

Five Periods of Representation in Artificial Intelligence

Murray Shanahan

DeepMind and Imperial College London

Trinity College Dublin Workshop on Representation

May 2022

The Five Periods

- The symbolic period
- The anti-representational period
- The deep learning period
- The data-centric period
- The neurosymbolic period

← Maybe there should be a probabilistic period here, starting around 2005. But I was busy, so I missed it

The Symbolic Period (1956 – 1991)

- Inaugurated by John McCarthy “Programs with Common Sense” (1956)
- Logic and model theory
- Objects, relations, propositions
- Interpretations, models, truth
- Valid inference
- Intelligent behaviours as the product of correct reasoning with correct representation

The Anti-Representational Period (1991 – 2012)

- Inaugurated by Rod Brooks “Intelligence Without Reason” (1991)
- Embodiment first
- The primacy of the sensorimotor loop
- Recapitulate evolution
- Inman Harvey (some time in the 1990s)
 - Audience member: “But I am consciously aware of representations that are before my mind”
 - Inman: “There is nothing I can do to help you”

The Deep Learning Period (2012 – 2020)

- Inaugurated by Krizhevsky, Sutskevar & Hinton “ImageNet Classification with Deep Convolutional Neural Networks” (2012)
- Deep learning
- Distributed representation
- Continuous rather than discrete structures
- Differentiability
- Representations have to be reverse engineered
- ICLR (International Conference on Learning Representations)

Representations in Deep Learning

Representation learning is a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification. Deep-learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level.

Le Cun, Bengio & Hinton, *Nature* (2015)

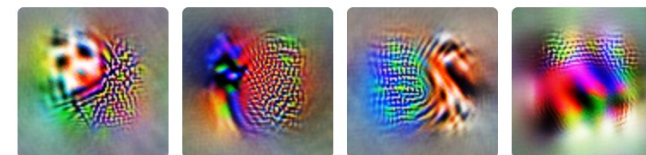
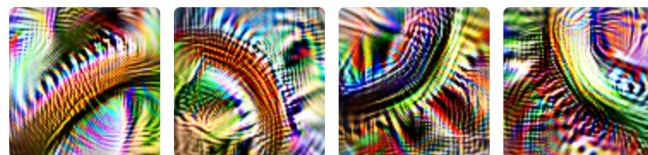
Representations Are Feature Vectors

Curve detectors

High-Low Frequency detectors

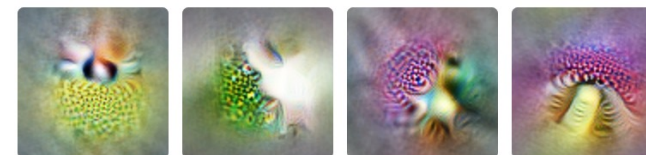
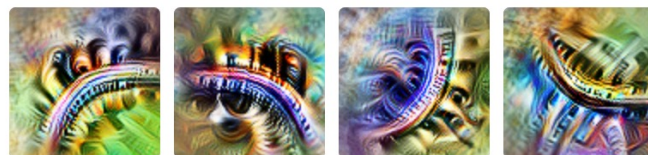
ALEXNET

Krizhevsky et al. [34]



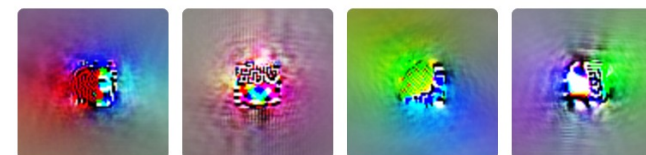
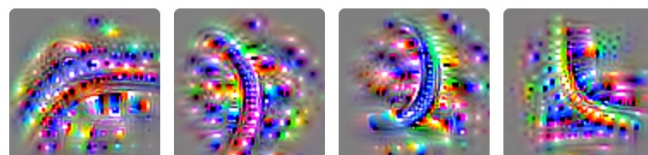
INCEPTIONV1

Szegedy et al. [26]



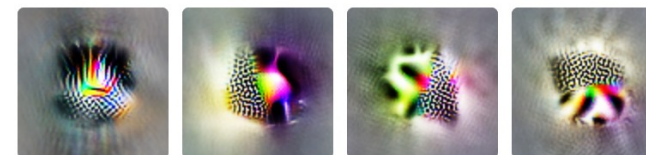
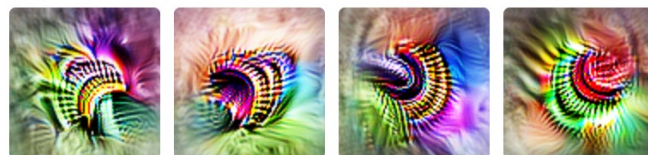
VGG19

Simonyan et al. [35]



RESNETV2-50

He et al. [36]



C.Olah, Zoom In: An Introduction to Circuits, *Distil*

The Data-Centric Period (2020 – present)

- Inaugurated by Brown et al. “Language Models are Few-Shot Learners” (2020) (GPT-3)
- Enormous numbers of parameters
- Enormous quantities of data
- Enormous amounts of compute
- Transformer architectures
- And that’s all you need

The NeuroSymbolic Period (the future)

- Inaugurated by who knows who
- Combines advantages of deep learning with symbolic-like representation
- Reconciling discrete and continuous structures
- Compositionality
- Abstraction
- And then ... AGI