image: flickr/kylemcdonald

Towards Visual Intelligence

Aykut Erdem

HACETTEPE UNIVERSITY COMPUTER VISION LAB

The International Symposium on Brain and Cognitive Science

May 6, 2018

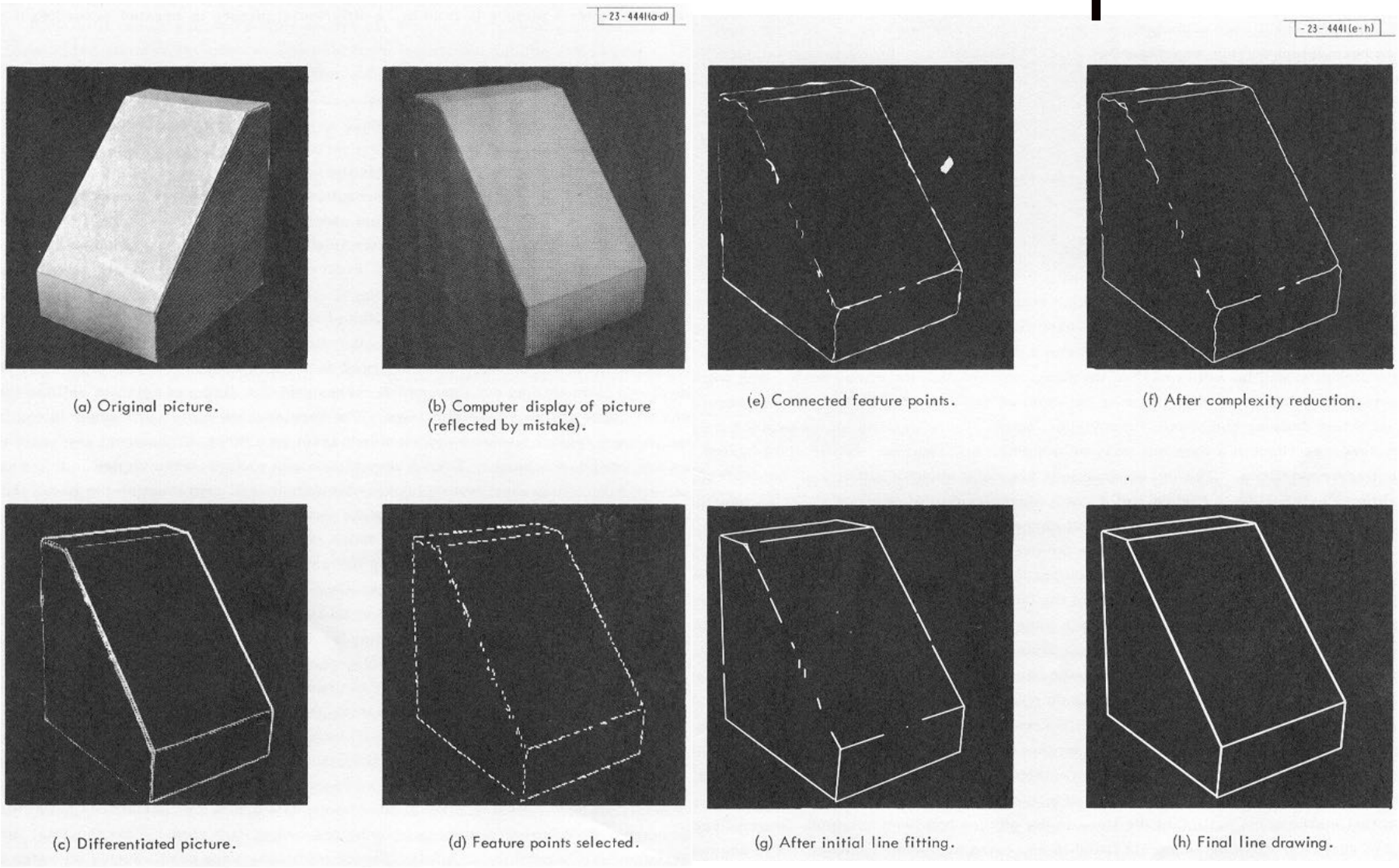
The Purpose of Vision

"What does it mean,to see? The plain man's answer (and Aristotle's too) would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is."

[Marr, 1982]

Image credit: The Sense of Sight (Annie Louisa Swynnerton, 1895)

The First PhD Thesis on Computer Vision



- Machine perception of three-dimensional solids [Roberts 1963]

The Summer Vision Project

General goals:

FIGURE-GROUND.
divide a vidisector picture into regions such as likely objects, likely background areas and chaos

REGION DESCRIPTION.
analysis of shape and surface properties

OBJECT IDENTIFICATION.
name objects by matching them with a vocabulary of known objects

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group — July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

[Papert 1966]

Why does vision appear easy to humans?

- Our brains are specialized to do vision.
- ~50% of the cortex in a human brain is devoted for visual processing

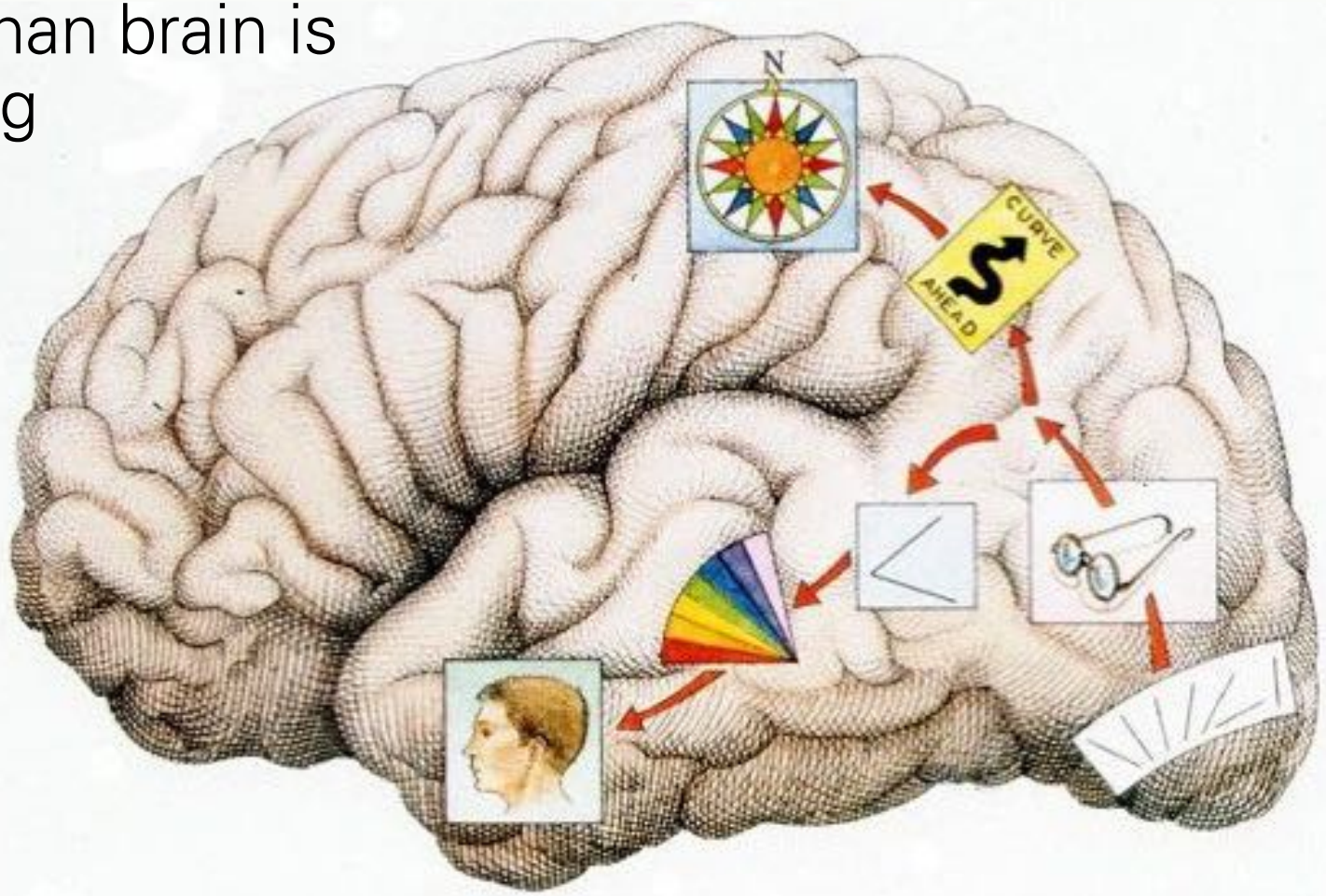
  (cf. motor control ~20-30%,
  language ~10-20%)

Visual perception*:
540,000,000 years of data

Bipedal movement:
230,000,000 years of data

Abstract thought:
100,000 years of data

*Color vision



Slide adapted from David Heeger

Fast Forward to 2012

IMAGENET Large Scale Visual Recognition Challenge (ILSVRC)

- **1.2M** training images, **1K** categories
- Measure top-5 classification error

The success of AlexNet, a deep convolutional network (CNN)

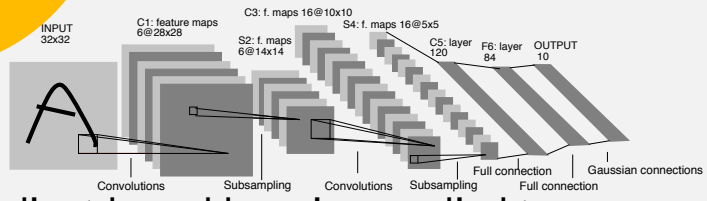
- 7 hidden layers (not counting some max pooling layers)
- 60M parameters

2012 Teams	%Error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

CNN based, non-CNN based

Cat

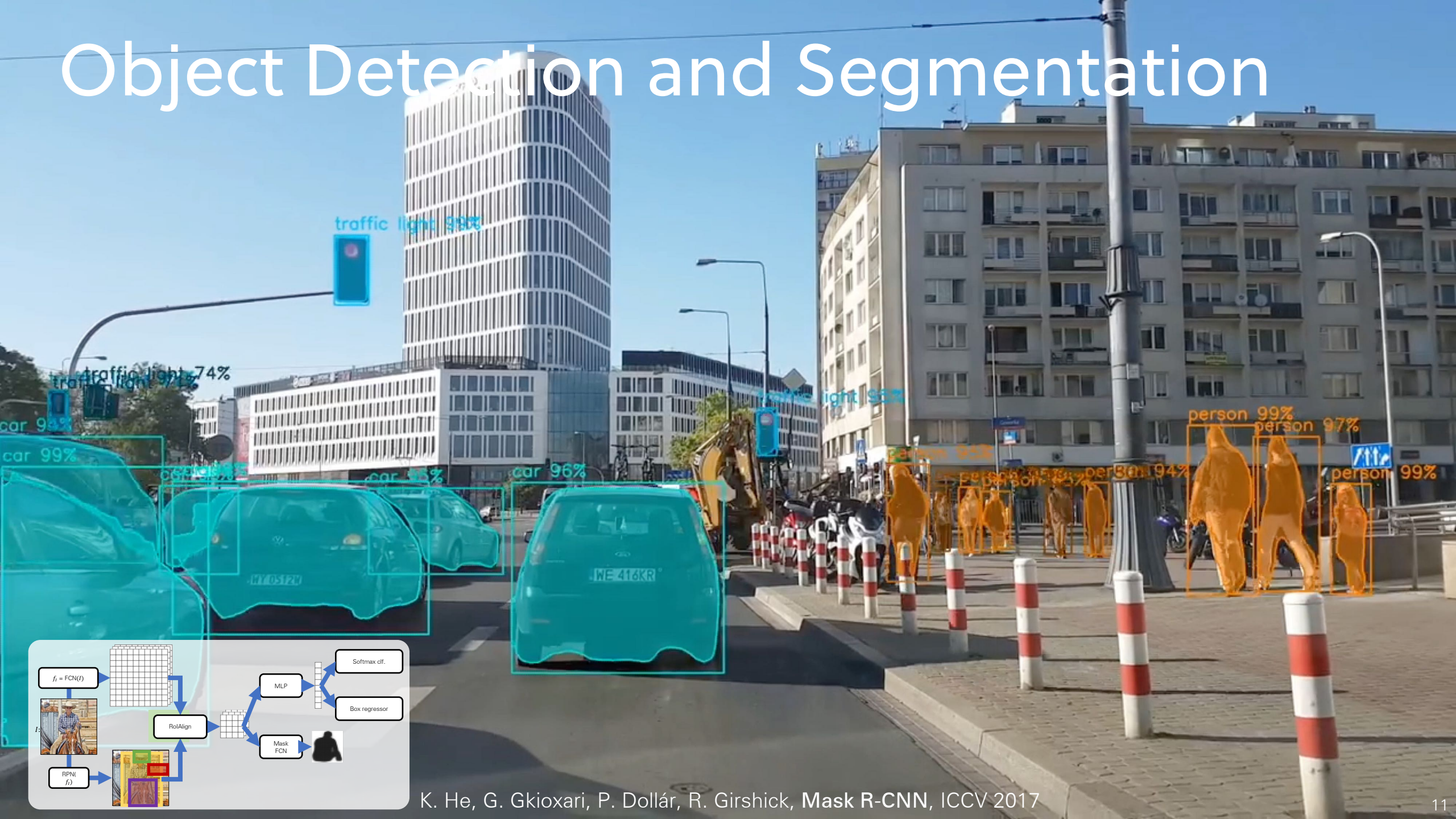
CNNs are biologically inspired by oriented cells in the visual cortex



Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. **Gradient-based learning applied to document recognition**. Proceedings of the IEEE. 86 (11): 2278–2324, 1998.

A. Krizhevsky, I. Sutskever, G.E. Hinton **ImageNet Classification with Deep Convolutional Neural Networks**. NIPS 2012

Object Detection and Segmentation



K. He, G. Gkioxari, P. Dollár, R. Girshick, **Mask R-CNN**, ICCV 2017

Pose Estimation

Z. Cao ,T. Simon, S.–E. Wei and Yaser Sheikhr. **Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields**. CVPR 2017

Photo Style Transfer

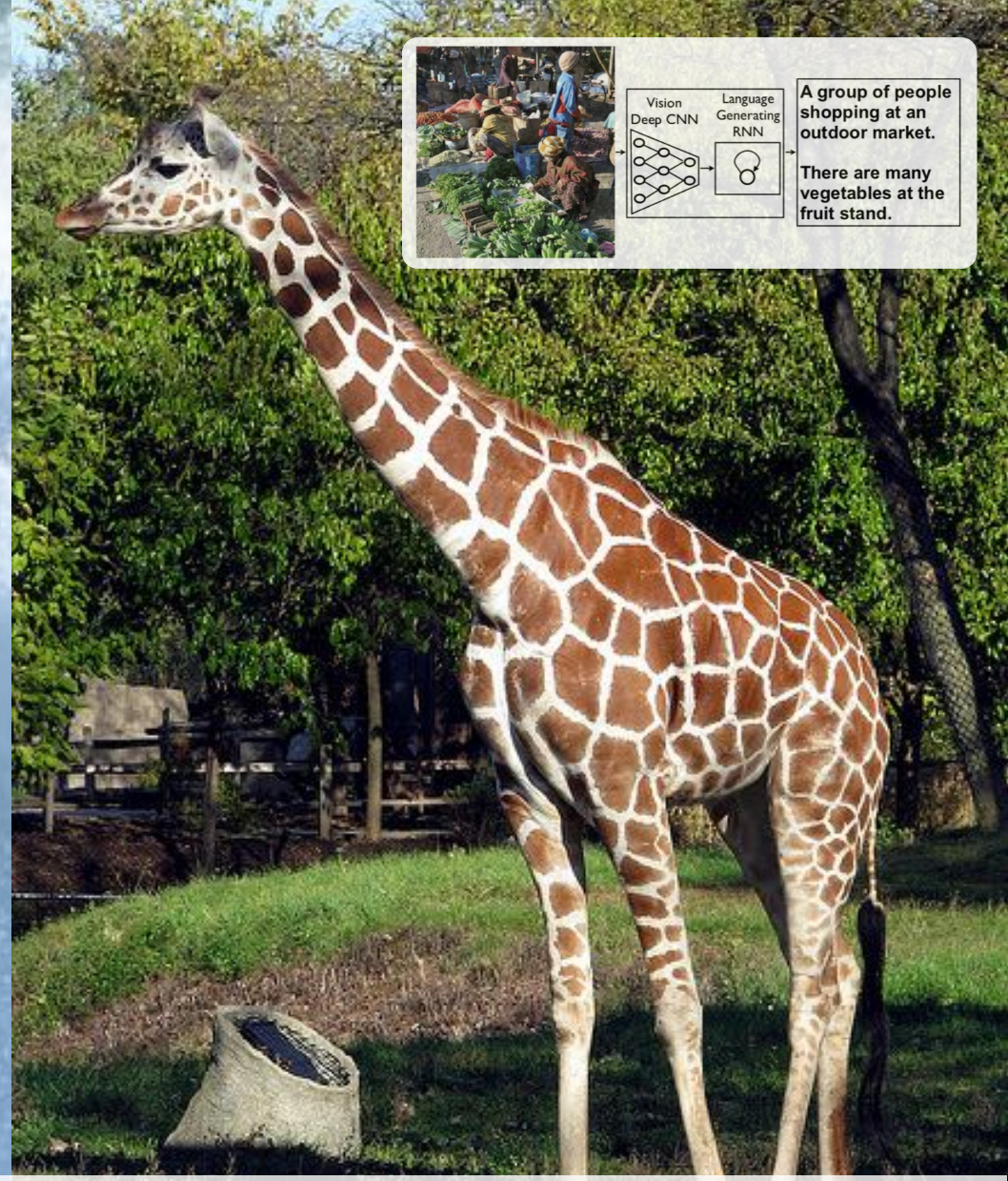
F. Luan, S. Paris, E. Shechtman & K. Bala. **Deep Photo Style Transfer**. CVPR 2017

Photo Style Transfer

F. Luan, S. Paris, E. Shechtman & K. Bala. **Deep Photo Style Transfer**. CVPR 2017

Image Captioning

A man riding a wave on a surfboard in the water.



A giraffe standing in the grass next to a tree.

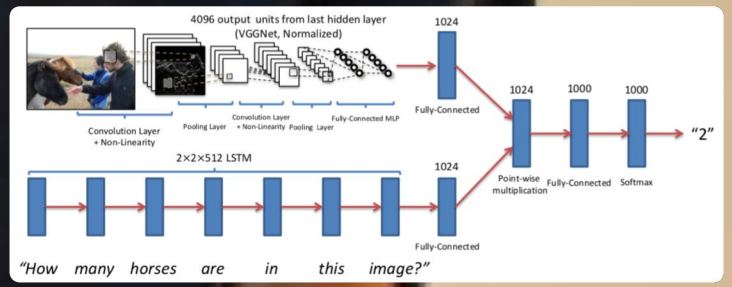
X. Chen and C. L. Zitnick. **Mind's Eye: A Recurrent Visual Representation for Image Caption Generation**. CVPR 2015.

Image Captioning

Yarış pistinde virajı almakta olan bir yarış arabası

M. Kuyu, A. Erdem & E. Erdem. Image Captioning in Turkish with Subword Units. SIU 2018

Visual Question Answering



Question: What is the girl reaching into?
Answer: apples

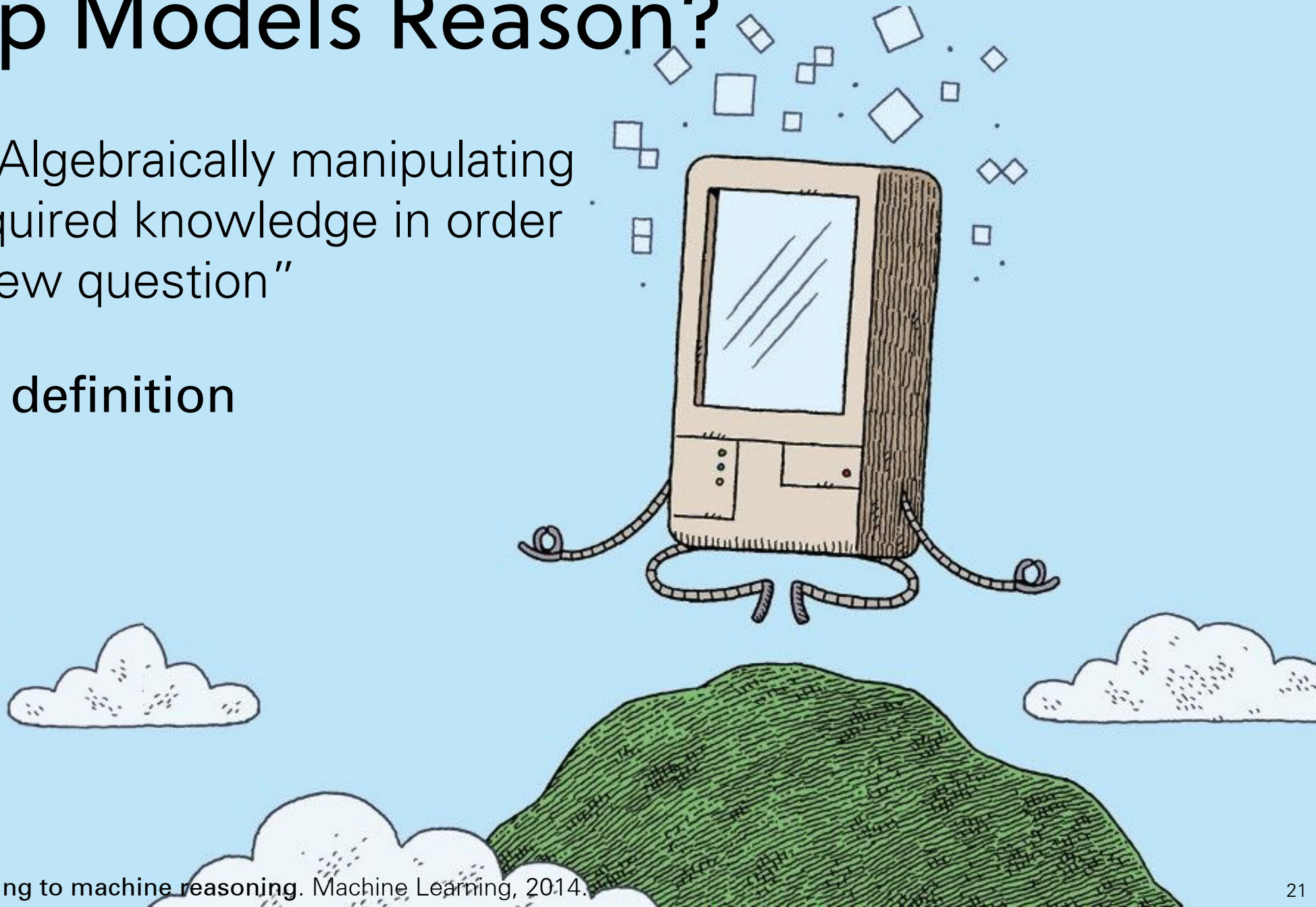
Y. Goyal, T. Khot, D. Summers-Stay, D. Batra, D. Parikh. **Making the V in VQA Matter: Elevating the Role of Image Understanding in Visual Question Answering.** CVPR 2017

cartoon: Tom Gauld

Can Deep Models Reason?

- **Reasoning:** "Algebraically manipulating previously acquired knowledge in order to answer a new question"

- **A very broad definition**



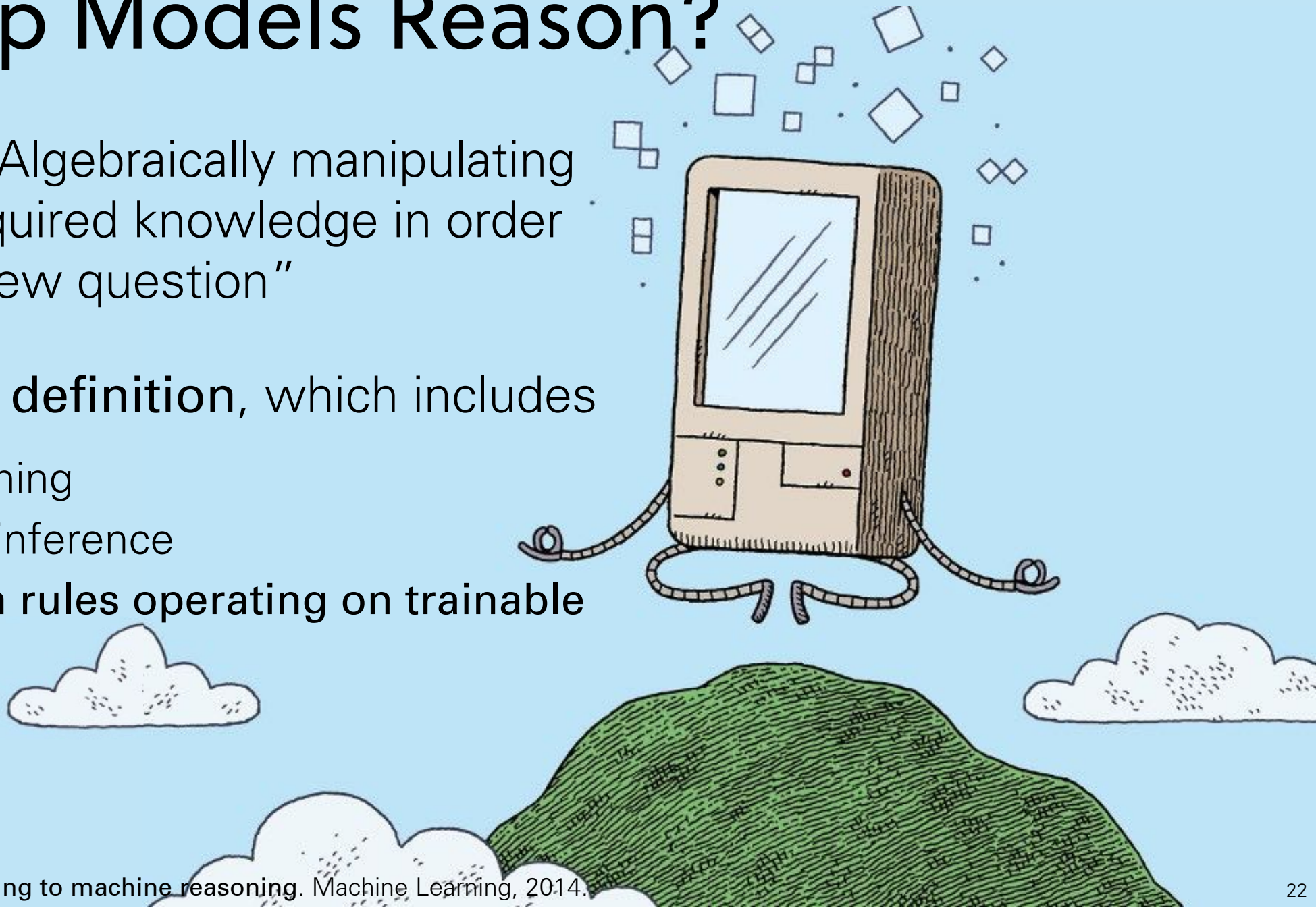
Léon Bottou. **From machine learning to machine reasoning**. Machine Learning, 2014.

cartoon: Tom Gauld

Can Deep Models Reason?

- **Reasoning:** "Algebraically manipulating previously acquired knowledge in order to answer a new question"

- **A very broad definition**, which includes
  - logical reasoning
  - probabilistic inference
  - **composition rules operating on trainable modules**

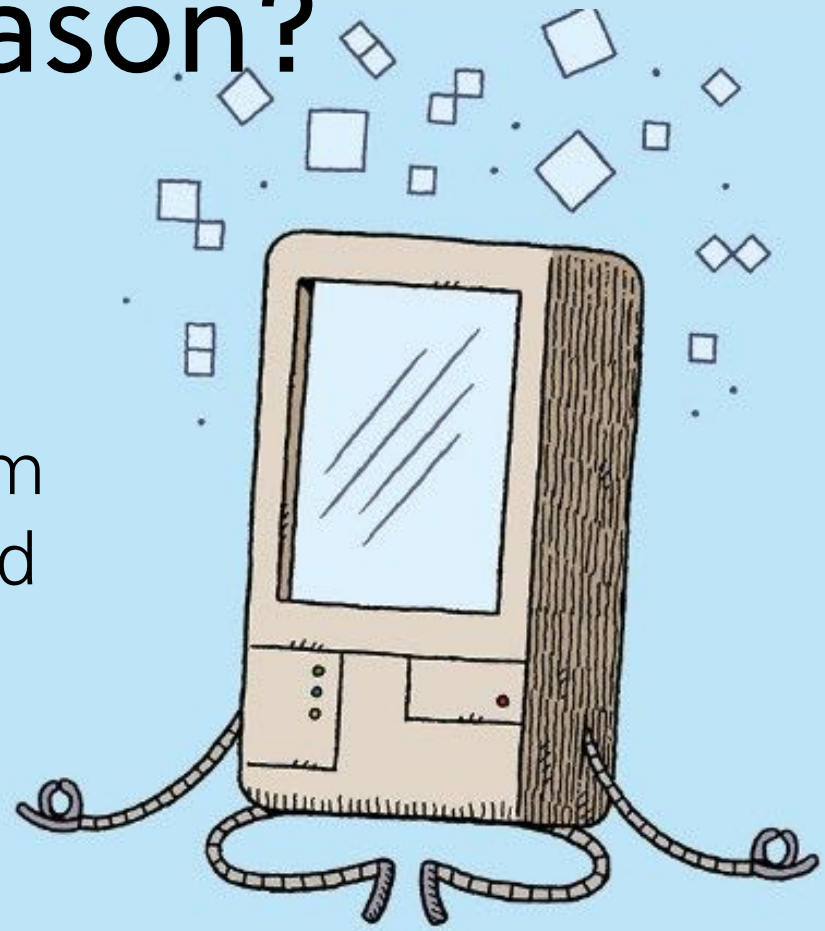


Léon Bottou. **From machine learning to machine reasoning**. Machine Learning, 2014.

cartoon: Tom Gauld

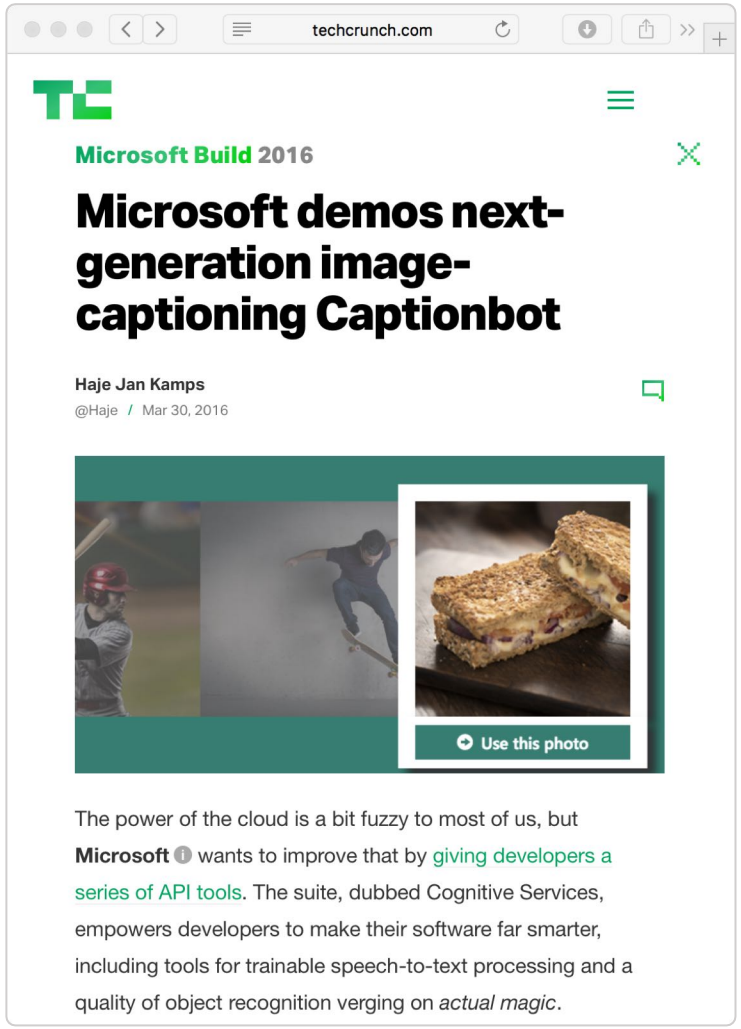
Can Deep Models Reason?

- Deep Learning models are **large correlation engines**
- They use **inductive bias** to learn from training data, which is a double-edged sword
  - Generalize well when **target** and **training** distributions are similar
  - Confuse **correlation** with **causation**



slide adapted from Christopher Manning

Take 2: Image Captioning

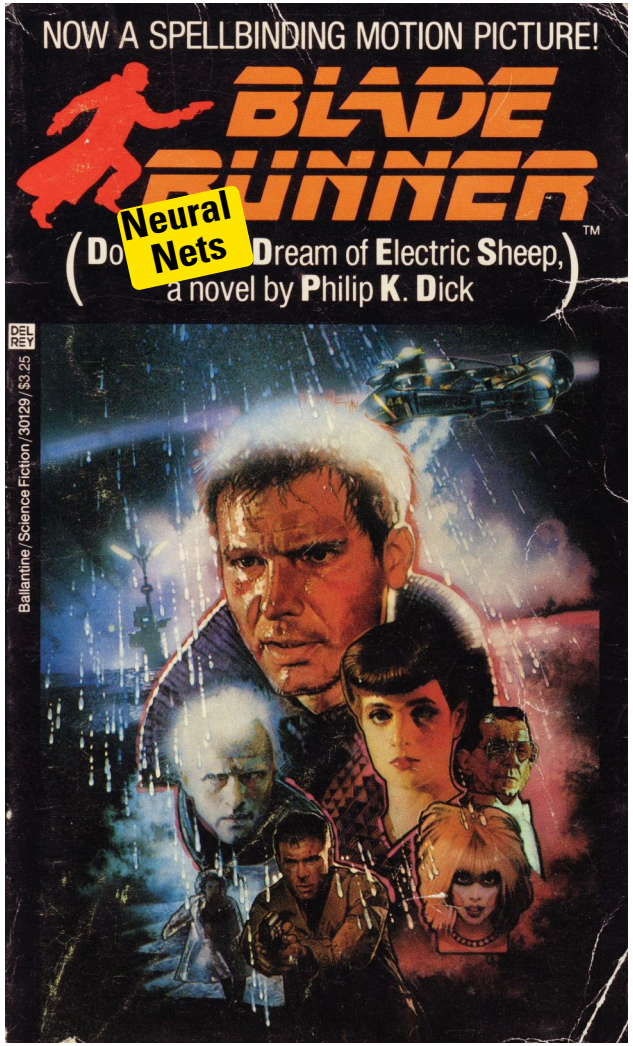


picdescbot @picdescbot · Feb 19

a herd of sheep grazing on a lush green field

6 3 18

Take 2: Image Captioning



picdescbot @picdescbot · Feb 19

a herd of sheep grazing on a lush green field

6 3 18

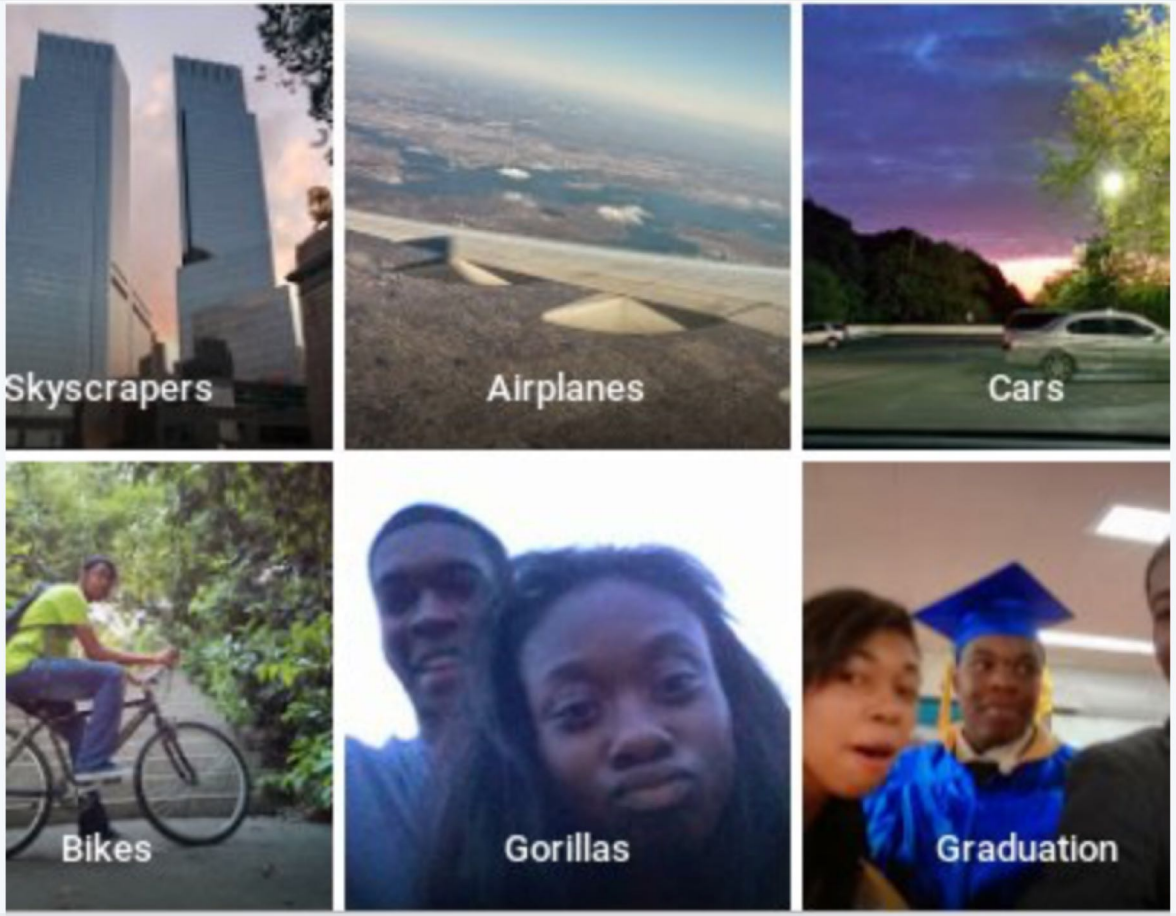
picdescbot @picdescbot · Mar 8

a yellow and orange flowers in a field

8 41 103

Suda yüzmekte olan bir köpek.

M. Kuyu, A. Erdem & E. Erdem. **Image Captioning in Turkish with Subword Units.** SIU 2018



jackyalciné is about 40% into the IndieWeb.
@jackyalcine

Google Photos, y'all fucked up. My friend's not a gorilla.

4:22 AM - Jun 29, 2015

2,280 3,592 people are talking about this

theguardian.com

Support The Guardian

The Guardian

News | Opinion | Sport | Culture | Lifestyle

World UK Science Cities Global development

Google says sorry for racist auto-tag in photo app

- **Google Photos labelled a picture of two black people as 'gorillas'**
- **Google Maps and Flickr have also suffered from race-related problems**

Jana Kasperkevic *in New York*

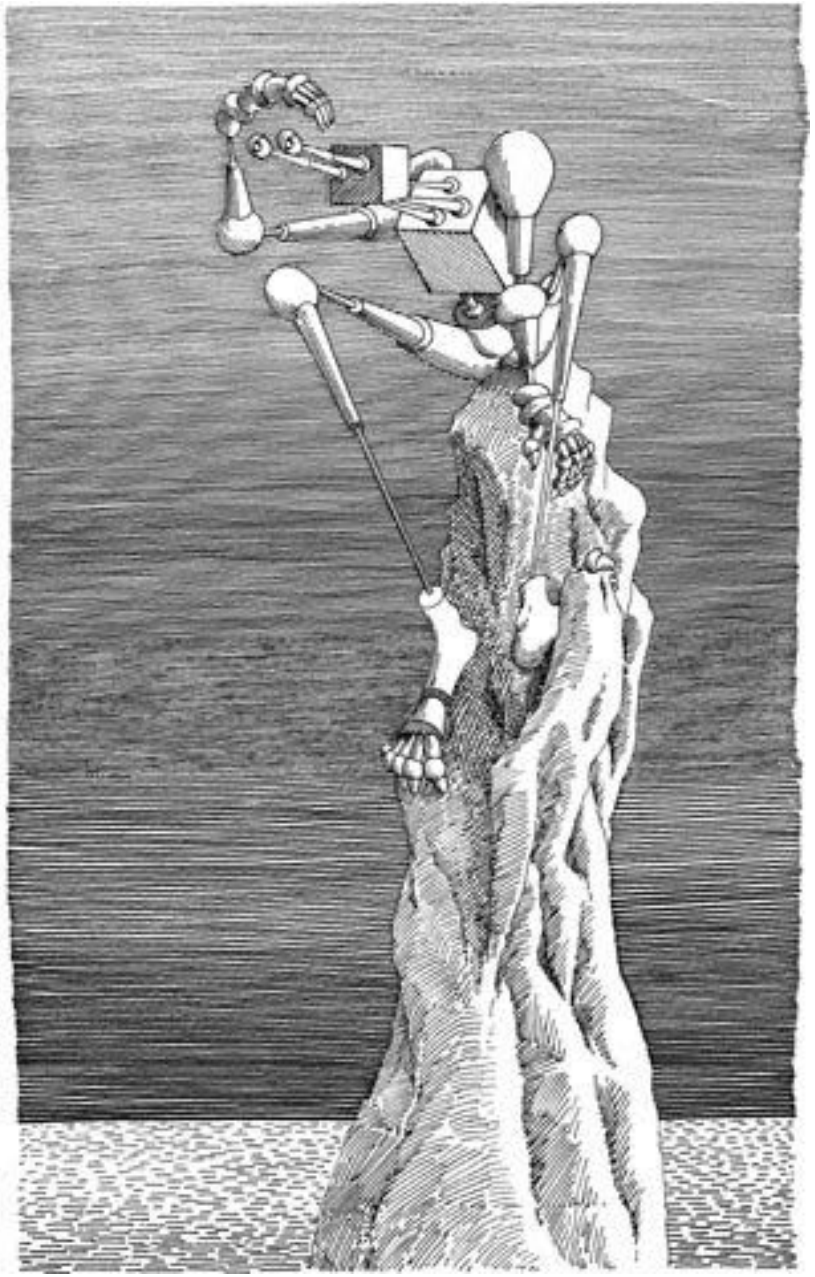
@kasperka Email

Wed 1 Jul 2015 18.52 BST

image: Daniel Mróz, Drawings for Stanislaw Lem's Cyberiad

Looking Forward

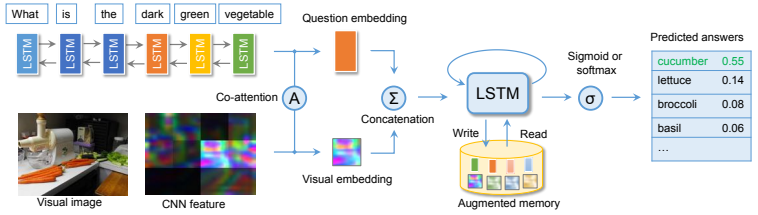
- Intelligence is not just about **Pattern Recognition**
- Learning is the process of **modeling the world**…
  - **explaining** and **understanding** what we see
  - **imagining** things we could see but haven't yet.
  - **problem solving** and **planning** actions to make things real.
  - **building new models** as we learn more about the world.
  - **sharing our models**, communicating to others, understanding their models, and learning from them and with them.



Lake, Ullman, Tenenbaum & Gershman. **Building machines that learn and think like people**. Behavioral and Brain Sciences, 2017

slide adapted from Josh Tenenbaum

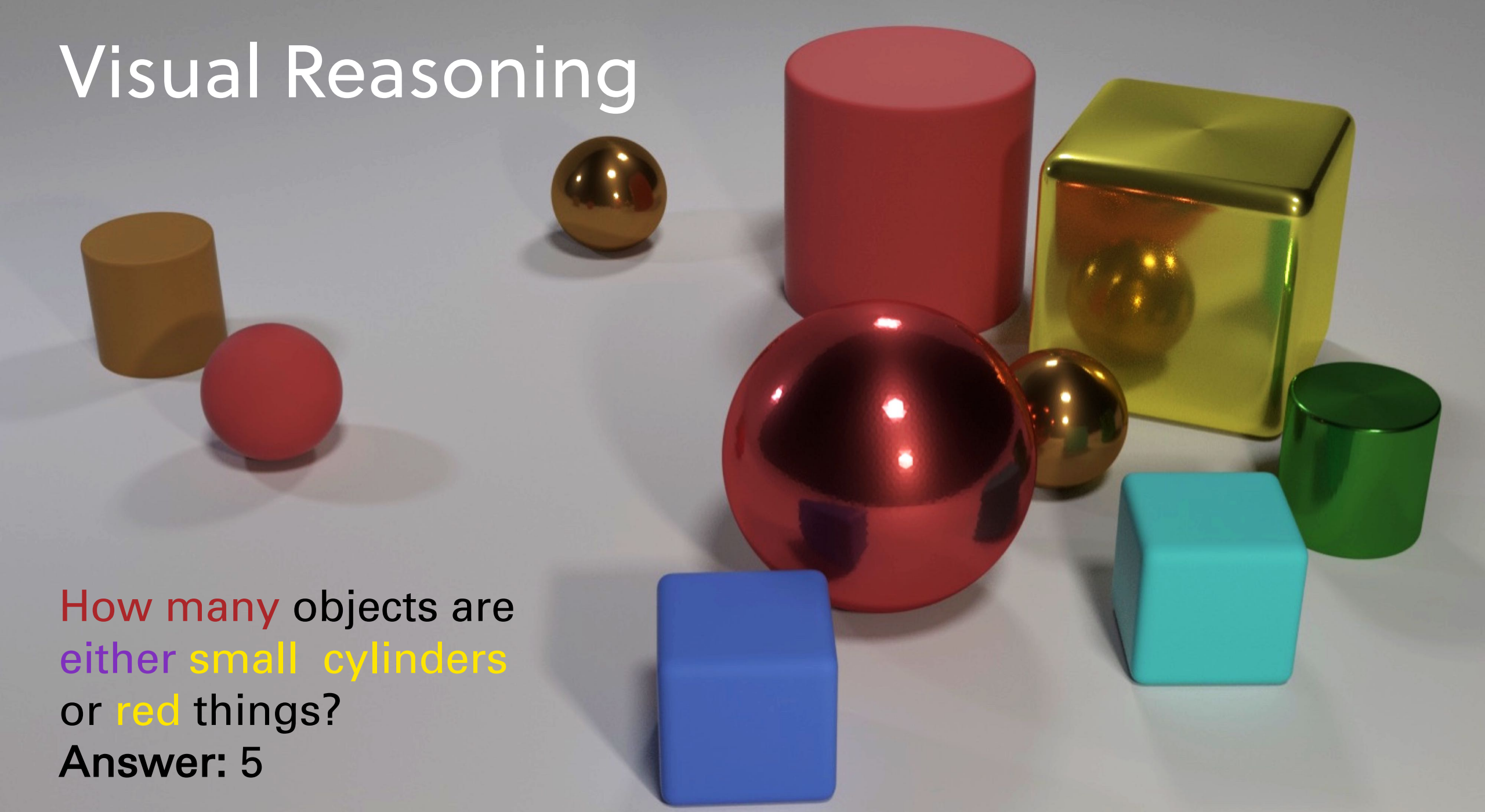
Explaining and understanding what we see



Q: What fruit is showing in this picture? **A:** Bananas

Kevin Xu et al. **Show, Attend and Tell: Neural Image Caption Generation with Visual Attention**. ICML 2015
Chao Ma et al. **Visual Question Answering with Memory-Augmented Networks**. CVPR 2018

Visual Reasoning

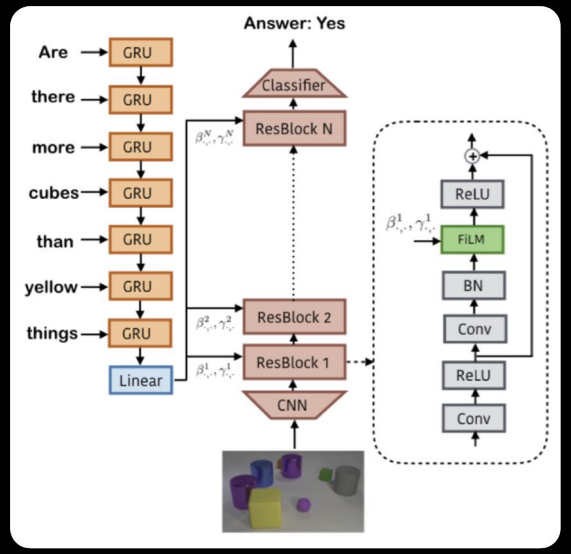


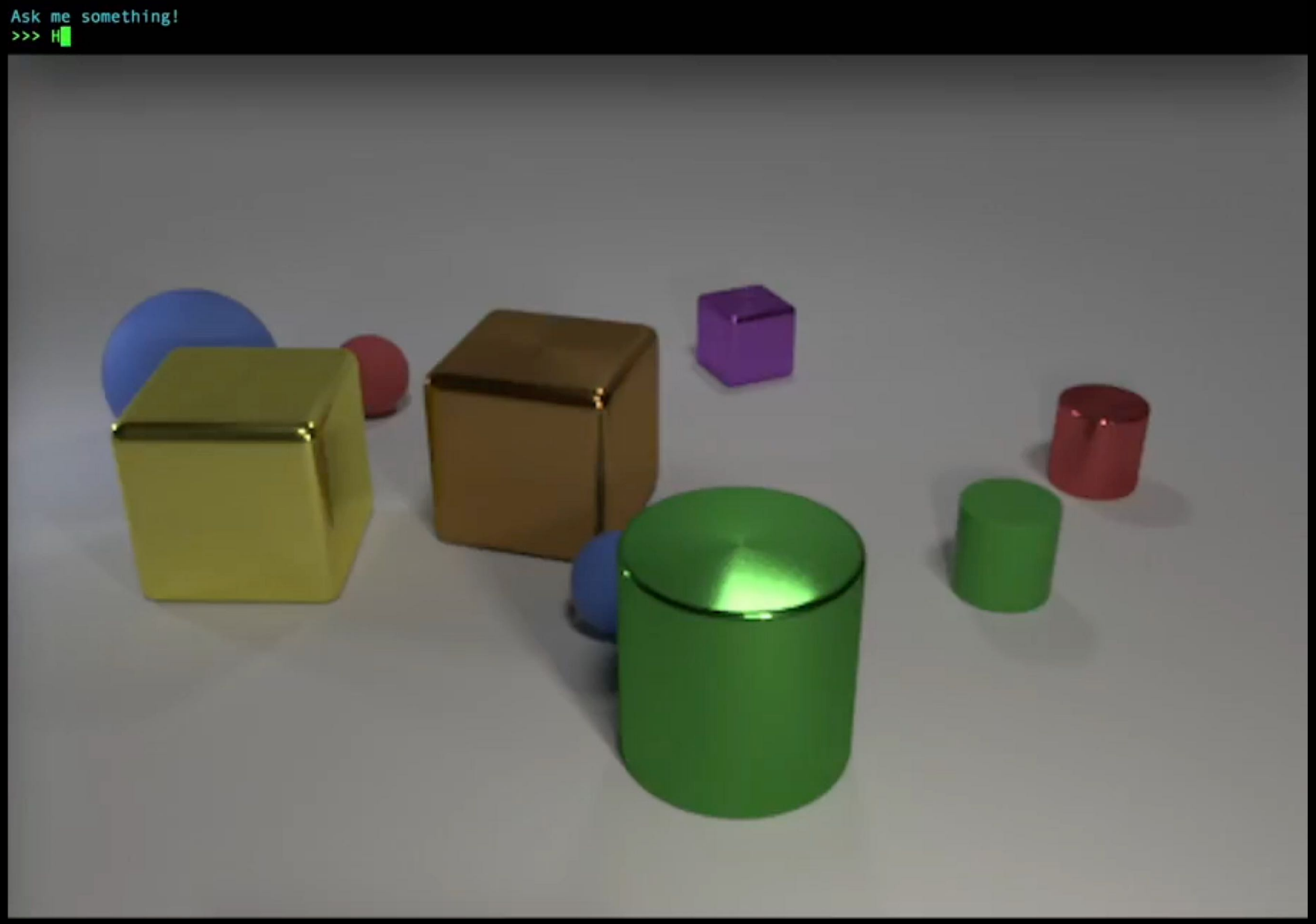
How many objects are either small cylinders or red things?

Answer: 5

Johnson et al. **CLEVR: A Diagnostic Dataset for Compositional Language and Elementary Visual Reasoning**. CVPR 2017

Visual Reasoning





E. Perez, F. Strub, H. de Vries, V. Dumoulin, A. Courville. **FiLM: Visual Reasoning with a General Conditioning Layer**. AAAI 2018

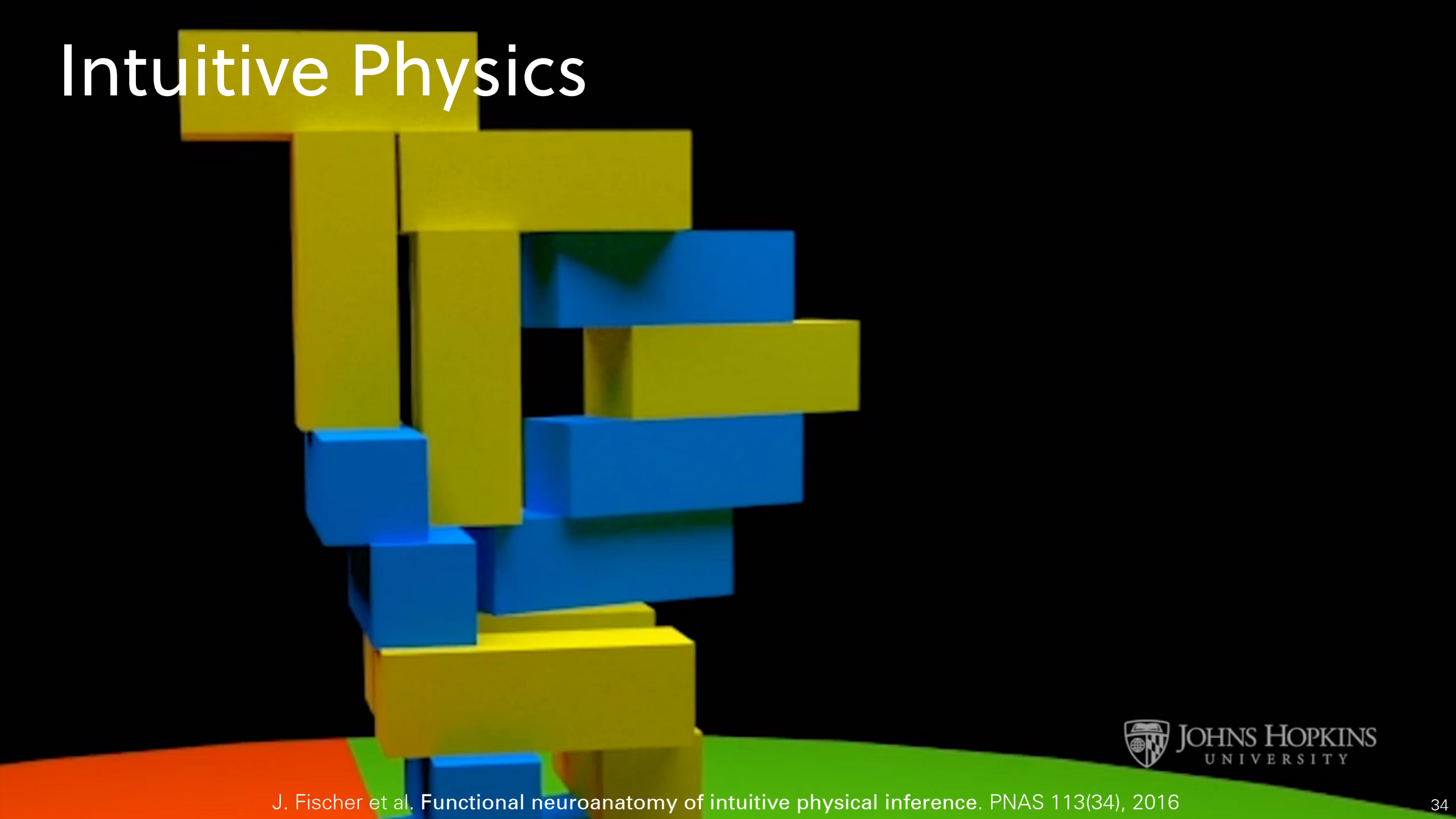
Intuitive Physics

- Common-sense understanding of **how the world operates at a physical level**
- Helps us to **perceive**, **understand** and **act** with our environment

Battaglia et al. **Simulation as an engine of physical scene understanding**. PNAS 110(45), 20

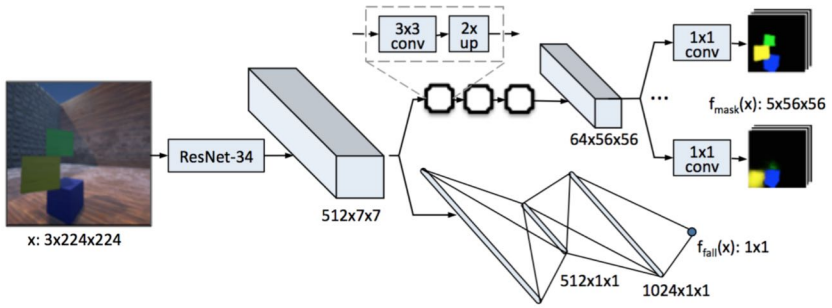
slide adapted from Peter Battaglia

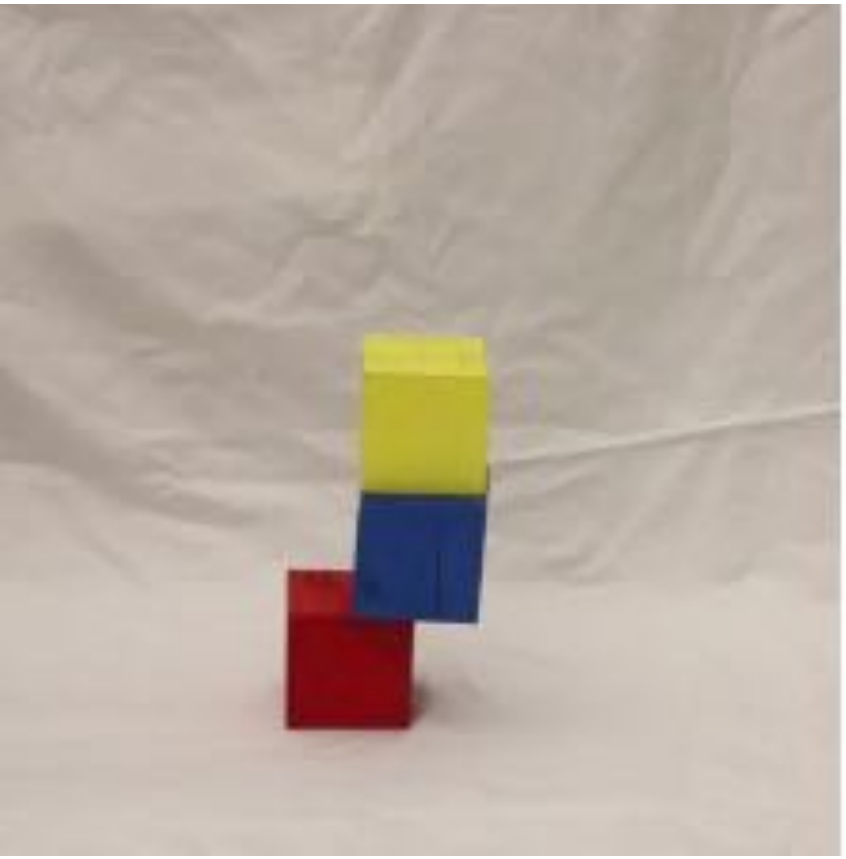
Intuitive Physics



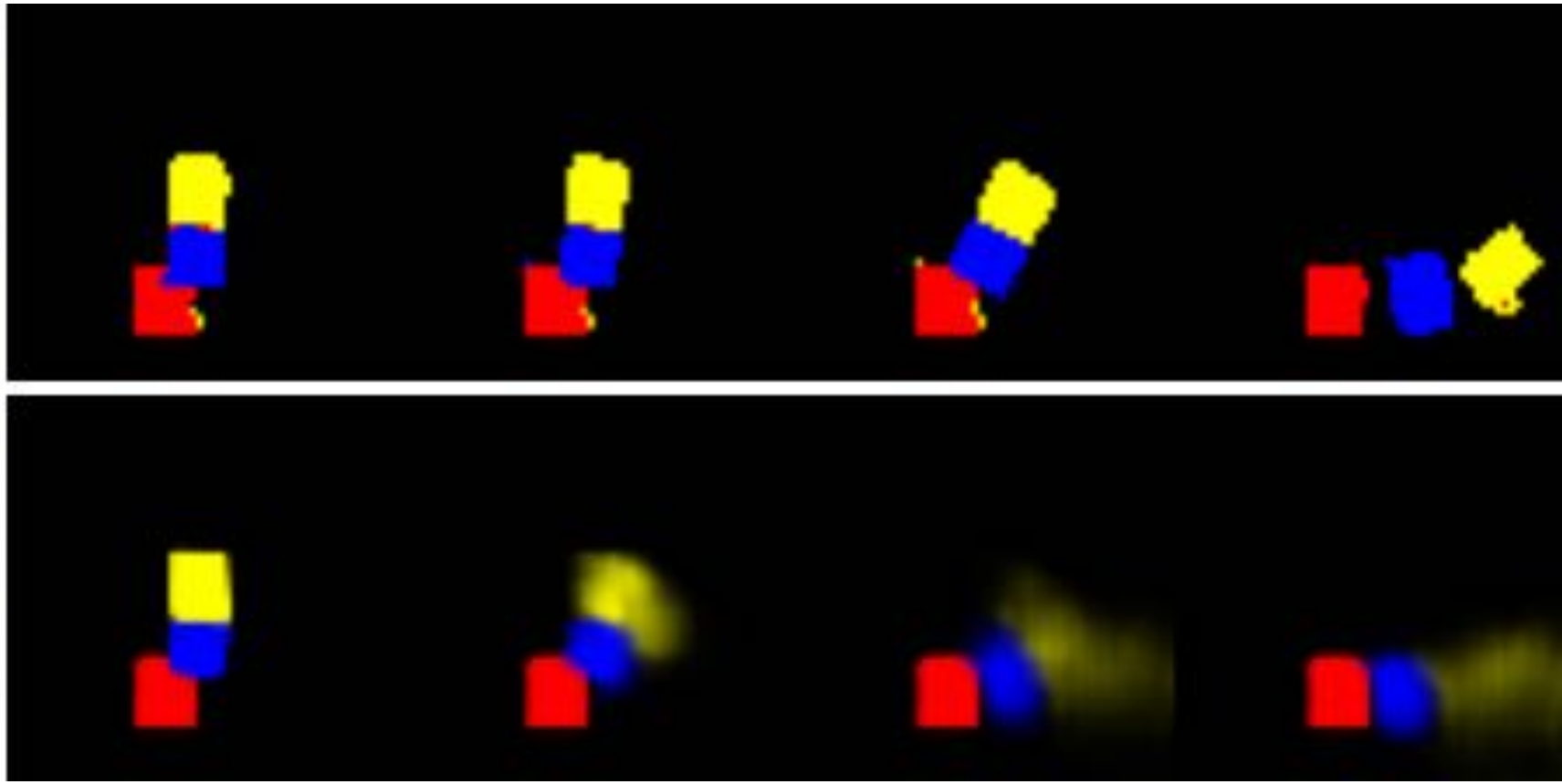
J. Fischer et al. **Functional neuroanatomy of intuitive physical inference.** PNAS 113(34), 2016

Intuitive Physics



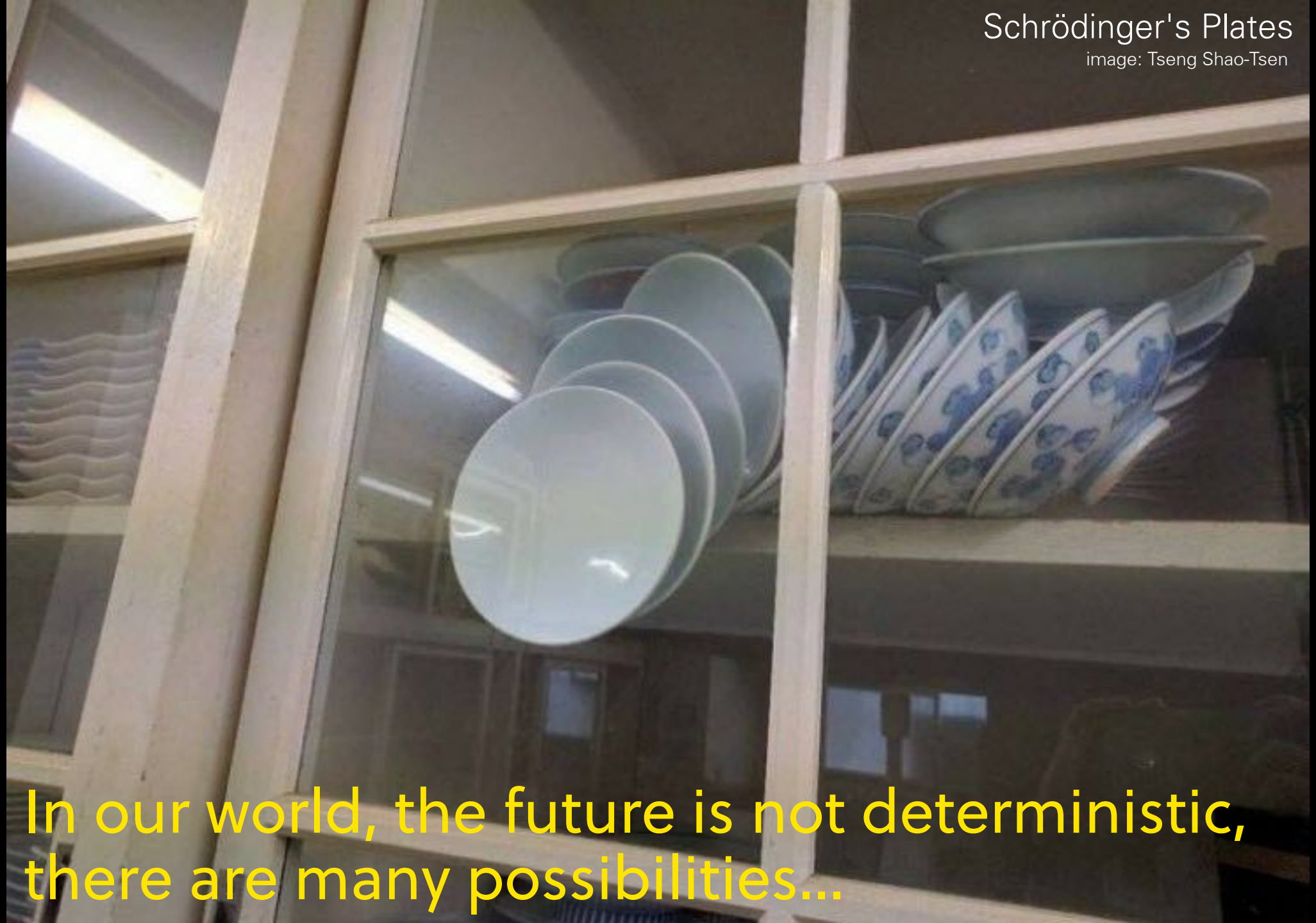


Initial frame



PhysNet predictions of the future

A. Lerer, S. Gross, R. Fergus. **Learning Physical Intuition of Block Towers by Example**. ICML 2016



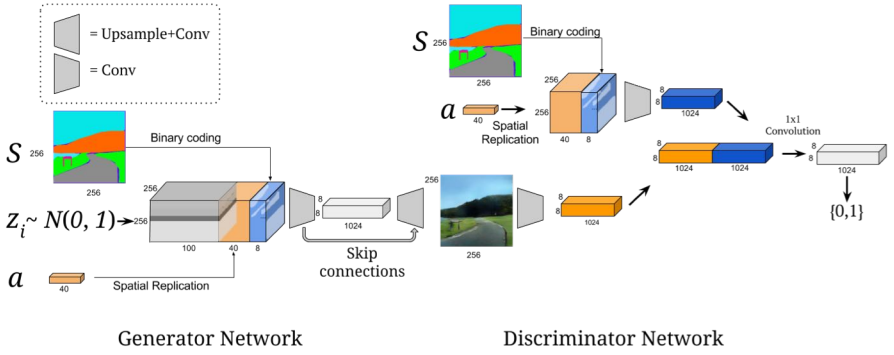
Schrödinger's Plates

image: Tseng Shao-Tsen

In our world, the future is not deterministic, there are many possibilities...

Imagining Things

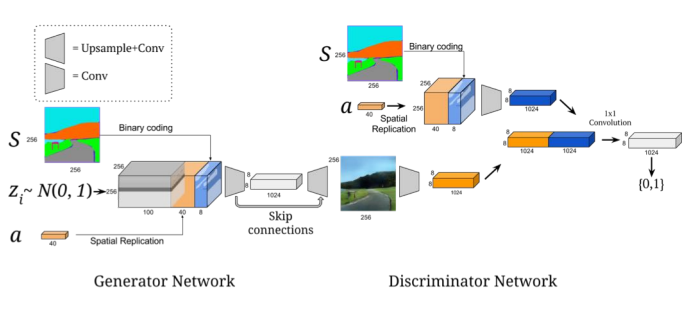
Imagining Things

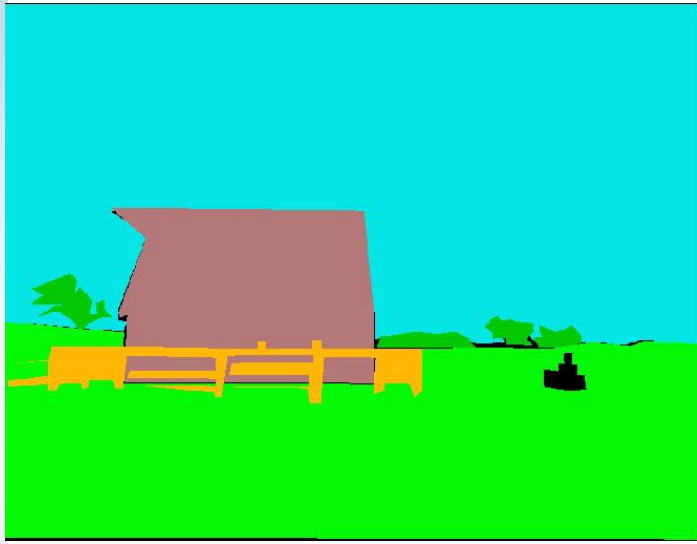


Input | Real | Imagined (w/ Original Attributes) | Imagined (Cloudy) | Imagined (Snowy) | Imagined (Sunset) | Imagined (Night)

Karacan, Akata, Erdem & Erdem. **Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts**. arXiv Preprint 2016

Imagining Things

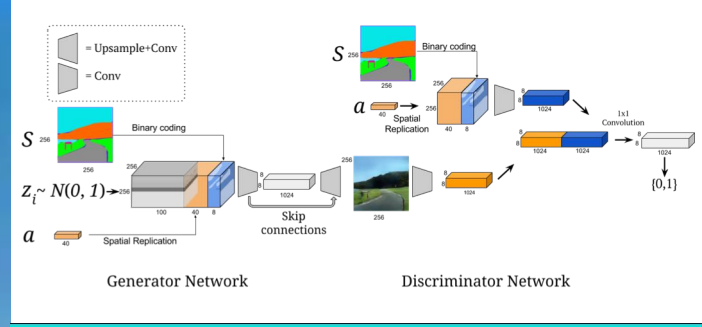


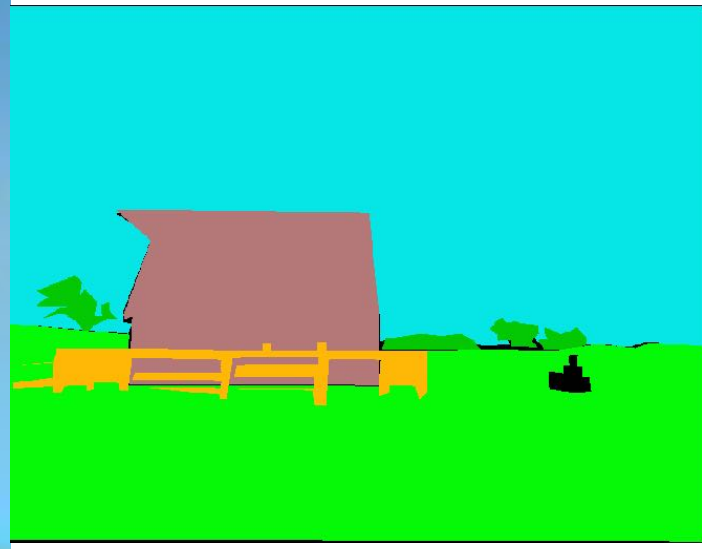


Semantic Layout

Karacan, Akata, Erdem & Erdem. **Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts**. arXiv Preprint 2016

Imagining Things



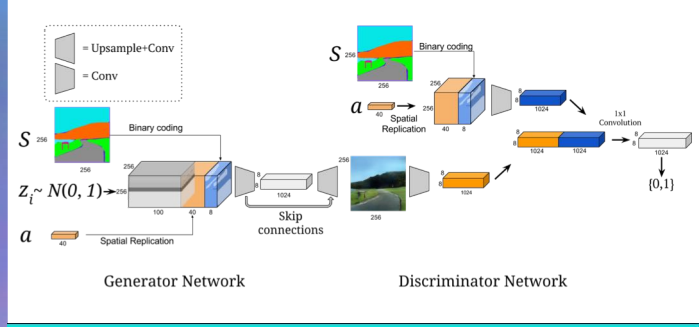


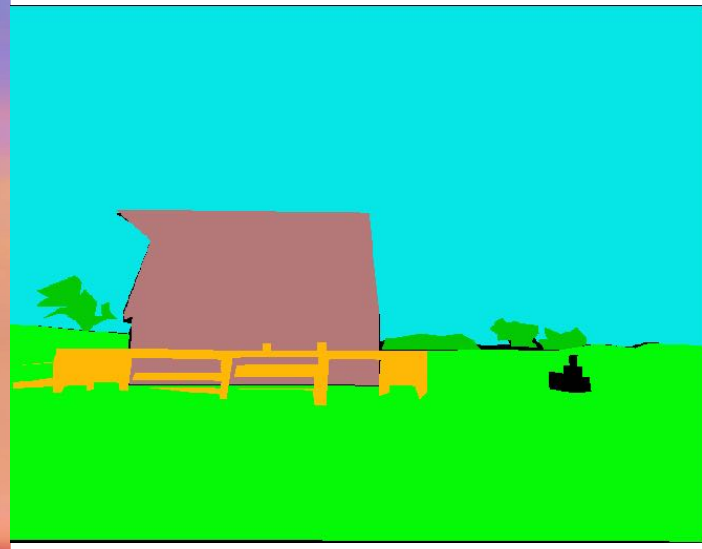
Semantic Layout

Clear sky + flowers

Karacan, Akata, Erdem & Erdem. **Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts**. arXiv Preprint 2016

Imagining Things



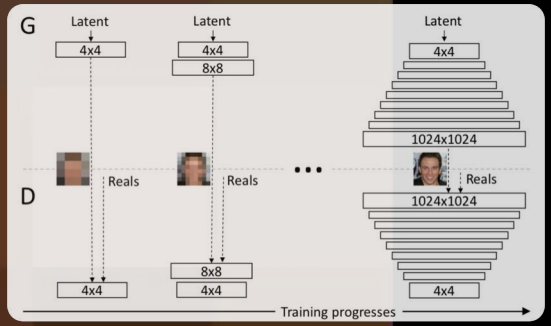


Semantic Layout

Sunset

Karacan, Akata, Erdem & Erdem. **Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts**. arXiv Preprint 2016

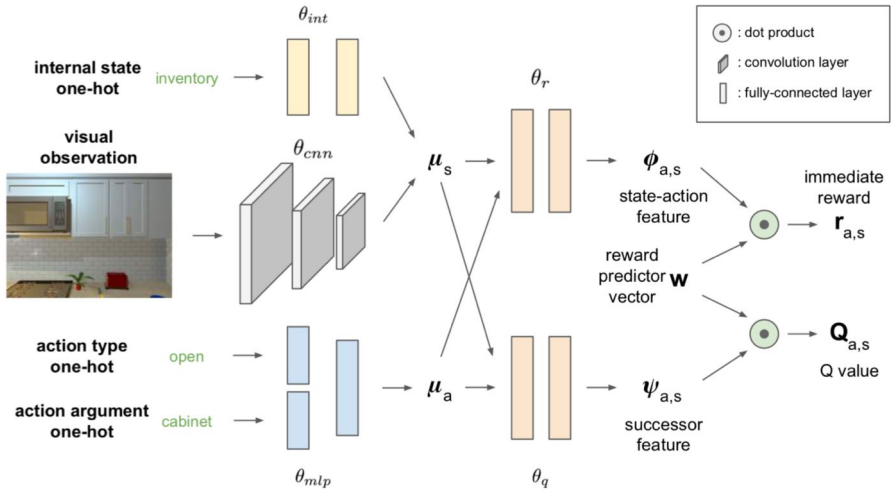
Imagining Things

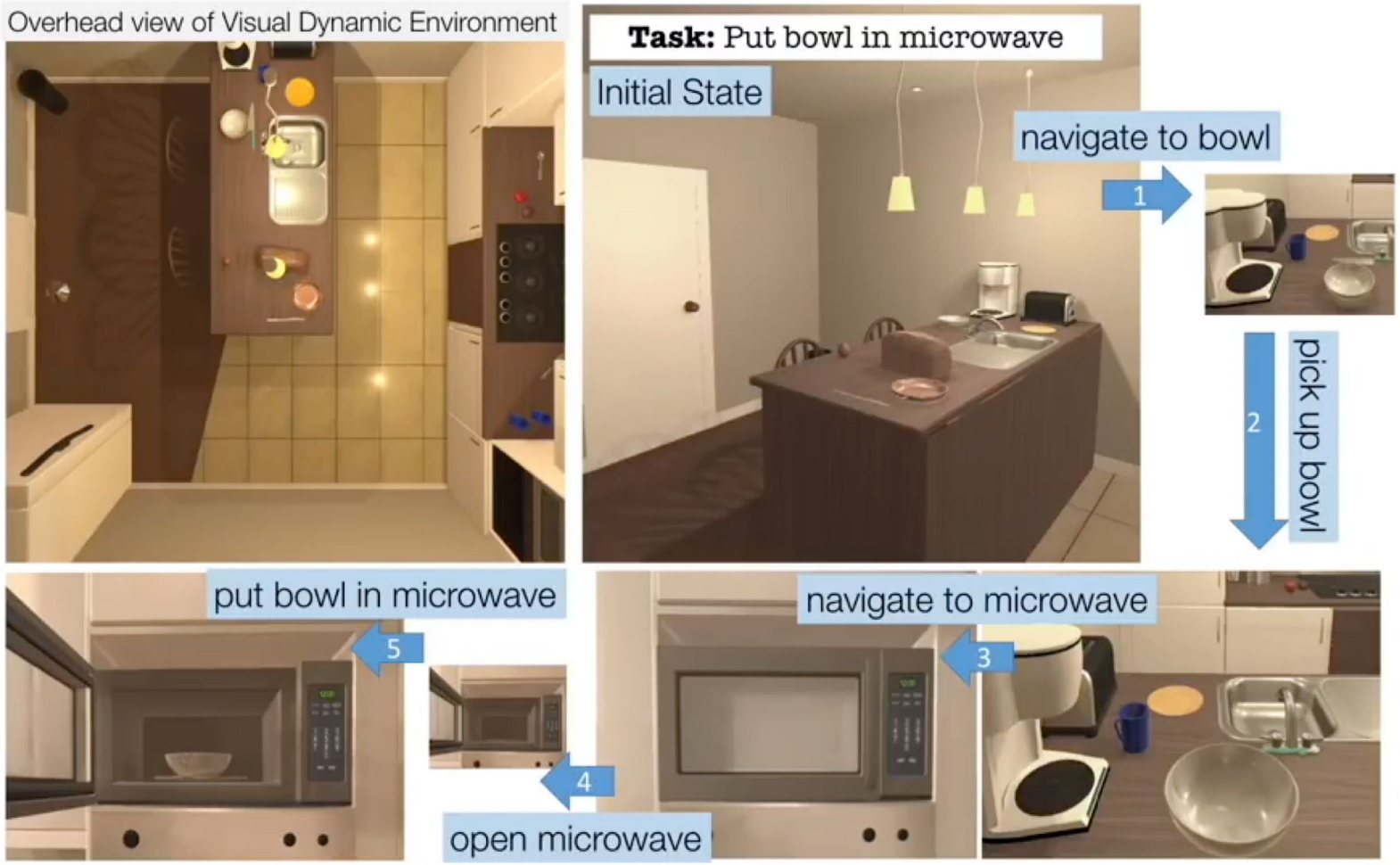


4x4

T.Karras, T.Aila, S.Laine and J, Lehtinen, "**Progressive Growing of GANs for Improved Quality, Stability, and Variation**", ICLR 2018

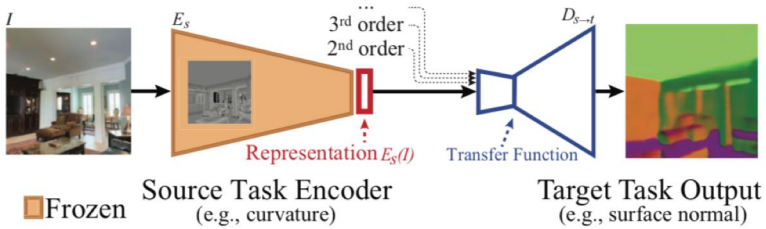
Planning

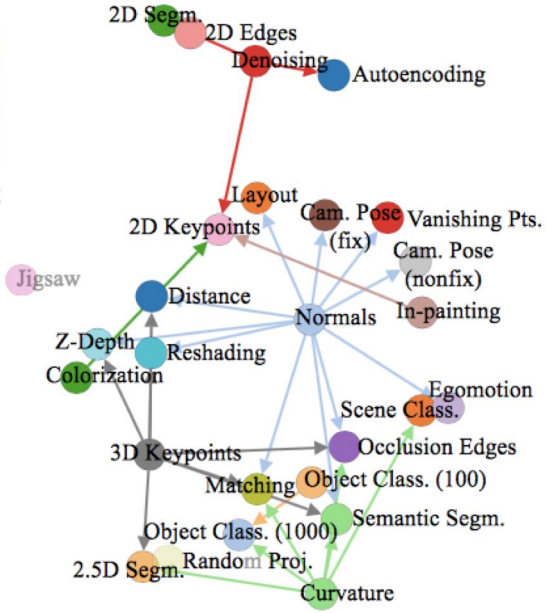


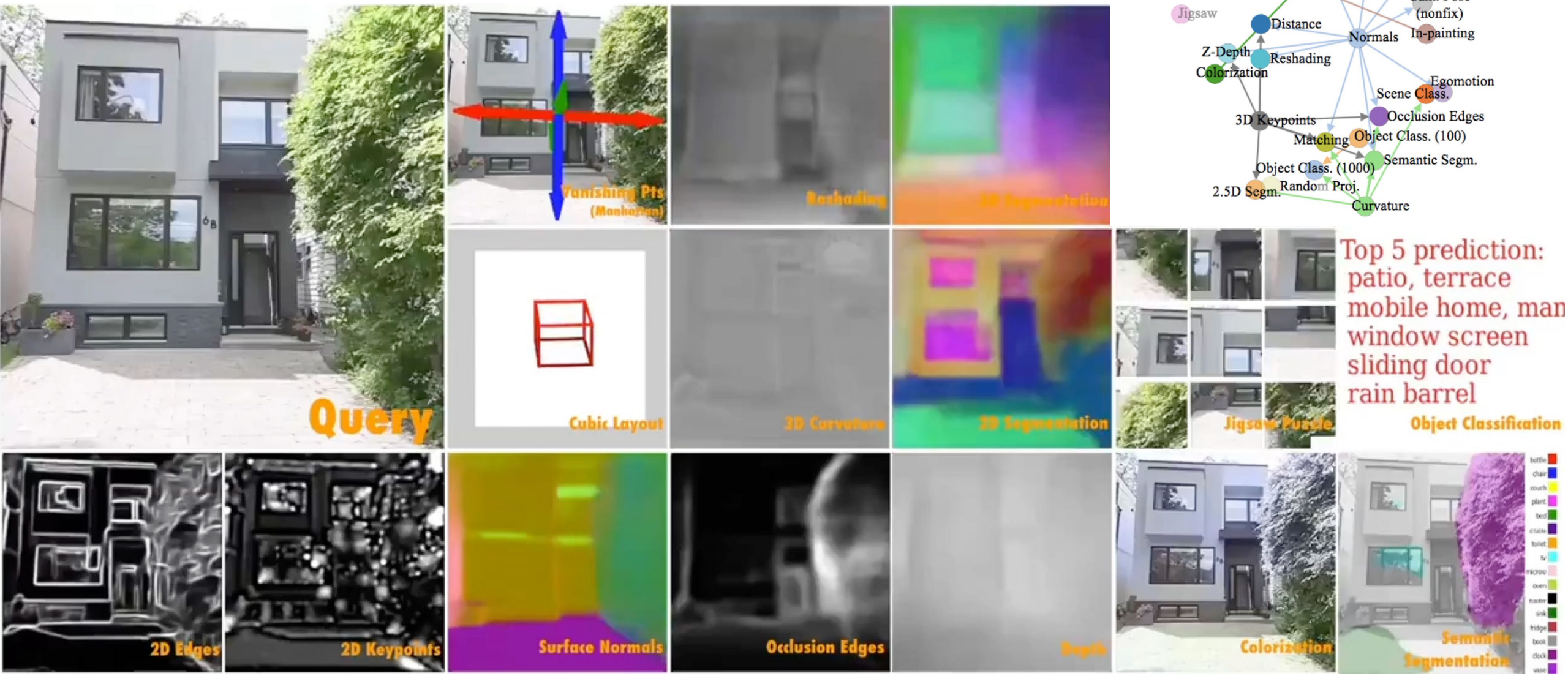


Y. Zhu, D. Gordon, E. Kolve, D. Fox, L. Fei-Fei, A. Gupta, R. Mottaghi & A. Farhadi. **Visual Semantic Planning using Deep Successor Representations**. ICCV 2017

Transfer Learning



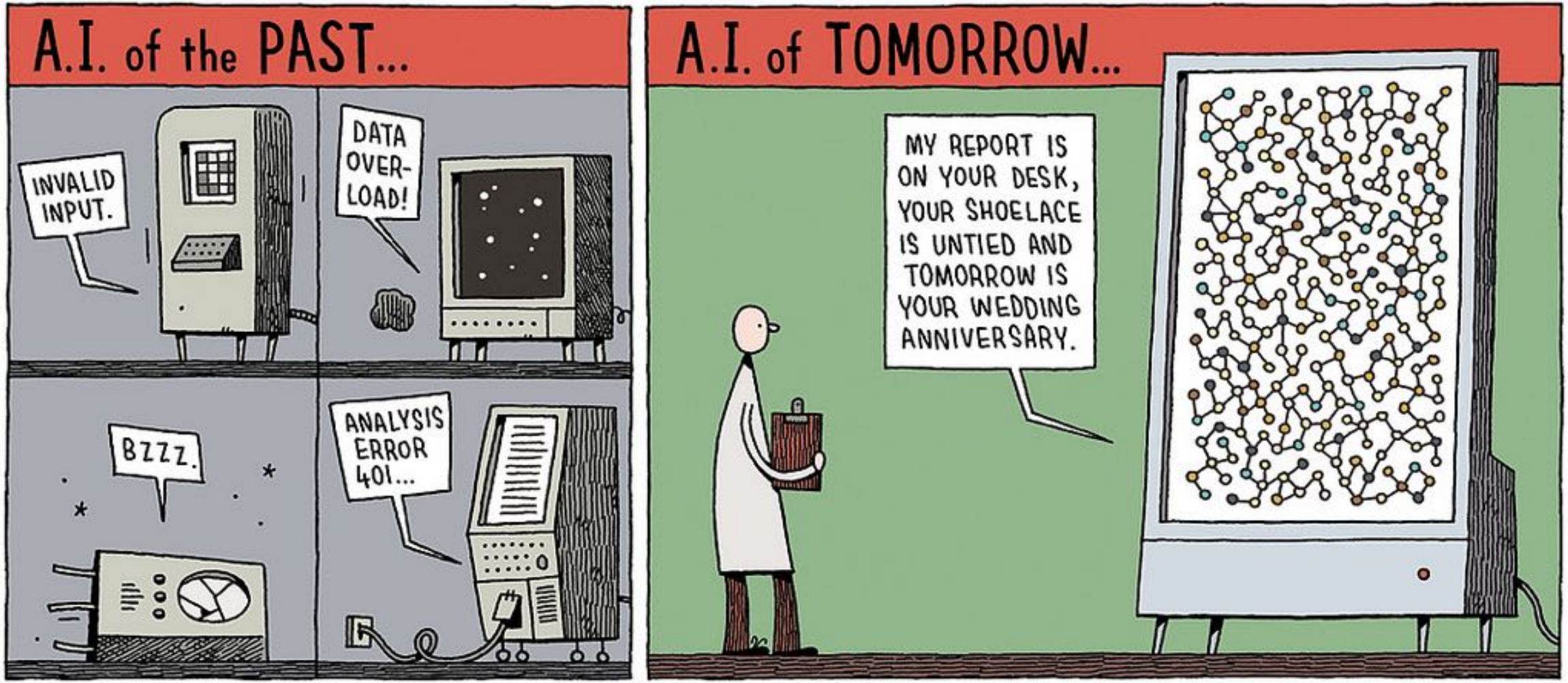




A.R. Zamir, A. Sax, W. Shen, L. Guibas, J. Malik & S. Saraves. **Taskonomy: Disentangling Task Transfer Learning**. CVPR 2018

cartoon: Tom Gauld

Can deep models reason?



- **We are not there yet!** But we can see real progress soon..