Learning from the Unexpected

Working summary · 6 August 2026

Somato-Dendritic Innovation Learning (SDIL)

> **The idea.** A neuron's teaching signal contains both useful feedback and ordinary activity-related bias. Predict the ordinary part from that neuron's own activity, subtract it, and learn only from what remains.

The whole method

For layer l, fit a local predictor during neutral observations and form the innovation:

$$\hat{a}_l = P_l(h_l), \quad r_l = a_l - \hat{a}_l.$$

Then update each forward weight with information available at that neuron:

$$\Delta W_l = \eta \, [r_l \odot \varphi'(u_l)] \, h_{l-1}^{T}.$$

There is no reverse-mode differentiation, backward computation graph, or transport of transposed forward weights. The feedback direction can come from causal perturbations or a reciprocal local-feedback network. The new operation is the per-neuron subtraction, not the feedback backbone.

Current flagship result: standard CIFAR-10 ResNets

Five paired seeds were run at every depth. SDIL improves as the network gets deeper and remains close to exact backpropagation (BP).

Method	ResNet-20	ResNet-32	ResNet-56	ResNet-56 cost / BP
SDIL	91.584	92.254	**92.760**	1.331×
BP	—	—	92.632	1.000×
Clean reciprocal feedback	—	—	92.670	—
DFA	—	—	30.850	—

Numbers are mean test accuracy in percent. Every SDIL seed improves from ResNet-20 to ResNet-56; the mean gain is 1.176 points. Its early-layer teaching direction has 0.999423 cosine agreement with the exact descent direction.

The result that isolates the subtraction

With strong activity-predictable interference, raw feedback and a magnitude-matched raw control reach 10.38% and 10.31% accuracy. SDIL reaches **97.35%**. Across five seeds, its paired gain over the magnitude-matched control is **87.038 ± 0.607 points**. Matching signal size does not explain the result; subtraction changes the direction used for learning.

Accuracy versus cost

In the matched author-code VGG16 comparison, SDIL reaches **90.70%** validation accuracy in **1.24 h**. Dual Prop reaches 92.38% in 5.97 h. Clean reciprocal feedback reaches 90.86% in 1.08 h. SDIL is much cheaper than Dual Prop at similar accuracy, but it does not beat every method on clean data.

What would make the paper stronger

- Show the same failure and recovery inside strong contrastive learners; the frozen Dual Prop bias screen is running now.
- Finish the matched CNN, ResNet, and Transformer comparison; report accuracy, runtime, memory, and update cost together.
- Test bias that drifts or is only partly predictable. Random noise is a negative control because this subtraction cannot remove it.

Inspired by neuron-specific somato-dendritic residuals reported by Francioni, Zhang, Ahrens and Harnett, *Nature* (2026). All numbers above come from frozen, retained runs in this repository; failed branches are not removed.