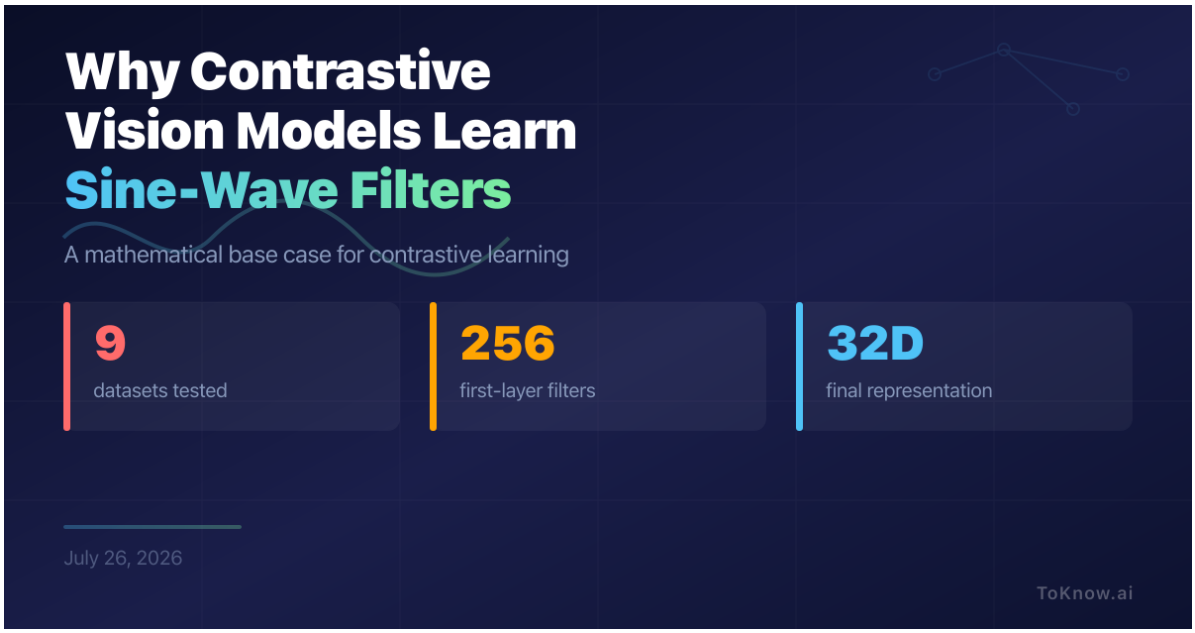
Why Contrastive Vision Models Learn Sine-Wave Filters

Kabui, Charles

2026-07-26

Read at ToKnow.ai



Antonio Torralba and Yair Weiss have worked out why simple vision models trained with contrastive learning, which teaches a model to keep altered views of the same image close and different images apart, often discover sine-wave filters. For image collections with stationary statistics, meaning the same broad patterns recur across locations, they derive an optimal one-layer convolutional neural network. Its 256 filters measure selected spatial frequencies, a

nonlinearity transforms them, global averaging removes position, and a 32-dimensional linear layer partially whitens the result so common patterns count less. A water-filling calculation chooses which frequencies matter from the dataset's average power spectrum. Tests covered six synthetic collections plus CIFAR-10, CIFAR-100, and ImageNet. With simple crops and several other augmentations, gradient descent learned the predicted filters. More modern augmentation recipes were the notable exception, producing more localized filters.

This gives a mechanism for an old puzzle. SimCLR showed that basic crops, colour changes, and blur can build useful visual representations, while earlier work found that training on structured noise can transfer to real-image recognition. Frequency statistics may explain part of both results. The theory does not explain today's full systems, though. Most experiments used grayscale images, a shallow network, and a custom loss that matches the standard InfoNCE loss only under a Gaussian assumption. The authors also report that these models are far from state of the art. This is an explanation of a clean base case, not a complete theory of deep networks or vision transformers.

Sources:

- A Theory of Contrastive Learning with Natural Images (arXiv)
- A Theory of Contrastive Learning with Natural Images (ICML 2026)
- SimCLR: A Simple Framework for Contrastive Learning
- Learning to See by Looking at Noise
