Brain-like replay for continual learning with artificial neural networks

Gido M van de Ven, Hava T Siegelmann, Andreas S Tolias

–

Bridging AI and Cognitive Science workshop (ICLR 2020)

Catastrophic forgetting in neural networks

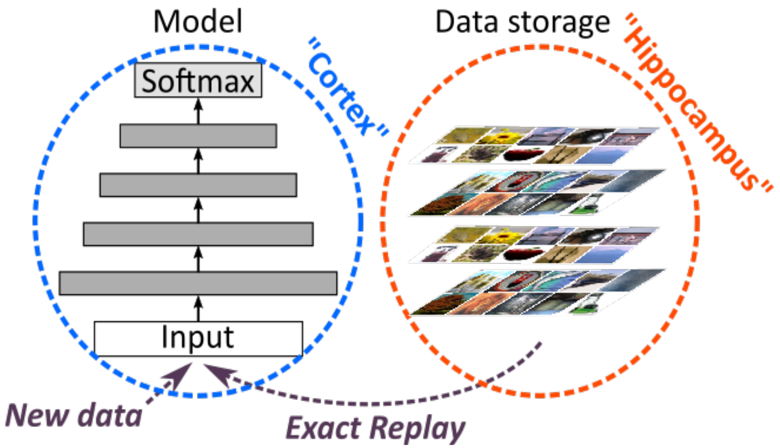
- When a neural network is trained on something new, it rapidly forgets what was learned before [McCloskey & Cohen, 1989 *Psych Learn Motiv*; Ratcliff, 1990 *Psych Rev*]

- 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
  - replay is orchestrated by the hippocampus, but also observed in cortex [Wilson & McNaughton, 1994 *Science;* O'Neill et al., 2010 *TINS*]

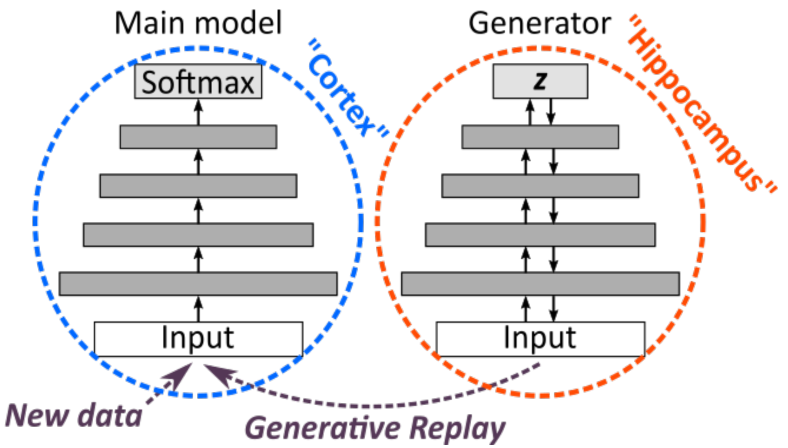
→ Could adding replay to artificial neural networks help protect them from catastrophic forgetting?

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* [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



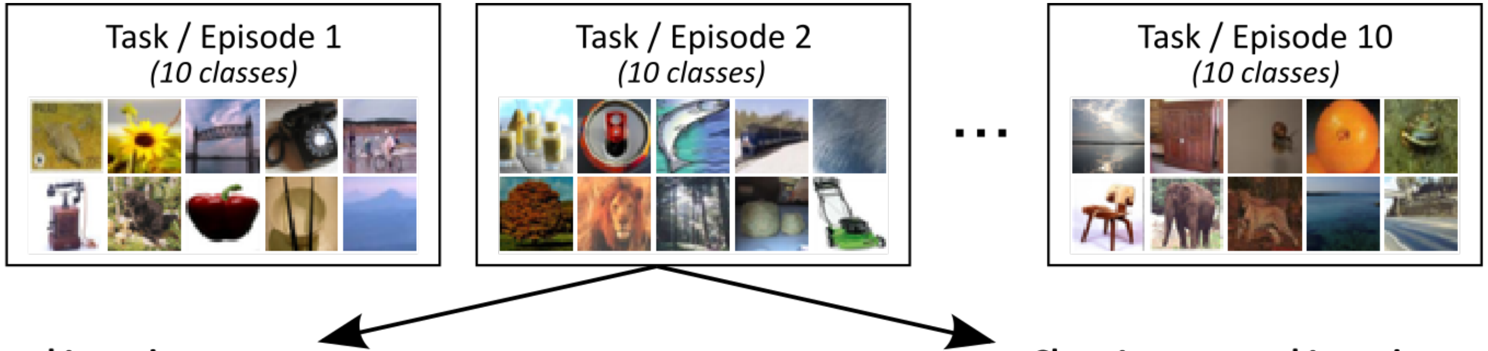
Does generative replay work?

- Generative replay works very well for MNIST-based continual learning problems [Shin et al., 2017 *NeurIPS*; van de Ven & Tolias, 2018 *arXiv*]
  - For class-incremental learning, generative replay is currently the only method capable of performing well without relying on stored data (even for MNIST!)
- Generative replay is reported to break down with more complex inputs (e.g., natural images) [Lesort et al., 2019 *IJCNN*; Aljundi et al., 2019 *NeurIPS*]

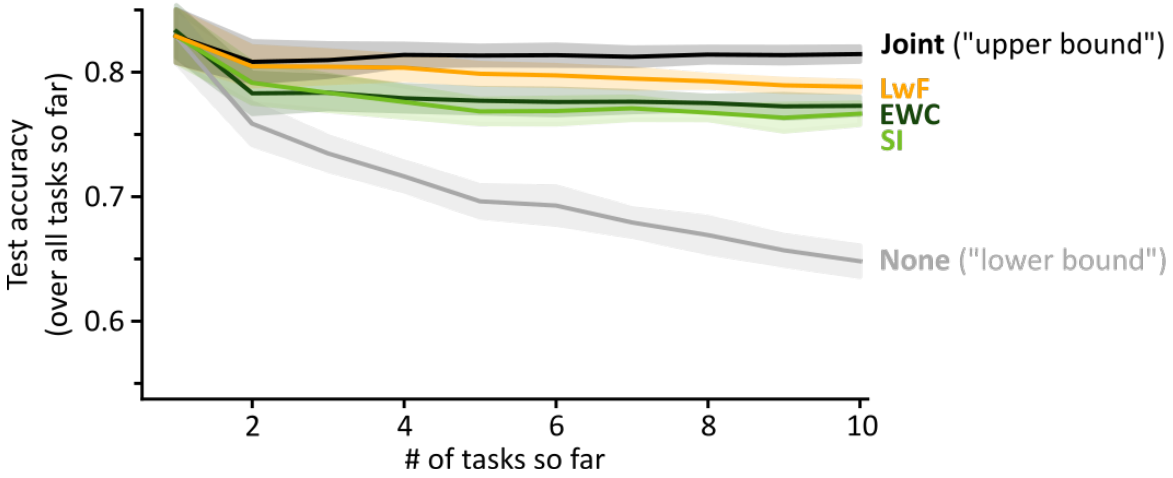
→ Two problems to be addressed:

  - This raises doubt as to whether or how replay could be used by the brain
  - Class-incremental learning with complex inputs (e.g., natural images) remains an unsolved problem in machine learning

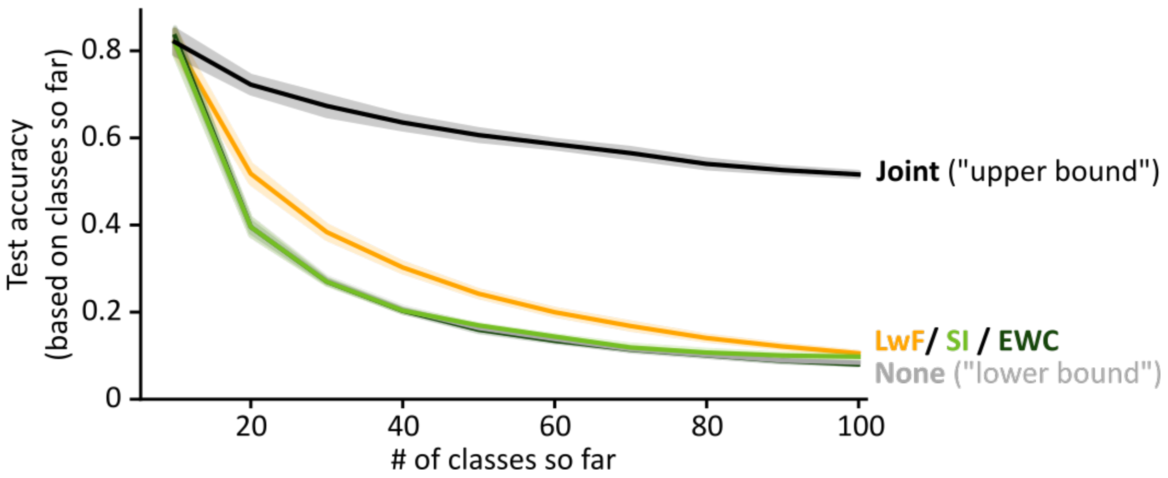
Generative replay on natural images



Task-Incremental Learning
Choice only between classes within given task



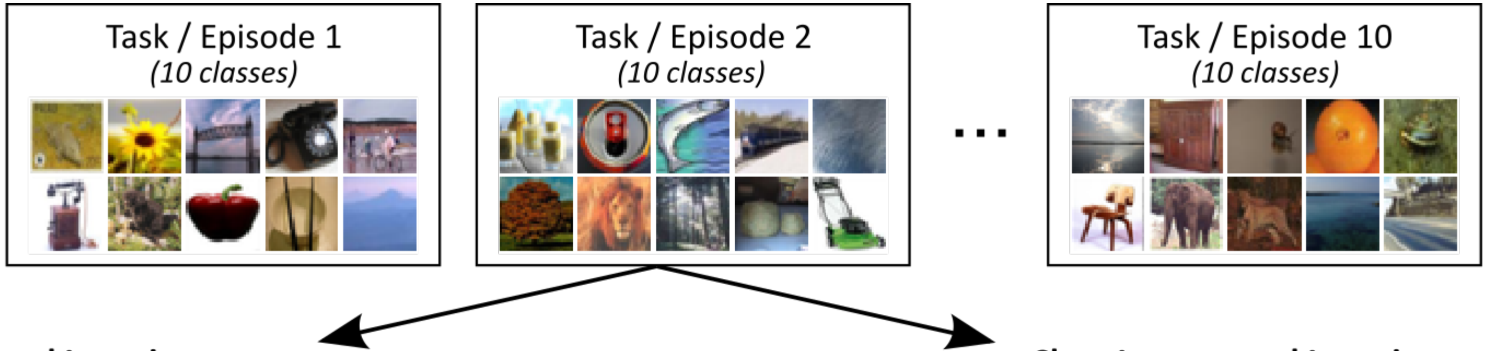
Class-Incremental Learning
Choice between all classes seen so far



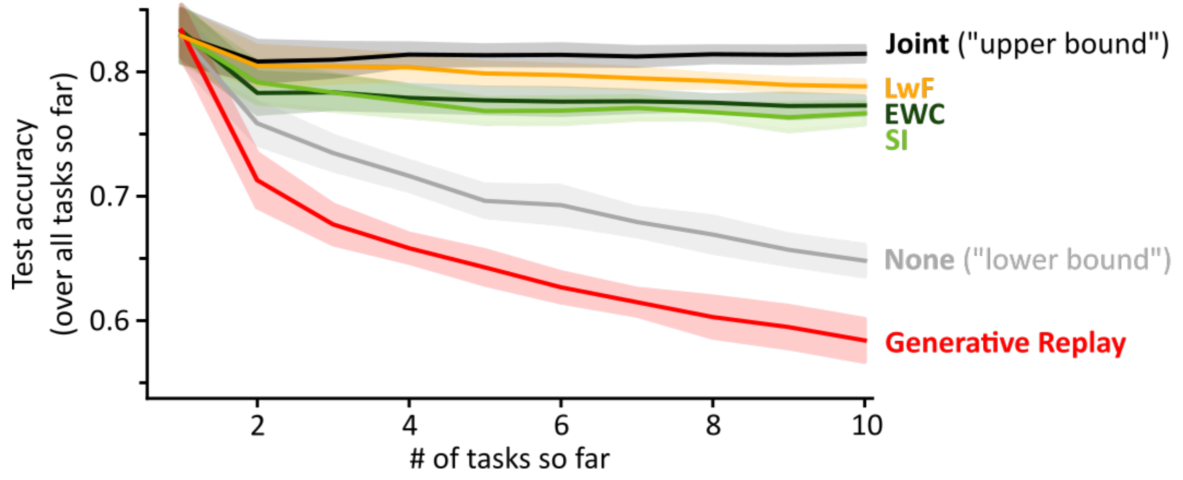
Synaptic Intelligence (SI): Zenke et al., 2017 *ICML*
Elastic Weight Consolidation (EWC): Kirckpatrick et al., 2017 *PNAS*
Learning without Forgetting (LwF): Li & Hoiem, 2017 *IEEE T Pattern Anal*

(all methods use pre-trained convolutional layers)

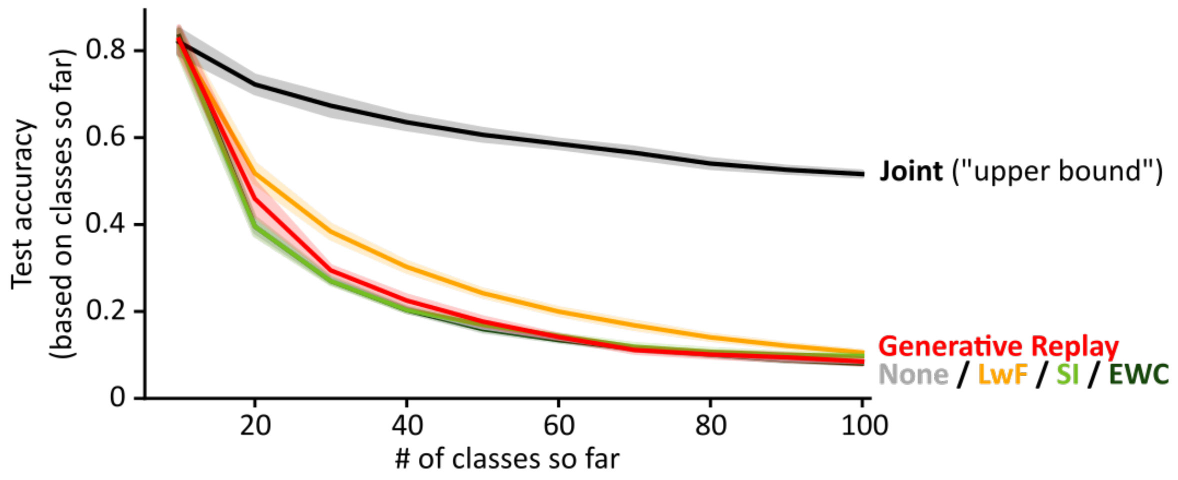
Generative replay on natural images



Task-Incremental Learning
Choice only between classes within given task



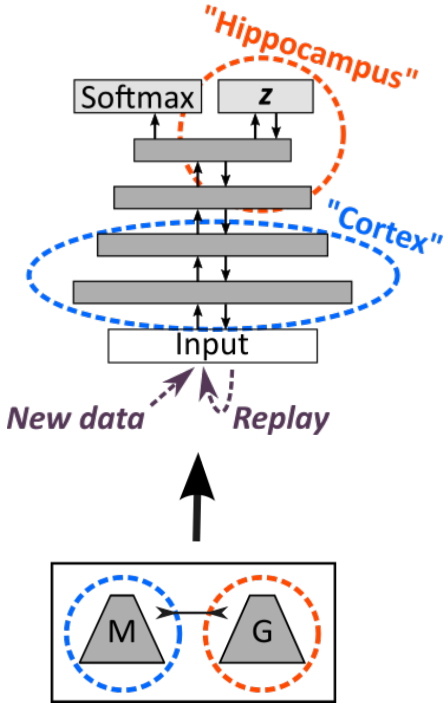
Class-Incremental Learning
Choice between all classes seen so far



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*

(all methods use pre-trained convolutional layers)

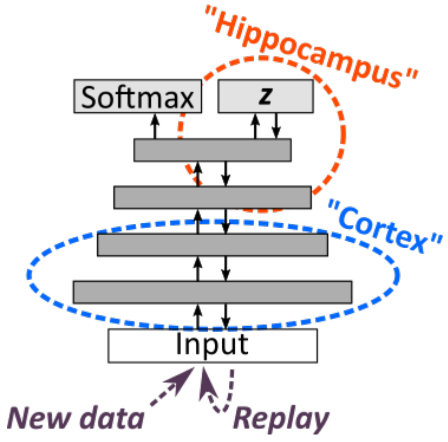
Brain-inspired Modifications to Generative Replay



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

Inspired by brain anatomy

Brain-inspired Modifications to Generative Replay

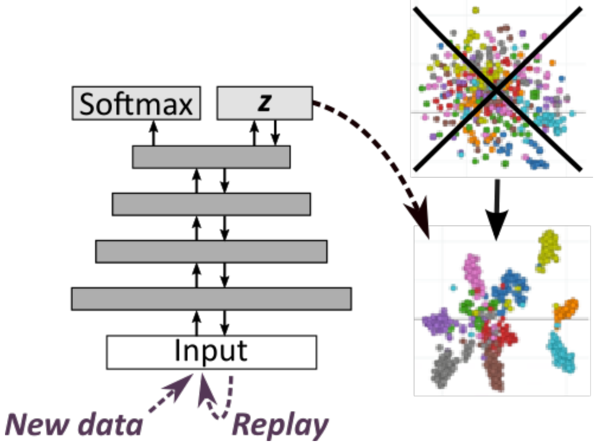


- ***Replay-through-Feedback:*** Merge generator into main model; replay is now generated by the feedback / backward connections

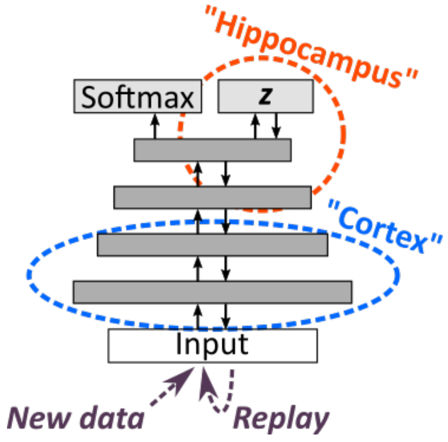
Inspired by brain anatomy

- ***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

Inspired by introspection



Brain-inspired Modifications to Generative Replay

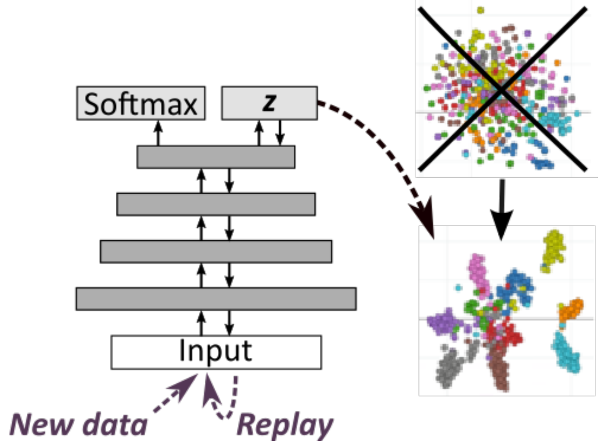


- ***Replay-through-Feedback:*** Merge generator into main model; replay is now generated by the feedback / backward connections

Inspired by brain anatomy

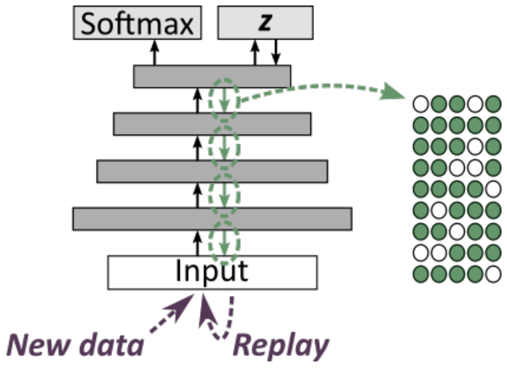
- ***Conditional Replay:*** Enable model to generate specific classes, by replacing the standard normal prior by a Gaussian mixture with a separate mode for each class

Inspired by introspection

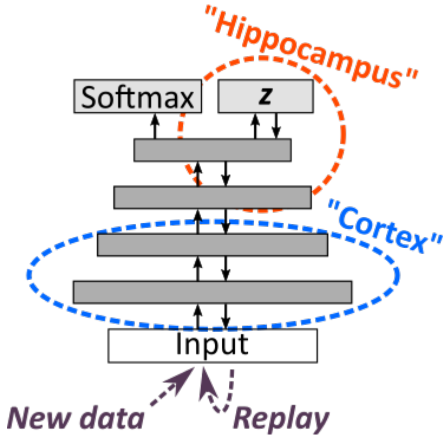


- ***Gating based on Internal Context:*** For each class, inhibit (or gate) a different subset of neurons during the generative backward pass

Inspired by inhibition & context-dependent processing

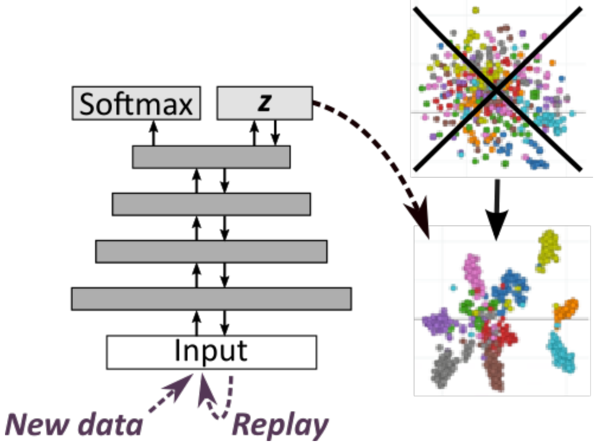


Brain-inspired Modifications to Generative Replay



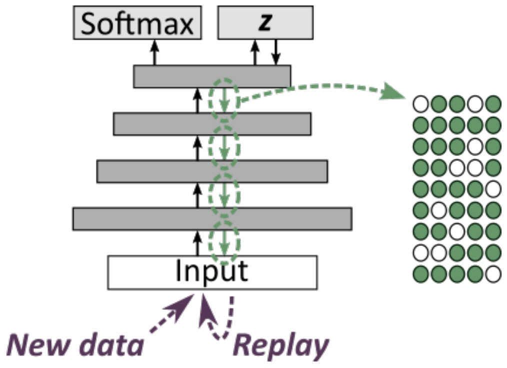
- ***Replay-through-Feedback:*** Merge generator into main model; replay is now generated by the feedback / backward connections

Inspired by brain anatomy



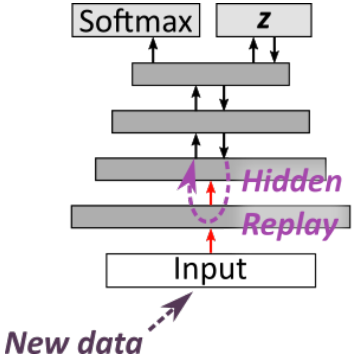
- ***Conditional Replay:*** Enable model to generate specific classes, by replacing the standard normal prior by a Gaussian mixture with a separate mode for each class

Inspired by introspection



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

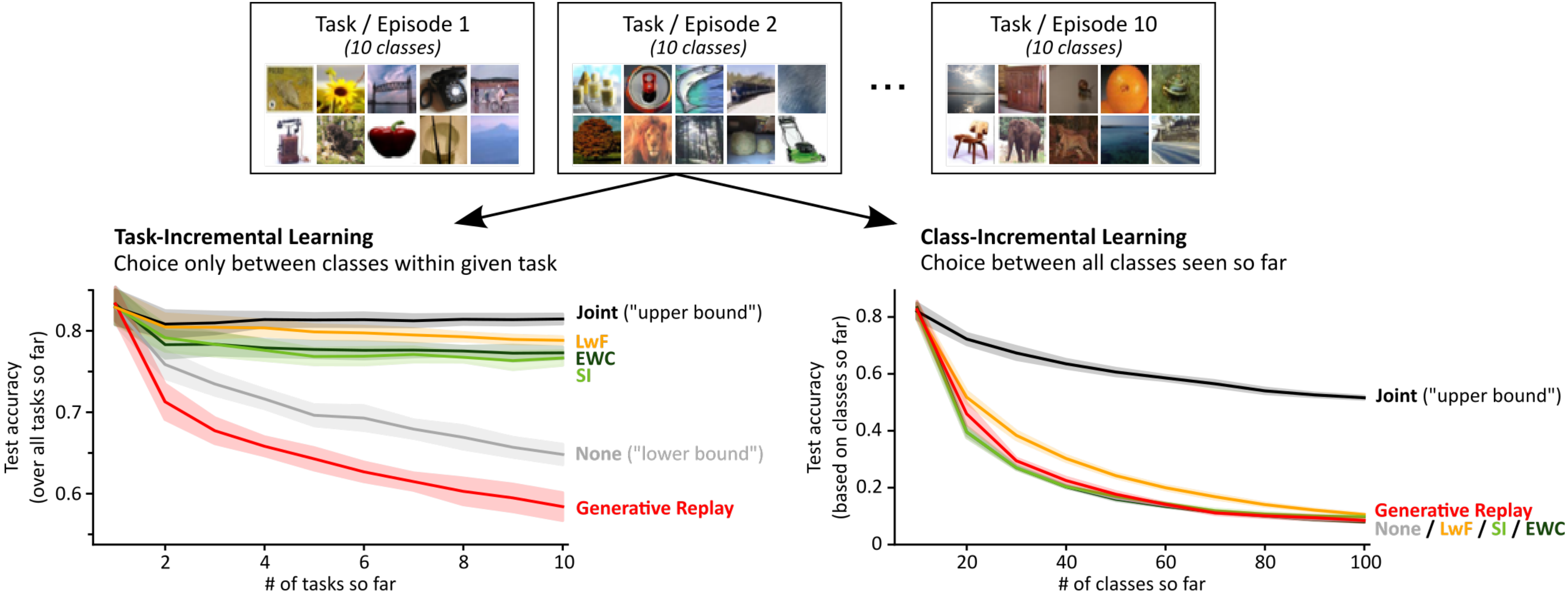
Inspired by inhibition & context-dependent processing



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

Inspired by developmental plasticity

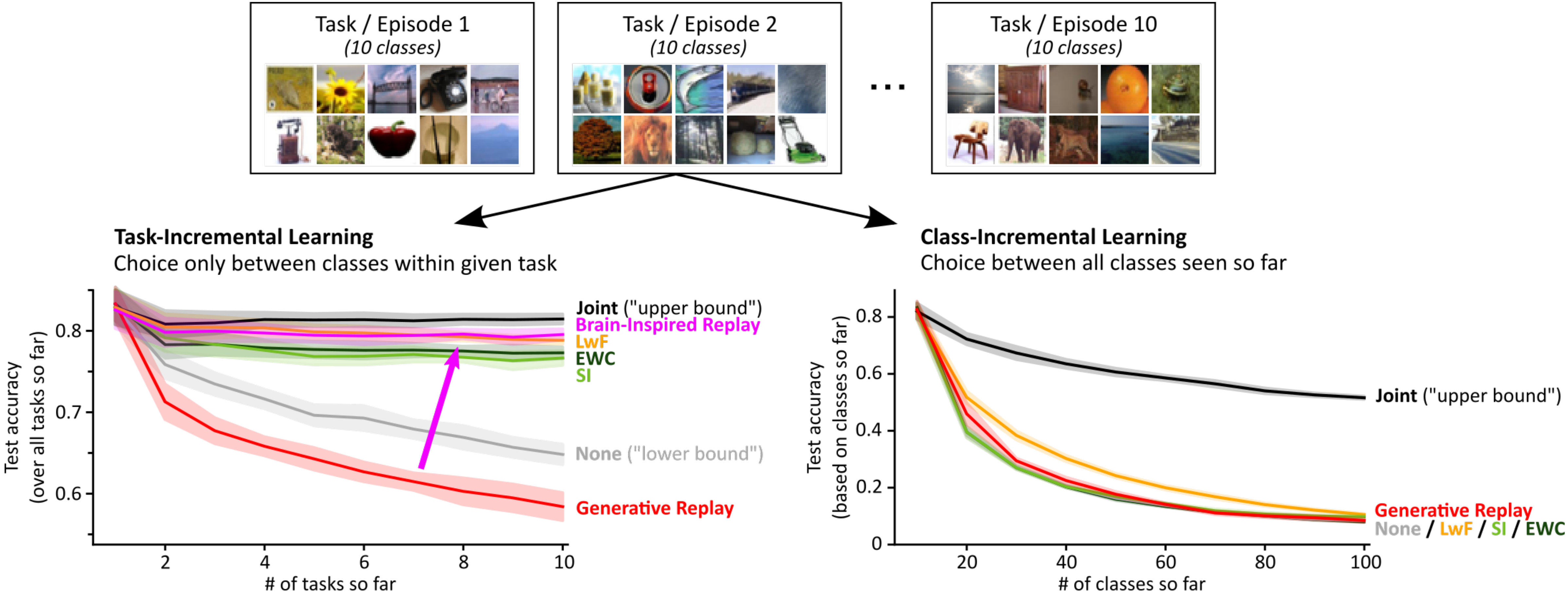
Brain-Inspired Replay on natural images



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*

(all methods use pre-trained convolutional layers)

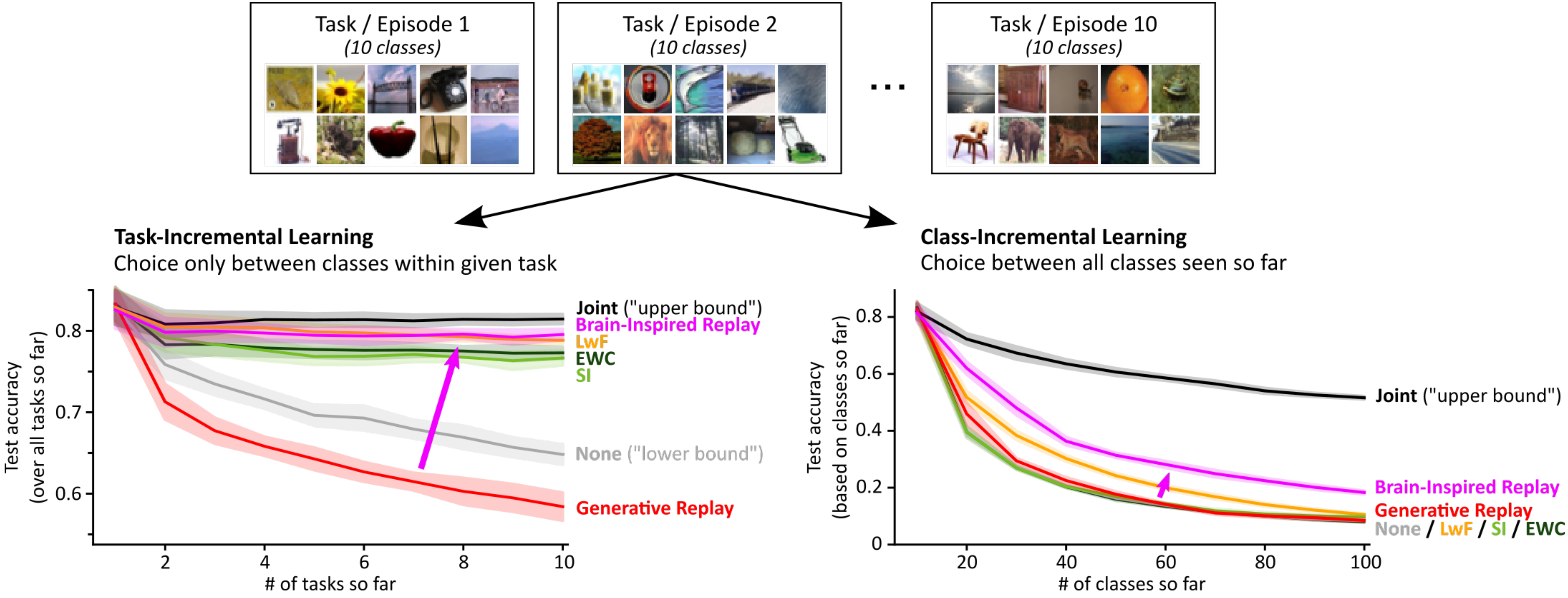
Brain-Inspired Replay on natural images



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*

(all methods use pre-trained convolutional layers)

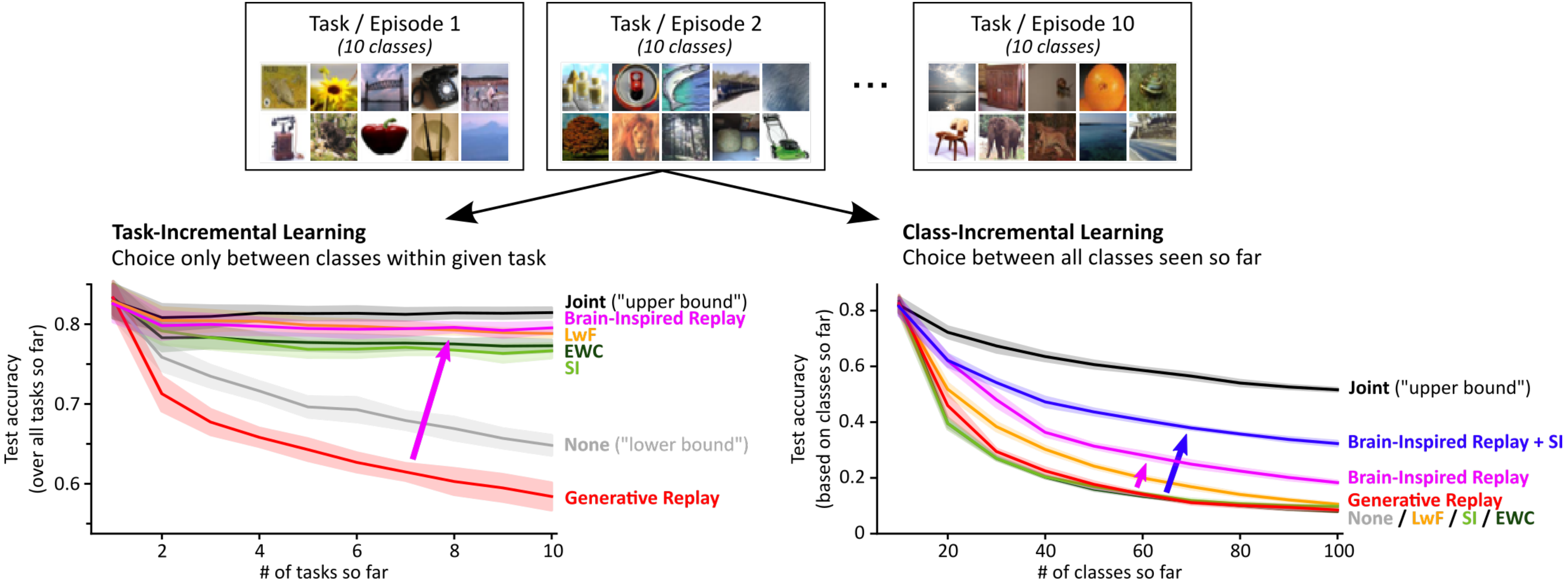
Brain-Inspired Replay on natural images



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*

(all methods use pre-trained convolutional layers)

Brain-Inspired Replay on natural images



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*

(all methods use pre-trained convolutional layers)

Summary

- We proposed a new, brain-inspired variant of generative replay in which internal or hidden representations are replayed that are generated by the network's own, context-modulated feedback connections

<u>Machine Learning contribution</u>

Our method is the first to perform well on the challenging problem of class-incremental learning with natural images without relying on stored data

<u>Cognitive Science contribution</u>

Our method provides evidence that replay could indeed be feasible way for the brain to combat catastrophic forgetting

I'm available to answer questions during Virtual Poster Session #2 (9-10pm GMT)

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.