SpaceNet: Make Free Space For Continual Learning

Elsevier Neurocomputing Journal

Ghada Sokar (PhD Student)

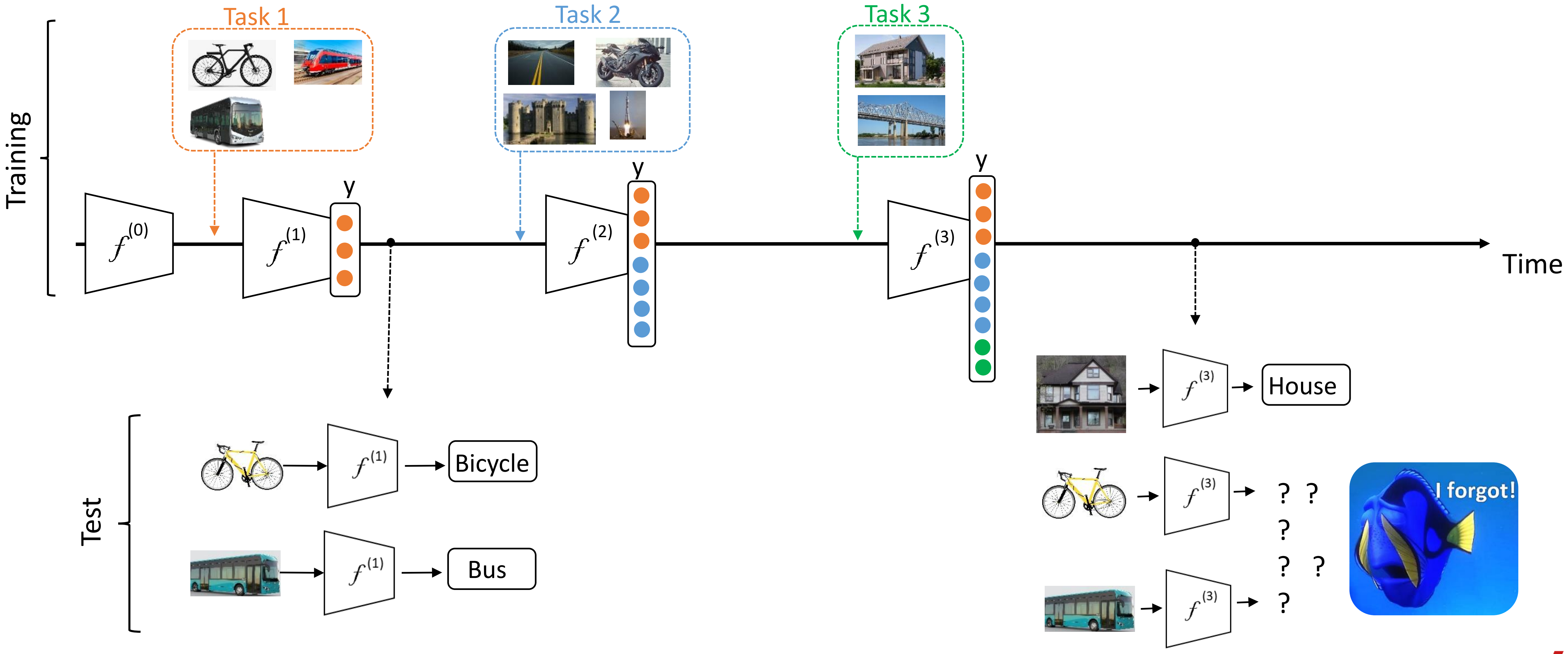
Eindhoven University of Technology, The Netherlands

g.a.z.n.sokar@tue.nl

ContinualAI Meetup. January 29th, 2021

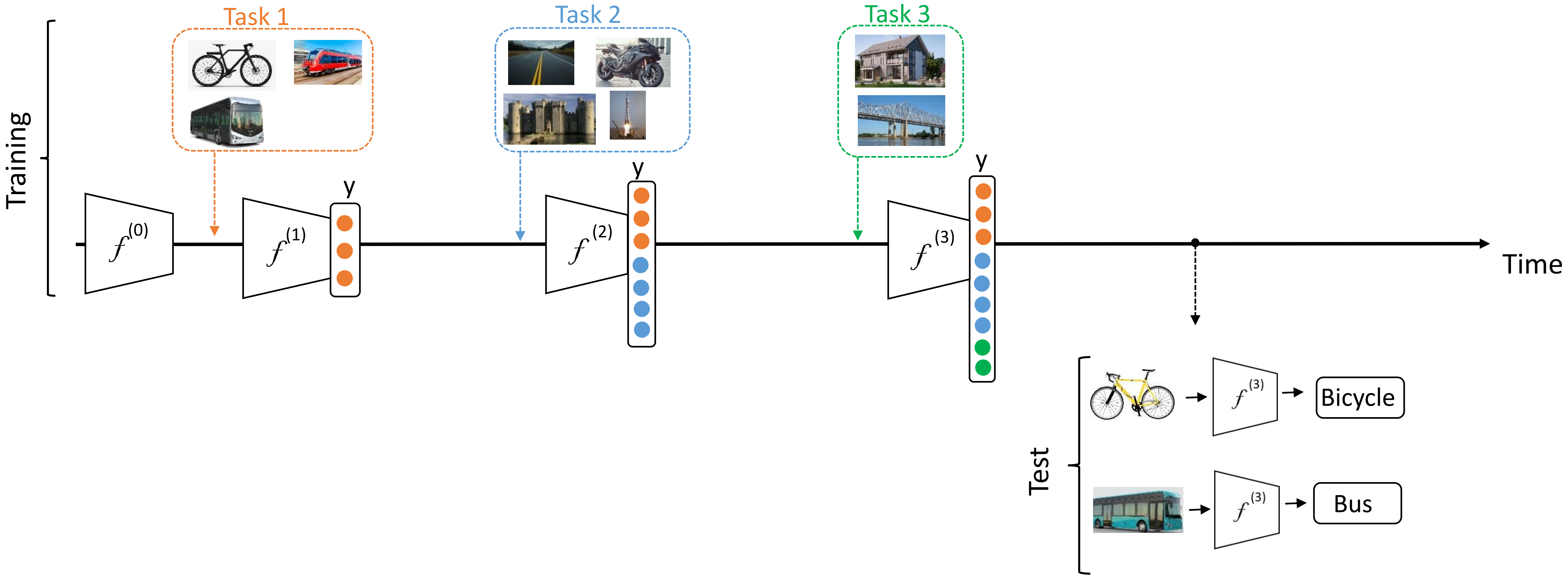
Continual Learning

Class Incremental Learning



Continual Learning

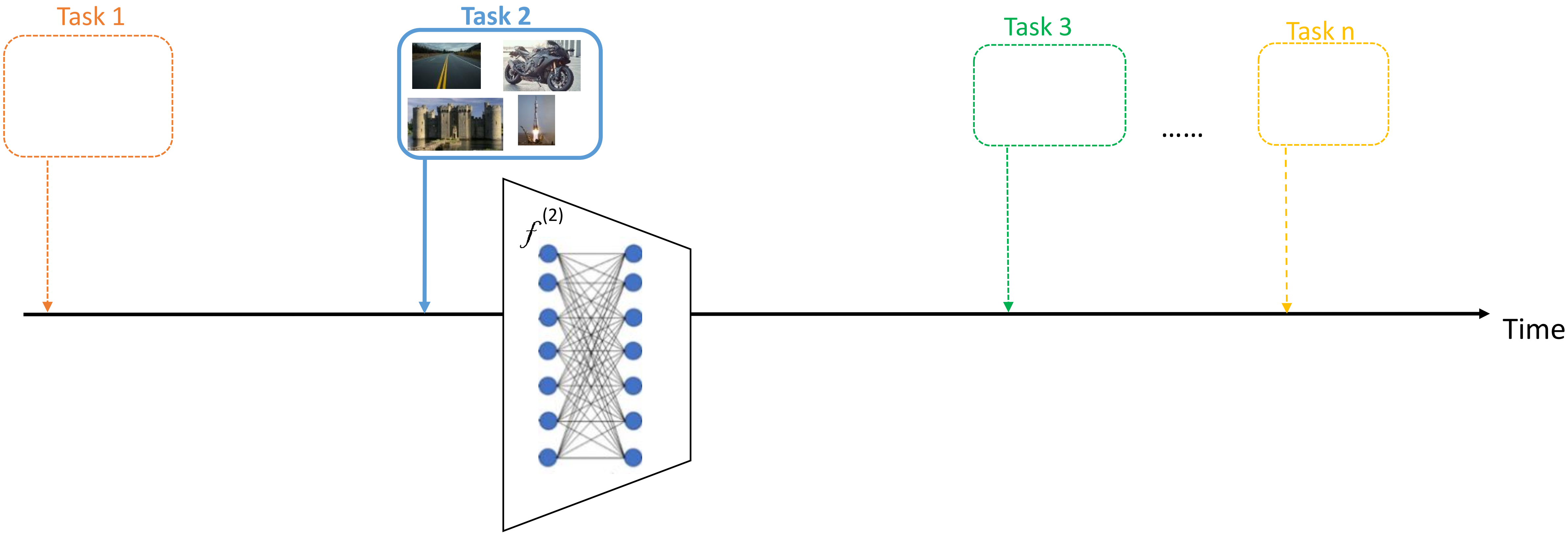
Class Incremental Learning



Why does the network forget?

Why does the network forget?

- When the task arrives, it utilizes all the available capacity (network parameters)
- It does not account for previous tasks or leave space for future tasks



Does each task really need the full capacity?

Does each task really need the full capacity?

- Deep neural networks are often over parameterized

Does each task really need the full capacity?

- Deep neural networks are often over parameterized
- Direct training of *sparse networks* achieves the same performance of dense network (Mocanu et al., 2018 Nature communication [1])

Test accuracy% (top-5, top-1) of Resnet-50 trained on Imagenet [2]

		top-5			top-1		
	Density level (# Parameters)	**20% (7.3M)**	**10% (5.1M)**	**100% (25.6M)**	**20% (7.3M)**	**10% (5.1M)**	**100% (25.6M)**
Adaptive Sparse Training	SET [1]	91.2	90.1	92.4	72.6	70.4	74.9
	Dynamic sparse [2]	92.4	90.5		73.3	71.6	
Static Sparse NN	Static sparse	90.4	88.4		71.6	67.8	

Key takeaway

Number of parameters can be reduced by 80~90% without degrading accuracy

[1] Mocanu, Decebal Constantin, et al. "Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science." *Nature communications* 9.1 (2018): 1-12.

[2] Mostafa, Hesham, and Xin Wang. "Parameter efficient training of deep convolutional neural networks by dynamic sparse reparameterization." International Conference on Machine Learning. PMLR, 2019.

How does the brain process information?

Neuroscience observations

- Neurons encode information in a sparse and distributed way (Attwell and Laughlin, 2001[1])
- The percentage of neurons active at the same time to be between 1% and 4% (Lennie, 2003[2])

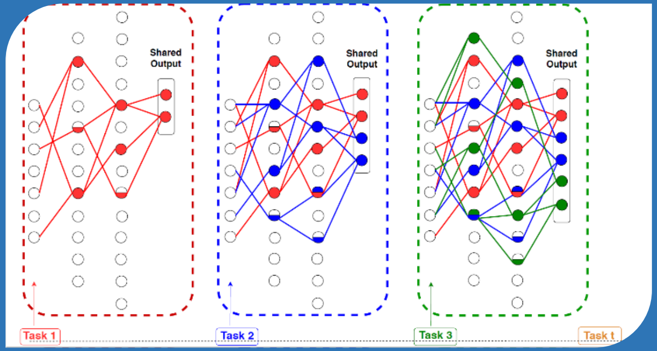
Key takeaway

The brain is so efficient even though the activity of neurons is highly sparse

[1] Attwell, D. and Laughlin, S. (2001). An energy budget for signaling in the grey matter of the brain. Journal of Cerebral Blood Flow and Metabolism, 21(10), 1133– 1145

[2] Lennie, P. (2003). The cost of cortical computation. Current Biology, 13, 493–497.

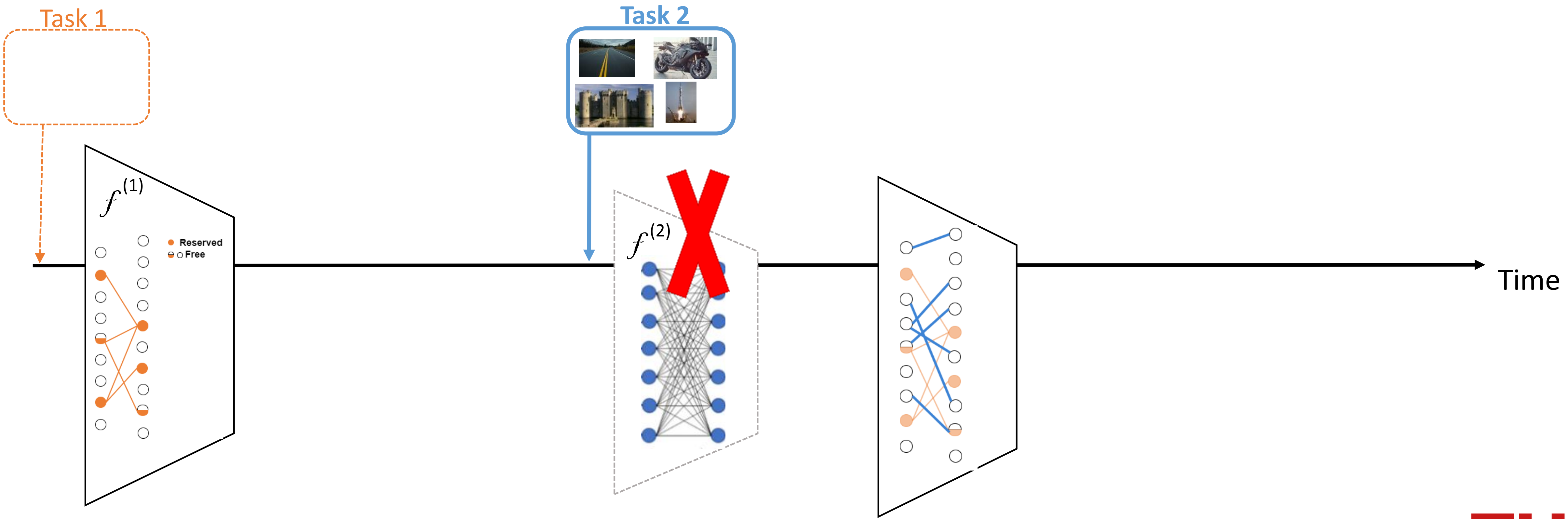
SpaceNet: Make Free Space For Continual Learning



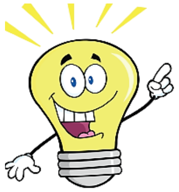
SpaceNet

Key Idea I

- Allocate *sparse connections* for each task between the free neurons



SpaceNet



- Key Idea II

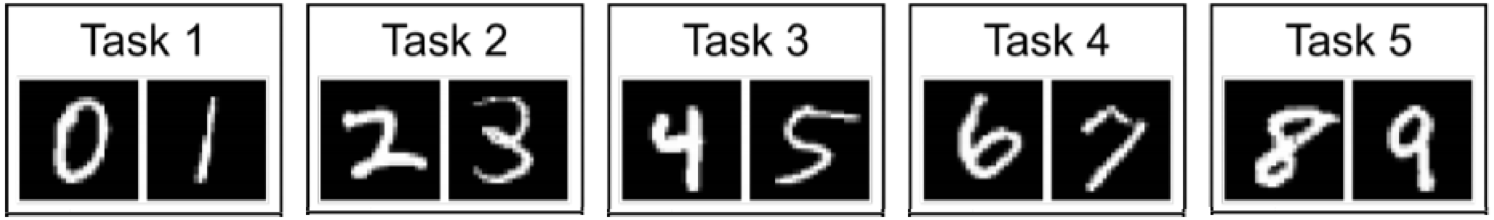
  Train each task using *adaptive sparse training* in which we *compact* the sparse connections in the most *important* neurons for that task

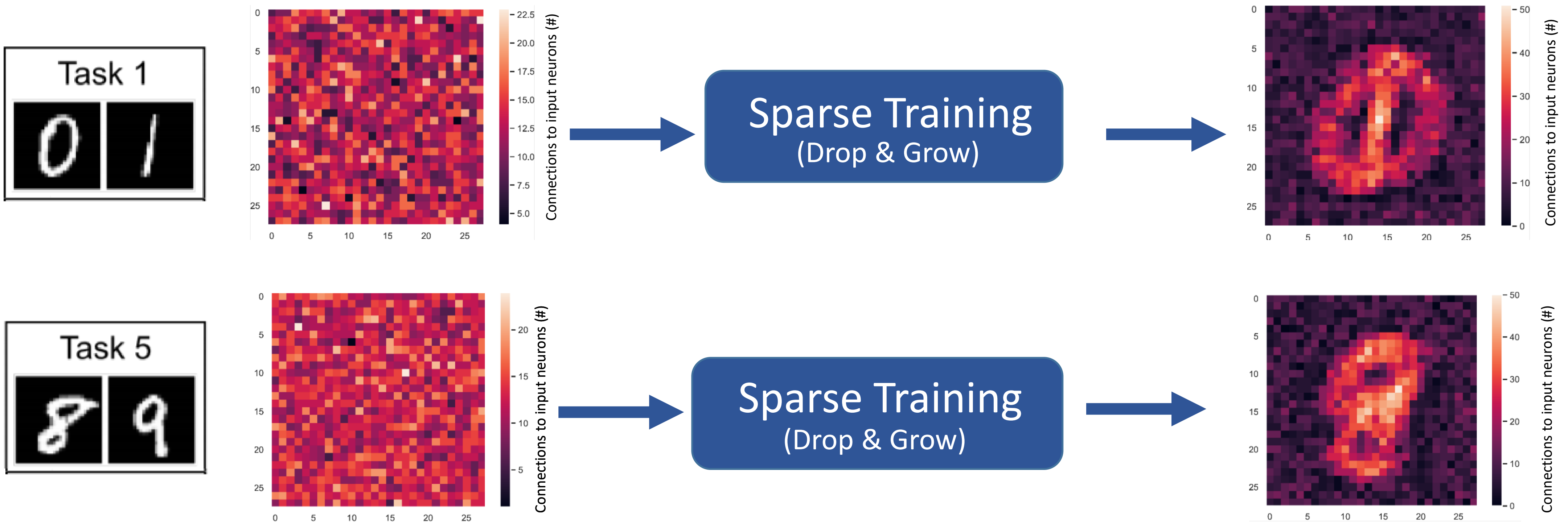
  Produce *sparse representation* and Leave *free space* for future tasks

Task 2

Results – Connections distribution

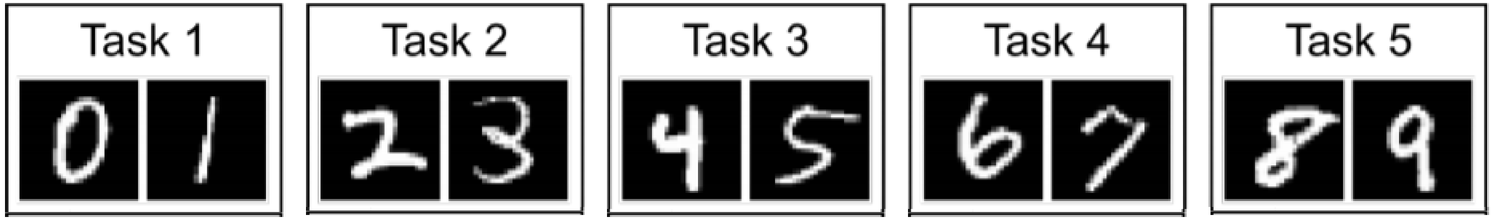
- Split MNIST





Results – Performance

- Split MNIST



Strategy	**Method**	**Accuracy (%)**
Regularization	EWC [1]	20.01
	SI [2]	19.99
Rehearsal	DGR [3]	90.79
	SpaceNet-Rehearsal	**95.08**
Architectural	**SpaceNet (ours)**	75.53
	Static-SparseNN	61.25

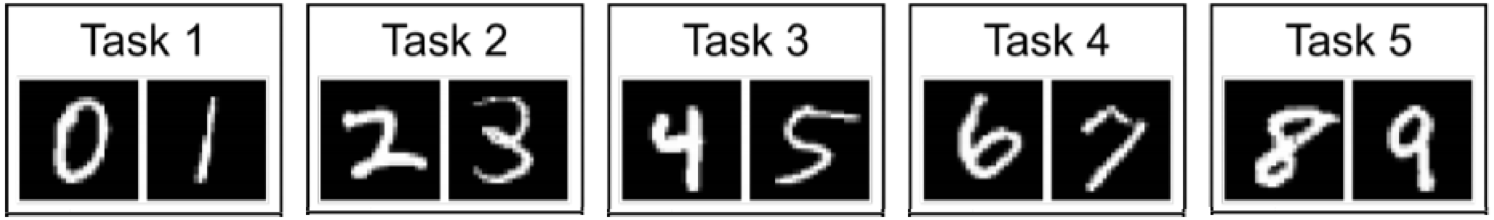
[1] Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks." *Proceedings of the national academy of sciences* 114.13 (2017): 3521-3526.

[2] Zenke, Friedemann, Ben Poole, and Surya Ganguli. "Continual learning through synaptic intelligence." *International Conference on Machine Learning*. PMLR, 2017.

[3] Shin, Hanul, et al. "Continual learning with deep generative replay." *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017.

Results – Performance

- Split MNIST



Strategy	Method	Accuracy (%)
Regularization	EWC	20.01 ± 0.01
	SI	19.99 ± 0.11
Rehearsal	DGR	90.79 ± 1.02
Architectural	DEN	56.95 ± 1.29
	SpaceNet (ours)	**75.53 ± 1.82**
	Static-SparseNN	61.25 ± 2.30

[1] Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks." *Proceedings of the national academy of sciences* 114.13 (2017): 3521-3526.

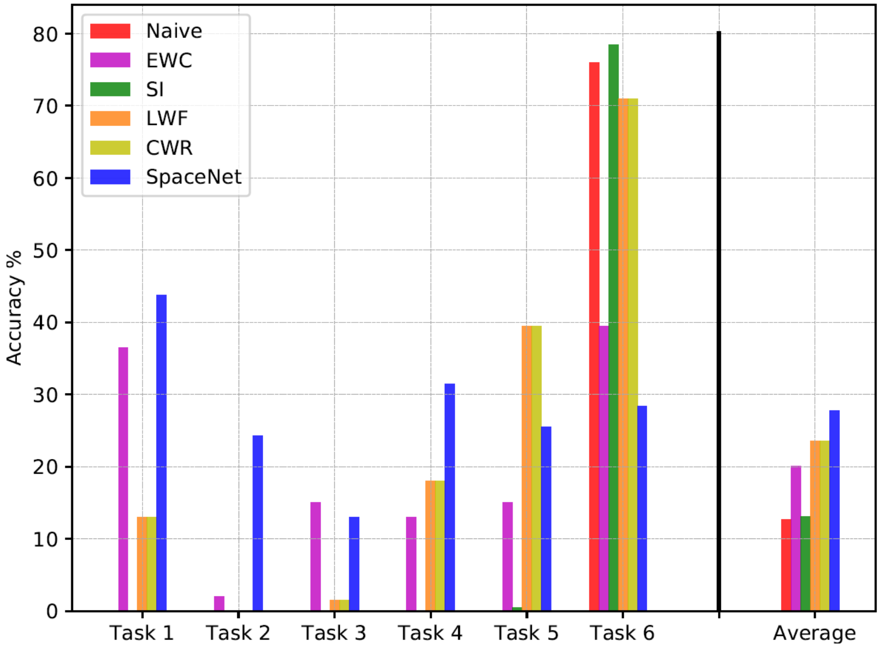
[2] Zenke, Friedemann, Ben Poole, and Surya Ganguli. "Continual learning through synaptic intelligence." *International Conference on Machine Learning*. PMLR, 2017.

[3] Shin, Hanul, et al. "Continual learning with deep generative replay." *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017.

Results

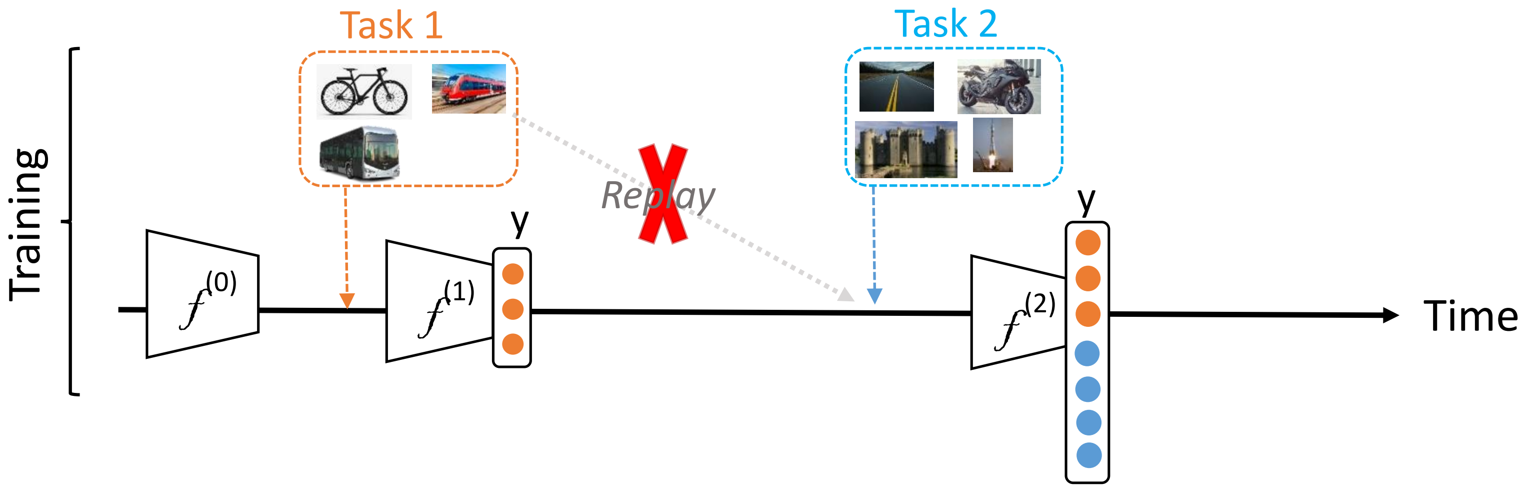
- Cifar-10/100

Task 1	**Task 2**	**Task 3**	**Task 4**	**Task 5**	**Task 6**
CIFAR 10 (All classes)	CIFAR 100 (Classes 1-10)	CIFAR 100 (Classes 11-20)	CIFAR 100 (Classes 21-30)	CIFAR 100 (Classes 31-40)	CIFAR 100 (Classes 41-50)



SpaceNet characteristics

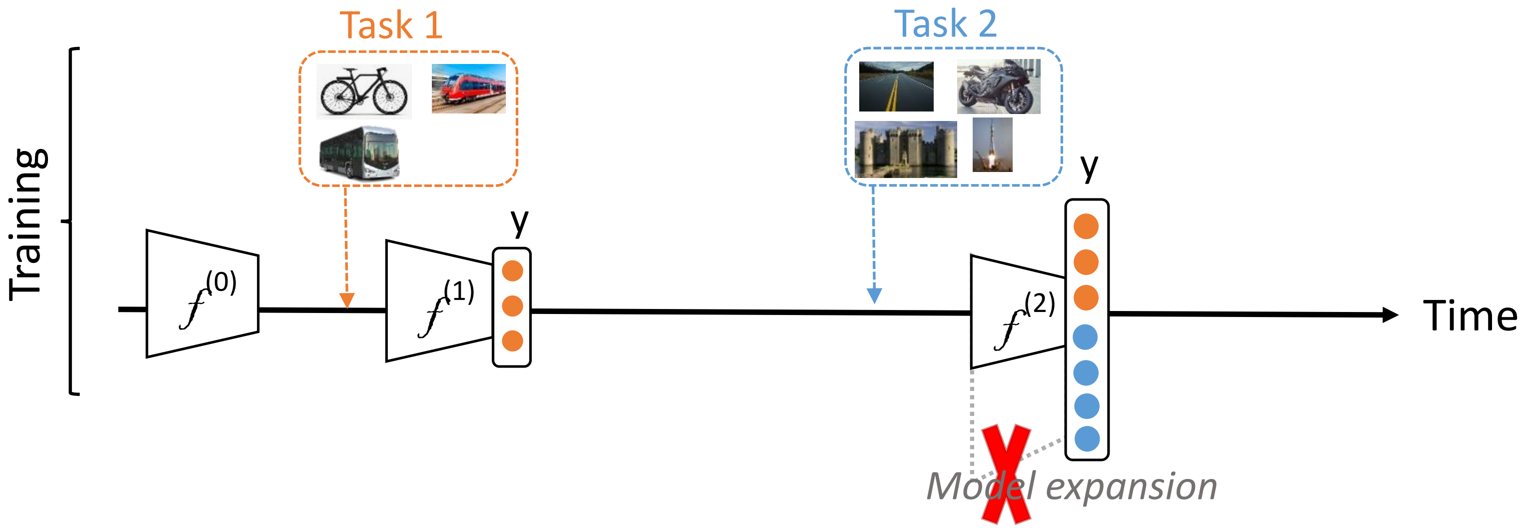
✓Rehearsal-Free



SpaceNet characteristics

✓Rehearsal-Free

✓Utilize the fixed capacity

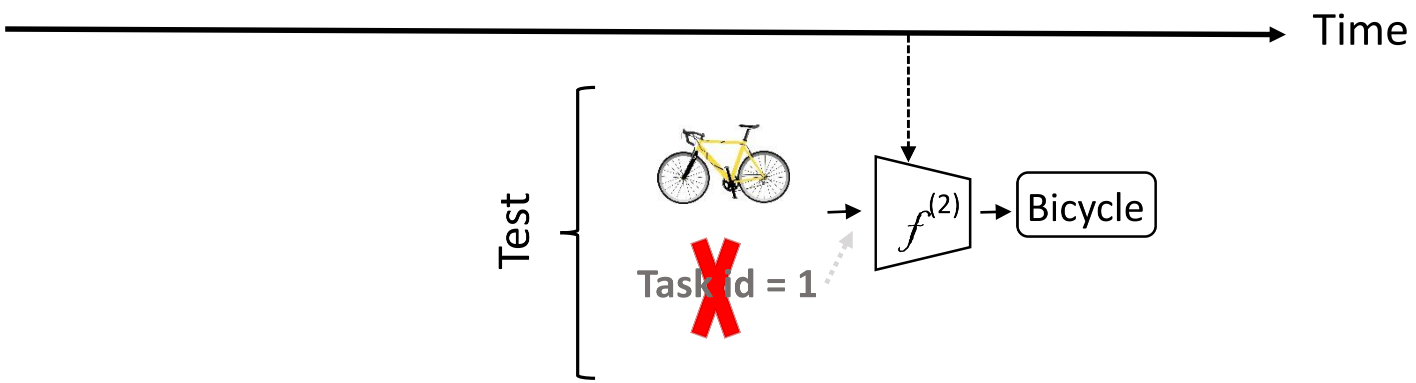


SpaceNet characteristics

✓Rehearsal-Free

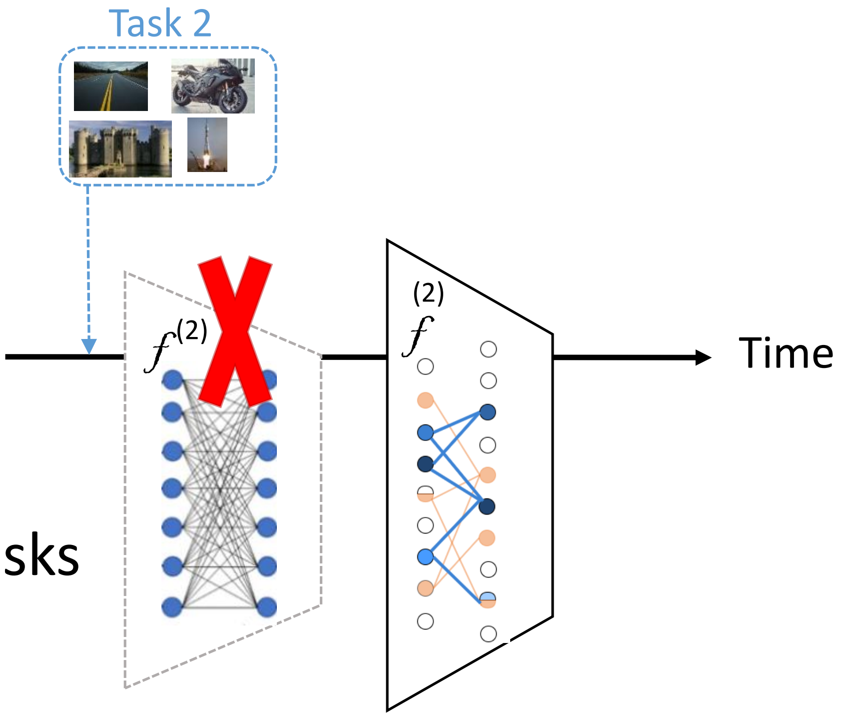
✓Utilize the fixed capacity

✓Task-agnostic inference



SpaceNet characteristics

- ✓Rehearsal-Free
- ✓Utilize the fixed capacity
- ✓Task-agnostic inference
- ✓Train each task using sparse training
  - ✓Sparse representations to reduce forgetting
  - ✓Compact space for each task, leaving room for future tasks



Future work and open questions

- How about a larger sequence of tasks?

- Can we allow for positive backward transfer as well?

- Can we use some of the previously allocated connections instead of allocating new ones?

Thank You!

Questions ?

Feel free to reach out!

g.a.z.n.sokar@tue.nl

SpaceNet: Make Free Space For Continual Learning



Ghada Sokar[1] Decebal Mocanu[1,2] Mykola Pechenizkiy[1]

[1] Eindhoven University of Technology, The Netherlands
[2] University of Twente, The Netherlands

Journal: Elsevier Neurocomputing Journal
Preprint: https://arxiv.org/pdf/2007.07617.pdf