Building invariances into neural decoding through adaptive self-alignment

Mehdi Azabou[1], Ran Liu[1], Mohammad Gheshlaghi Azar[2], Chi-Heng Lin[1], Kiran Bhaskaran Nair[3], Erik Johnson[4], Max Dabagia[1], Will Gray Roncal[4], Michal Valko[2], Keith Hengen[3], Eva Dyer[1,5]
[1]Georgia Institute of Technology, [2]DeepMind, [3]Washington University in St. Louis, [4]Johns Hopkins University Applied Physics Laboratory, [5]Empry University

Abstract

Traditionally, neural decoding has been performed through supervised approaches that aim to map specific behaviors or stimuli to specific neural activity patterns through labeled data. However, supervised methods often fail to generalize to new animals, new time points or as the animal shifts across different levels of engagement or learning. To break free from the chains of supervision, a number of unsupervised learning perspectives have been introduced, many of them essentially aiming to "reconstruct" or generate new neural activity patterns after transforming them down into a lower-dimensional latent space. However, these reconstruction-based approaches often suffer from similar challenges when faced with domain shift and thus do not explicitly account for or try to build invariance to temporal or other shifts.

Here, we ask whether we can build in invariances into representations more directly. To do this, we explore the utility of "self supervision" for learning representations for multi-unit neural datasets. Rather than using labels to guide learning, we essentially ask the network to build a representation that makes it easy to predict across nearby points in time, as well as across adaptively "mined" samples that are nonlocal but close in terms of their representations in the network.

Learning representations from neural activity

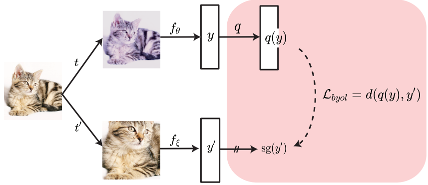
Our aim is to form a representation that captures the latent brain state. This is met with multiple challenges:
- High-dimensionality of the neural space
- Lack of well established augmentations in the neural domain
- Hidden background states that lead to further domain shift
- Lack of structural bias that can be injected in the architecture

Domain shift during data collection makes the uncovering of the true latent manifold harder.

Self-supervised Learning

Self-supervised methods leverage augmentations to bake certain invariances into the representation. These methods generalize surprisingly well, in spite of the absence of labels.

BYOL [2] does this by using a predictor tasked to predict across representations of two independently augmented views of the same sample.



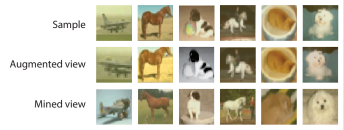
Approach

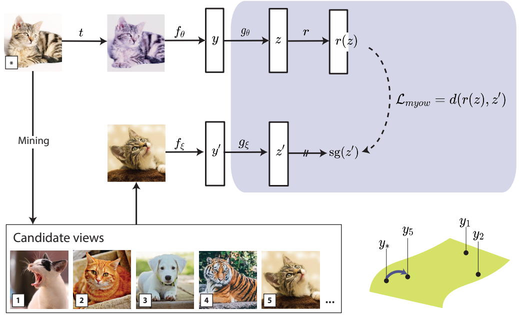
Challenge: Current self-supervised learning (SSL) methods depend on handcrafted augmentations to perform well.
Solution: Instead of augmenting the data, mine the dataset.

Find other samples that are close in the representation space and use them as positive views.

Both augmented and mined views are used to learn the representation.

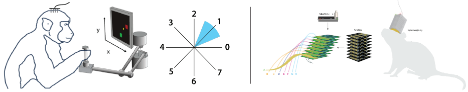
$$\mathcal{L} = \mathcal{L}_{byol} + \lambda \mathcal{L}_{myow}$$



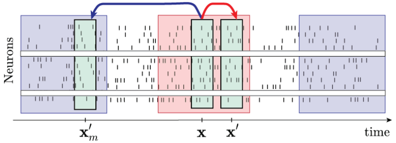


MYOW for neural decoding

We tested our approach on datasets from trial-based reaching datasets from non-human primates, and on free behavior in rodent visual cortex and hippocampus. We considered the following decoding tasks:
- Reach direction decoding: Predict one of eight reach targets in the reaching task.
- Arousal state decoding: Predict Rem, nRem, or Wake.



We consider augmentations including: Temporal Shift, Randomized Dropout, Gaussian Noise and Sparse additive noise (Pepper).



Relative positioning: when mining for a view that has a similar latent neural state, we restrict the set of candidates to be temporally distant from the sample itself.

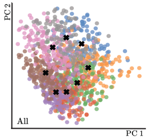
The MYOW predictor thus bridges the domain shift by aligning different instanciations of the same neural state across time.

State-of-the-art representation quality on neural activities

We show that by incorporating nonlocal but "similar" time points into the system, and predicting across these distinct time points, the network can build time-invariant representations that allow for more faithful decoding on downstream tasks and resilience to domain shift.

	Reach								Sleep	
	Chewie-Day 1		Chewie-Day 2		Mihi-Day 1		Mihi-Day 2		Rat-V1	Mouse-CA1
	Acc	δ-Acc	Acc	δ-Acc	Acc	δ-Acc	Acc	δ-Acc	F1-score	F1-score
Supervised	63.29	77.22	**72.29**	**81.51**	63.64	79.02	61.49	68.44	86.34	93.01
Autoencoder	48.40	67.51	46.79	65.84	50.94	68.03	55.19	74.98	34.17	57.73
SimCLR	59.02	78.65	50.39	64.75	59.55	77.52	54.47	71.65	80.28	80.23
BYOL	63.80	81.90	57.17	77.36	59.50	79.78	60.82	78.30	85.42	93.24
MYOW	**70.41**	**86.24**	60.95	81.36	**70.48**	**83.24**	**64.35**	**80.58**	**88.01**	**93.70**

Table 1 Accuracy in the prediction of brain states from spiking neural activity



When tested on these diverse neural datasets, we found robust results using similar augmentations (i.e., temporal shifts, neuron dropout), suggesting that this idea can beused widely in different systems using the same basic principles.

Ablations

We train BYOL and MYOW on the Chewie-Day 1 dataset and use different sets of augmentations. We show the importance of augmentations in building good representations. We also show that, in all these settings, MYOW brings in a learning signal that forms consistently better representations.

On the right, we visualize the latent space when it is learned through Temporal Shift only.

	TS	RDrop	Noise	Pepper	Acc
BYOL	✓				41.75
		✓	✓	✓	55.70
	✓	✓			61.39
	✓	✓	✓	✓	63.80
MYOW	✓				46.61
		✓	✓	✓	53.15
	✓	✓			67.97
	✓	✓	✓	✓	70.41

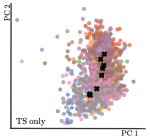


Table 2 Accuracy in the prediction of brain states from spiking neural activity

Conclusion & Future work

1. We introduce a new method for SSL that mines the dataset to find positive examples and uses them for across-sample prediction. We show that our approach can be used to learn meaningful representations for brain activity datasets, and demonstrate the promise of this method in settings where augmentations alone may be insufficient to drive learning.

2. In our application to spiking neural data, we demonstrate that both dropout and temporal augmentations are necessary for building meaningful representations of different brain states.

For future work, we plan to develop generative models that simultanesouly produce high-quality latent representation and generate realistic neural firing rates. The framework in question would leverage components from both reconstruction-based methods and predictive self-supervised methods, and learn a multi-objective representation.

https://nerdslab.github.io/myow/

References

[1] Azabou, et al., Mine Your Own vieW: Self-Supervised Learning Through Across-Sample Prediction, 2021
[2] Grill et al., Bootstrap your own latent: A new approach to self-supervised learning, 2020

Funding

NIH BRAIN R01: Theory, Models, Methods Program (NIH- 1R01EB029852)