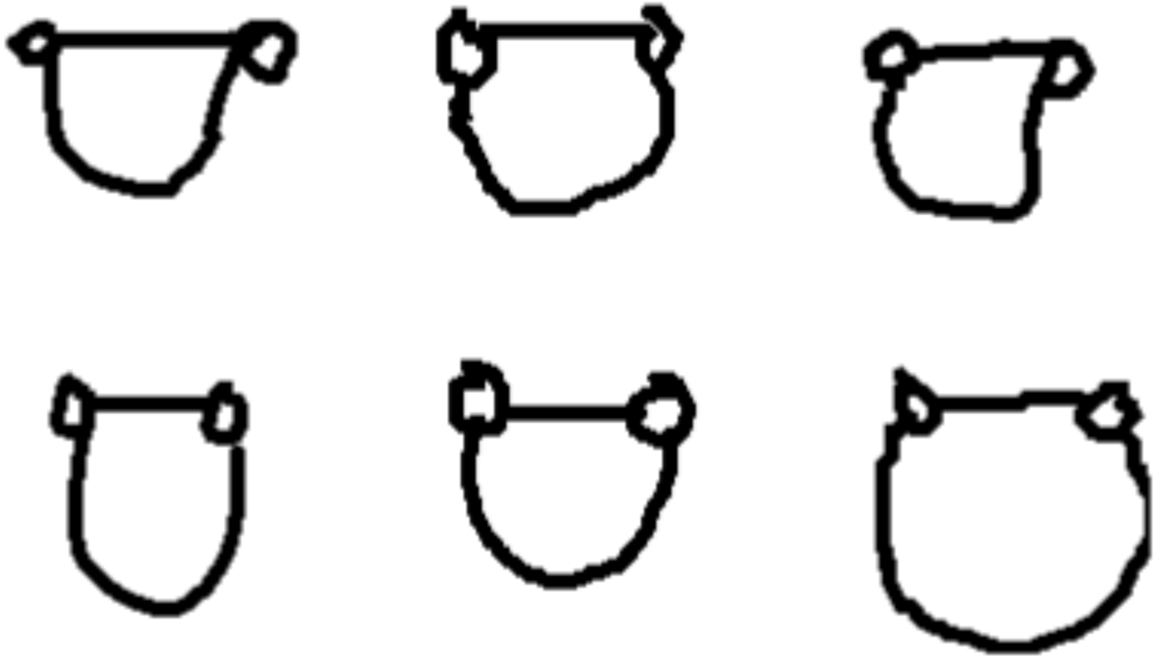


Illustrative Problem

– Break –

Illustrative Problem

Class 2



Illustrative Problem

– Break –

Illustrative Problem

To which class does this example belong?

→ Class 1!

Overview of this lecture

- Introduction to continual learning
- Let’s be clear about the problem: three continual learning scenarios
- **Strategy 1:** Generative replay
- **Strategy 2:** Generative classifiers
- Which strategy do *we* use?

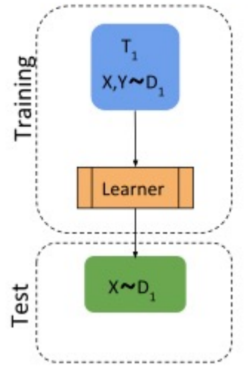
What is continual learning?

- In *classical machine learning*, an algorithm has access to all training data at the same time

- With *continual learning*, two key differences are:
  - the training data arrives incrementally
  - the distribution from which the training data is sampled changes over time

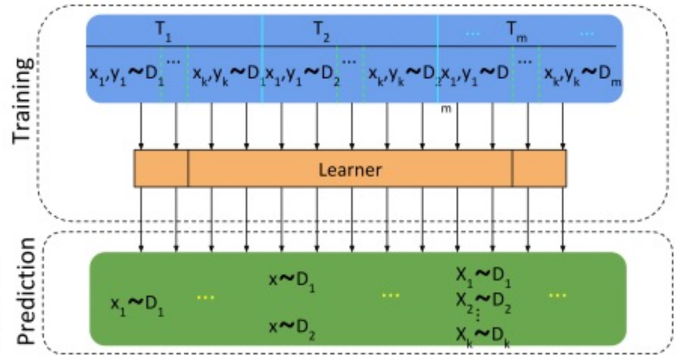
Continual learning in relation to other fields

Standard ML



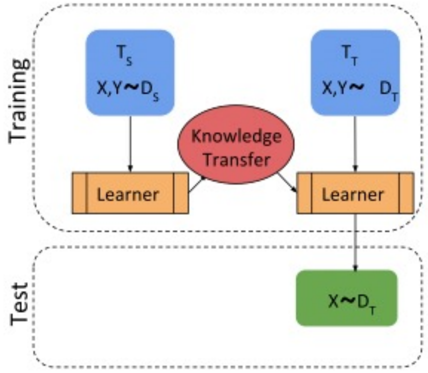
- One task
- Data available at same time

Continual Learning



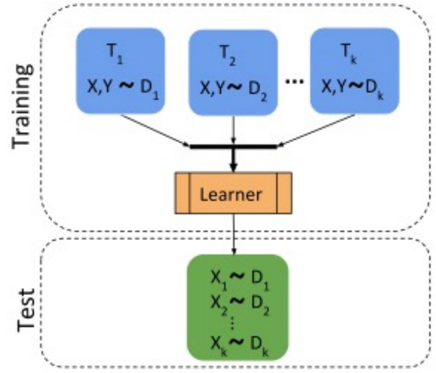
- Multiple tasks
- Data arrive incrementally — non-stationarity
- Goal: all tasks

Transfer Learning



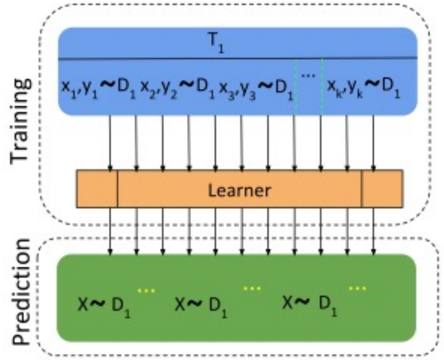
- Multiple tasks
- Data arrive incrementally
- Goal: last task

Multitask Learning



- Multiple tasks
- Data available at same time
- Goal: all tasks

Online Learning

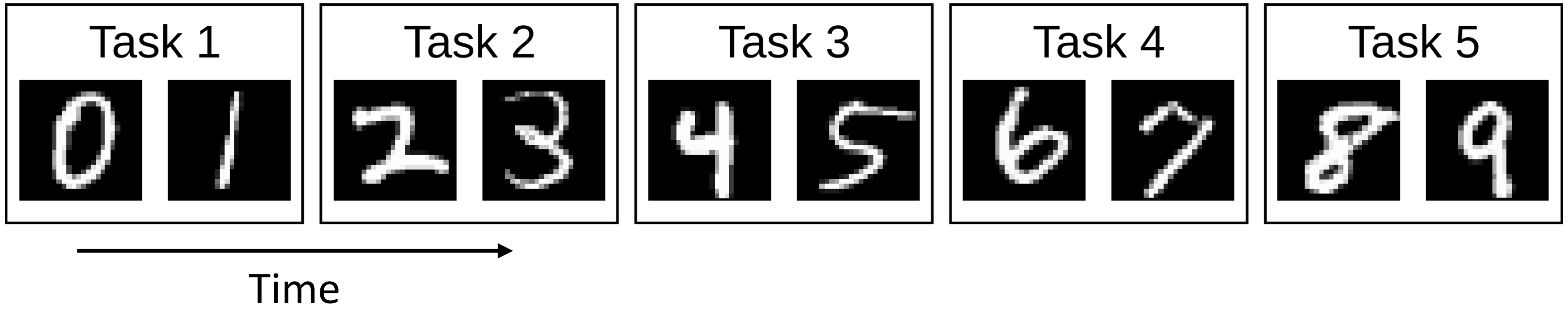


- One task
- Data arrive incrementally

Source: [De Lange et al. (2022, TPAMI)]

The canonical continual learning example: Split MNIST

- MNIST dataset is split in multiple parts/episodes/tasks that must be learned sequentially
- After all tasks have been learned, the model should be good at all tasks
- Typically, no or only a small amount of data from past tasks can be stored

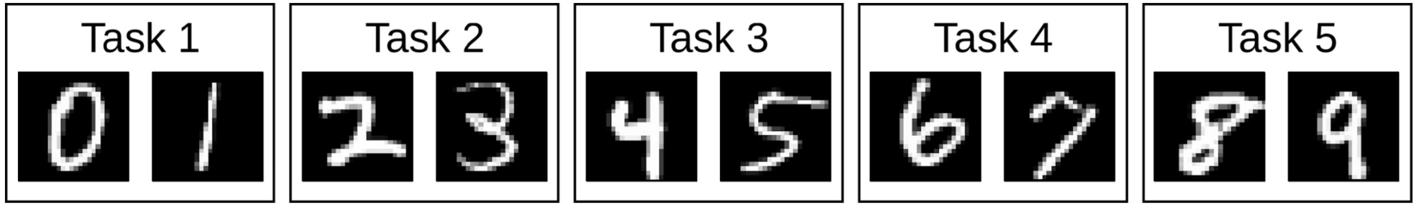


Important problem: ***catastrophic forgetting***

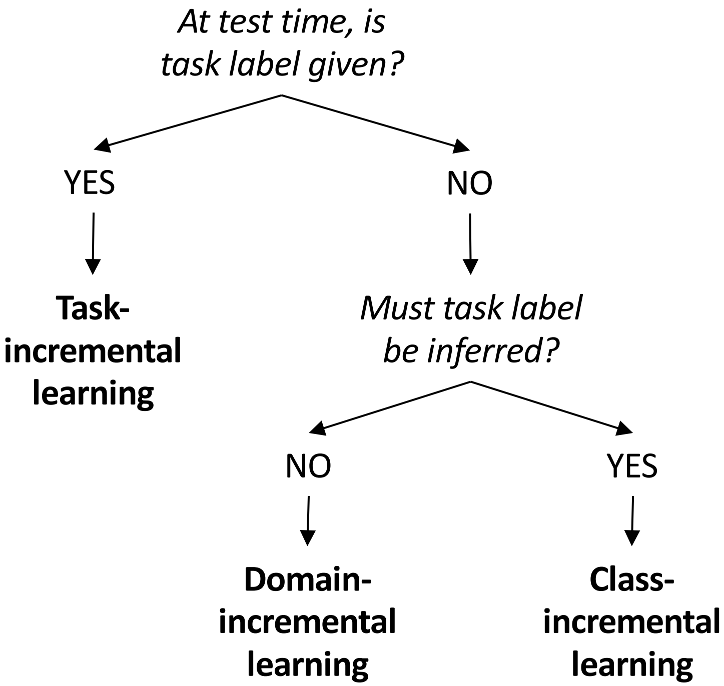
➢ When learning a new task, deep neural networks tend to rapidly forget past tasks

Let’s be clear about the continual learning problem …

Split MNIST:



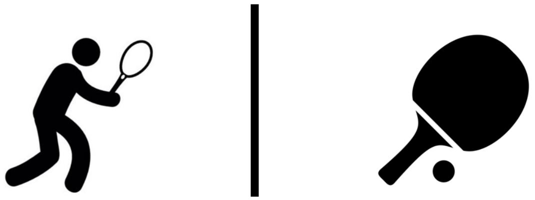
	Type of choice
Task-incremental	Choice between the two digits of the task
Domain-incremental	Is the digit odd or even?
Class-incremental	Choice between all ten digits



Van de Ven & Tolias (2018) *NeurIPS Workshop*; **van de Ven et al (2022)** *Nature Machine Intelligence*

Three continual learning scenarios

- Task-incremental learning *(Task-IL)*
  - Incrementally learn a set of clearly distinguishable tasks

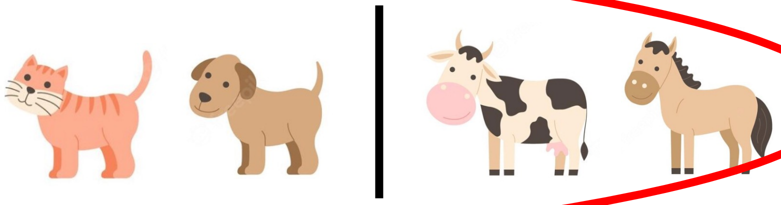


Important challenge: achieve positive transfer between tasks

- Domain-incremental learning *(Domain-IL)*
  - Learn the same type of problem in different contexts

Important challenge: alleviate catastrophic forgetting

- Class-incremental learning *(Class-IL)*
  - Incrementally learn a growing number of classes



Important challenge: learn to discriminate between objects not observed together

Images designed by Freepik

Strategy 1: Generative replay

Deep Learning:

- Interleaved learning prevents catastrophic forgetting

[McCloskey & Cohen, 1989 *Psych Learn Motiv*; Ratcliff, 1990 *Psych Rev*]

⟷

Neuroscience:

- Replay is hypothesized to have an important role in memory consolidation

[Wilson & McNaughton, 1994 *Science;* O'Neill et al., 2010 *TINS*; van de Ven et al., 2016 *Neuron*]

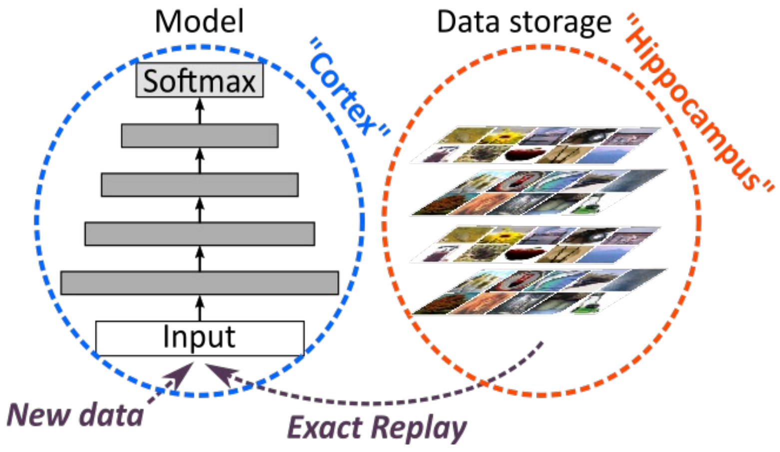
Motivation:

➢ **Use replay to enable deep neural networks to do 'continual learning'**

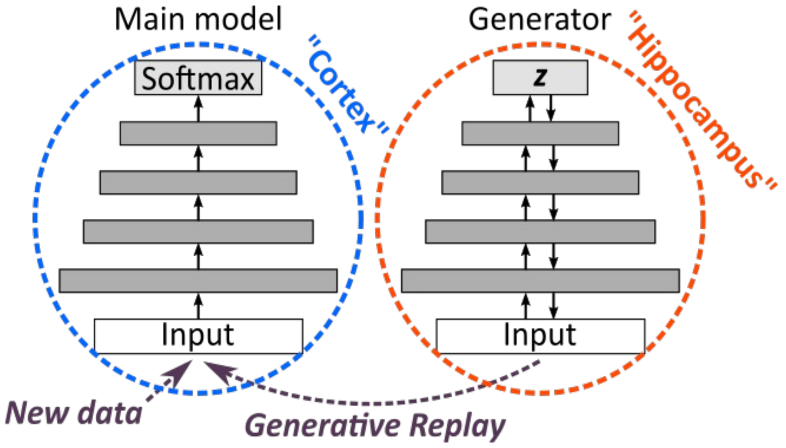
➢ **Use artificial neural networks as a computational model for replay in the brain**

How to add replay to artificial neural networks?

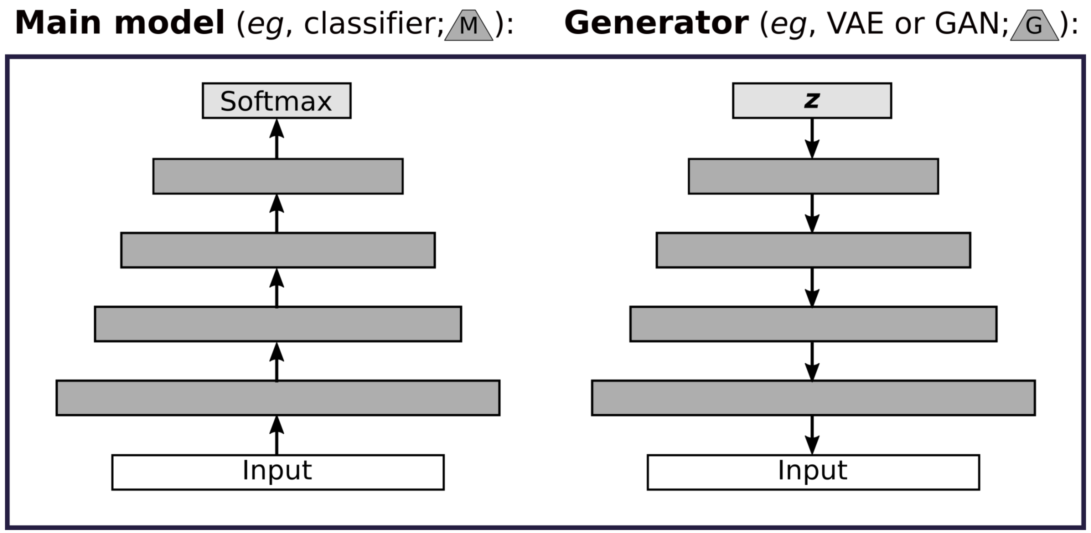
- Store data and interleave – *"exact"* or *"experience replay"*
  - *Initial argument for role of replay in memory consolidation in the brain* [McClelland et al., 1995 *Psych Rev*]
  - *Unclear how the brain could do directly store data*
  - Not always possible (e.g., privacy concerns, limited storage)
  - Problematic when scaling up to true lifelong learning



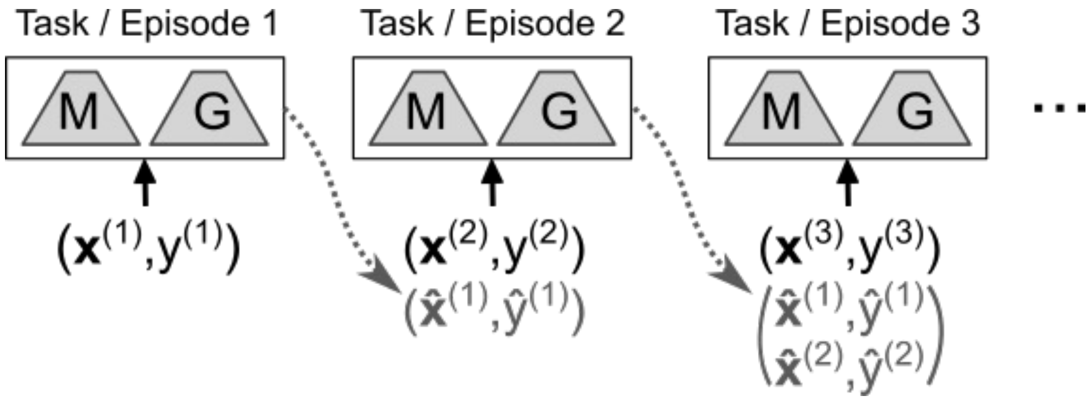
- Use a generative model – *"generative replay"*
  - *More realistic from neuroscience point of view*
  - *Views hippocampus as a generative neural network and replay as a generative process; see also* [Liu et al., 2018 *Neuron*; Liu et al., 2019 *Cell*]
  - Learning a generative model as a more scalable, privacy-preserving way of remembering previous seen data



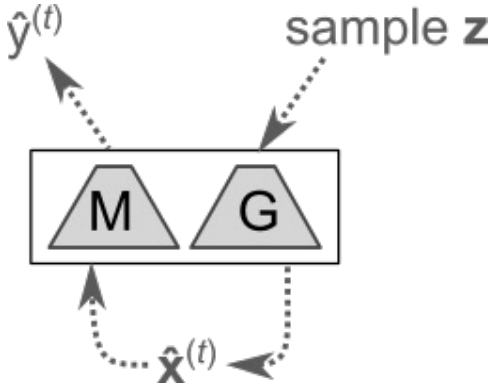
How to implement generative replay?



Incremental training protocol:

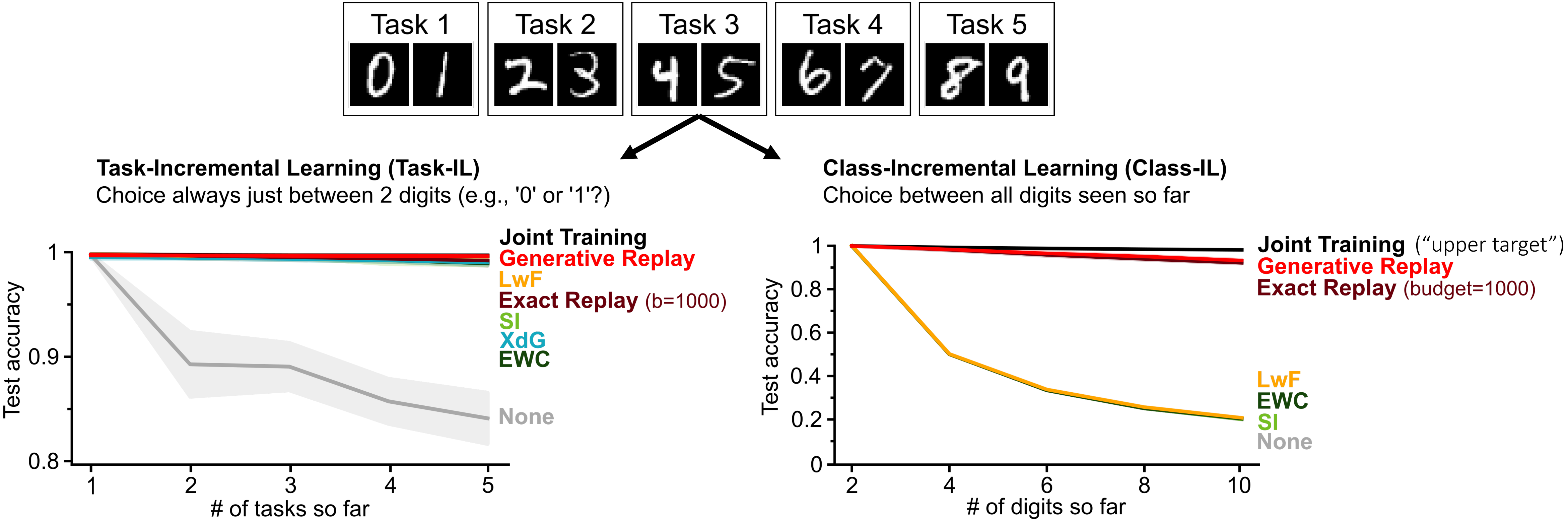


Generation of a sample to be replayed:



Shin et al. (2017) *NeurIPS*

Does generative replay work?



But… (1) MNIST digits are relatively easy to generate

(2) constantly retraining on all previous tasks seems very inefficient

PyTorch code for these experiments: https://github.com/GMvandeVen/continual-learning

Efficiency: How much replay is needed?

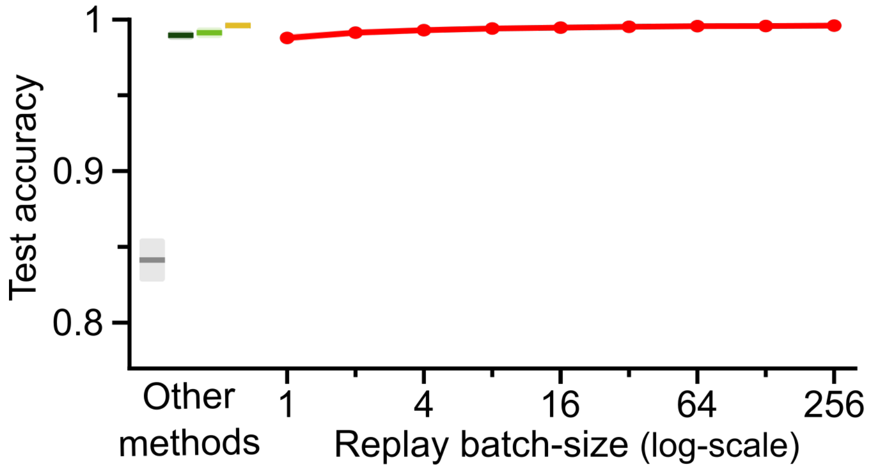
- Previous tasks' datasets do not need to be replayed "fully"
- How far could the number of replayed sampled per batch be reduced?

$$\mathcal{L}_{\text{total}} = \frac{1}{N_{\text{tasks so far}}}\mathcal{L}_{\text{current}} + \left(1 - \frac{1}{N_{\text{tasks so far}}}\right)\mathcal{L}_{\text{replay}}$$

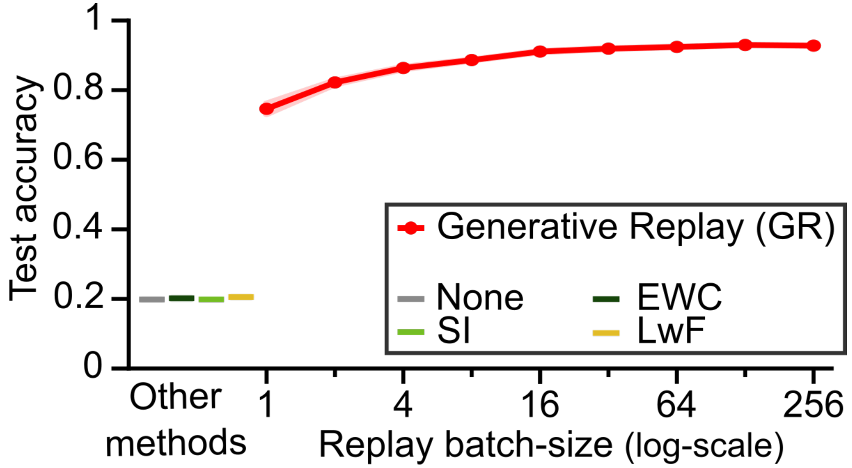
$N_{\text{current-batch}} = 128$

$N_{\text{replay-batch}} = ?$

Task-incremental learning



Class-incremental learning



→ ***Fully replaying previous tasks is not needed, replaying only a few examples could suffice***

(these experiments are on Split MNIST)

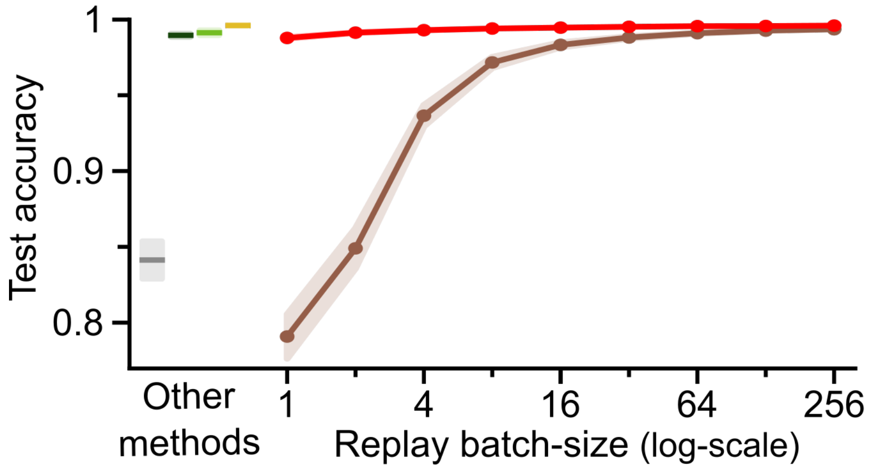
Efficiency: How much replay is needed?

- Previous tasks' datasets do not need to be replayed "fully"
- How far could the number of replayed sampled per batch be reduced?

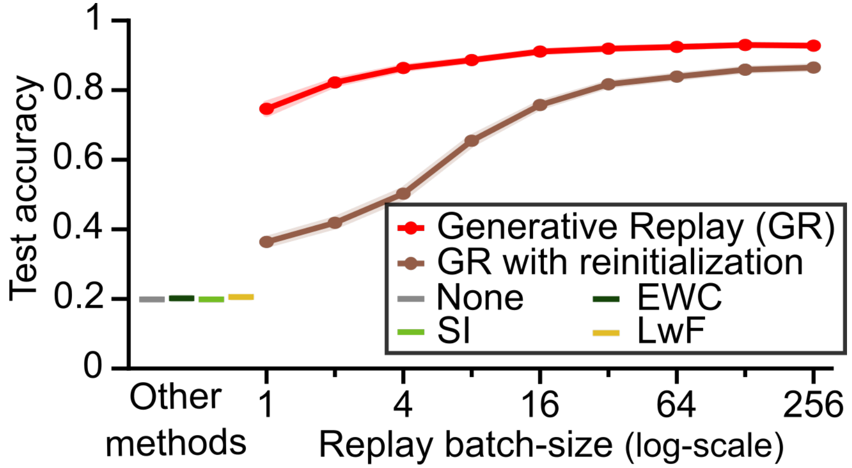
$$\mathcal{L}_{\text{total}} = \frac{1}{N_{\text{tasks so far}}}\mathcal{L}_{\text{current}} + \left(1 - \frac{1}{N_{\text{tasks so far}}}\right)\mathcal{L}_{\text{replay}}$$

$N_{\text{current-batch}} = 128$ $\qquad$ $N_{\text{replay-batch}} = ?$

Task-incremental learning



Class-incremental learning

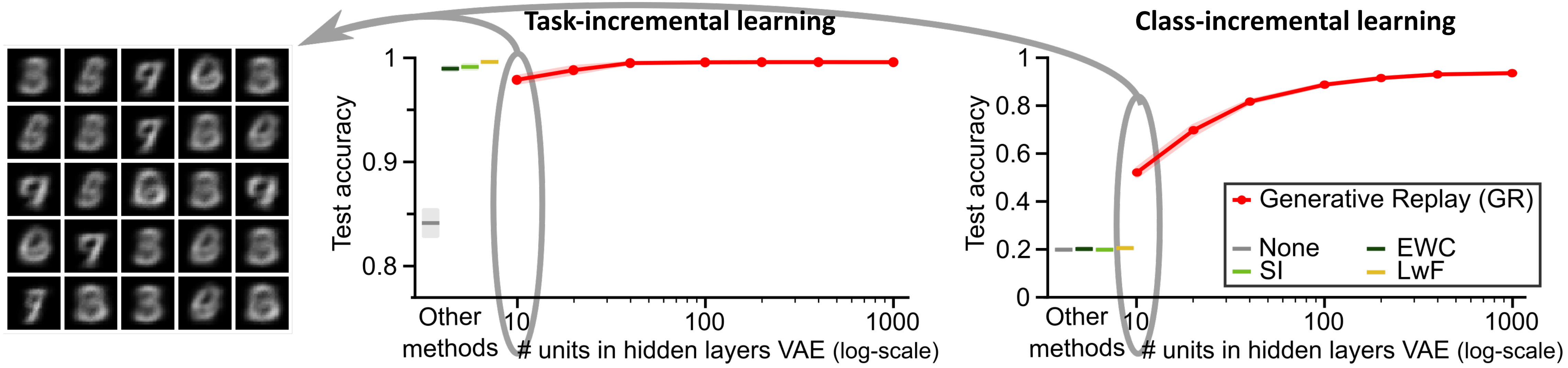


→ ***Fully replaying previous tasks is not needed, replaying only a few examples could suffice***

(these experiments are on Split MNIST)

Robustness: How good does the replay need to be?

- Generating MNIST-digits is relatively easy; could this scale to more complicated inputs?
- Performance of generative replay is evaluated as function of the size of the generator

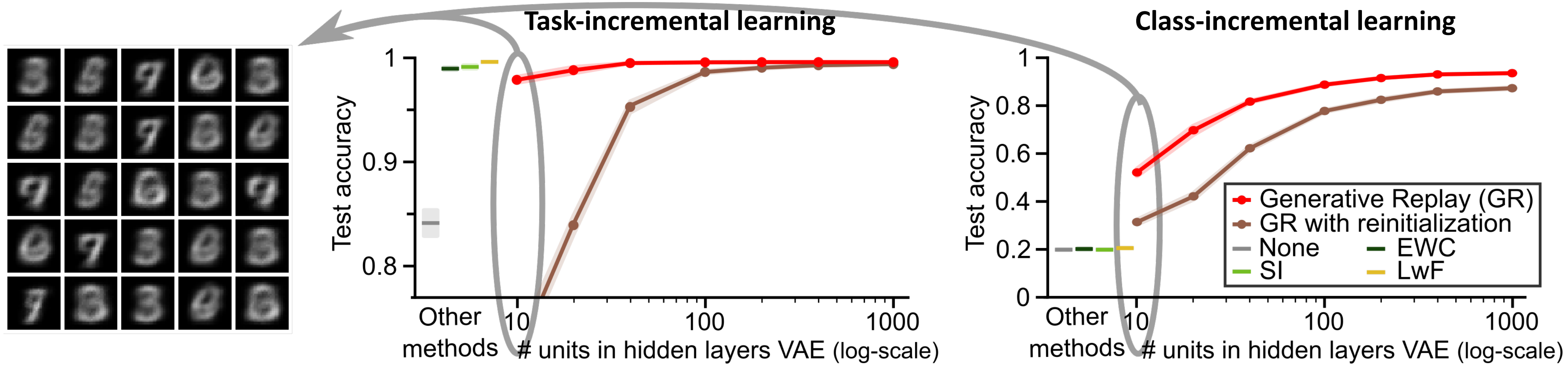


→ *A perfect memory is not needed, a low-quality generative model could suffice*

(these experiments are on Split MNIST)

Robustness: How good does the replay need to be?

- Generating MNIST-digits is relatively easy; could this scale to more complicated inputs?
- Performance of generative replay is evaluated as function of the size of the generator



→ ***A perfect memory is not needed, a low-quality generative model could suffice***

(these experiments are on Split MNIST)

Robustness and efficiency of replay

<u>INTERIM SUMMARY:</u>

- Even replaying a few or poor-quality samples can substantially boost continual learning performance
- "Not forgetting" is easier for a network than "learning"

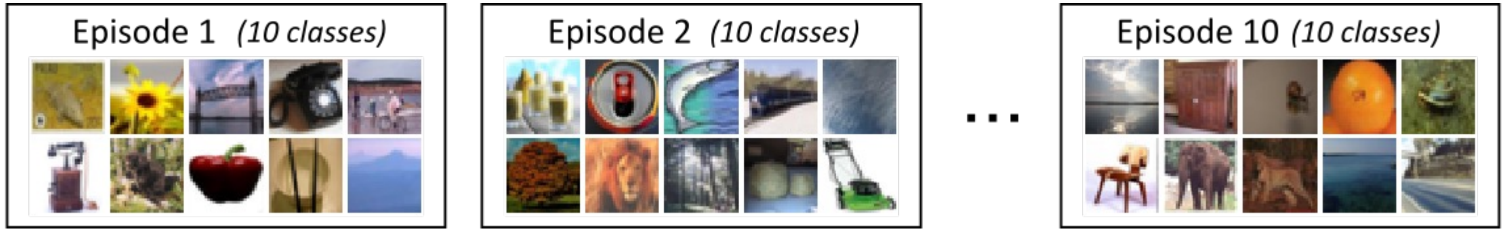
Further details: - van de Ven et al. (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications* **11**: 4069

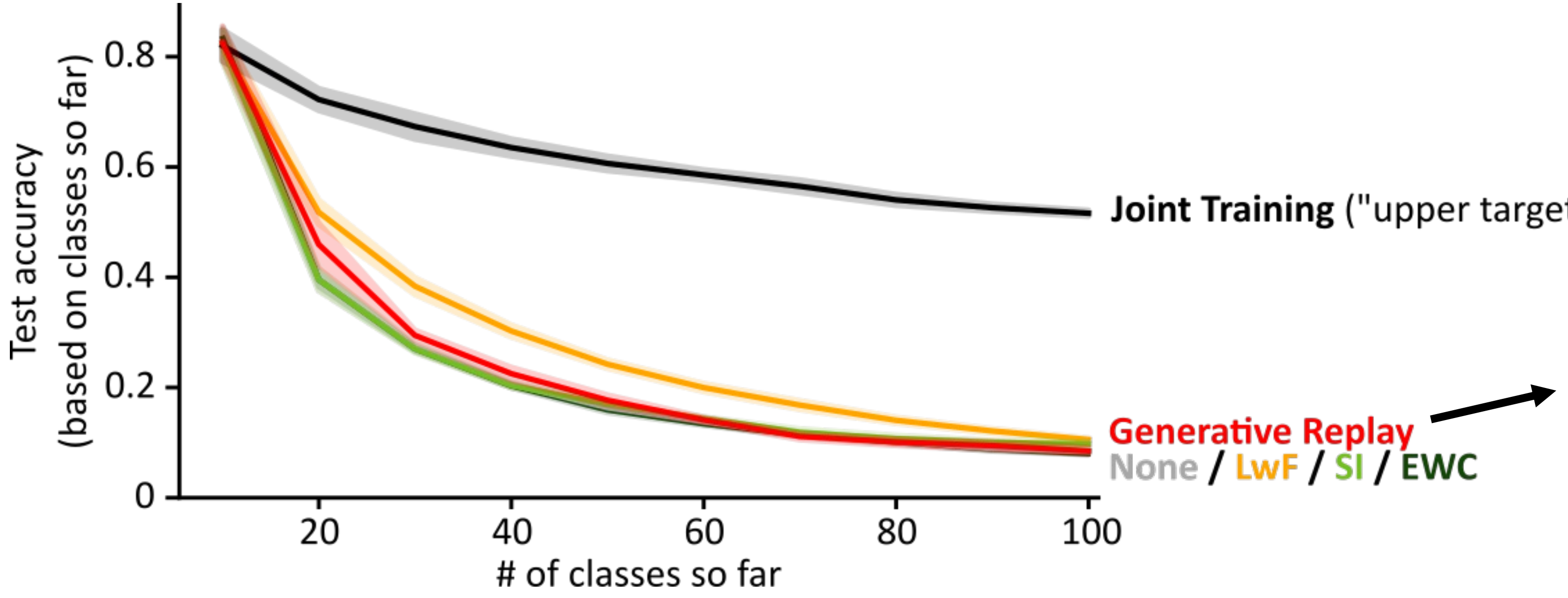
Next step:

→ Scale up generative replay to problems with more complex inputs

What about natural images?

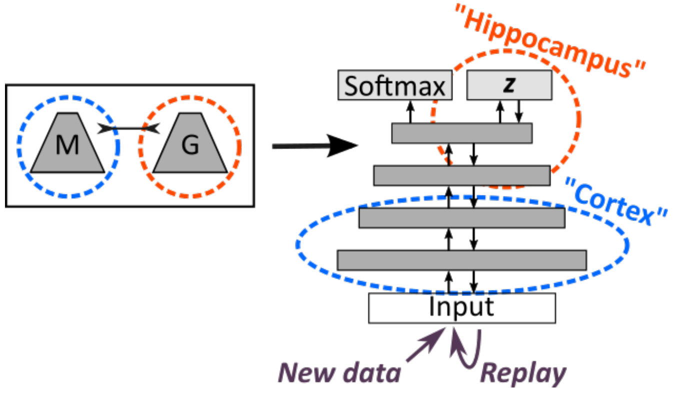
Class-incremental CIFAR-100:





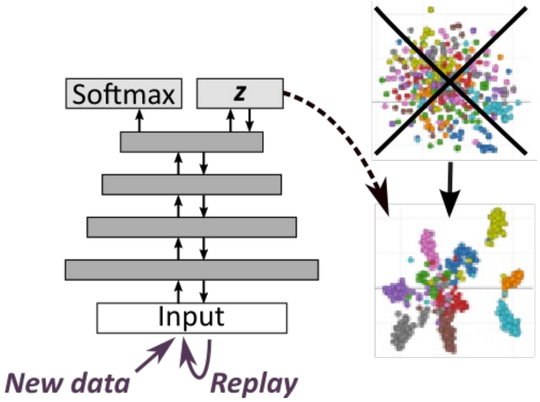
Standard versions of generative replay break down on problems with more complex inputs (e.g., natural images) [see also Lesort et al., 2019 *IJCNN*; Aljundi et al., 2019 *NeurIPS*]

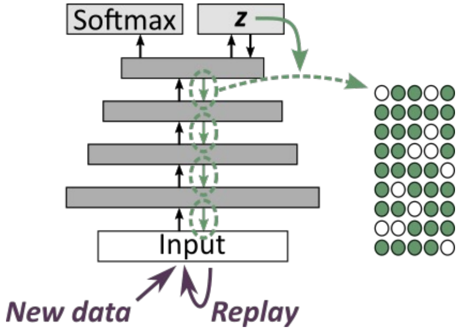
Brain-inspired modifications to Generative Replay



- ***Replay-through-Feedback:*** Merge generator into the classifier; replay is now generated by the feedback / backward connections

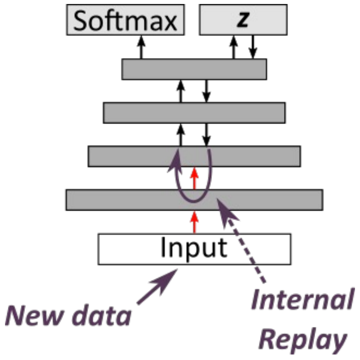
- ***Conditional Replay:*** Enable model to generate specific classes, by replacing the standard normal prior by a Gaussian mixture with a separate mode for each class





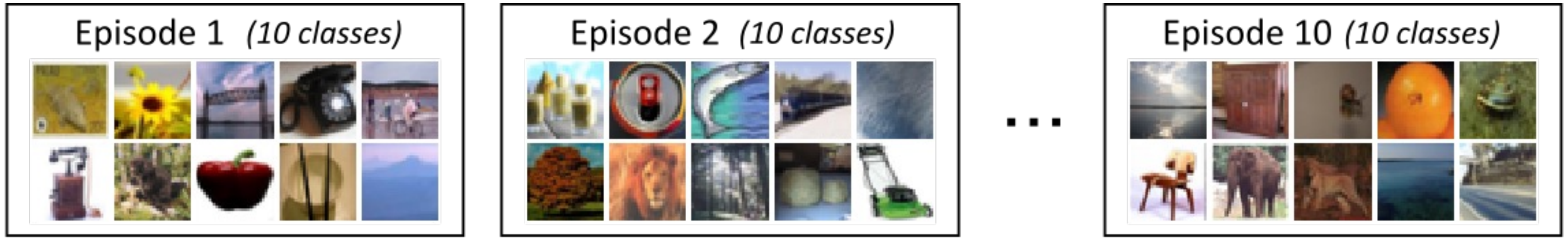
- ***Gating based on Internal Context:*** For each class, inhibit (or gate) a different subset of neurons during the generative backward pass

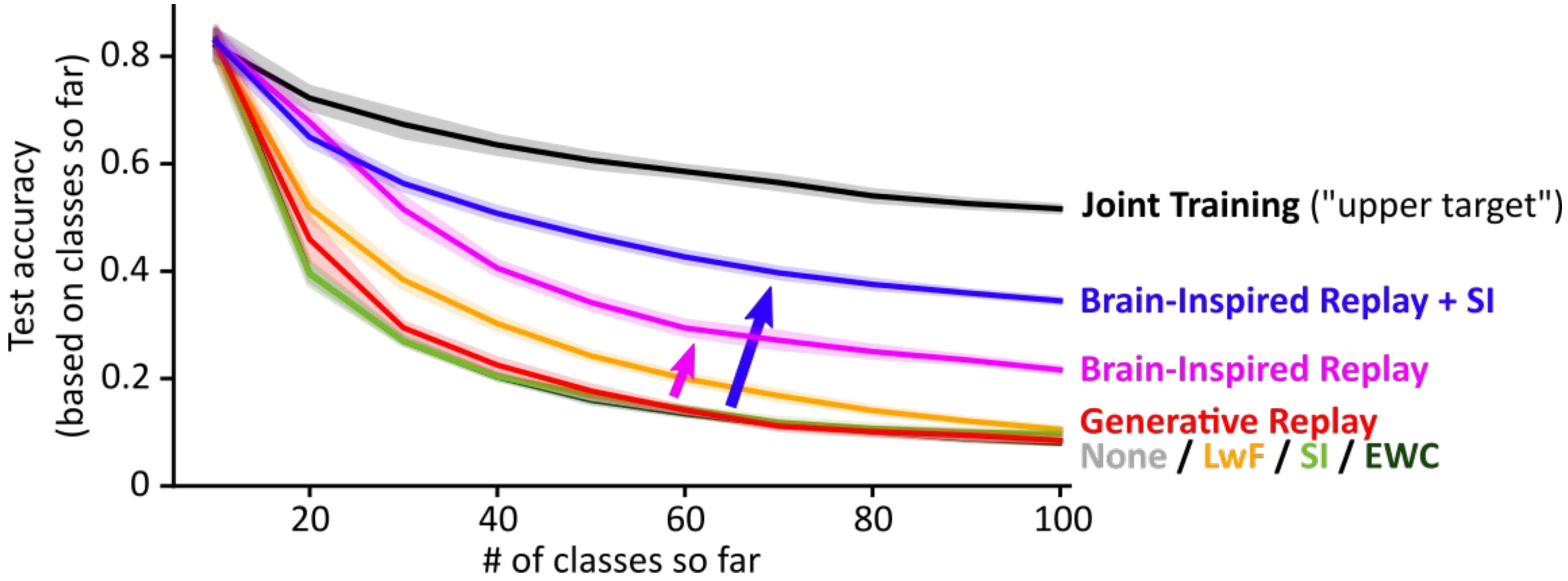
- ***Internal Replay:*** Replay internal or hidden representations, instead of at the input level (e.g., pixel level)



Brain-inspired replay on natural images

Class-incremental CIFAR-100:





(all methods use pre-trained convolutional layers)

PyTorch code for these experiments: https://github.com/GMvandeVen/brain-inspired-replay

Scaling generative replay to more complex problems

<u>INTERIM SUMMARY:</u>

- Scaling up generative replay to problems with more complex inputs is not straight-forward
- Brain-inspired modifications help to scale up generative replay
- In particular, replaying abstract, high-level representations increases performance while lowering computational costs (see also: Liu et al., 2020 *CVPR-W*; Pellegrini et al., 2020 *IROS*)

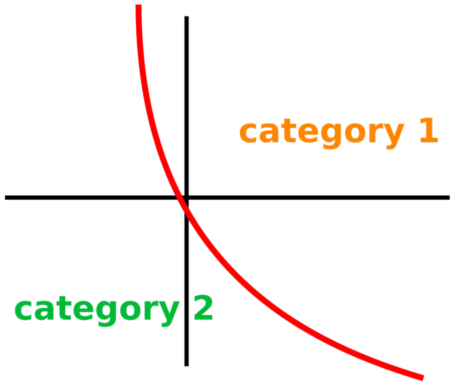
Further details: - van de Ven et al. (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications* **11**: 4069

But despite improvements, a substantial performance gap remains relative to the upper target of jointly training on all classes …

Strategy 2: Generative Classification

Discriminative classifiers

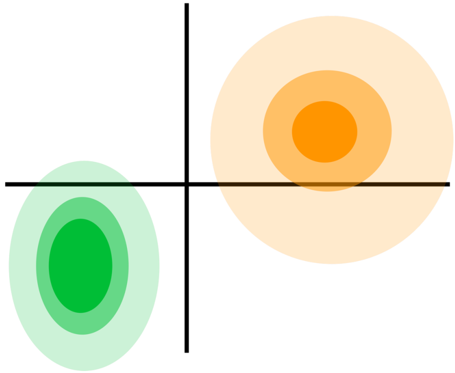
- Directly learn $p(y|\boldsymbol{x})$, or $\underset{y}{\operatorname{argmax}}\, p(y|\boldsymbol{x})$.



- Learn rules / shortcuts / features to distinguish between the classes to be learned
- Comparison between classes is during *training*

Generative classifiers

- Learn $p(\boldsymbol{x}, y)$, factorized as $p(\boldsymbol{x}|y)p(y)$, and classify using Bayes' rule



- Learn a model / template / representation for each class to be learned
- Comparison between classes is during *inference*

Generative classification ***rephrases a class-incremental problem as a task-incremental problem***, whereby each 'task' is to learn a class-conditional generative model.

van de Ven et al (2021) *CVPR-W proceedings*

Naïve implementation for proof-of-principle

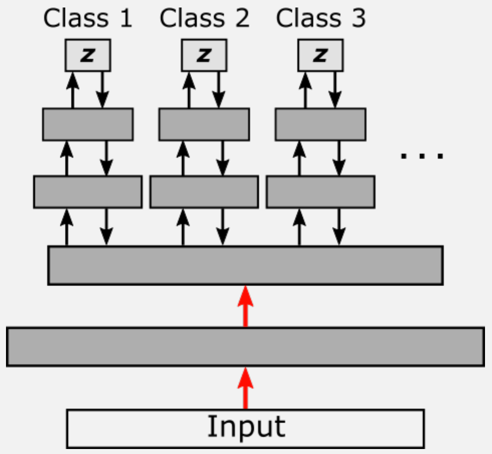
- Separate VAE model for each class
  (this is the naïve solution for a task-incremental learning problem, upon which successful task-incremental learning methods should be able to improve)

- If a pretrained network is available, the VAE models are trained on the latent features

- Class-conditional likelihoods are estimated using importance sampling

- The *total* number of parameters is similar to that of brain-inspired replay

Schematic



Making a classification decision

(1) Estimate class-conditional likelihoods

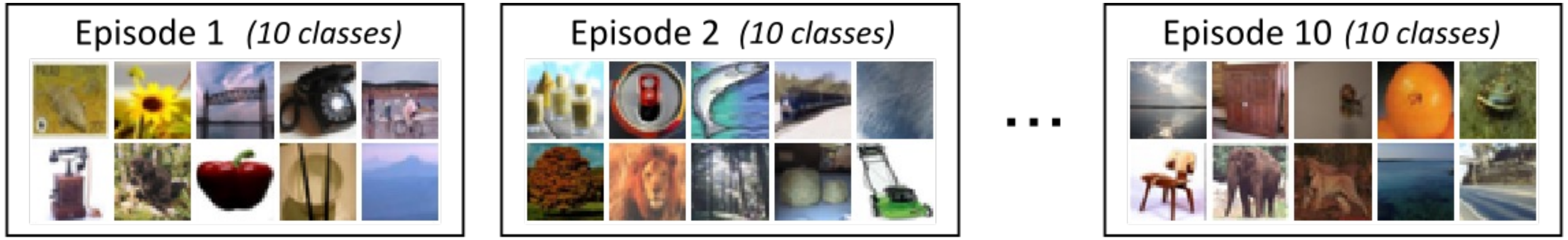
$$p(\boldsymbol{x}_{\text{test}}|y=1) \;\; p(\boldsymbol{x}_{\text{test}}|y=2) \;\; p(\boldsymbol{x}_{\text{test}}|y=3) \;\; \ldots$$

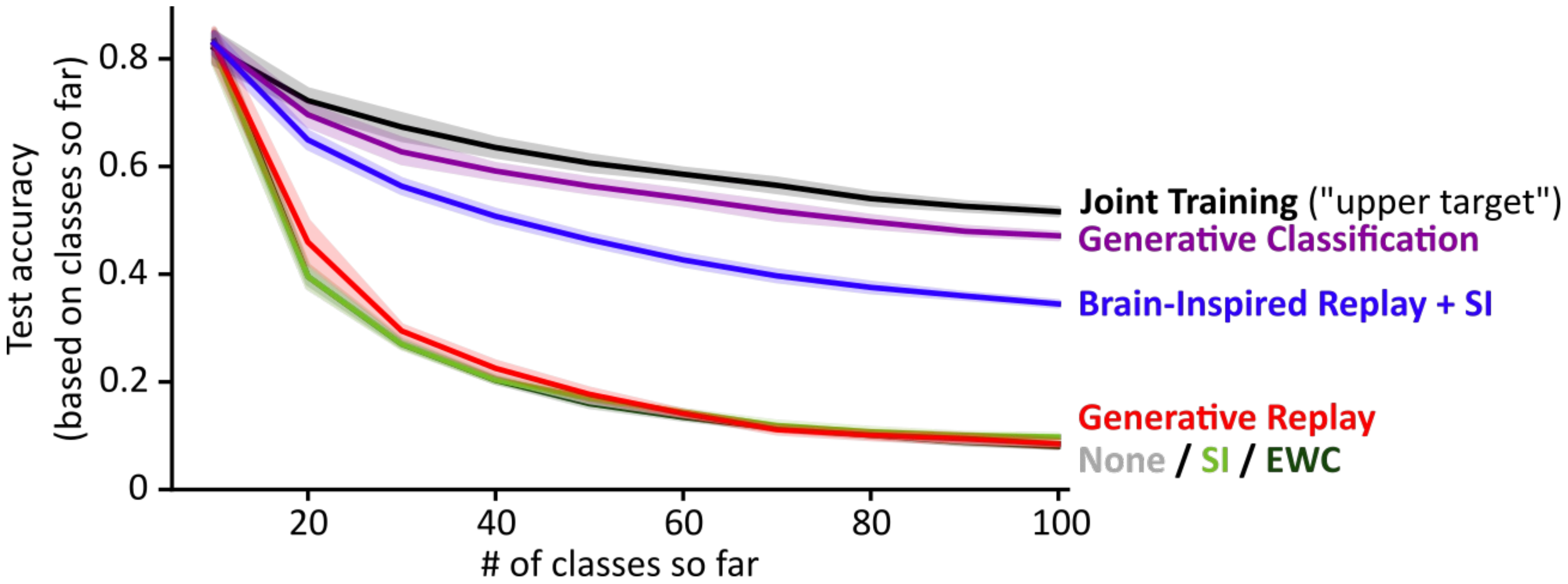
(2) Classify using Bayes' rule

$$\hat{y}^{(\boldsymbol{x}_{\text{test}})} = \underset{i}{\operatorname{argmax}}\, p(\boldsymbol{x}_{\text{test}}|y=i)$$

Naïve implementation for proof-of-principle

Class-incremental CIFAR-100:





(all methods use pre-trained convolutional layers)

PyTorch code for these experiments: https://github.com/GMvandeVen/class-incremental-learning

Generative classification

INTERIM SUMMARY:

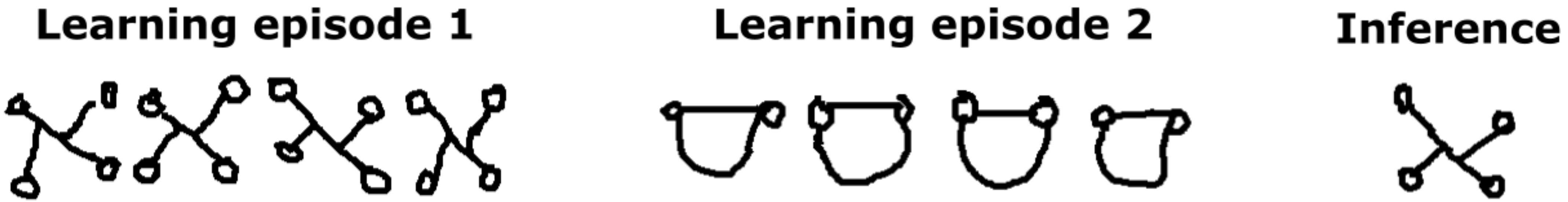
- Generative classification is a promising, “rehearsal-free” strategy for class-incremental learning
- Generative classification rephrases a class-incremental learning problem as a task-incremental learning problem
- How to use a generative model? At least in some settings, using it directly for generative classification outperforms using it indirectly for generative replay

Further details: - van de Ven et al. (2021) Class-incremental learning with generative classifiers. *CVPR-W proceedings*: 3611-3620

Limitations / future work:

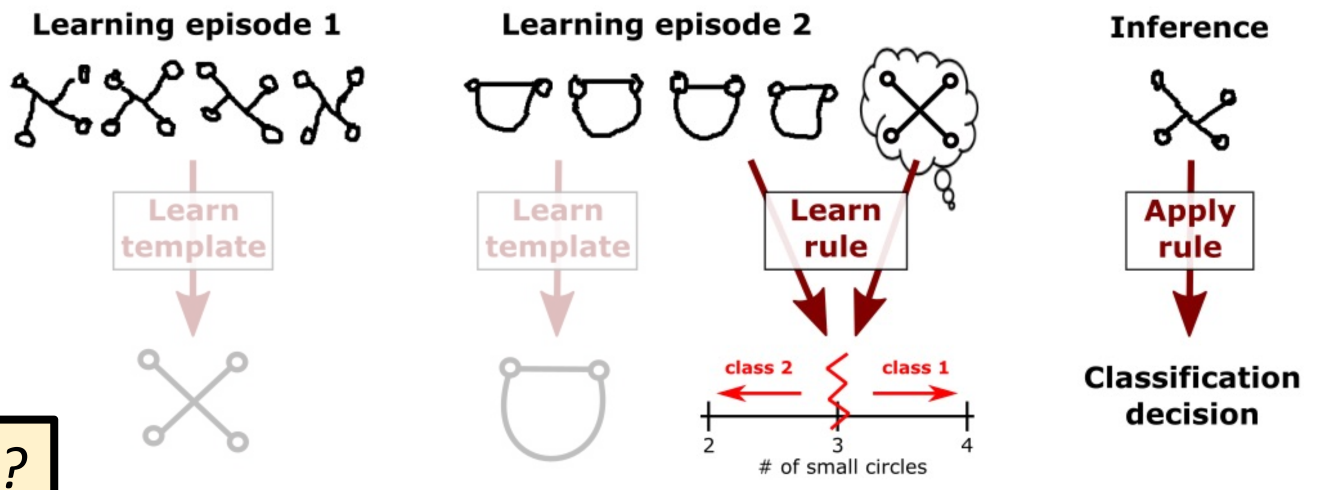
- How to share parts of the different generative models remains an open question
- Inference is slow, as likelihood must be computed/estimated for each possible class

Which strategy do *we* use?



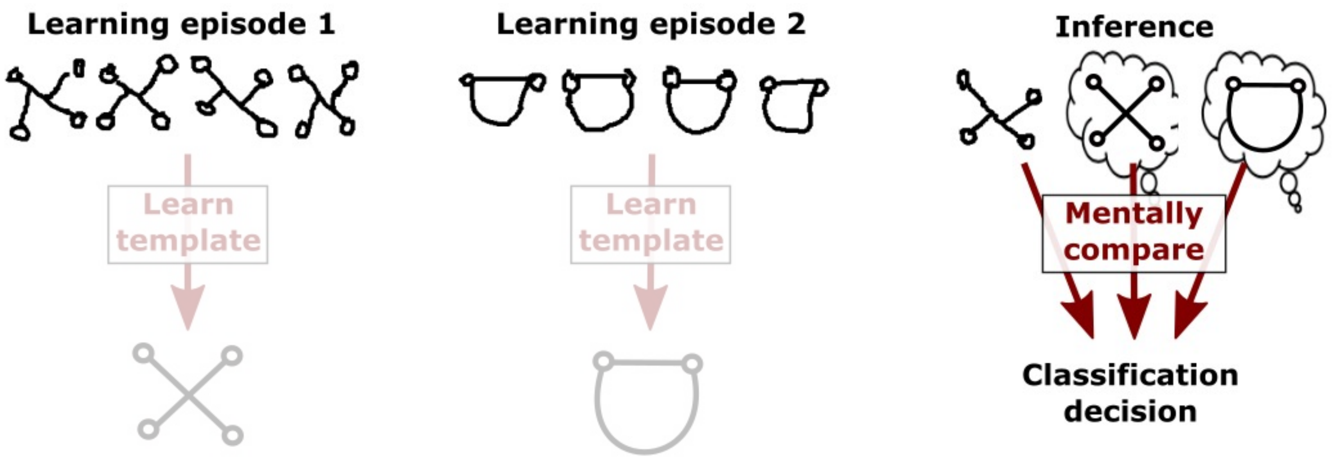
Which strategy do *we* use?

Generative Replay:



Or did you use another strategy?

Generative Classification:



Overall summary

- Continual learning is not a unitary problem: there are three different scenarios, each with their own challenges
- Class-incremental learning requires learning to distinguish classes that are not observed together

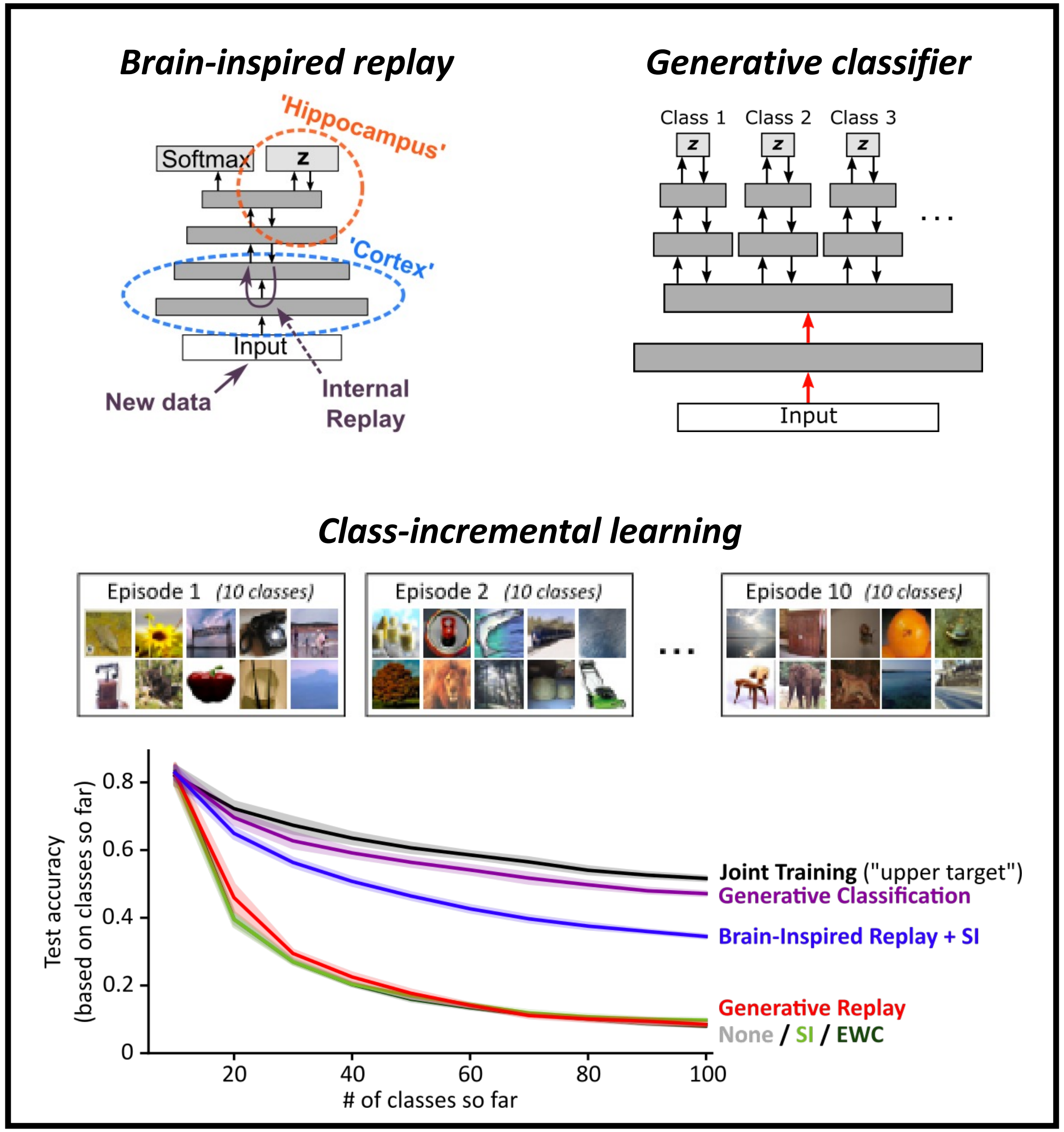
van de Ven et al (2022) *Nature Machine Intelligence*

- Even replaying a few or poor-quality samples can substantially boost continual learning performance
- Scaling generative replay up to problems with more complex inputs is nevertheless not straight-forward
- Replaying abstract, high-level representations increases performance while lowering computational costs

van de Ven et al (2020) *Nature Communications*

- How to use a generative model? Directly (for generative classification) or indirectly (for generative replay)?
- Generative classification is a promising, "rehearsal-free" strategy for class-incremental learning

van de Ven et al (2021) *CVPR-W proceedings*



All code is available: https://github.com/GMvandeVen

Acknowledgements

Mentors:

- Andreas Tolias (Baylor)
- Tinne Tuytelaars (KU Leuven)

Collaborators:

- Stefano Fusi (Columbia)
- Máté Lengyel (Cambridge)
- Hava Siegelmann (UMass)
- Matthias De Lange (KU Leuven)
- Kimberly Tolias (Baylor)
- Alexander Reyes (NYU)
- Zhe Li (Baylor)

Funding:

MSCA erc DARPA IARPA IBRO

References

- Aljundi R, Lin M, Goujaud B, Bengio Y (2019) Gradient based sample selection for online continual learning. *NeurIPS:* 11816-11825.
- De Lange M, Aljundi R, Masana M, Parisot S, Jia X, Leonardis A, Slabaugh G, Tuytelaars T (2022) A Continual Learning Survey: Defying Forgetting in Classification Tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**: 3366-3385.
- Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, …, Hadsell R (2017) Overcoming catastrophic forgetting in neural networks. *PNAS* **114**: 3521-3526.
- Lesort T, Caselles-Dupré H, Garcia-Ortiz M, Stoian A, Filliat D (2019). Generative models from the perspective of continual learning. *IJCNN:* 1-8.
- Li Z, Hoiem D (2017) Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40***:* 2935-2947.
- Liu K, Sibille J, Dragoi G (2018) Generative predictive codes by multiplexed hippocampal neuronal tuplets. *Neuron* **99**: 1329-1341.
- Liu Y, Dolan RJ, Kurth-Nelson Z, Behrens TE (2019). Human replay spontaneously reorganizes experience. *Cell* **178**: 640-652.
- Liu X, Wu C, Menta M, Herranz L, Raducanu B, Bagdanov AD, ..., van de Weijer J (2020). Generative feature replay for class-incremental learning. *CVPR-W proceedings*: 226-227.
- Masse NY, Grant GD, Freedman DJ (2018) Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization. *PNAS* **115**: E10467-E10475.
- McClelland JL, McNaughton BL, O'Reilly RC (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* **102**: 419-457.
- McCloskey M, Cohen NJ (1989) Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation* **24**: 109-165.
- O’Neill J, Pleydell-Bouverie B, Dupret D, Csicsvari J (2010) Play it again: reactivation of waking experience and memory. *Trends in Neurosciences* **33**: 220-229.
- Pellegrini L, Graffieti G, Lomonaco V, Maltoni D (2020) Latent replay for real-time continual learning. *IROS*: 10203-10209.
- Ratcliff R (1990) Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychological Review* **97**: 285-308.
- Shin H, Lee JK, Kim J, Kim J (2017) Continual learning with deep generative replay. *NeurIPS*: 2994-3003.
- van de Ven GM, Siegelmann HT, Tolias AS (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications* **11**: 4069.
- van de Ven GM, Trouche S, McNamara CG, Allen K, Dupret D (2016) Hippocampal offline reactivation consolidates recently formed cell assembly patterns during sharp wave-ripples. *Neuron* **92**: 968-974.
- van de Ven GM, Tuytelaars T, Tolias AS (2022) Three types of incremental learning. *Nature Machine Intelligence* **4**: 1185-1197.
- van de Ven GM, Li Z, Tolias AS (2021) Class-incremental learning with generative classifiers. *CVPR-W proceedings*: 3611-3620.
- Wilson MA, McNaughton BL (1994) Reactivation of hippocampal ensemble memories during sleep. *Science* **265**: 676-679.
- Zenke F, Poole B, Ganguli S (2017) Continual learning through synaptic intelligence. *ICML*: 3987-3995.