

VisionHOPE Rewrites Its Own Learning Rule for Every Image

Kabui, Charles

2026-10-24

[Read at ToKnow.ai](#)



Most image models learn once, in training, and then apply the same fixed rules to every photo. VisionHOPE changes one rule. It is a vision backbone, the part of an image model that turns pixels into features, and inside each block five small memories track what the model has seen and set how big each learning step is and how much of the past to keep. Because the memories also drive their own updates, the learning step shifts as the model scans an image. It ports

[Nested Learning](#), a way of stacking learning loops inside one network, from language to images. At 27 million parameters it names the right category for 84.1% of ImageNet photos on its first guess, best or joint best at all three sizes tested.

On ADE20K, where a model labels every pixel of a scene as wall, floor or chair, the smallest version averages 49.4% overlap with the correct labels, against 48.0% for the best model of its size. On COCO, which asks a model to box each object in a photo, it scores 47.9% average precision. Its cost grows in step with the number of image tiles, where a transformer's grows with the square of it, so a 1280 pixel input stays affordable.

The closest rival that also adapts while it looks at an image scores 84.0% on ImageNet, so the lead there is 0.1%. The authors have not yet tried it as the vision part of a model that reads both pictures and text.

Read More: [DINOv3: Label-Free Vision Pretraining on 1.7 Billion Images](#), which changes what goes into a vision backbone rather than how it updates.

Sources:

- [VisionHOPE paper \(arXiv abstract\)](#)
- [VisionHOPE full paper with result tables](#)
- [VisionHOPE code and checkpoints](#)
- [Nested Learning: The Illusion of Deep Learning Architectures](#)

Disclaimer: For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. **Read more:** [/terms-of-service](#)