Brain-inspired replay for continual learning with artificial neural networks

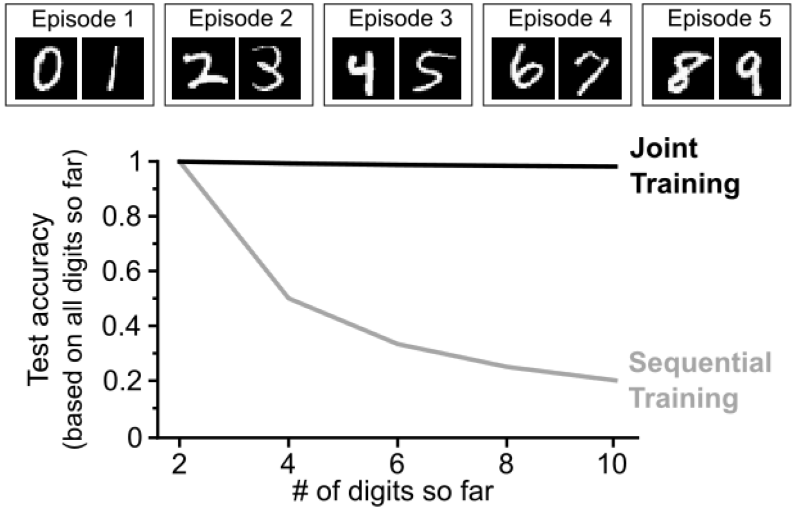
Gido van de Ven, Hava Siegelmann & Andreas Tolias

For full details:

van de Ven GM, Siegelmann HT, Tolias AS (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications*, **11**: 4069.

Code: https://github.com/GMvandeVen/brain-inspired-replay

Introduction



- When a neural network is trained on something new, it 'catastrophically' forgets what was learned before

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

- Humans continually accumulate information throughout their lifetime
- A brain mechanism thought to underlie this ability is the replay of neuronal activity patterns that represent previous experiences

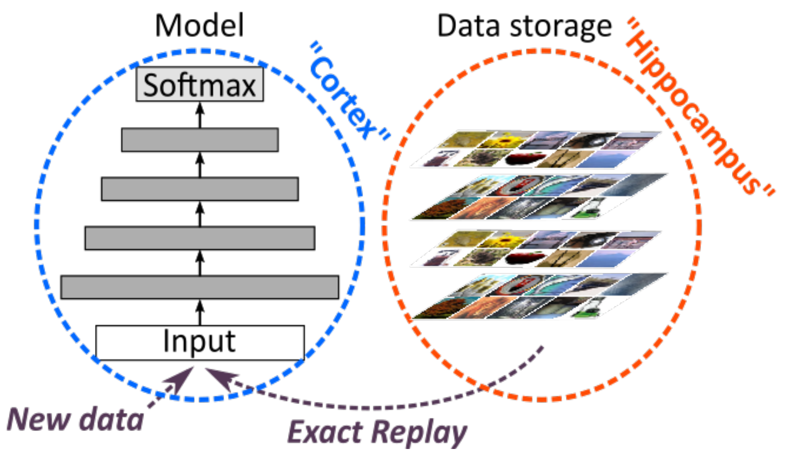
  [Wilson & McNaughton, 1994 *Science;* O'Neill et al., 2010 *TINS*; van de Ven et al., 2016 *Neuron*]

Motivation:

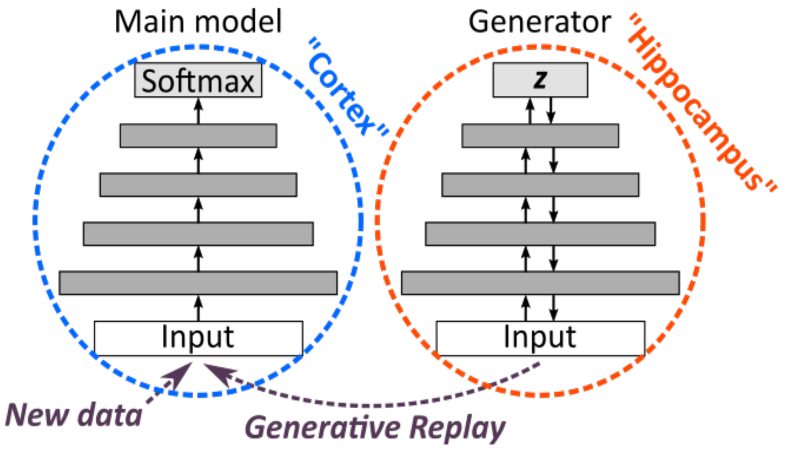
- ➢ **Use replay to enable artificial neural networks to do 'continual learning'**
- ➢ **Use artificial neural networks as a computational model for replay in the brain**

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

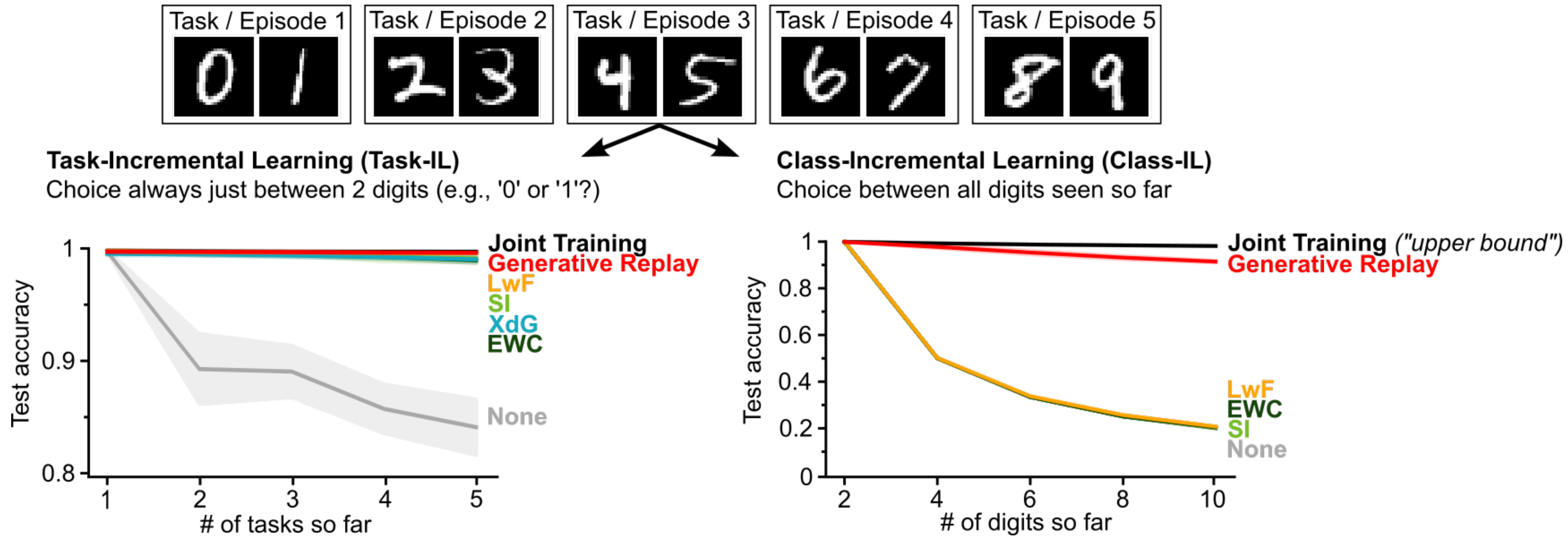


- Use a generative model – *"generative replay"*
  - *More realistic from neuroscience point of view*
  - *Views hippocampus as a generative neural network and replay as a generative process; see also* [Liu et al., 2018 *Neuron*; Liu et al., 2019 *Cell*]
  - Learning a generative model as a more scalable,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



Does (generative) replay work?

- Generative replay works very well for MNIST-based continual learning problems
- For class-incremental learning, some form of replay even seems to be required



Generative Replay: Shin et al., 2017 *NeurIPS*
Synaptic Intelligence (SI): Zenke et al., 2017 *ICML*
Elastic Weight Consolidation (EWC): Kirckpatrick et al., 2017 *PNAS*
Learning without Forgetting (LwF): Li & Hoiem, 2017 *IEEE T Pattern Anal*
Context-dependent Gating (XdG): Masse et al., 2018 *PNAS*

A similar point was made it previous work:
- van de Ven & Tolias (2018) *arXiv: 1809.10635*
- van de Ven & Tolias (2019) *NeurIPS Continual Learning workshop*

Robustness and efficiency of replay

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

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

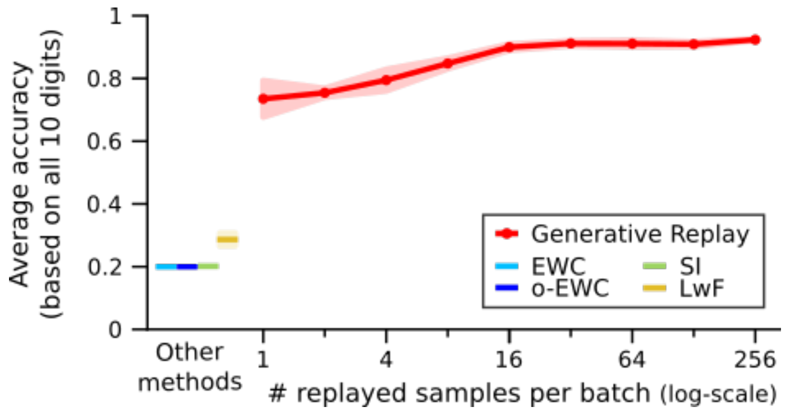
Robustness and efficiency of replay

But… (1) MNIST digits are relatively easy to generate
(2) constantly retraining on all previous tasks seems very inefficient

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

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

How much replay is needed?



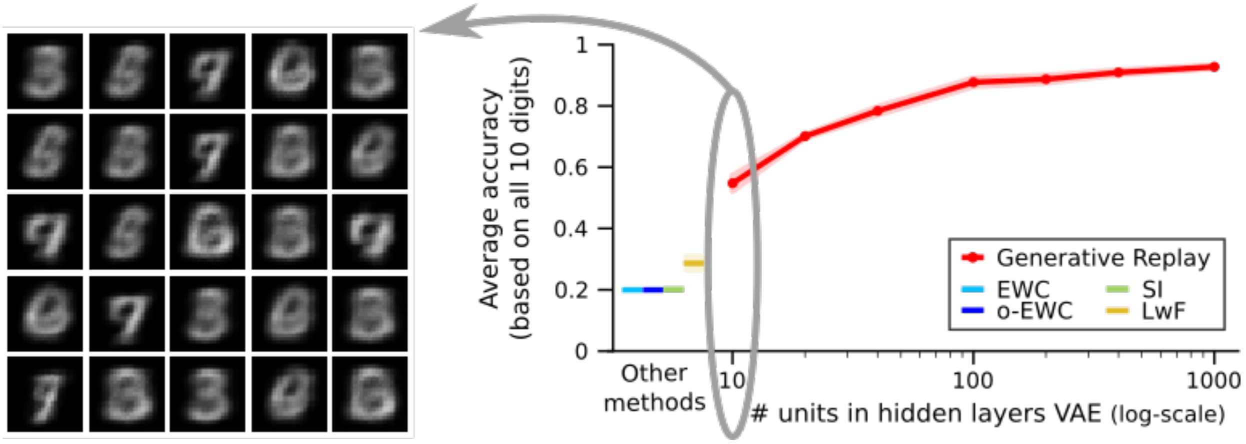
→ *Fully replaying previous tasks is not needed, replaying only a few examples could suffice*

Robustness and efficiency of replay

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

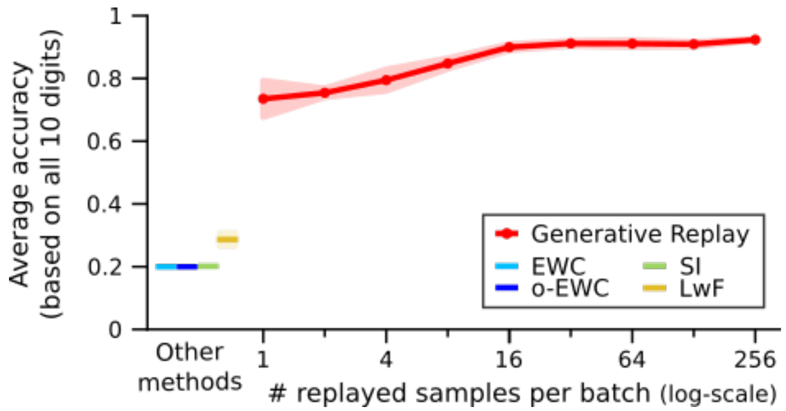
(2) constantly retraining on all previous tasks seems very inefficient

How good does replay need to be?



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

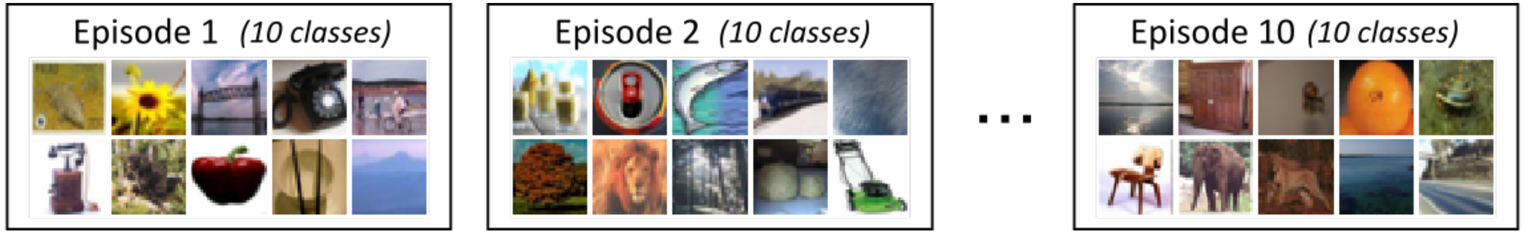
How much replay is needed?

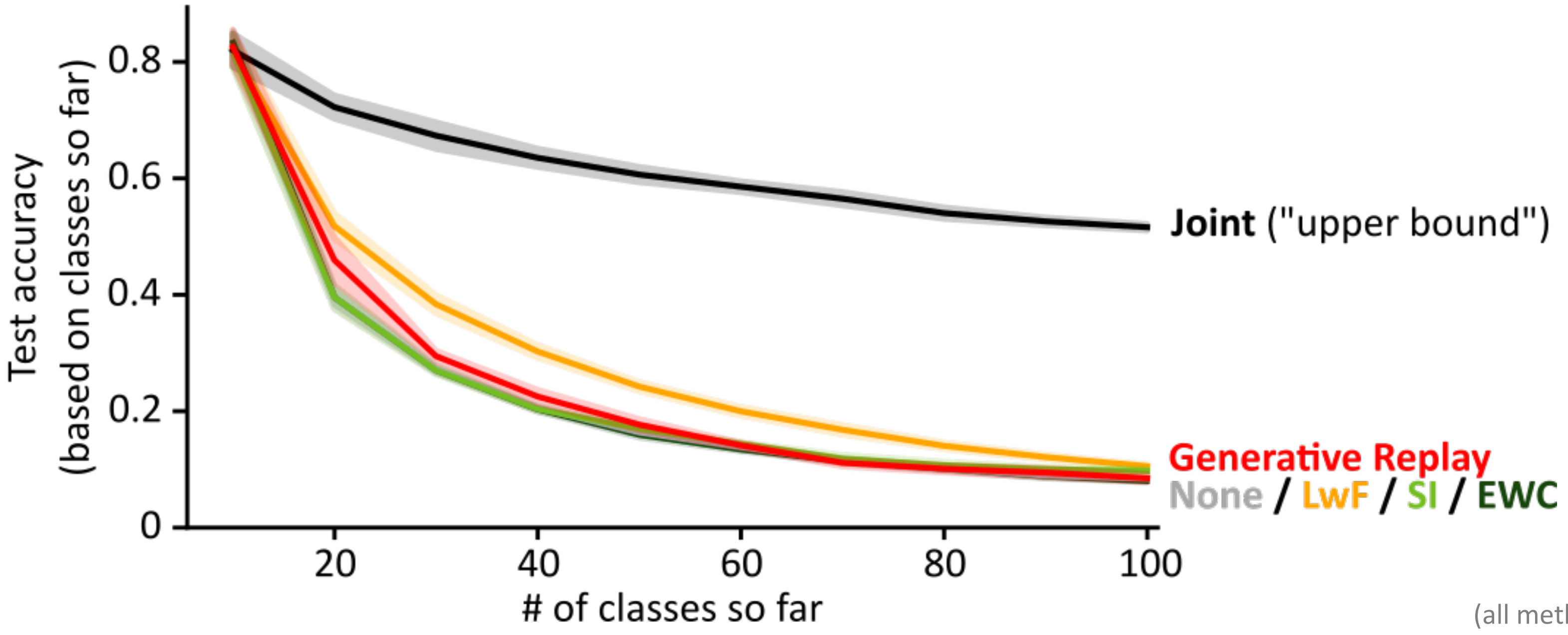


→ *Fully replaying previous tasks is not needed, replaying only a few examples could suffice*

What about more complex inputs?

Class-incremental CIFAR-100:

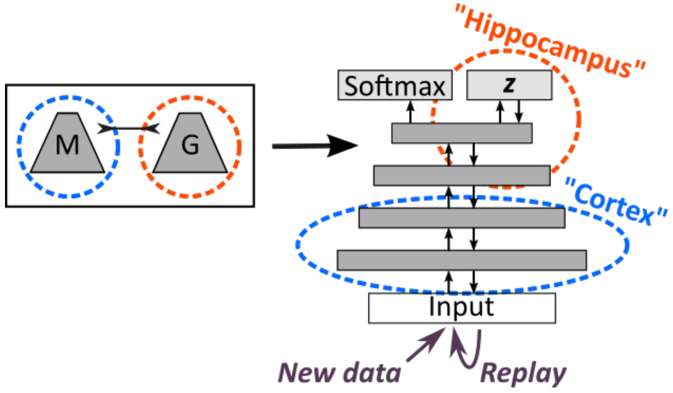




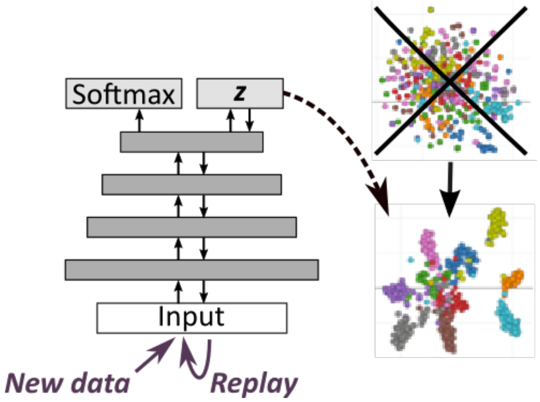
(all methods use pre-trained convolutional layers)

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*]

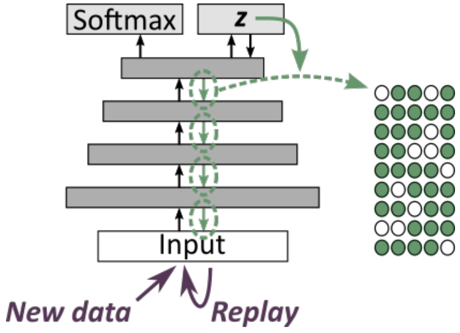
Brain-inspired modifications to Generative Replay



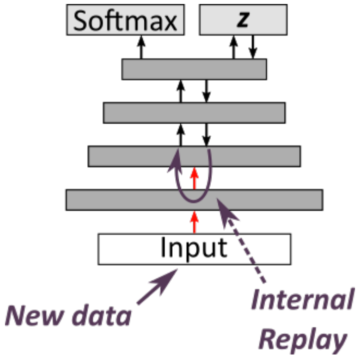
- ***Replay-through-Feedback:*** Merge generator into main model; replay is now generated by the feedback / backward connections



- ***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



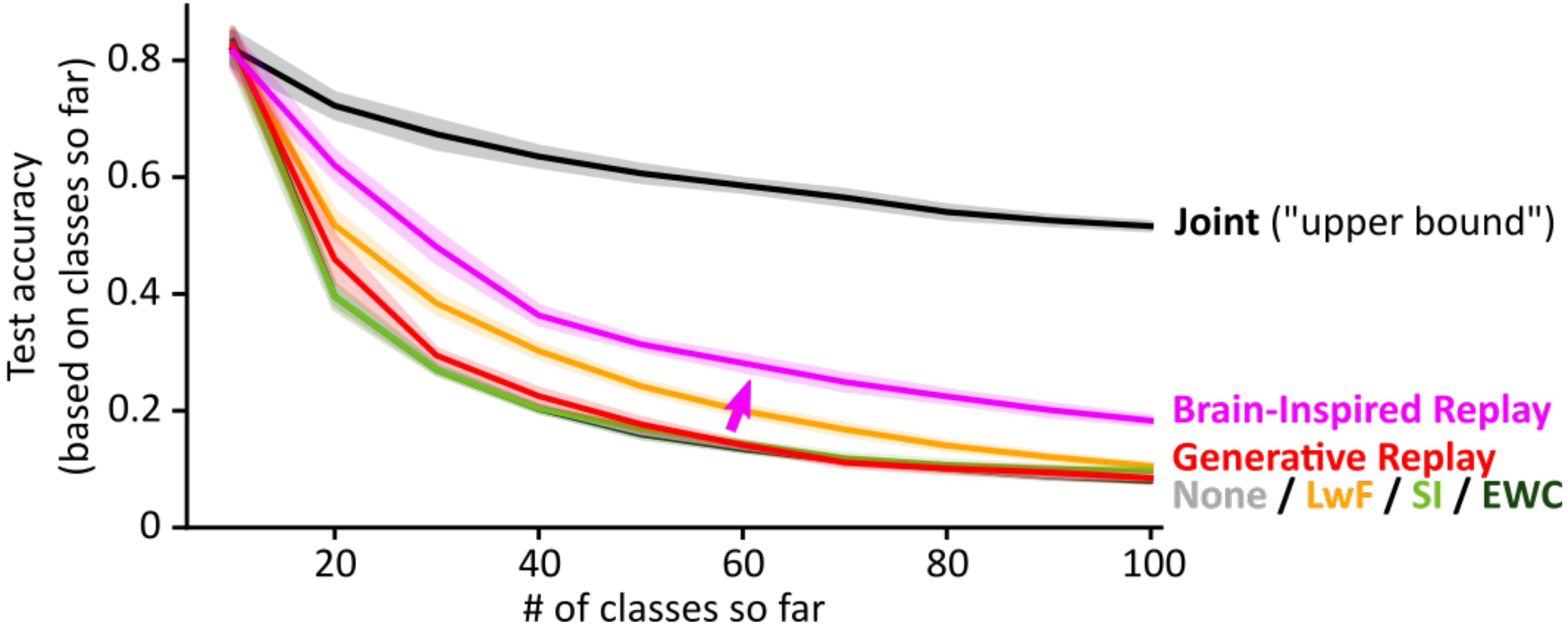
- ***Gating based on Internal Context:*** For each class, inhibit (or gate) a different subset of neurons during the generative backward pass



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

Brain-inspired replay on natural images

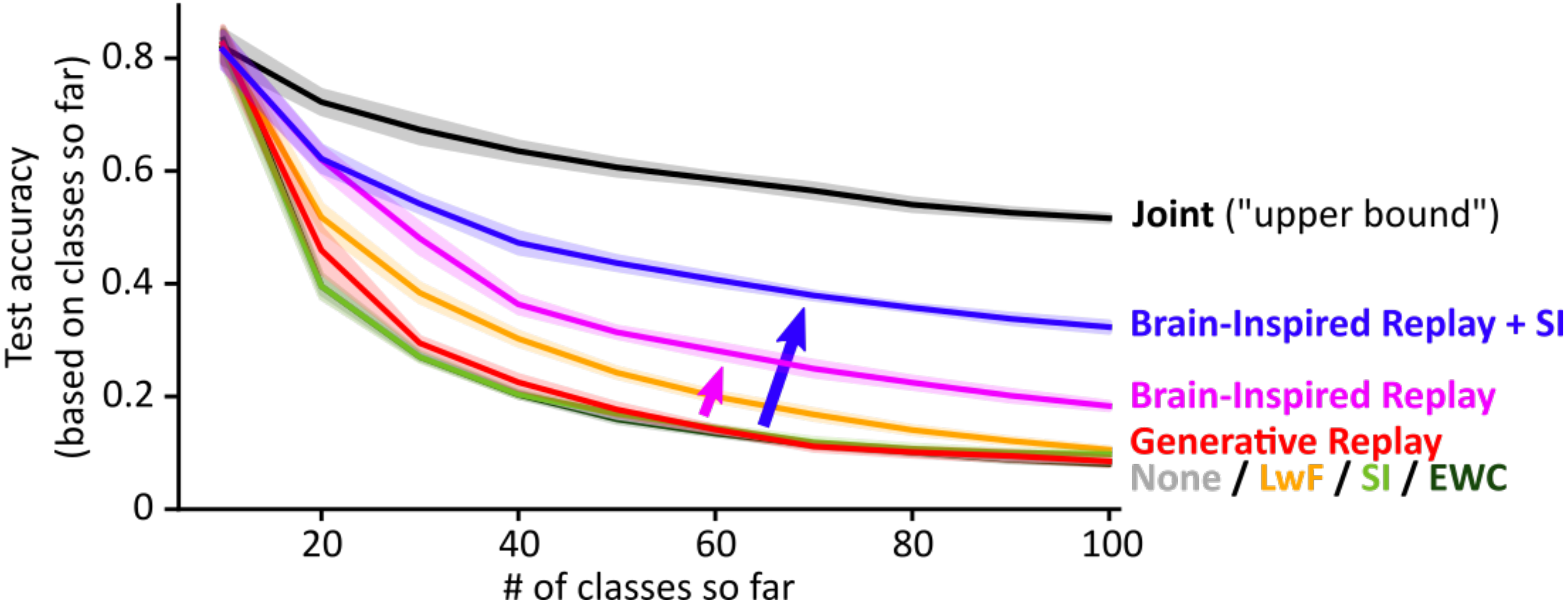
Class-incremental CIFAR-100:



(all methods use pre-trained convolutional layers)

Brain-inspired replay on natural images

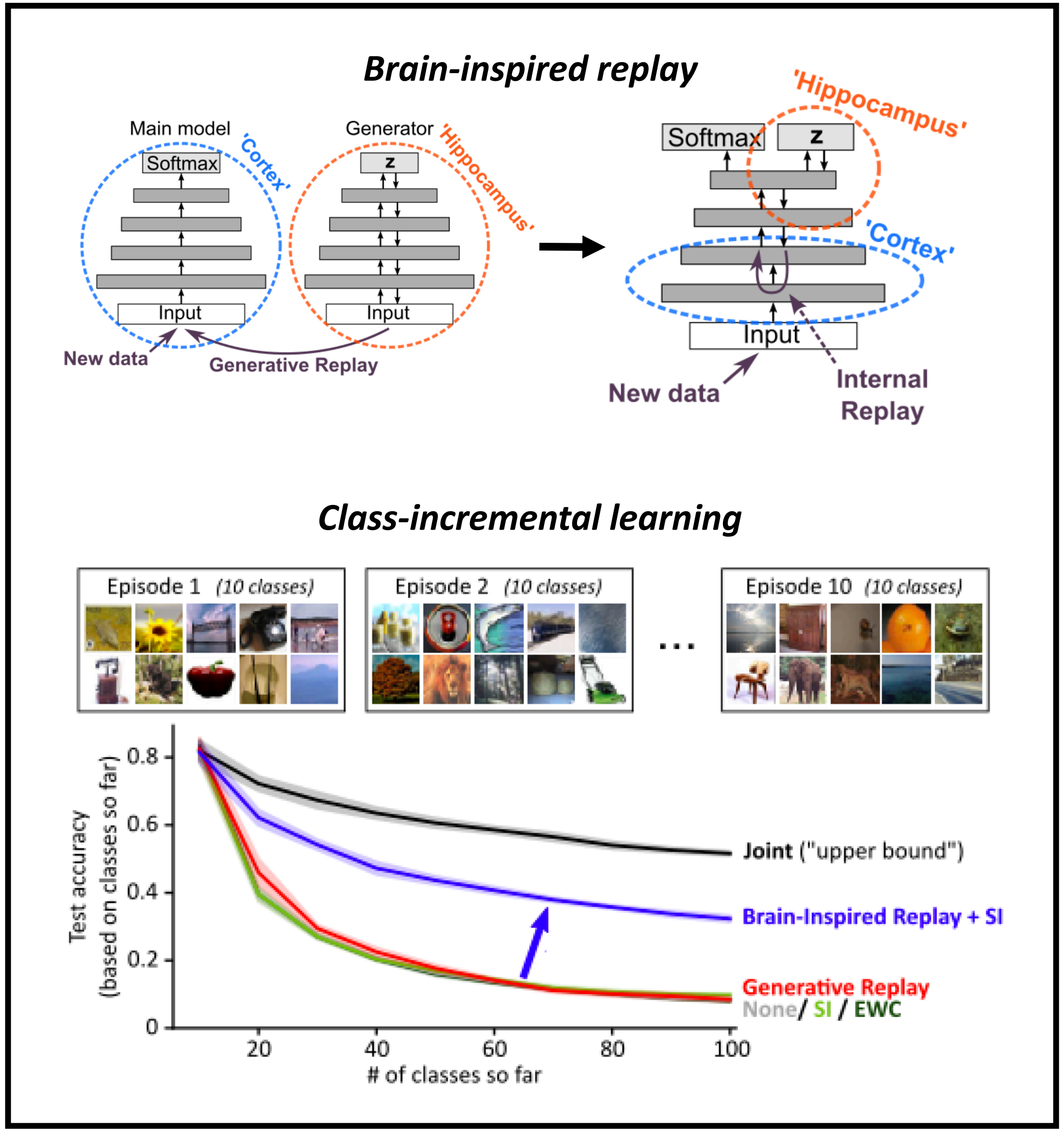
Class-incremental CIFAR-100:



(all methods use pre-trained convolutional layers)

Summary

- Replay is especially important for class-incremental learning (i.e., learning to distinguish between classes that are not observed together)
- Even replaying a few or poor-quality samples can substantially boost continual learning performance
- Scaling generative replay up to problems with more complicated inputs is nevertheless not straight-forward
- Modelling generative replay after the brain can substantially increase performance while lowering computational costs
- Our brain-inspired replay method replays internal or hidden representations that are generated by the network's own, context-modulated feedback connections



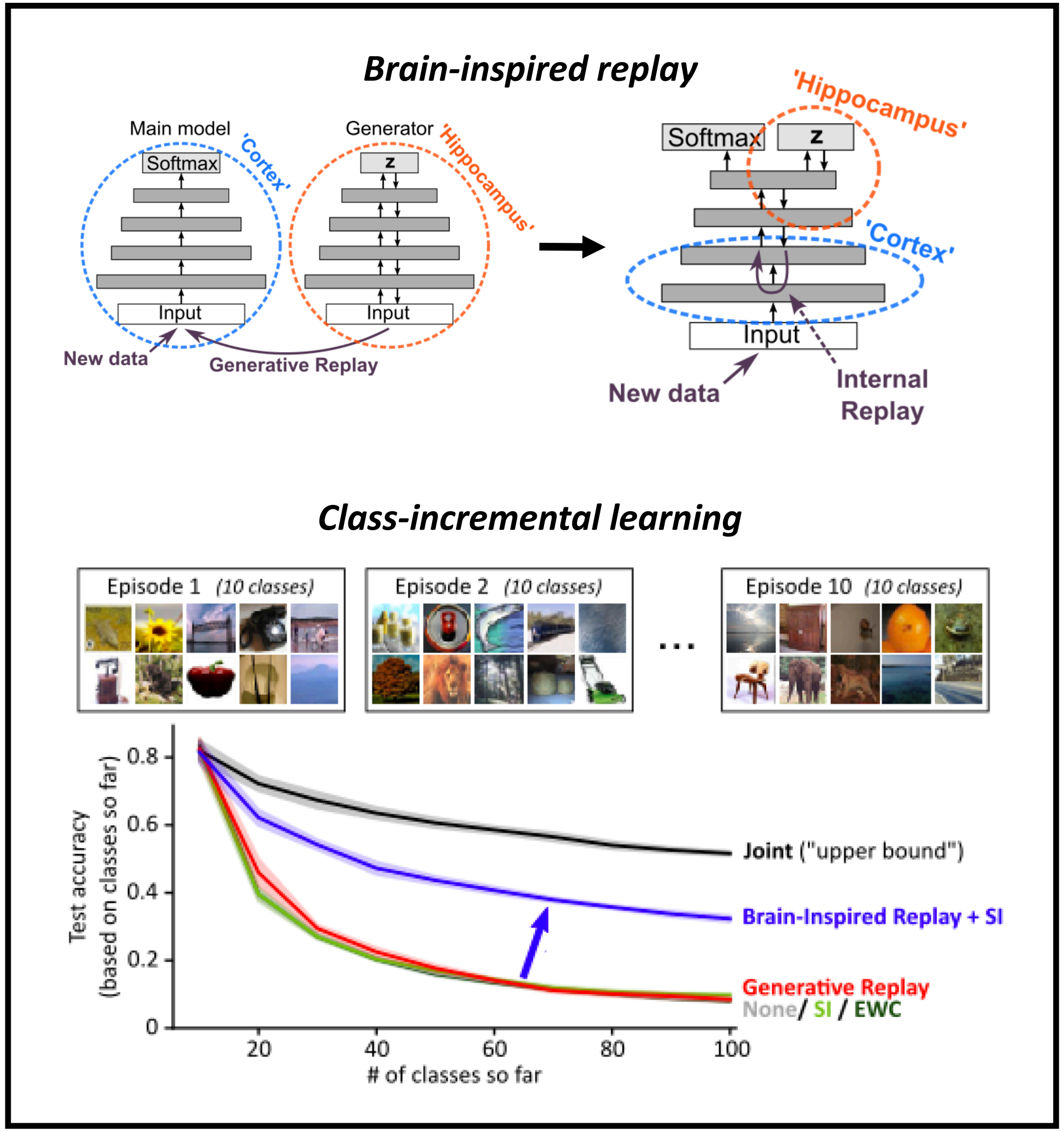
For full details:

van de Ven GM, Siegelmann HT, Tolias AS (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications*, **11**: 4069.

Code: https://github.com/GMvandeVen/brain-inspired-replay

Acknowledgements

We thank Mengye Ren, Zhe Li and Máté Lengyel for comments on various parts of this work, and Johannes Oswald and Zhengwen Zeng for useful suggestions. This research project has been supported by an IBRO-ISN Research Fellowship, by the Lifelong Learning Machines (L2M) program of the Defence Advanced Research Projects Agency (DARPA) via contract number HR0011-18-2-0025 and by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/Interior Business Center (DoI/IBC) contract number D16PC00003. Disclaimer: The views and conclusions contained in this presentation were those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA, IARPA, DoI/IBC, or the U.S. Government.



IBRO DARPA IARPA

References

- Aljundi R, Lin M, Goujaud B, Bengio Y (2019) Gradient based sample selection for online continual learning. *NeurIPS:* 11816-11825.
- Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, Hassabis D, Clopath C, Kumaran D, 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.
- 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.
- 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, 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, Siegelmann HT, Tolias AS (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications* **11**: 4069.
- van de Ven GM, Tolias AS (2019) Three scenarios for continual learning. *ArXiv preprint*: 1904.07734.
- van de Ven GM, Tolias AS (2018) Generative replay with feedback connections as a general strategy for continual learning. *ArXiv preprint*: 1809.10635.
- 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.