Incrementally learning new classes: generative replay *versus* generative classification

Gido M. van de Ven[1,2], Zhe Li[1] & Andreas S. Tolias[1,3]

[1] Center for Neuroscience and Artificial Intelligence, Department of Neuroscience, Baylor College of Medicine, Houston, US
[2] Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, UK
[3] Department of Electrical and Computer Engineering, Rice University, Houston, US

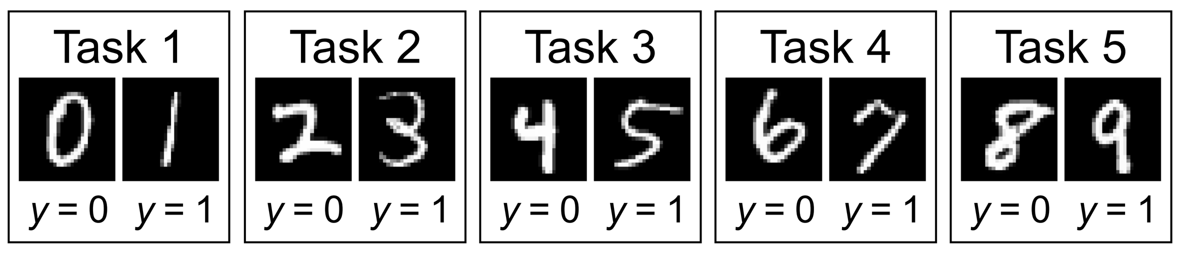
Baylor College of Medicine

Introduction

Incrementally learning from non-stationary data, referred to as 'continual learning', is a key feature of natural intelligence, but an unsolved problem in deep learning. Particularly challenging for deep neural networks is the problem of 'class-incremental learning', whereby a network must learn to distinguish between classes that are not observed together.

Three types of incremental learning

Continual learning is not a unitary problem. In previous work, we identified these three distinct types, or "scenarios", of incremental learning, based on how the mapping that must be learned relates to the aspect of the data that changes over time [1].



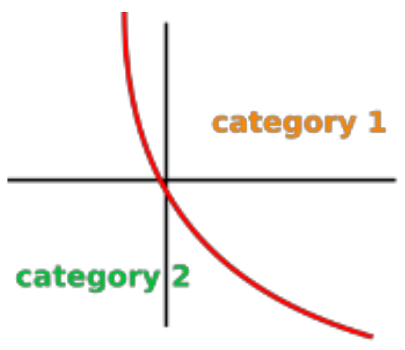
$\mathcal{X}$ = input pixel space
$\mathcal{Y}$ = {0,1}
$\mathcal{T}$ = {1,2,3,4,5}

	Type of choice	*Mapping to learn*
Task-incremental learning	Choice between the two digits of the task	$f: \mathcal{X} \times \mathcal{T} \rightarrow \mathcal{Y}$
Domain-incremental learning	Is the digit odd or even?	$f: \mathcal{X} \rightarrow \mathcal{Y}$
Class-incremental learning	Choice between all ten digits	$f: \mathcal{X} \rightarrow \mathcal{T} \times \mathcal{Y}$

Discriminative vs. generative classification

Discriminative classifiers

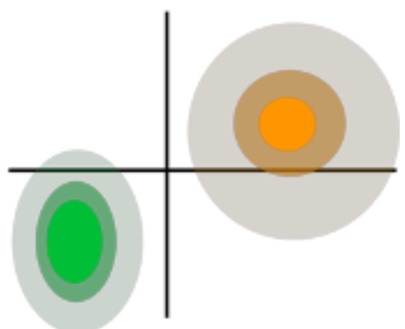
- Directly learn $p(y|\boldsymbol{x})$



- Learn rules / shortcuts / features to distinguish between classes
- Compare classes during *training*

Generative classifiers

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



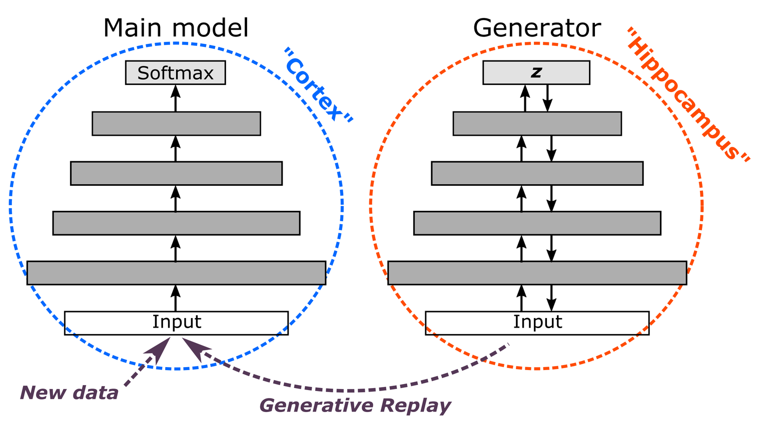
- Learn a model / template / representation for each class
- Compare classes during *inference*

Generative classification rephrases an often challenging class-incremental problem as a typically easier addressable task-incremental problem, whereby each 'task' is to learn a class-conditional generative model.

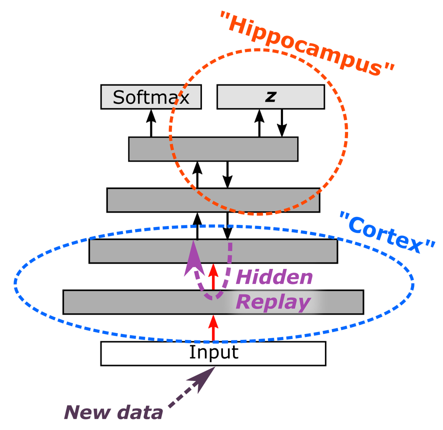
Workarounds for discriminative classifiers

Elastic Weight Consolidation (EWC) / Synaptic Intelligence (SI):
These parameter regularization methods slow down learning for parameters important for past task [2,3]. These methods can be interpreted as performing sequential approximate Bayesian inference on the parameters of a neural network.

Generative Replay (GR):
Both a generative model and a classifier are learned [4]. When training on new classes, samples from the generative model representing previously learned classes are interleaved with the training data of the new classes.

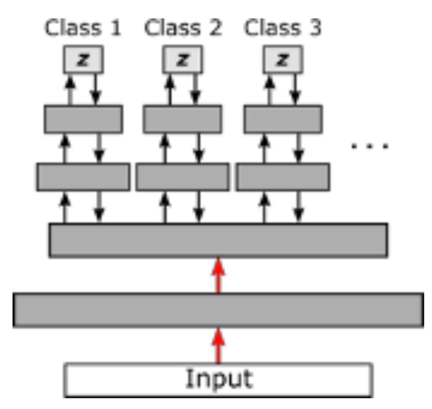


Brain-Inspired Replay (BI-R):
A brain-inspired variant of generative replay in which abstract, high-level representations are replayed that are generated by the network's own feedback connections [5].

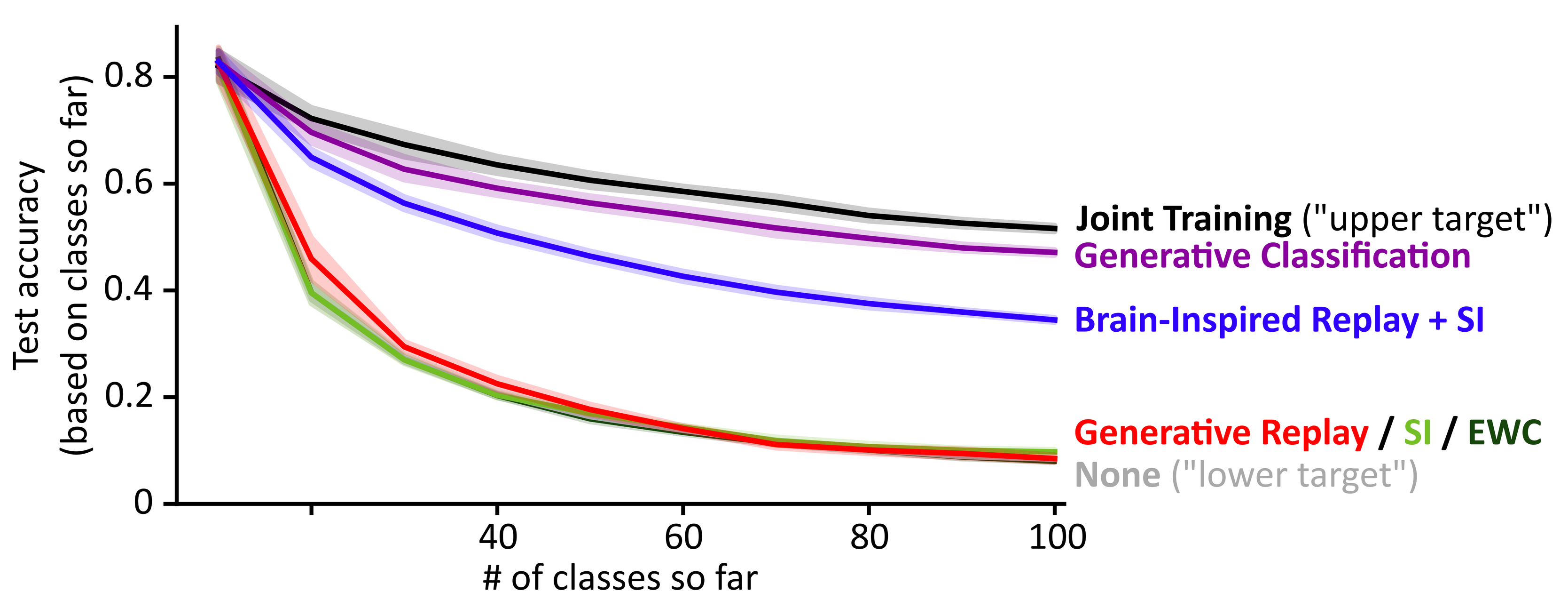


Proof-of-principle implementation

- Separate VAE model for each class
- If a pretrained network is available, the VAE models are trained on the latent features
- Likelihoods are estimated using importance sampling
- The *total* number of parameters is similar to that of generative replay

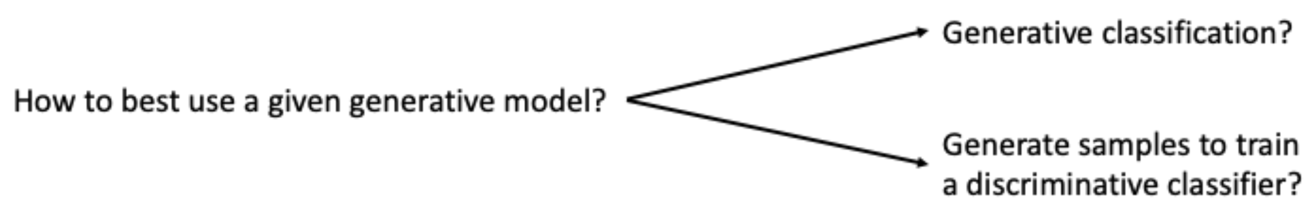


Results on CIFAR-100



Direct comparison: How to use a generative model?

These results indicate that directly using a generative model for generative classification leads to better class-incremental learning performance than using those models indirectly to generate replay for discriminatively training a classifier.



Here we test this more directly, by training a discriminative classifier on the samples of the generative models used by the generative classifier.

	MNIST	CIFAR-10	CIFAR-100	CORe50
Generative classifier	**93.79** (± **0.08**)	**56.03** (± **0.04**)	**49.55** (± **0.06**)	**70.81** (± **0.11**)
Discriminative classifier trained on generated samples	85.93 (± 0.43)	13.71 (± 0.61)	33.84 (± 0.14)	47.86 (± 1.77)

Shown is test accuracy (in %) over all classes. All experiments were performed 10 times with different random seeds, reported are the mean (+/- SEMs) over these runs.

Highlights

- Generative classification is a very promising strategy for incrementally learning new classes.
- This study provides a fresh perspective to the ongoing debate in computational cognitive science whether, or when, the brain performs inference using discriminative or generative computations [6].
- More details: **van de Ven GM, Li Z, Tolias AS (2021) Class-incremental learning with generative classifiers. *CVPR workshop proceedings*: 3611-3620.**

References

[1] van de Ven & Tolias (2019) Three scenarios for continual learning. *arXiv*: 1904.07734.
[2] Kirkpatrick et al. (2017) Overcoming catastrophic forgetting in neural networks. *PNAS* **114**(13): 3521-3526.
[3] Zenke et al. (2017) Improved multitask learning through synaptic intelligence. *arXiv*: 1703.04200.
[4] Shin et al. (2017) Continual learning with deep generative replay. *NeurIPS*: 2994-3003.
[5] van de Ven et al. (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications* **11**(1): 1-14.
[6] DiCarlo et al. (2021) How does the brain combined generative models and direct discriminative computations in high-level vision? *GAC proposal*.

Acknowledgements

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 herein are 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.