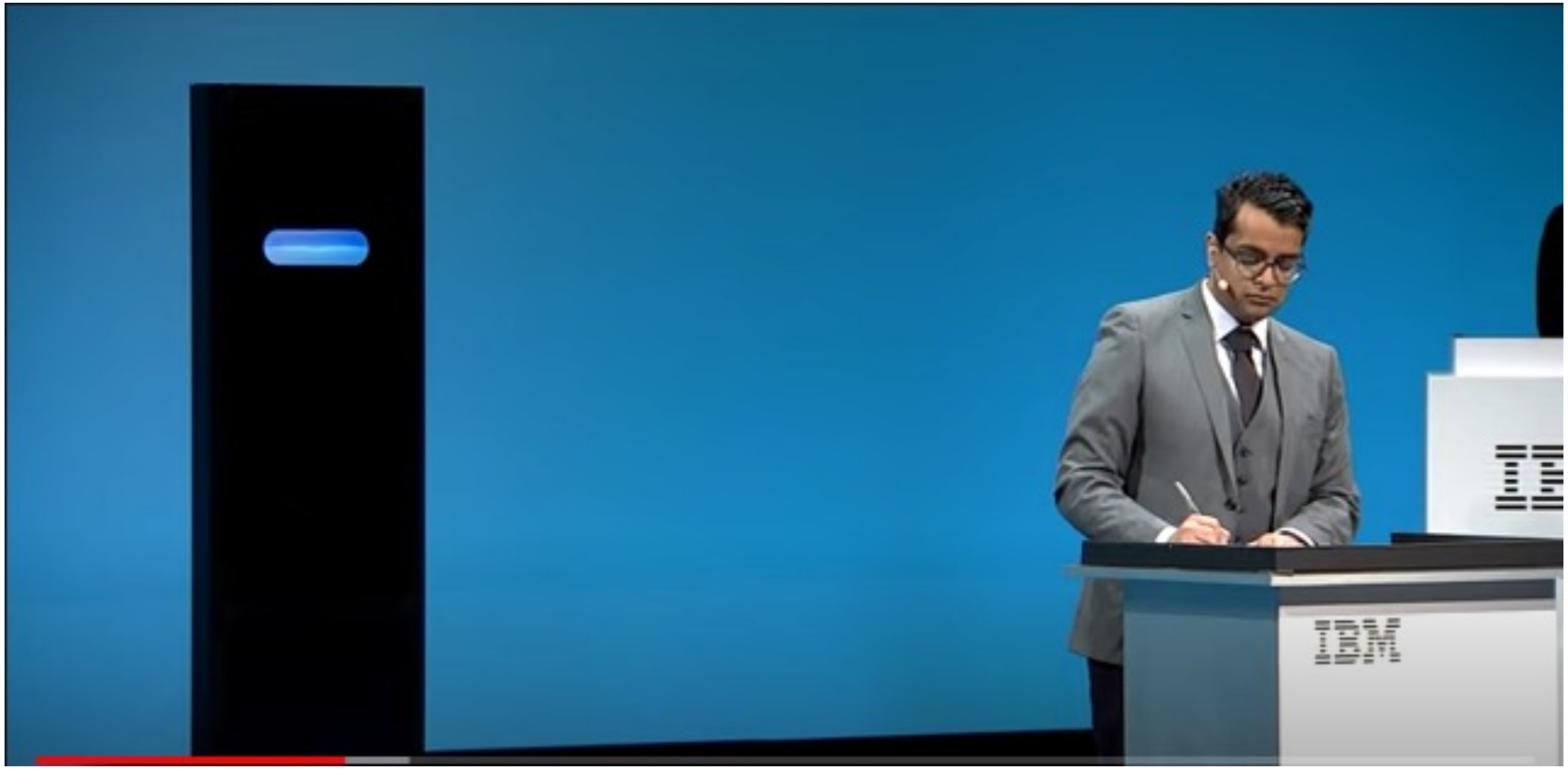
Human Learning vs. ANN Learning

ZHAW CENTRE FOR ARTIFICIAL INTELLIGENCE

NOVEMBER 17, 2021

Christoph von der Malsburg
Frankfurt Institute for Advanced Studies
malsburg@fias.uni-frankfurt.de

IBM Debator



https://youtu.be/3_yy0dnIc58?t=508

I ANN Learning

Encoder-Decoder Architectures

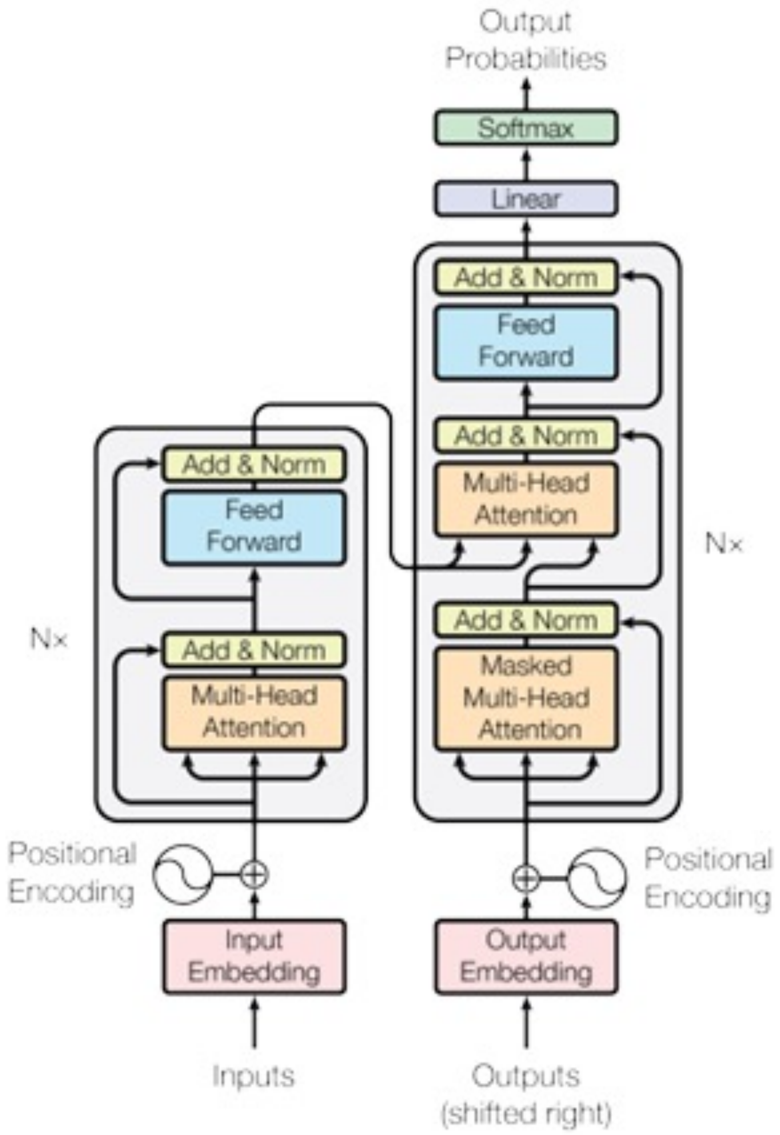


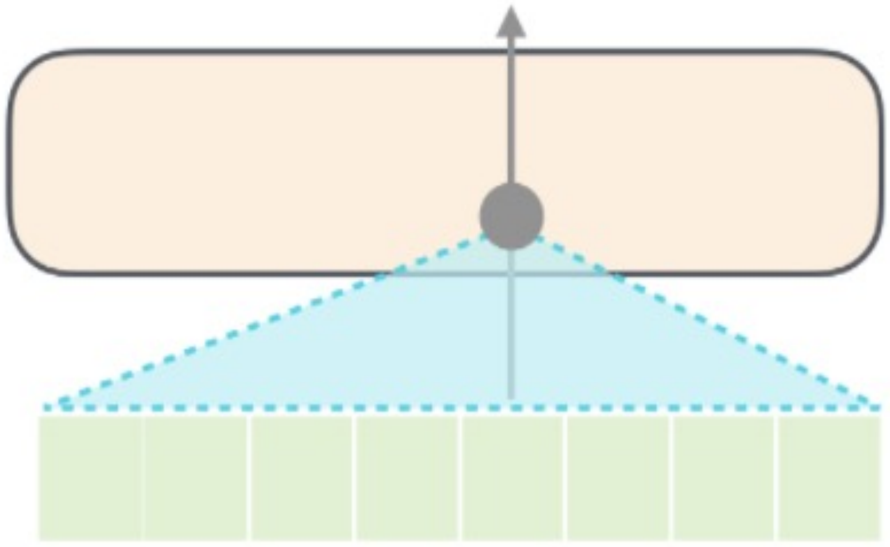
Figure 1: The Transformer - model architecture.

Attention Is All You Need. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin. https://arxiv.org/abs/1706.03762

FIAS Frankfurt Institute for Advanced Studies

Token Embedding

Self-Attention



Observations

- Chomski proved wrong
- Searl proved right
- Overcoming supervised learning
- Overcoming the single-neuron dogma
- Beyond the benchmark stage
- Neural circuits can represent language

FIAS Frankfurt Institute for Advanced Studies

Observations

- Generation of intelligence akin to evolution
- R. Sutton's "bitter lesson"

  concentrate on meta methods

  structure (objects, space, transformation properties)
  should develop on its own, not be put in manually

Changed Landscape

- Human-brain - scale computing within reach
- Engineering over Science
- Industry over Academia

AI: The Future?

Scale it up to get to human-level intelligence!

or

does it take further scientific breakthroughs?

Problem Fronts

- Behavior
- Semantics
- Generalization

II Human Learning

The Nursery

Children learn from a simple environment

and generalize to others

After having seen a cat, a dog and a horse,

the child needs only one image to know the okapi

Quantities of Information

Virtual Reality simulation: < 10 Gigabytes

That's all the information the child absorbs

Quantities of Information

But

the child sees the output of VR, not the program

1 Mbit/sensation, 5 sensations/sec, 8h/day, 3 years =

160 Tbit

GPT-3: 20 bit/token, 300 billion tokens =

6 Tbit

Humans absorb 10^9 bits over their lifetime

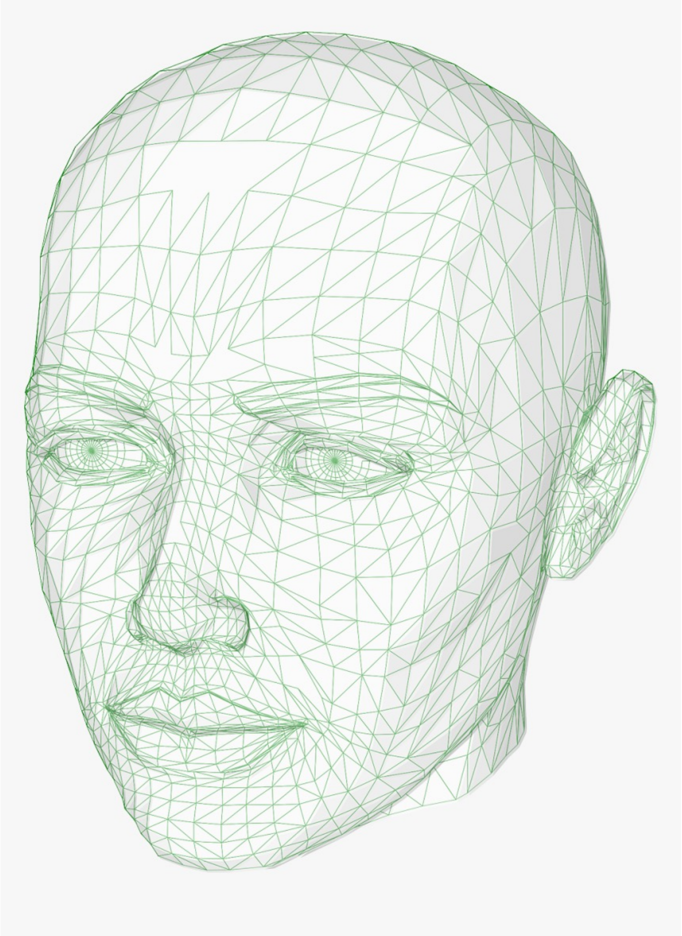
TK Landauer CogSci 10, 477-493 (1986)

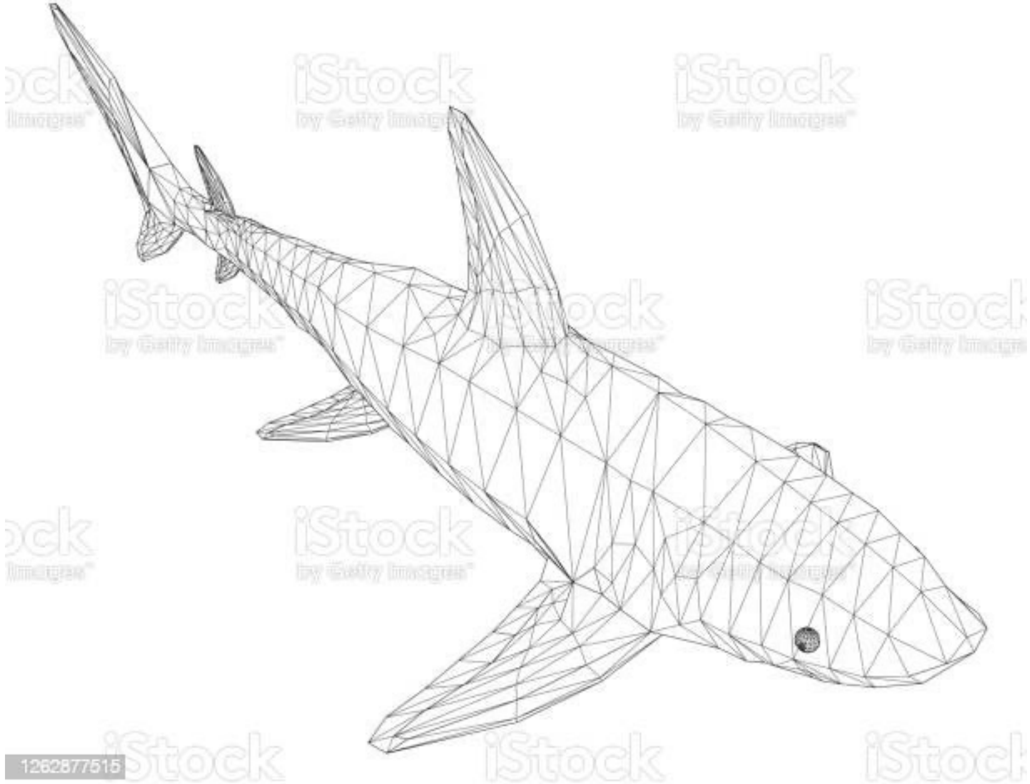
Information Gap: 160 Tbit vs 10 Gbyte

Kolmogorov algorithm:
shortest algorithm to create the structure

Computer Graphics

Computer Graphics





Computer Graphics

Computer Graphics

http://www.oyonale.com/modeles.php?lang=en&page=40

Computer Graphics as ontological model of the environment

- Scenes are the product of aspects, of subsystems

  Shapes, geometry, physics, behavior, social patterns

- Generalization by compositional structure

Vision as Inverse Computer Graphics

http://www.oyonale.com/modeles.php?lang=en&page=40

Vision needs System Integration

- Scene gist recognition
- Action recognition
- Depth perception
- Illumination (shadow) modeling
- Motion extraction
- Figure-ground separation
- Invariant object recognition

Learning comes after Perception

- Explaining the scene in terms of components
- Learning the structure of components

III Ontogenesis

Information Gap

- One Gbyte of genetic information

  DNA: 3.3 billion nuclear basis @ 2 bits each

- One Pbyte do describe the brain's wiring

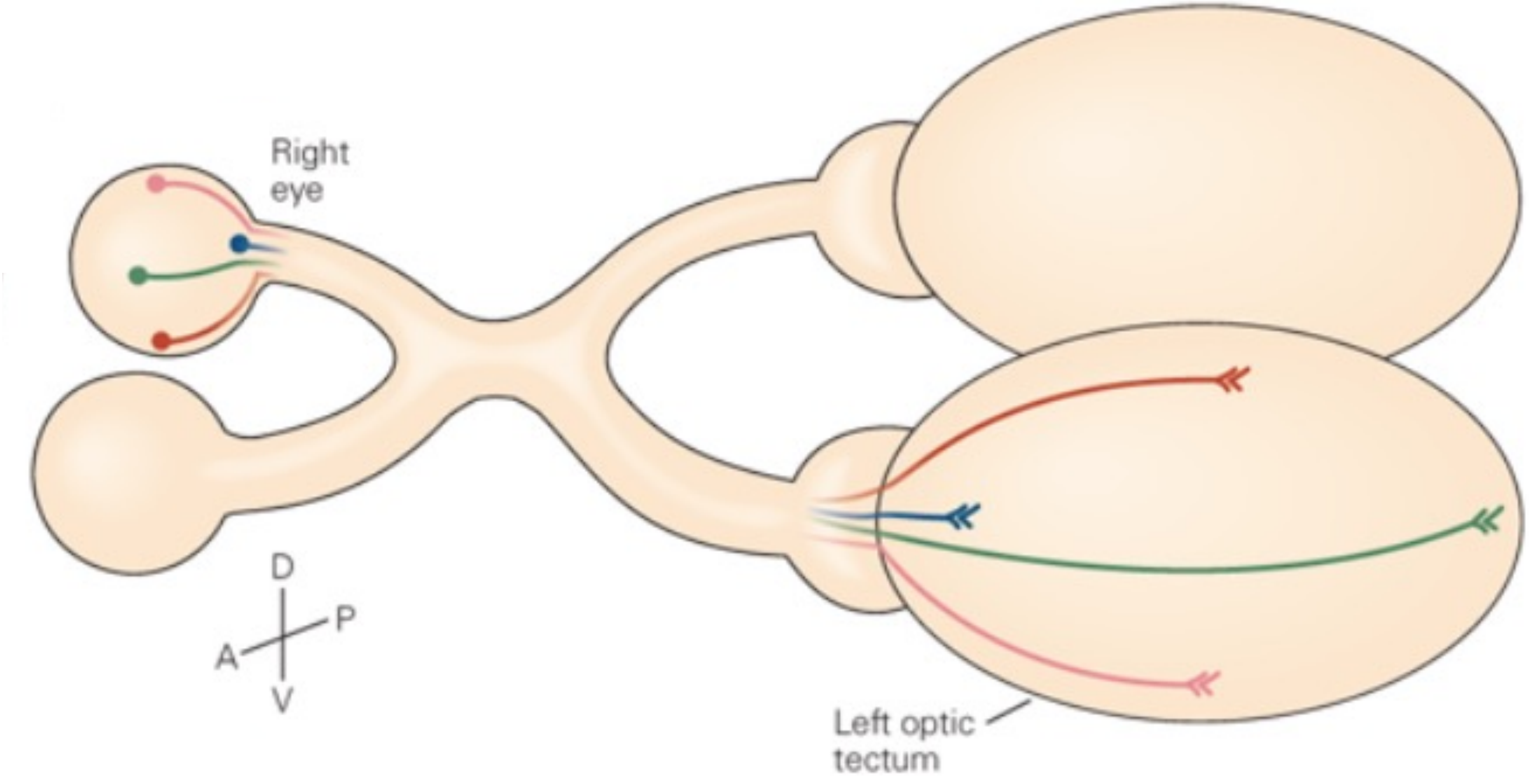
  10^{14} synapses, each taking 33 bits
  to address one of the 10^{10} neurons

Filling this Information Gap

- Volume needed as working space
- Filled by construction, self-organization
- Learning: mere Gbits

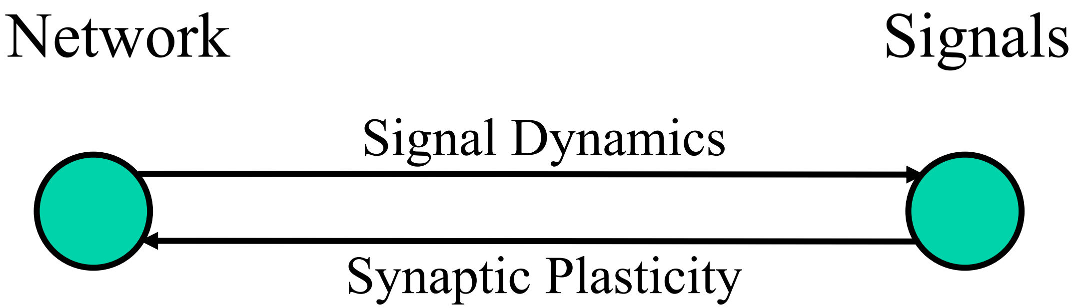
Retinotopy

Paradigm of Network Self-Organization



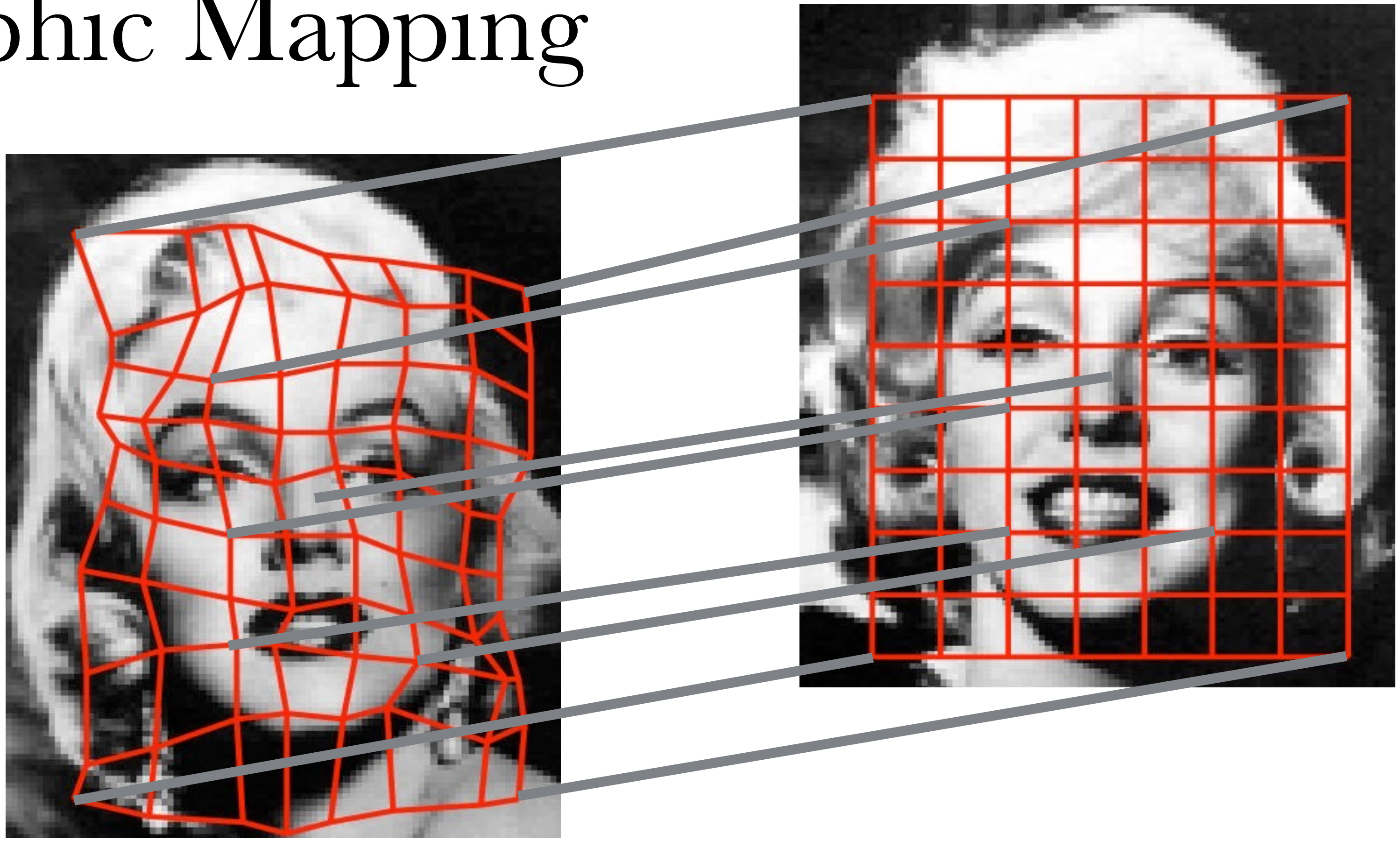
FIAS Frankfurt Institute for Advanced Studies

Network Self-Organization
as Kolmogorov Algorithm



Network Self-Organization

Homeomorphic Mapping

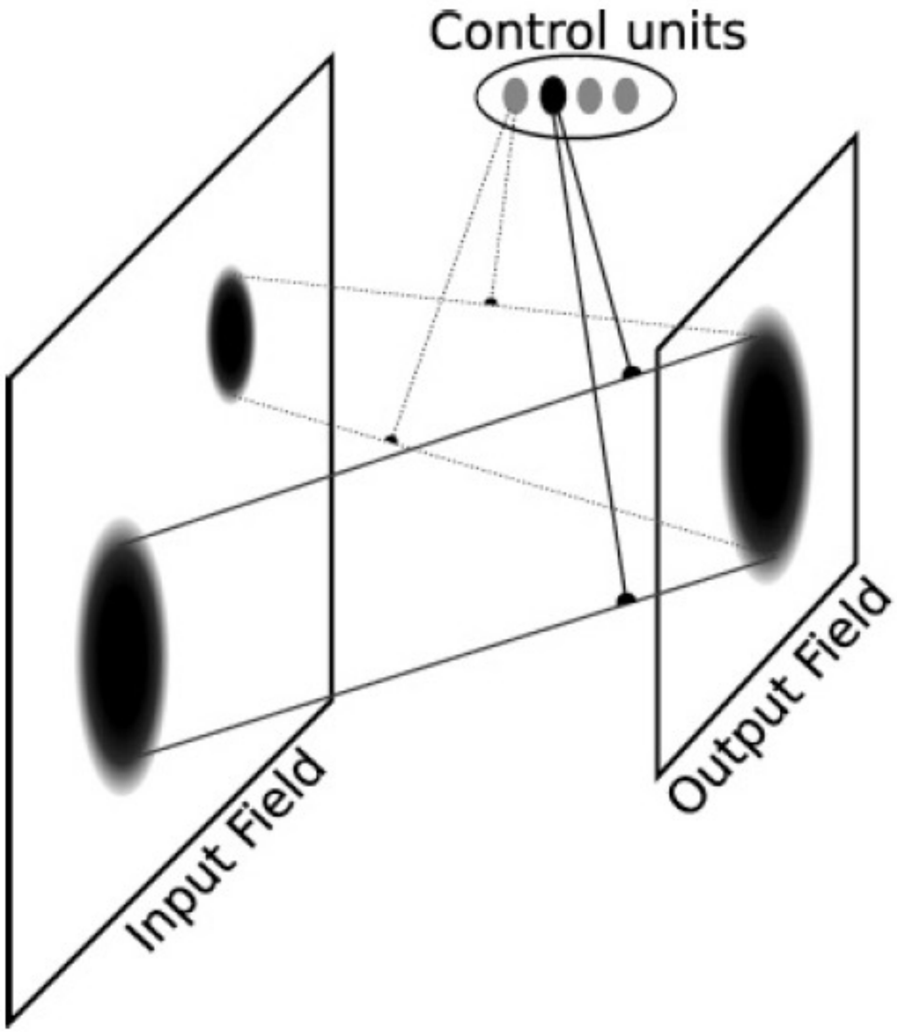


Two rules:

- Cooperation
- Sparsity

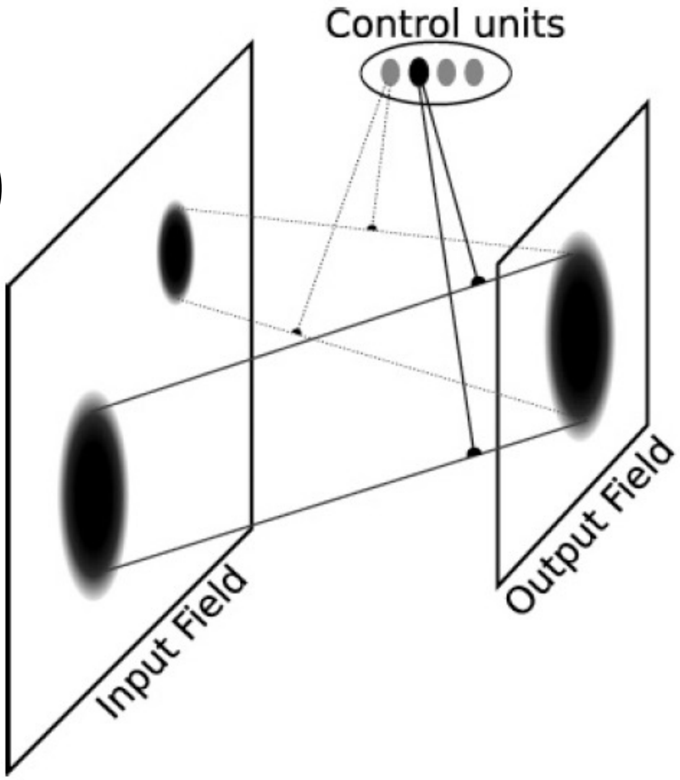
Shifter Circuits

C Anderson, D VanEssen, B Olshausen



Generalization by separation of aspects

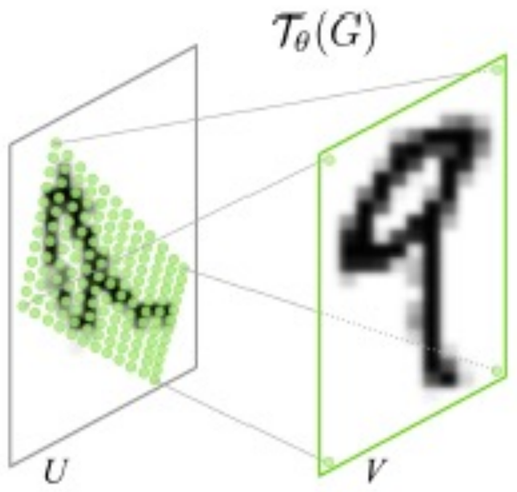
- Inner structure of the object (“what”)
- Projection from image (“where”)

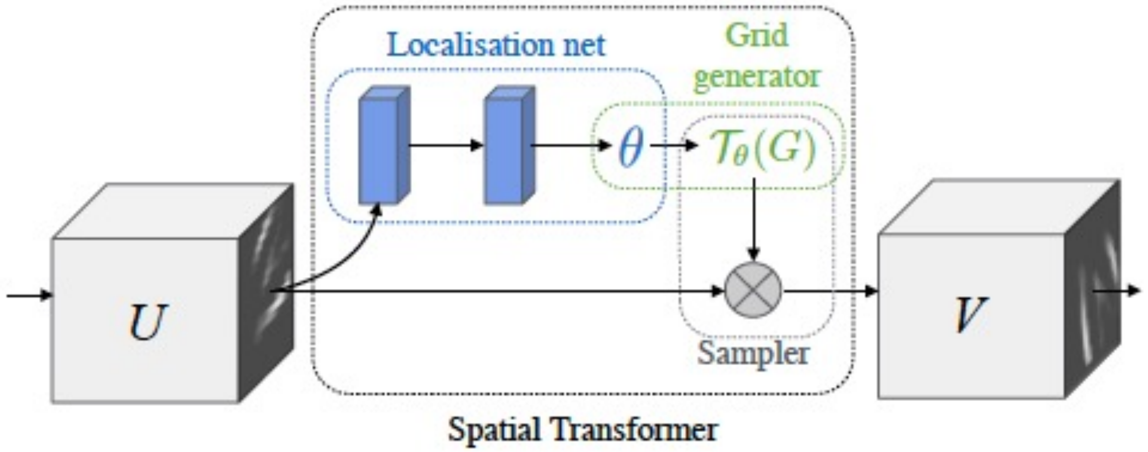


Spatial Transformer Networks

Max Jaderberg **Karen Simonyan** **Andrew Zisserman** **Koray Kavukcuoglu**

Google DeepMind, London, UK





$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$

IV Conclusions

Central Issue: Architecture!

Basis for the Integration of Subsystems

Data Structure

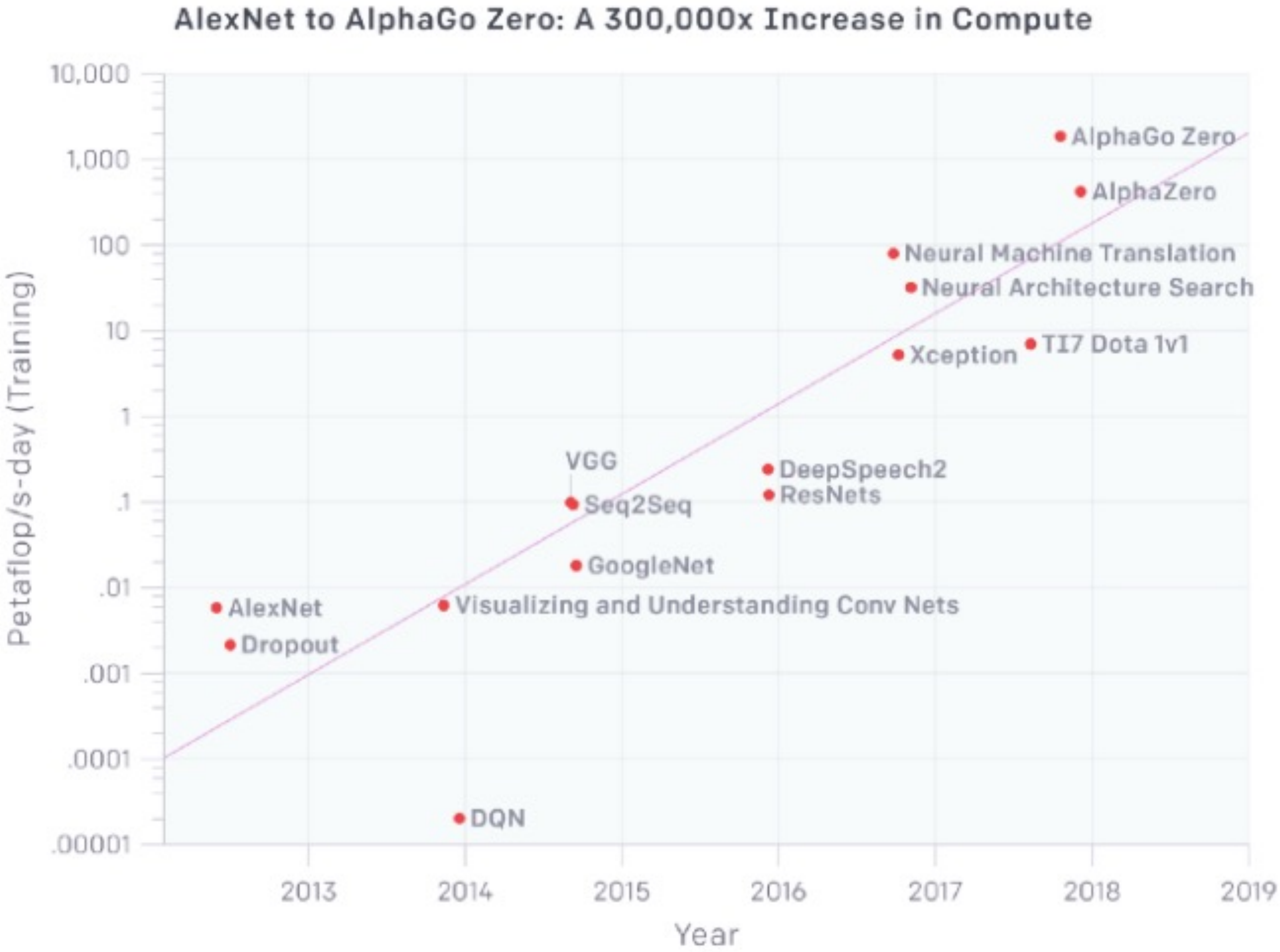
attractor nets as
representation of mental content

Mechanism of Organization

network self-organization

The Unreasonable Effectiveness of the Brain

Computing Power Doubles every 3.4 Months



FIAS Frankfu
for Advance

GPT-3: The Pinnacle of Machine Learning

OpenAI (Microsoft)

Model	Total train compute (PF-days)	Total train compute (flops)	Params (M)	Training tokens (billions)
GPT-3 Small	2.60E+00	2.25E+20	125	300
GPT-3 Medium	7.42E+00	6.41E+20	356	300
GPT-3 Large	1.58E+01	1.37E+21	760	300
GPT-3 XL	2.75E+01	2.38E+21	1,320	300
GPT-3 2.7B	5.52E+01	4.77E+21	2,650	300
GPT-3 6.7B	1.39E+02	1.20E+22	6,660	300
GPT-3 13B	2.68E+02	2.31E+22	12,850	300
GPT-3 175B	3.64E+03	3.14E+23	174,600	300

https://arxiv.org/pdf/2005.14165.pdf