Neuroscience-Inspired Encoding and Learning: A Path to Robust Representation Learning

We present a neuroscience-inspired strategy to improve the robustness of encoder models by combining Artificial Kuramoto Oscillatory Neurons (AKOrN) [1] with PhiNet [2], a non-contrastive self-supervised learning approach. AKOrN models neurons as interacting oscillators using the Kuramoto framework, while PhiNet draws inspiration from the hippocampus and temporal prediction theory to build stable, generalizable representations without relying on negative pairs.

Our motivation stems from the challenge of learning robust representations in domains such as brain imaging, where datasets are small, noisy, and highly variable. Conventional vision models tend to overfit and are vulnerable to adversarial perturbations. By integrating the oscillatory dynamics of AKOrN with the predictive structure of PhiNet, we aim to obtain models that are inherently more robust without the need for adversarial training, synthetic augmentation, or heavy post-training procedures.

We pretrain AKOrN with different SSL methods and evaluate performance on CIFAR-10 and CIFAR-100 under the AutoAttack benchmark. With 200 pretraining epochs, SimSiam [3] and PhiNet already improved robustness over contrastive approaches such as SimCLR [4], achieving robust accuracies around 69% while maintaining clean accuracies of about 84%. Extending pretraining to 400 epochs further amplified these differences. On CIFAR-10, AKOrN combined with PhiNet reached a robust accuracy of **76.56%**, surpassing the current state of the art of 75.28%. On CIFAR-100, the same framework achieved **46.62%**, again outperforming the best reported result of 44.78%. These results are particularly significant given that the leading methods in RobustBench rely on adversarial training or synthetic data, whereas our approach does not.

Beyond performance, this work suggests that drawing on neuroscientific principles can make representation learning both more general and more resilient to perturbations. The oscillatory dynamics in AKOrN naturally encode interactions across pixels, while PhiNet's dual learning objectives encourage both rapid adaptation and stable long-term representation. Together, these mechanisms yield encoders that are competitive with or superior to existing methods in adversarial robustness.

Our findings open promising directions for future work, particularly in applying neuroscience-inspired models to domains such as brain imaging, where robustness is essential for reliable interpretation and hypothesis testing. More extensive experiments on larger and more complex datasets will be required to fully assess the potential of this framework, as well as exploring integrations with architectures such as ResNet to improve clean accuracy further. Nonetheless, these results demonstrate that careful architectural design and biologically motivated pretraining can achieve state-of-the-art robustness without additional training tricks, advancing both machine learning and its applications to neuroscience.

[1] Takeru Miyato, Sindy Löwe, Andreas Geiger, and Max Welling. Artificial kuramoto oscillatory neurons, 2025.

[2] Satoki Ishikawa, Makoto Yamada, Han Bao, and Yuki Takezawa. Phinets: Brain-inspired non-contrastive learning based on temporal prediction hypothesis, 2025.

[3] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 15750–15758,2021.165

[4] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations.