Rethinking Hebbian Principle: Low-Dimensional Structural Projection for Unsupervised Learning

Shikuang Deng, Jiayuan Zhang, et al.

The Hebbian Update Problem

- Hebbian learning is biologically plausible but struggles in modern deep networks.

- **Key issues:** unstable updates and a lack of a clear, scalable objective.

- This limits performance and explainability.

Key idea and intuition

- **Our Insight:** Preserve Structure
  Reformulate "Neurons that fire together, wire together" as a mathematical objective:

$$\text{Loss} = |Similarity_{output} - Similarity_{input}|$$

- The goal: Make relationships between samples in the output match the input.
- Theoretically, this finds the **optimal low-dimensional projection** of the input data.

Our method

$$Y = f(X, W), \quad Z = \phi(Y, \theta)$$
$$\hat{X} = X/\|X\|_2, \quad \hat{Z} = Z/\|Z\|_2$$
$$(K_X) = XX^\top, \quad (K_Z) = ZZ^\top$$
$$\mathcal{L}_{\text{SPHeRe}} = \|K_Z - K_X\|_F^2, \quad \mathcal{L}_{\text{orth}} = \|Z^\top Z - I\|_F^2$$
$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{SPHeRe}} + \lambda \mathcal{L}_{\text{orth}}$$

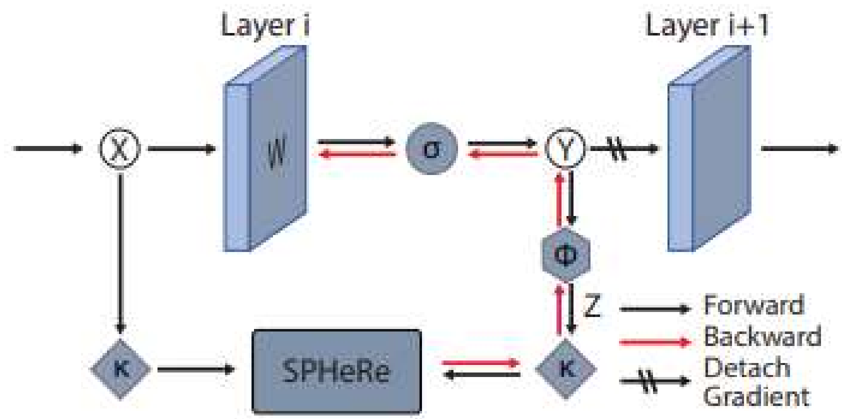


Figure 1: The concept of SPHeRe.

Results: State-of-the-Art Performance

SPHeRe outperforms existing Hebbian methods on standard benchmarks.

Approaches	CIFAR-10	CIFAR-100	Tiny-ImageNet
D-CSNN [30]	73.7	45.17	14.36
Hard WTA [31]	72.2	32.56	–
Hard WTA [11]	74.6	–	–
SoftHebb [12]*	80.3	56.0	–
SoftHebb [12]†	78.86	54.18	34.12
SPHeRe	**81.11 ± 0.11**	**56.79 ± 0.69**	**40.33 ± 0.24**

Results: Superior Reconstruction

SPHeRe-encoded features best reconstruct original images, retaining more complete information.

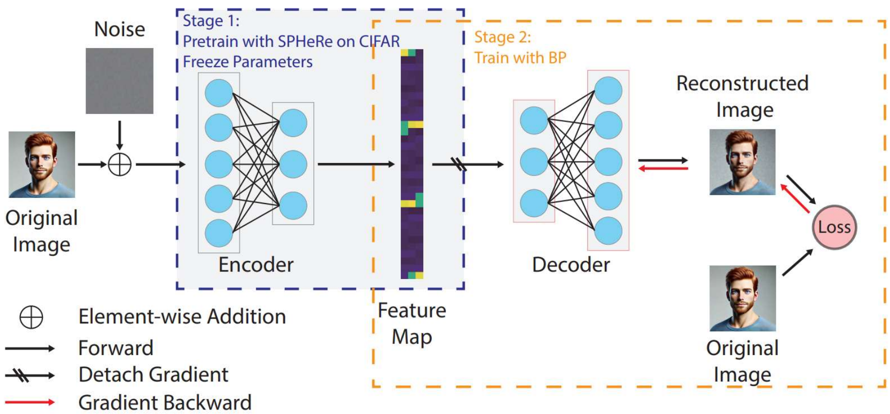
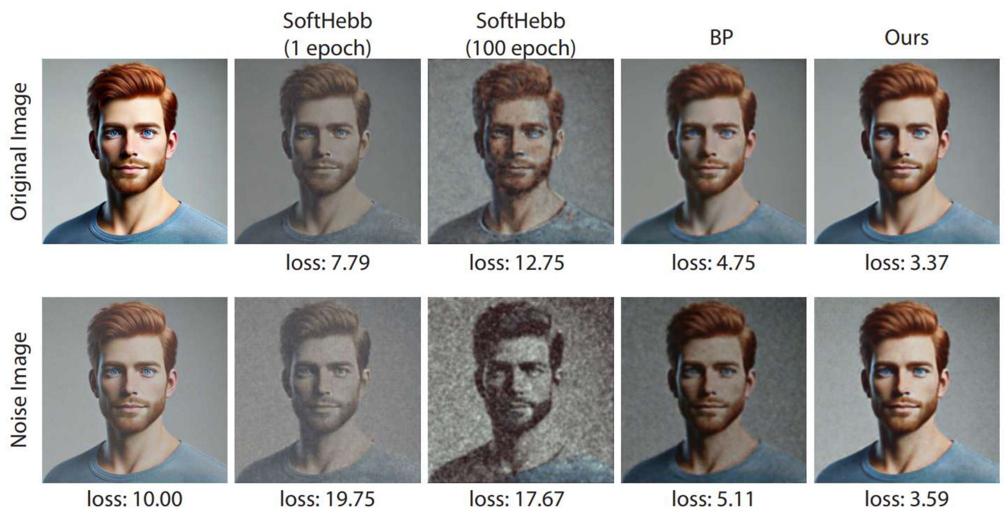


Figure 4: Auto-Encoder framework.



Results: Effective in Continual Learning

Resists "Catastrophic Forgetting“:

Our features are general and robust to distribution shifts.

Approaches	Split-CIFAR100	Split-TinyImageNet
EWC [32]	71.96± 0.37	62.87± 0.31
HAT [33]	75.70± 0.50	54.97± 0.84
OWM [34]	70.89± 0.13	–
GPM [35]	74.99 ± 0.12	66.00 ± 0.24
SoftHebb [12]*	51.1	–
SPHeRe	72.72± 1.14	63.46± 0.79
SPHeRe-EWC	**76.53**± **0.64**	**67.05**± **1.16**

Future works

Some current flaws and possible future direction:

Current SPHeRe and all other Hebb-like methods:

1. lack scalability to deeper networks.
2. require high channel width to replace depth