SDU Summer School

Deep Learning

Summer 2022

Welcome to the Summer School

About Myself

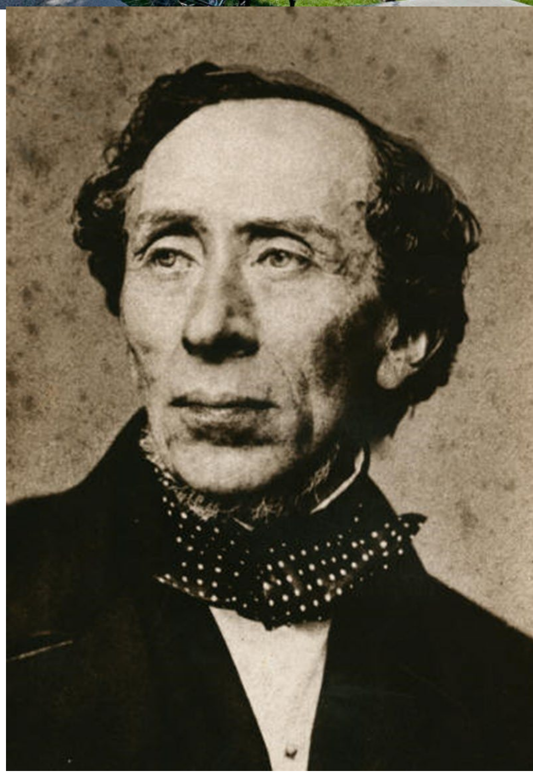
Richard Röttger

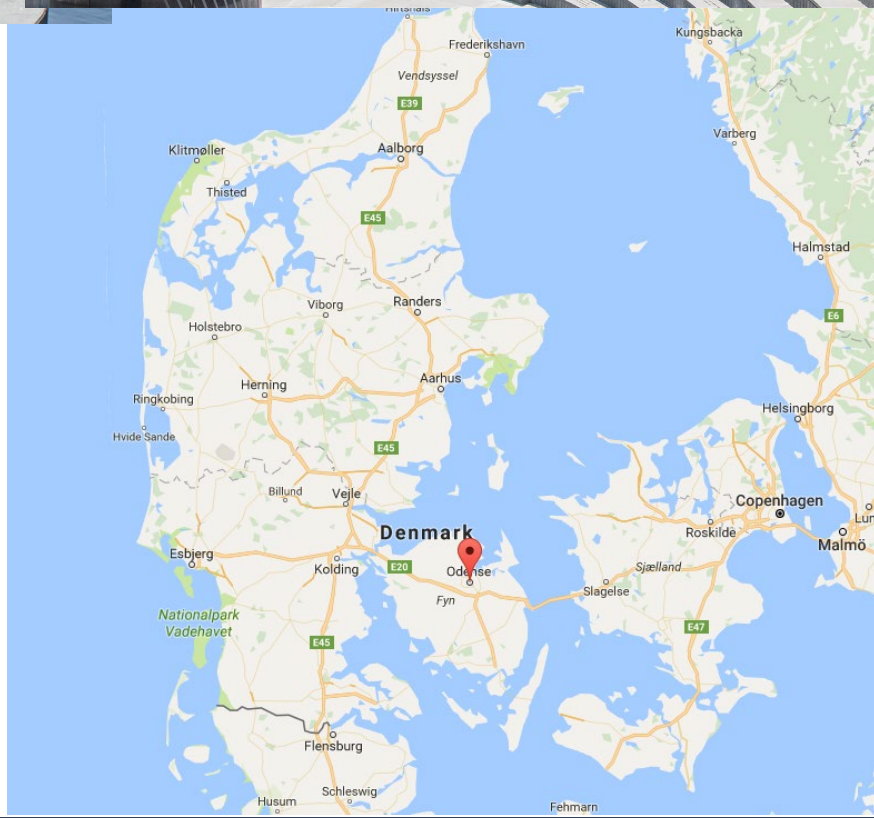
- Computer Scientist, TUM & Stay at UC Berkeley
- PhD at Max Planck Institute for Informatics
- At SDU since 2014

Main Research Interests

- Machine Learning
  - Interest in various kinds of machine learning approaches
- Clustering
  - Development of novel algorithms
  - (De novo) phenotyping is basically clustering
- Systems Medicine
  - Don’t look at the various entities in isolation
  - Understand and incorporate the interplay between the biological components

Welcome to SDU





Some Facts about SDU

- Founded in September 1966
- Now around 30k students, almost 20% of whom are from abroad, and more than 4,000 employees
- Campus in Odense and regional campuses in Slagelse, Kolding, Esbjerg and Sønderborg.

FAKULTETER	2013	2014	2015	2016
Sundhedsvidenskab	4.340	4.606	5.090	4.950
Naturvidenskab	2.036	2.150	2.308	2.317
Teknik	2.918	3.018	3.377	3.556
Samfundsvidenskab	11.406	12.023	12.543	11.722
Humaniora	8.029	8.216	7.981	7.129
SDU i alt	28.729	30.013	31.299	29.674

BYER	2013	2014	2015	2016
Odense	20.675	21.894	23.352	22.558
Kolding	2.634	2.686	2.797	2.594
Esbjerg	1.321	1.371	1.378	1.098
Sønderborg	1.942	1.836	1.552	1.283
Slagelse	2.157	2.226	2.220	2.141
SDU i alt	28.729	30.013	31.299	29.674

- Around 115 different study programmes in the fields of the humanities, social sciences, natural sciences, health sciences and engineering

But who are you?

Course Organization

- **Lectures**
  - We will cover the theory in lectures
  - Exact sequence not fixed yet

- **Exercises/Lab**
  - Then you will get exercises / programming tasks
  - You solve these tasks in teams of up to 5 students
  - Then you present your solutions to the class

- **Exam**
  - **Project Hand-In, pass/fail**
  - Work in a group, up to 5 people max
  - Heavily comment your code AND clearly explain who was responsible for what code.

Theory vs. Practice

- It is fundamental to understand the theory
  - When should you apply deep learning?
  - What are the benefits what are the downsides?
  - Why does the model not learn/what is going wrong?

- You can only answer these questions when you understand the theory!

- I try to cover both, but at the end it is a University class

Introduction to the Course

- Course web-page:
  - You will find all relevant material on itslearning
  - Slides will be uploaded the day before the lecture.

- Exercises:
  - Same goes for the exercises
  - Solutions will also be uploaded after we have discussed the exercises

Book and Slides

- **Deep Learning**
  Ian Goodfellow and Yoshua Bengio and Aaron Courville
  MIT Press, 2016

http://www.deeplearningbook.org/

- An Introduction to Statistical Learning:

http://www-bcf.usc.edu/~gareth/ISL/index.html

Content of the Course

- What will we cover in this course

1. Mathematical Foundations
   - Linear Algebra
   - Some Statistics
2. Linear Regression
   - They form the basic operations for Neural Networks
3. Feed Forward Networks
   - The simplest Form of DNN
4. Convolutional Neural Networks
5. Recurrent Neural Networks

Introduction

What is Deep Learning

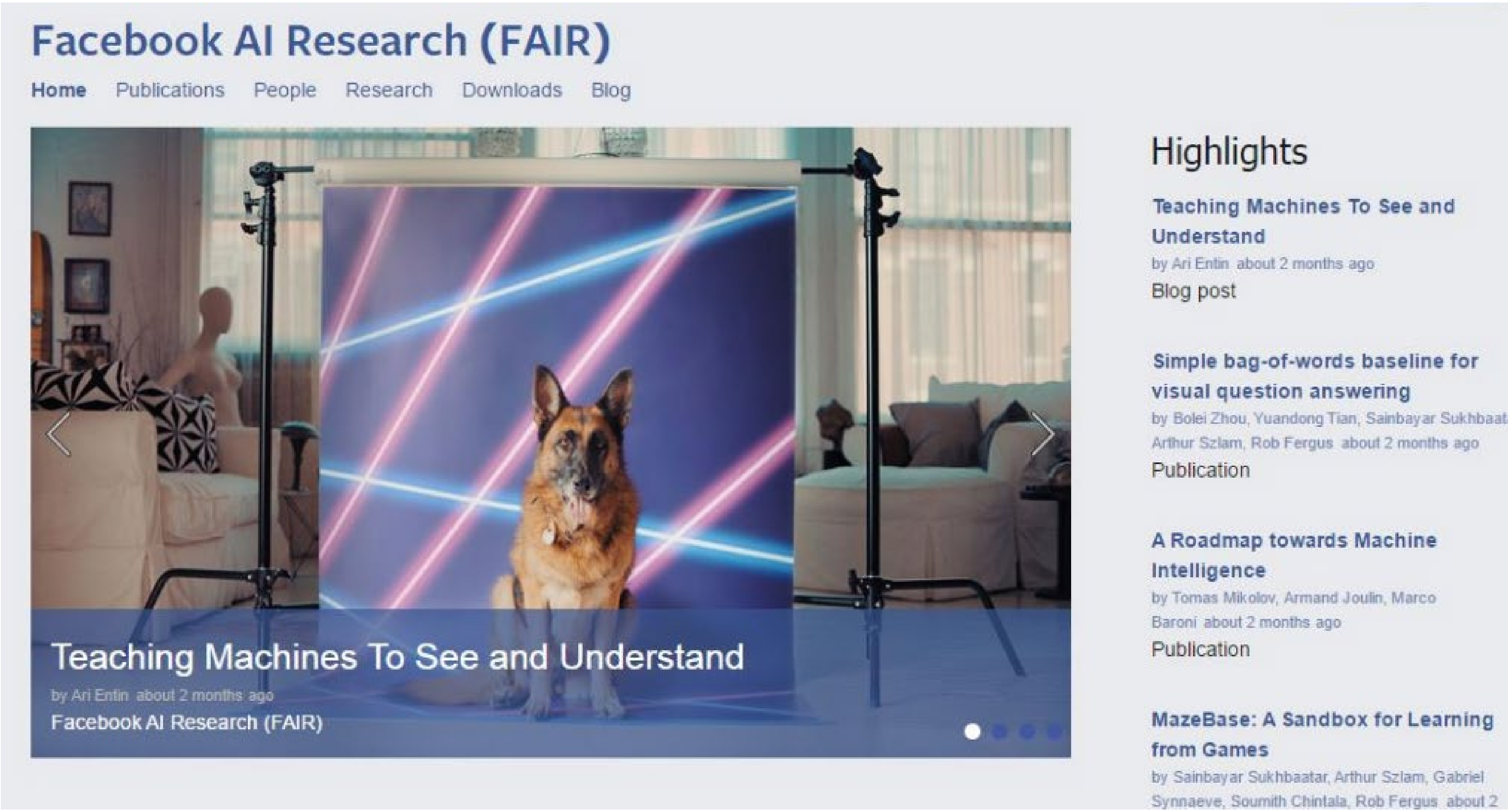
- **Short Answer:**

A Recent Buzzword ;-)

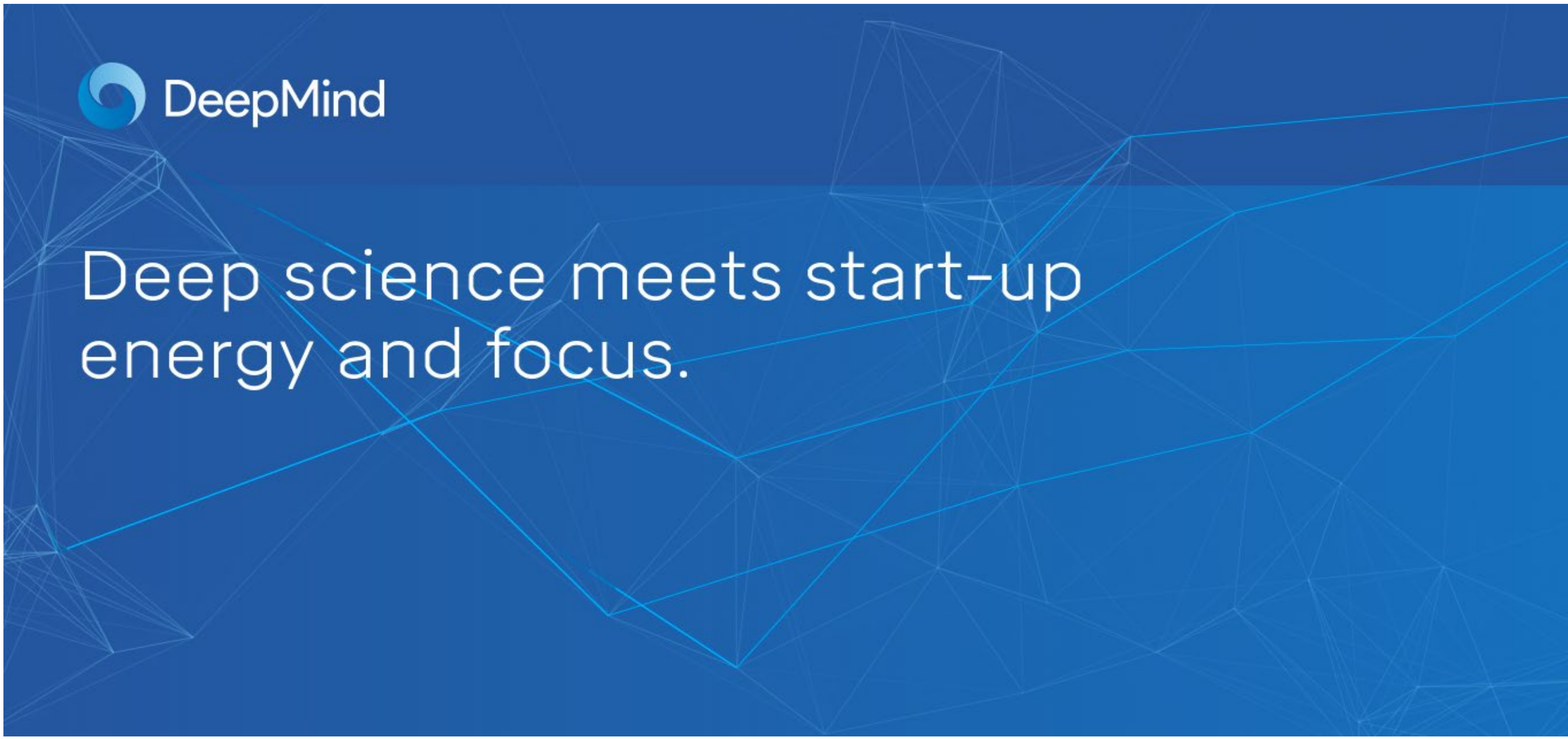
And a Money Printing Machine

Money Printing Machine

Examples: Facebook

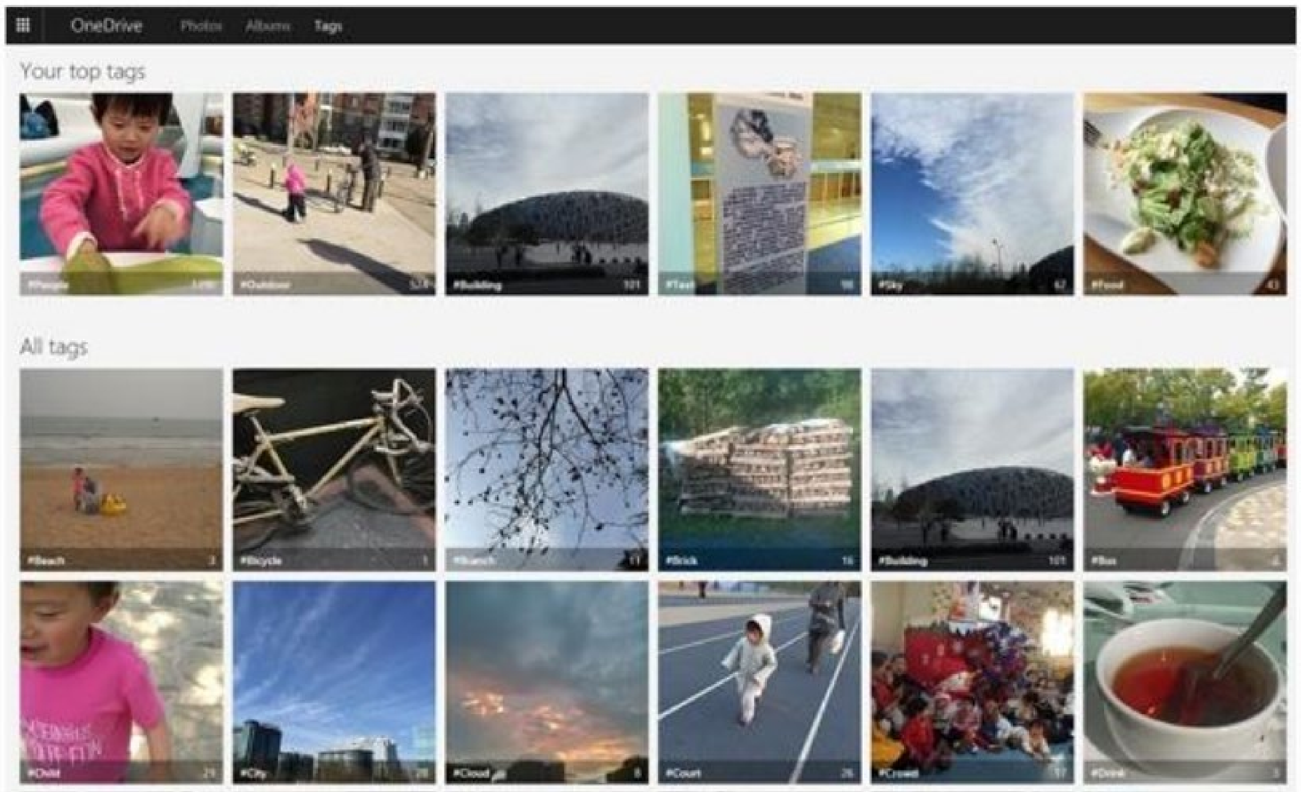


Examples: Google



Examples: Microsoft

Microsoft Researchers' Algorithm Sets ImageNet Challenge Milestone



VCs go bonkers on Start-Ups

This Week In AI Stats: $7.4 Billion Invested In AI Startups In Q2

Gil Press Contributor

Enterprise & Cloud

I write about technology, entrepreneurs and innovation.

The surveys, studies, forecasts and other quantitative assessments of the health and progress of AI for this week find increased VC excitement over AI startups, low trust in AI advice by U.S. consumers and global business executives, China and Germany ahead of the US in the use of AI in healthcare while the US spends more, and the inaccuracy of facial recognition.

Medical Science

Artificial Intelligence (AI) in Healthcare Market to Reach $27.6 billion by 2025- Exclusive Report by Meticulous Research®

In 2018, NLP holds the largest share among all the healthcare AI technologies. This segment is also expected to grow at a highest CAGR during the forecast period.

Email | Print Friendly | Share

July 30, 2019 09:33 ET | **Source:** Meticulous Market Research Pvt. Ltd.

London, July 30, 2019 (GLOBE NEWSWIRE) -- According to a new market research report **"Artificial Intelligence in Healthcare Market by Product (Hardware, Software, Services), Technology (Machine Learning, Context-Aware Computing, NLP), Application (Drug Discovery, Precision Medicine), End User, and Geography – Global Forecast to 2025"**, published by Meticulous Research®, the global artificial intelligence in healthcare market is expected to grow at a CAGR of 43.5% from 2018 to reach $27.6 billion by 2025.

Artificial intelligence (AI) is utilized by the healthcare industry in various applications such as patient data and risk analytics, medical imaging and diagnosis, drug discovery, precision medicine, hospital workflow management, and patient management as it applies various human intelligence-based functions such as reasoning, learning, and problem-solving skills on different disciplines such as biology, computer science, mathematics, linguistics, psychology, and engineering.

What is Deep Learning?

- **Some longer Answer:**

Some of the most exciting developments in Machine Learning, Vision, NLP, Speech, Robotics & AI that shows the most impressive performance!

Classification: Labradoodle or fried chicken?

- https://www.cs.tau.ac.il/~dcor/Graphics/pdf.slides/YY-Deep%20Learning.pdf

Classification: Puppy or bagel?

- https://www.cs.tau.ac.il/~dcor/Graphics/pdf.slides/YY-Deep%20Learning.pdf

Classification: Sheepdog or mop?

- https://www.cs.tau.ac.il/~dcor/Graphics/pdf.slides/YY-Deep%20Learning.pdf

Classification: Barn owl or apple?

- https://www.cs.tau.ac.il/~dcor/Graphics/pdf.slides/YY-Deep%20Learning.pdf

Classification: Raw chicken or Donald Trump?

➢ https://www.cs.tau.ac.il/~dcor/Graphics/pdf.slides/YY-Deep%20Learning.pdf

Surprising Results

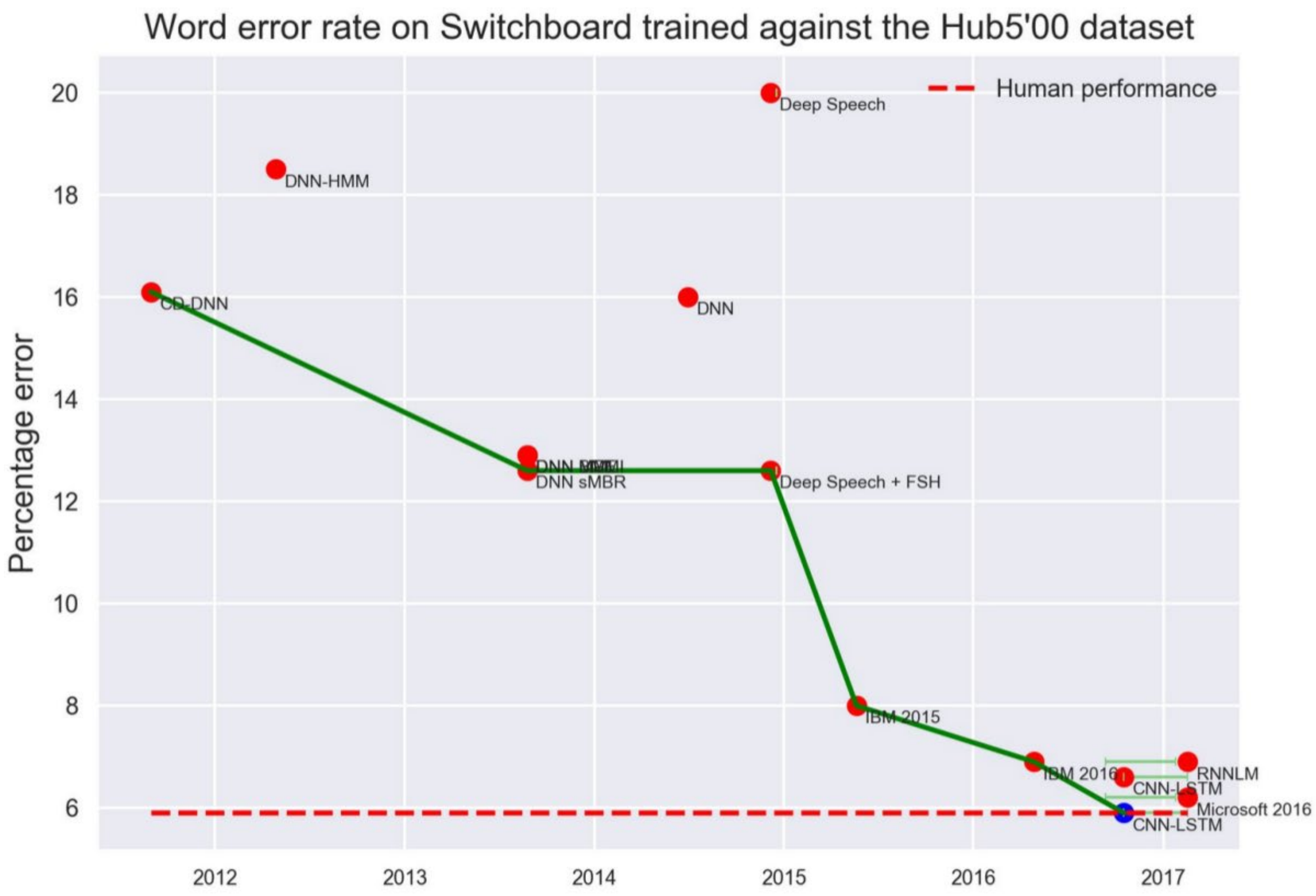
- This is a tremendously difficult task for a computer
  - A picture represents a huge feature space
  - The classification must be robust with respect to rotation, light conditions etc.

Surprising Results

- This is a tremendously difficult task for a computer
  - A picture represents a huge feature space
  - The classification must be robust with respect to rotation, light conditions etc.

Deep Neural Networks have surpassed human performance

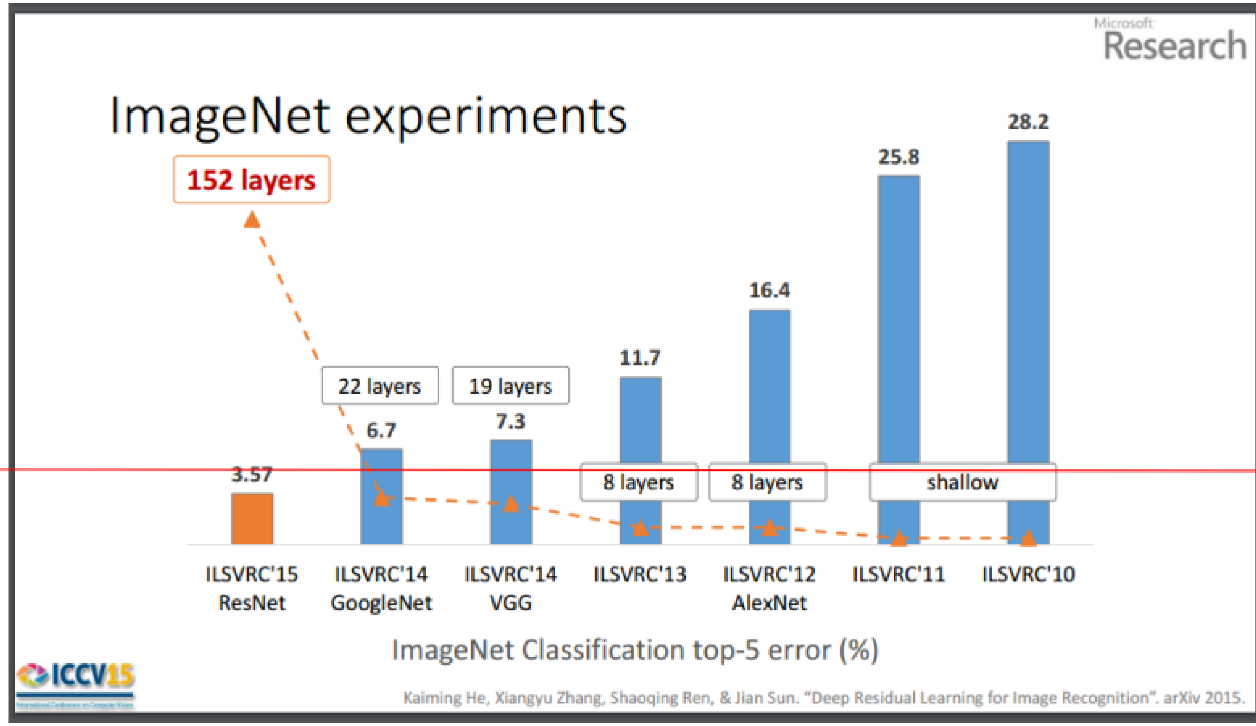
Transcribing Phone Calls



ImageNet Classification

- Image classification
  - 1000 classes

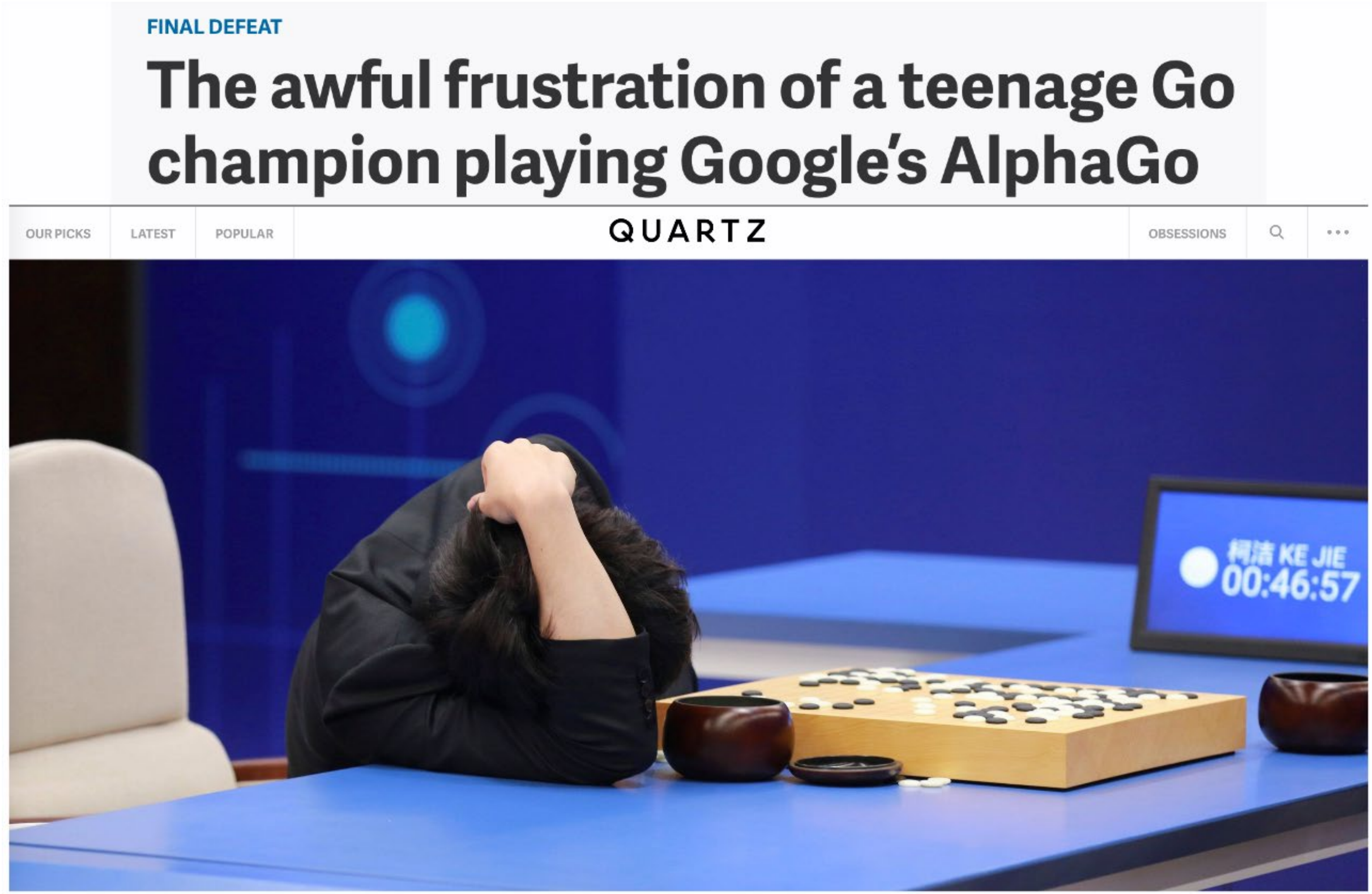
Human performance: ~5%



Slides from Kaimin He, MSRA

AlphaGo

AlphaGo



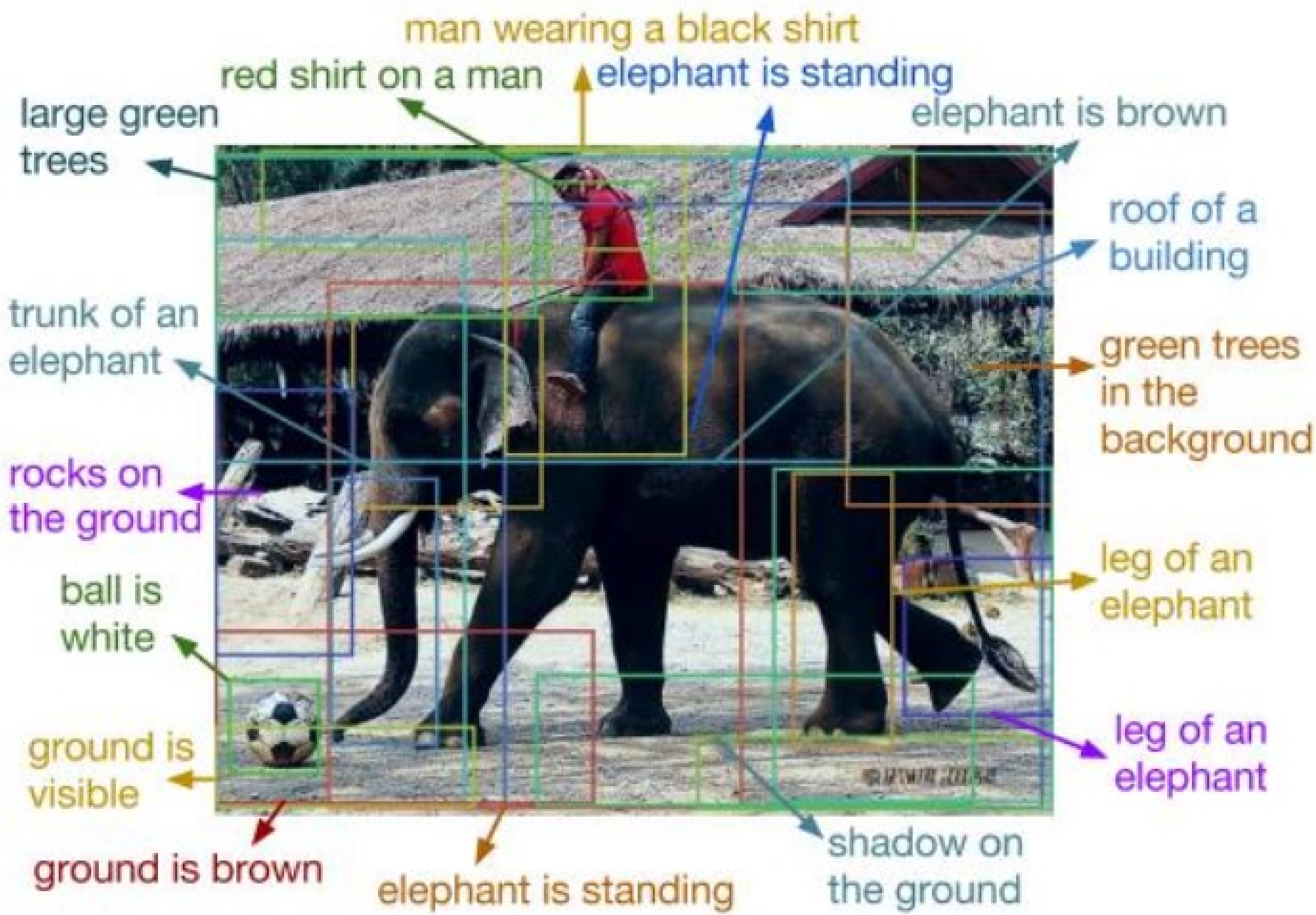
FINAL DEFEAT
The awful frustration of a teenage Go champion playing Google's AlphaGo
OUR PICKS
LATEST
POPULAR
QUARTZ
OBSESSIONS
KE JIE
00:46:57

More Difficult Tasks

- Object location

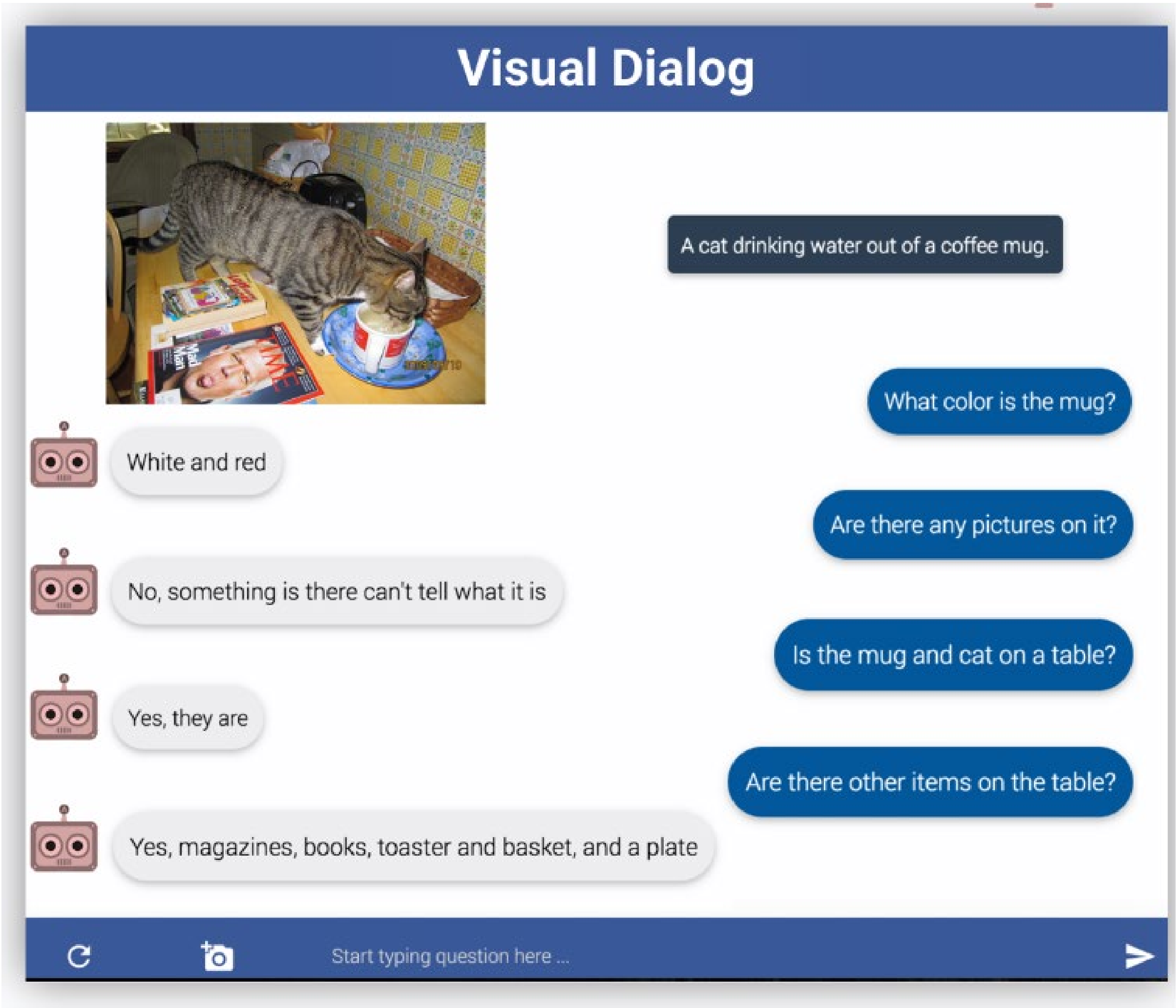
Slides from Kaimin He, MSRA

Even more Difficult



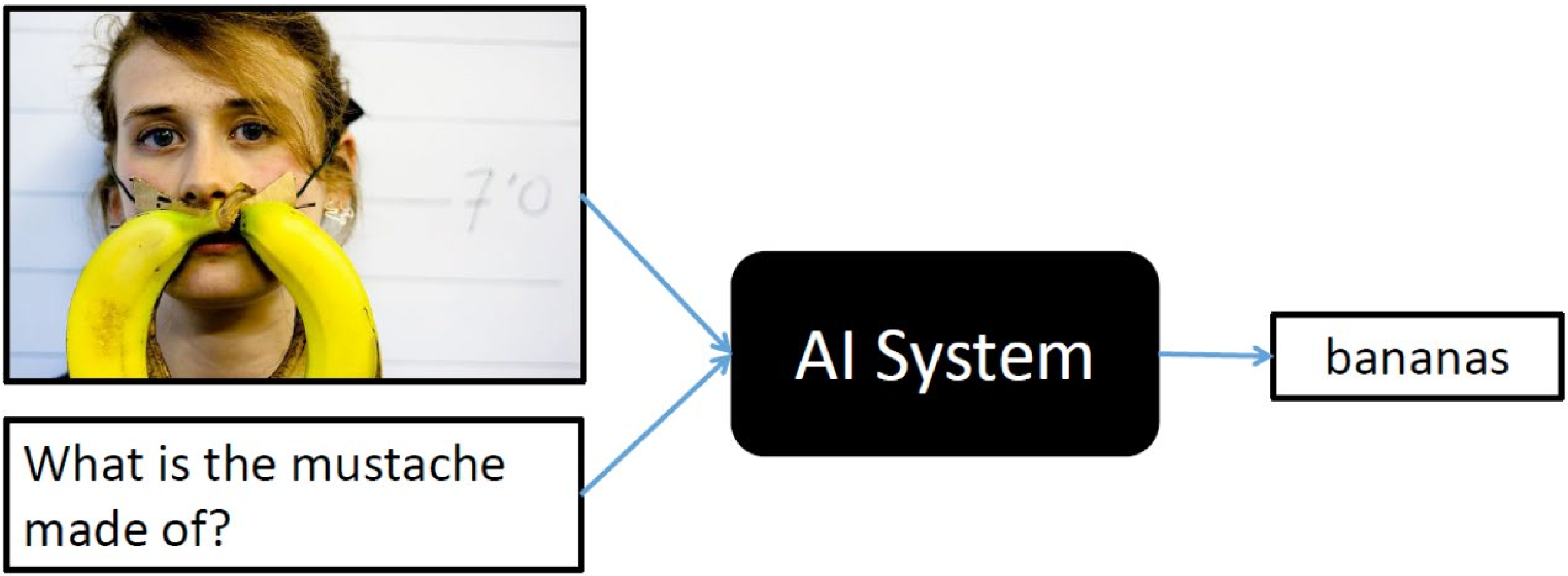
- Figure from the paper “DenseCap: Fully Convolutional Localization Networks for Dense Captioning”, by Justin Johnson, Andrej Karpathy, Li Fei-Fei

Increasingly more difficult tasks



- Das et al., 2017

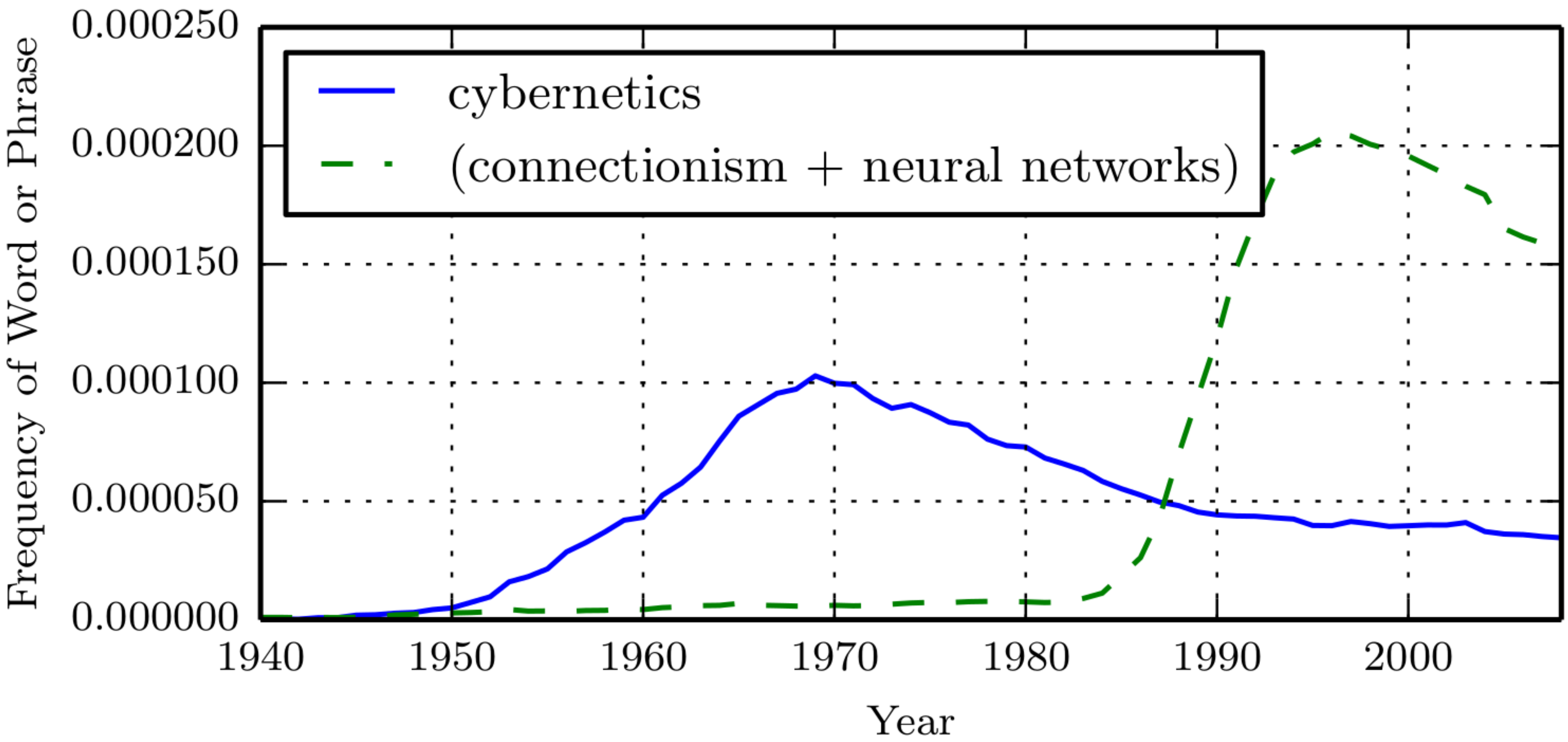
Visual Question Answering (VQA)



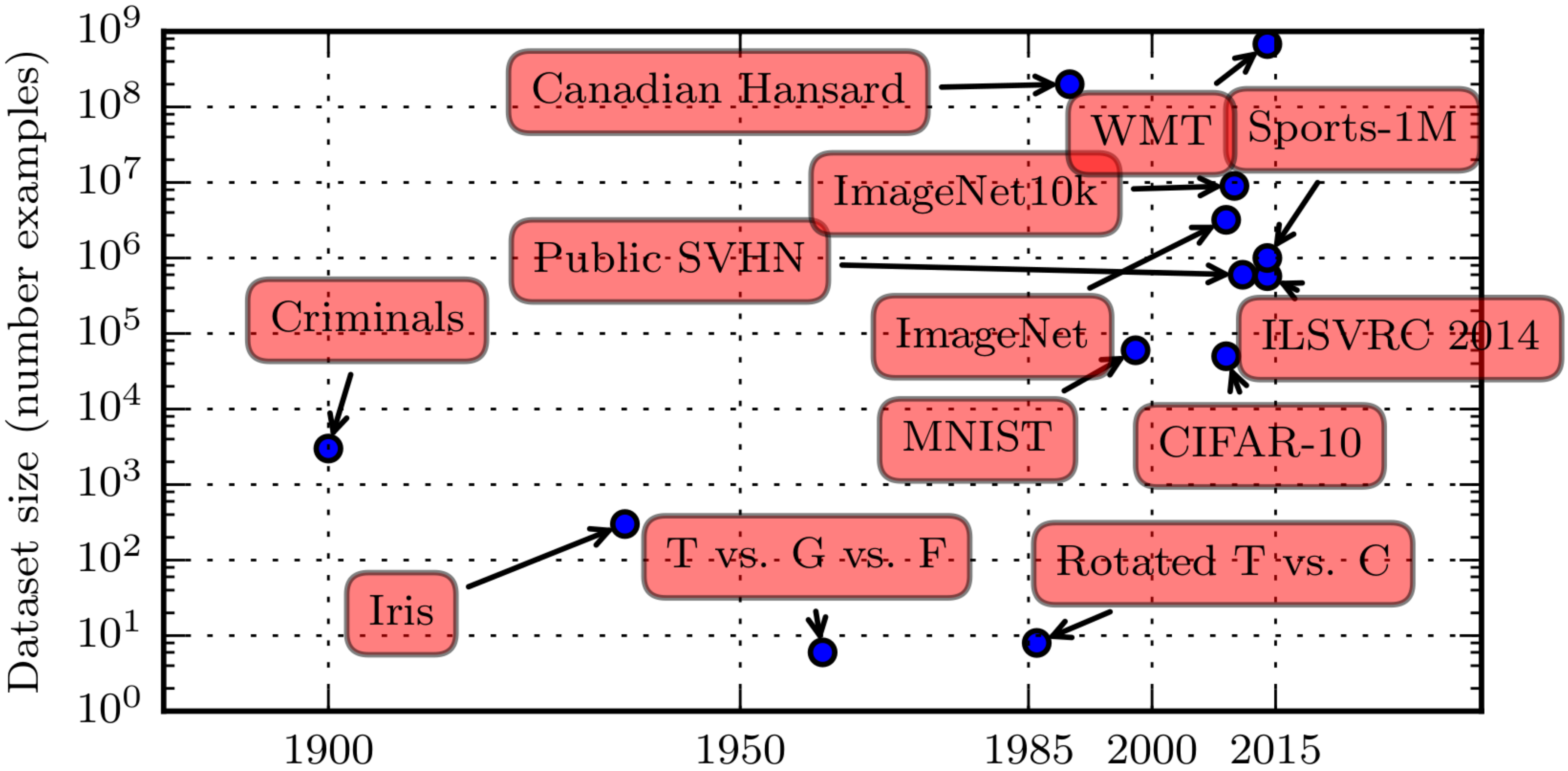
All recent developments … really?

The fundamental ideas used in deep learning were established in the 50s and before.

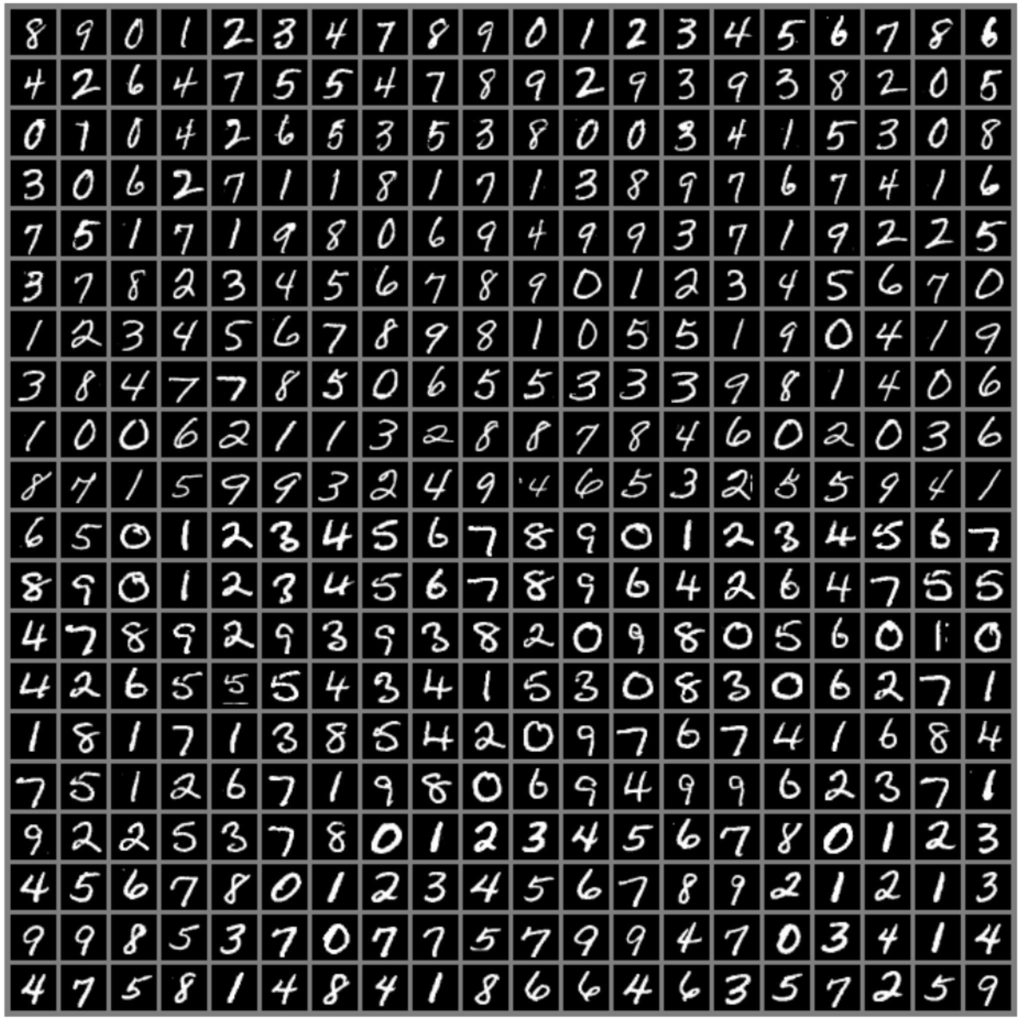
Historical Waves



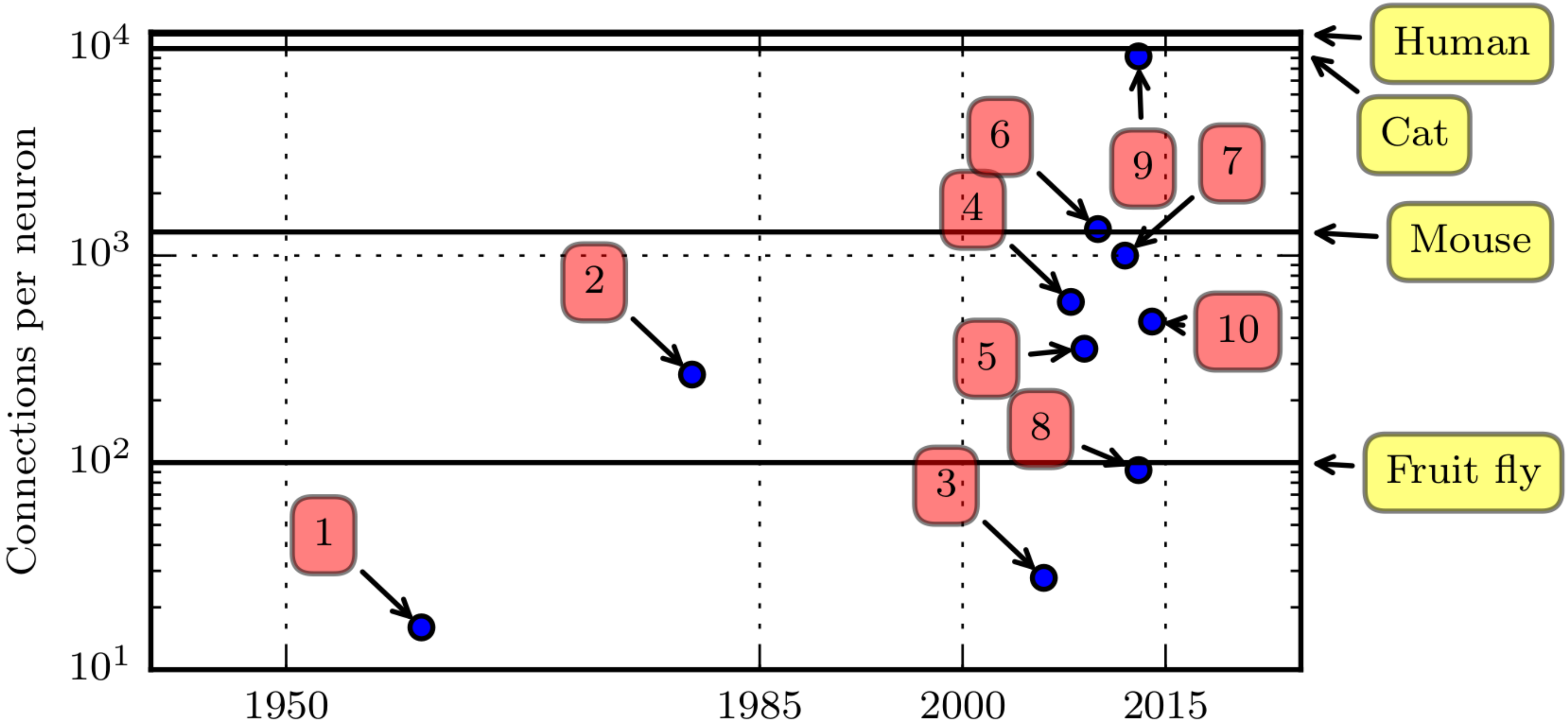
Enablers of Deep Learning: Growing Dataset Sizes



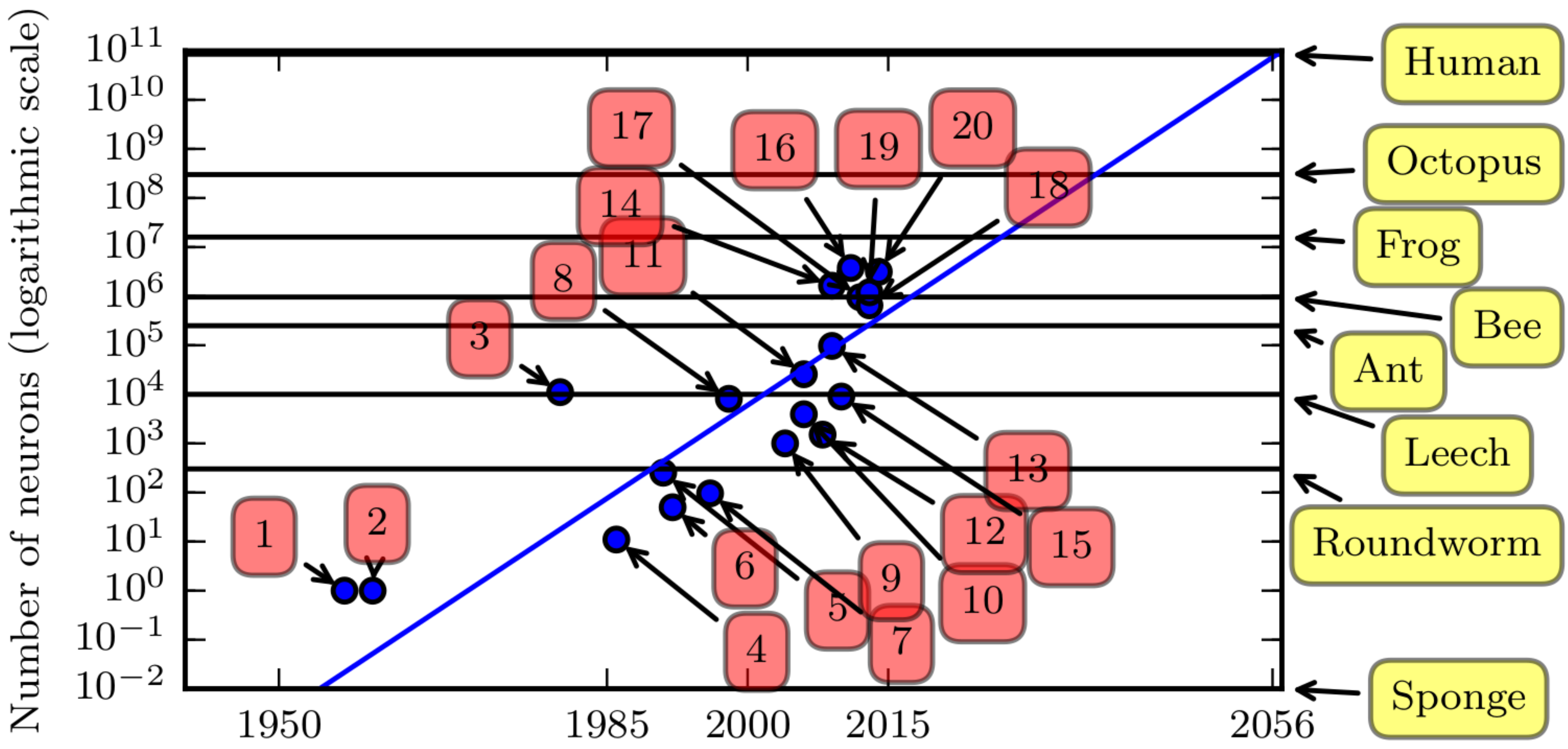
The MNIST Dataset



Enablers of Deep Learning: Compute Power
Connections per Neuron



Enablers of Deep Learning: Compute Power
Number of Neurons



What is Deep Learning?
(for real now)

Let’s start: What is Machine Learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at task T, as measured by P, improves with experience E

Okay, and what does that mean?

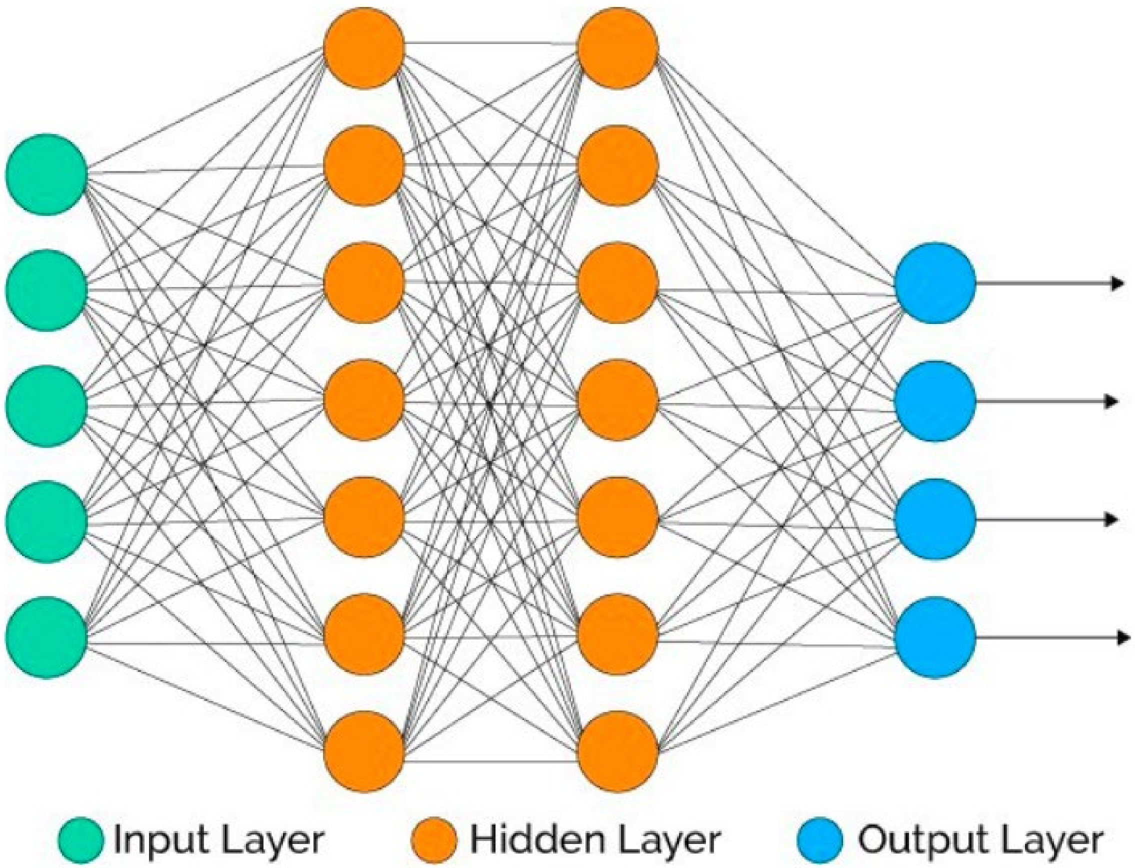
- **The Task T:**
  - The task can have a multitude of different forms:
    - Classification
    - Prediction
    - Or even to drive a car or to beat the living crap out of GO players
- **Performance P:**
  - For example:
    - Number of misclassifications
    - Number of won games
    - Number of accidents or rule infringements while driving the car
- **Experience E:**
  - Just means, the program should improve with an increasing number of examples

How does a Computer Learn?

- A common method for machine learning is fitting a function to the dataset at hand:
  - A function defining the decision boundary
  - A function describing the output in dependence of the input
  - etc.
- What function to use:
  - In principle anything goes
  - In practice, we choose a family of functions
    - I.e., linear functions, polynomial functions, etc.
- What is learning then:
  - Each function family has certain degrees of freedom
  - Learning is then to find the best parametrization for your function family

What has this to do with Deep Learning?

- Deep learning is based on neural networks
- Neural networks are an elegant way of describing complicated functions!



The Driving Ideas of Deep Learning

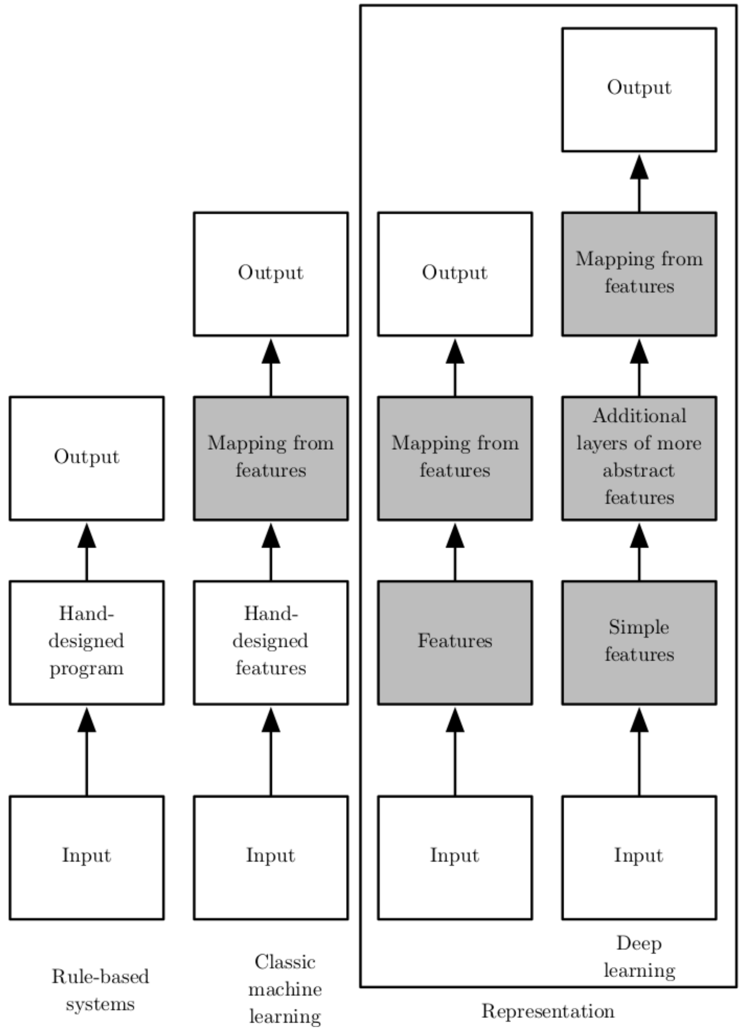
- Representation Learning using deep neural networks

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

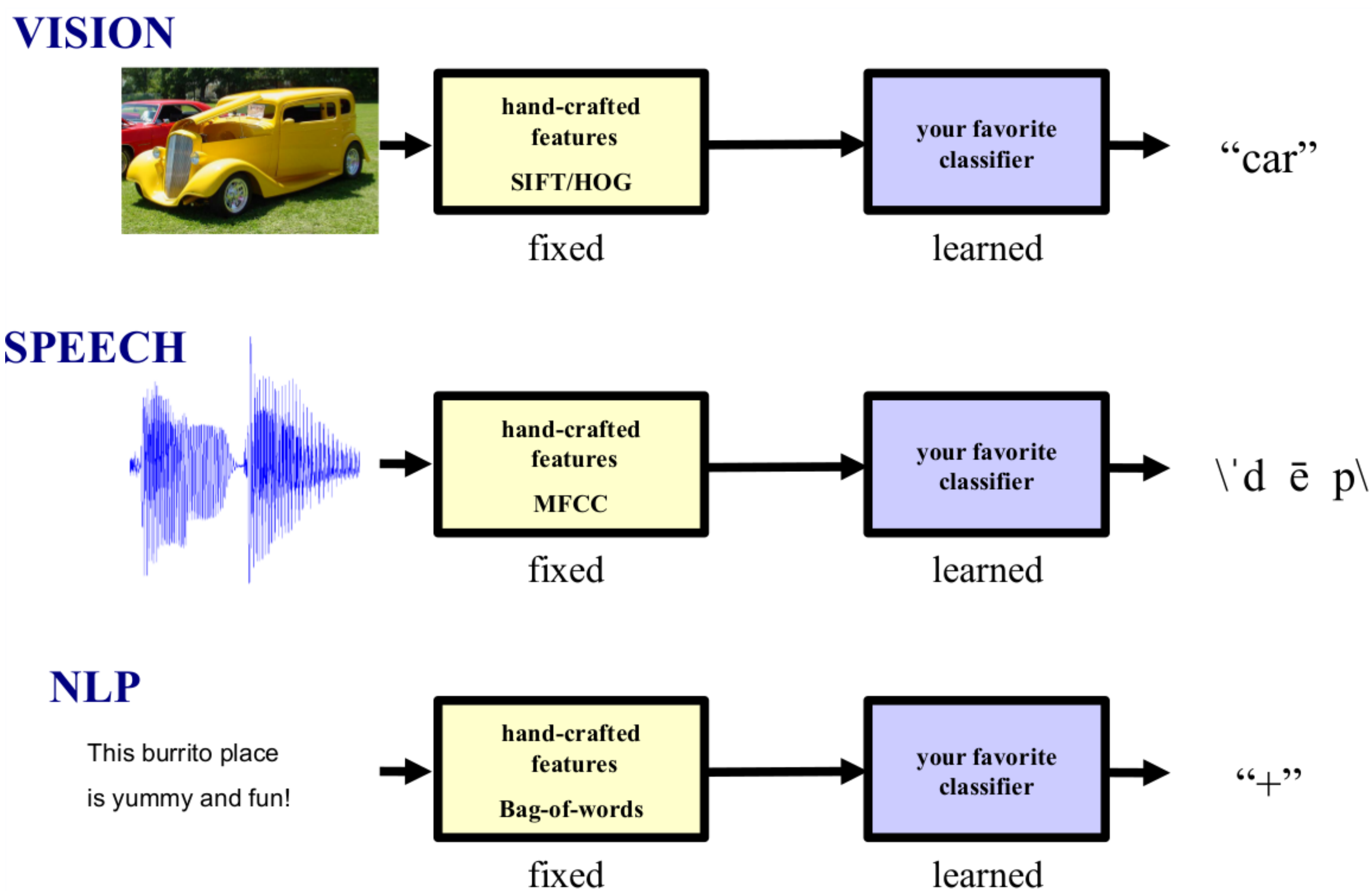
- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- Distributed Representations
  - No single neuron "encodes" everything
  - Groups of neurons work together

Different Stages of Machine Learning

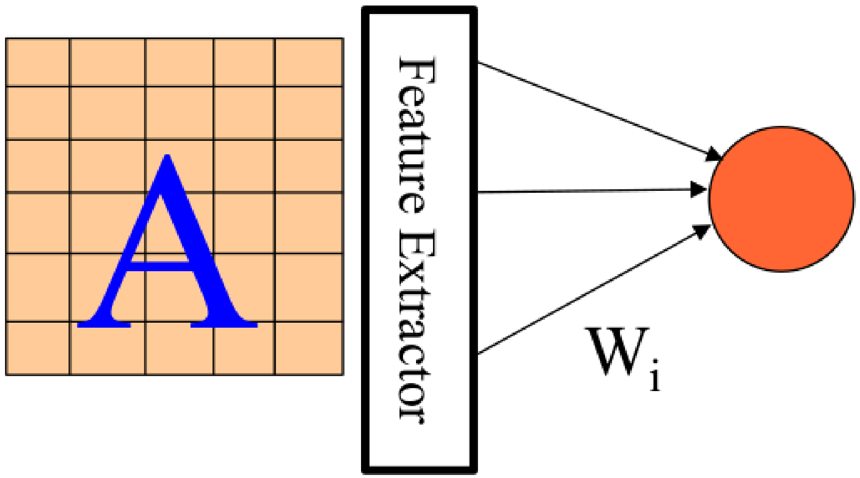


Classic Machine Learning



Old Paradigm

- The first learning machine: the Perceptron (~1960)
- The Perceptron was a linear classifier on top of a simple feature extractor
- The vast majority of practical applications of ML today use glorified linear classifiers or glorified template matching.
- **Designing a feature extractor requires considerable efforts by experts.**



$$y = sign\left(\sum_{i=1}^{N} W_i F_i(X) + b\right)$$

The Driving Ideas of Deep Learning

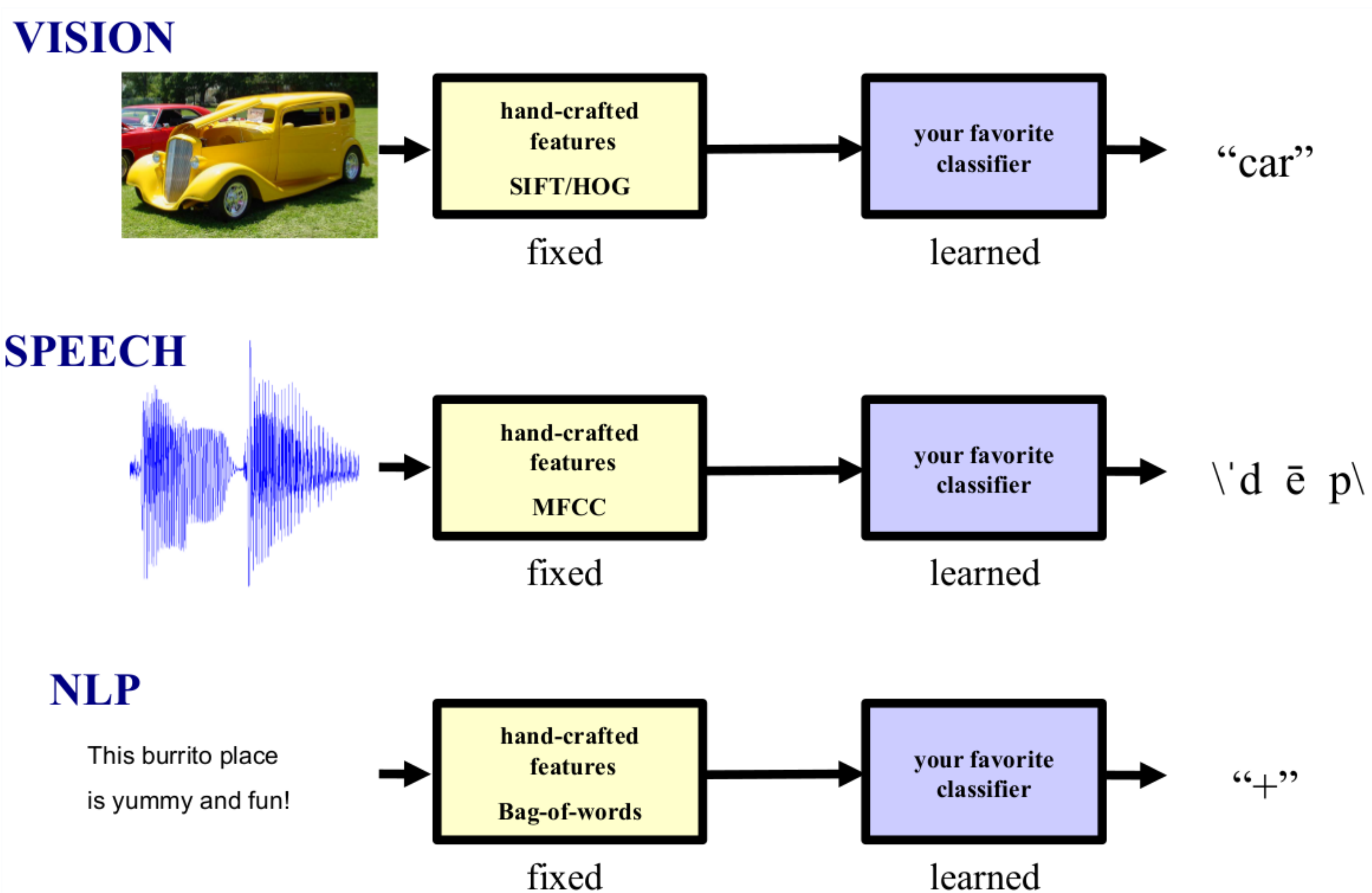
- Representation Learning using deep neural networks

- **End-to-End Learning**
  - **Learning (goal-driven) representations**
  - **Learning to feature extraction**

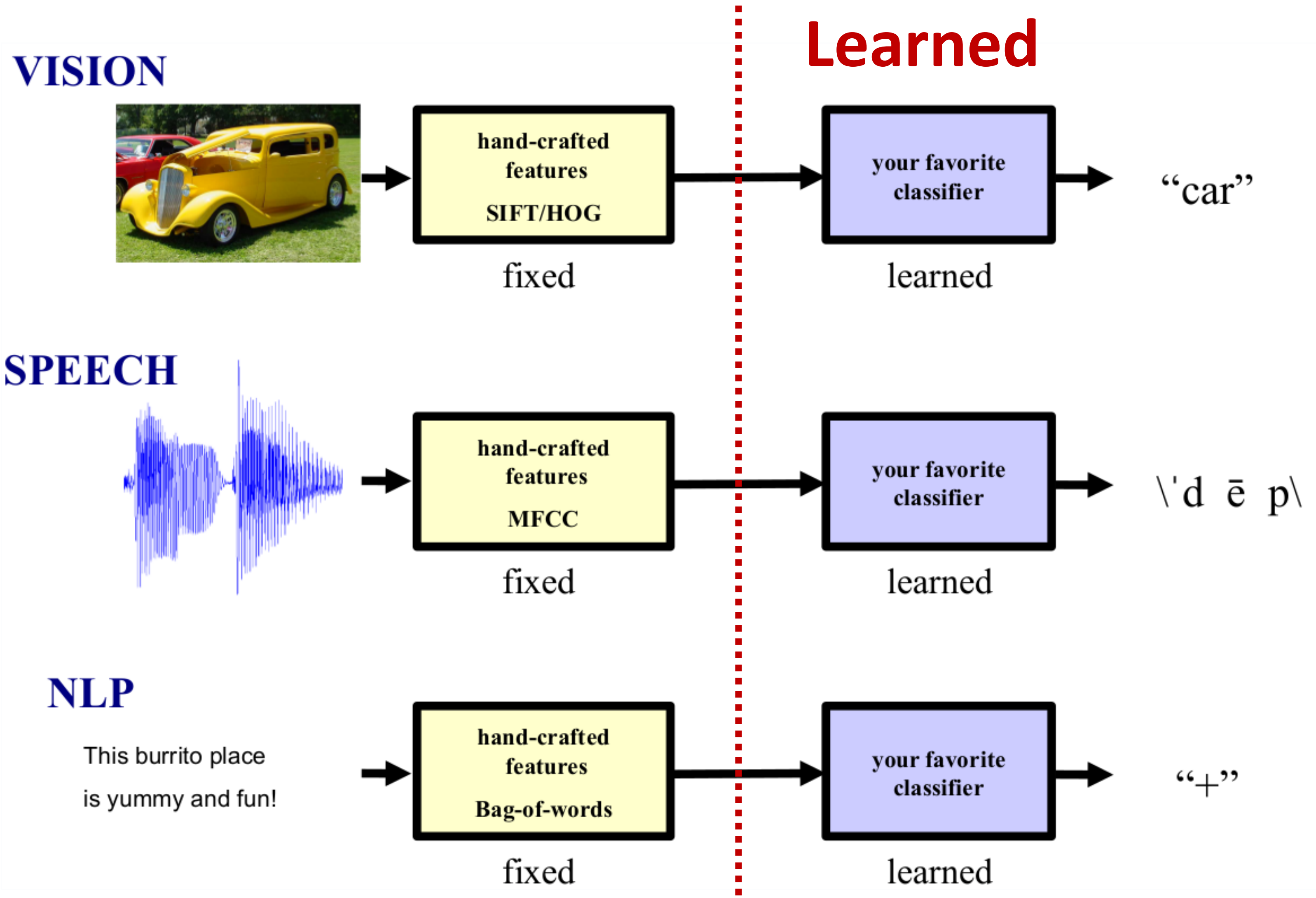
- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- Distributed Representations
  - No single neuron “encodes” everything
  - Groups of neurons work together

Traditional Machine Learning

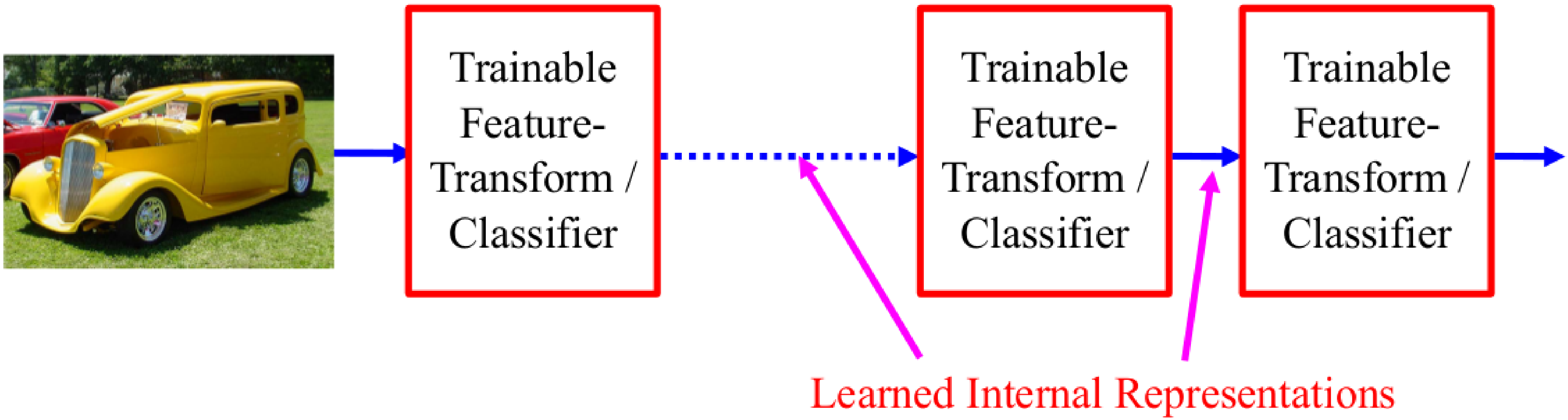


Traditional Machine Learning



Deep Learning = End-to-End Learning

- A hierarchy of trainable feature transforms
  - Each module transforms its input representation into a higher-level one.
  - High-level features are more global and more invariant.
  - Low-level features are shared among categories.



The Driving Ideas of Deep Learning

- Representation Learning using deep neural networks

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

- **(Hierarchical) Compositionality**
  - **Cascade of non-linear transformations**
  - **Multiple layers of representations**

- Distributed Representations
  - No single neuron "encodes" everything
  - Groups of neurons work together

Hierarchical Compositionality

VISION

pixels → edge → texton → motif → part → object

SPEECH

sample → spectral band → formant → motif → phone → word

NLP

character → word → NP/VP/.. → clause → sentence → story

Building A Complicated Function

Library of simple functions

$\boldsymbol{cos(x)}$

$\boldsymbol{x^n}$

$\boldsymbol{sin(x)}$

$\boldsymbol{log(x)}$

$\boldsymbol{e^x}$

Compose →

Building A Complicated Function

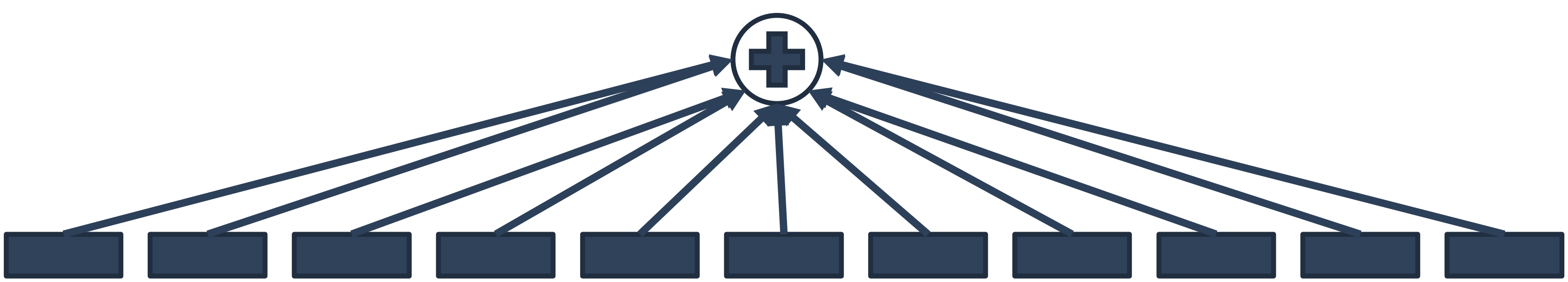
Library of simple functions

$\cos(x)$ x^n $\sin(x)$ $\log(x)$ e^x

Compose

- **Linear Combinations**
  - Boosting
  - Kernels
  - …

$$f(x) = \sum_i \alpha_i g_i(x)$$



Building A Complicated Function

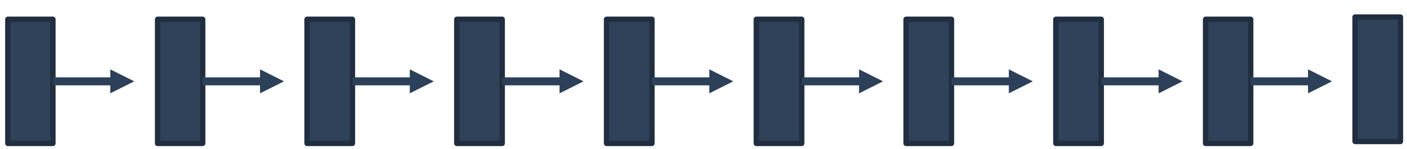
Library of
simple functions

x^n $cos(x)$ $sin(x)$ $log(x)$ e^x

Compose

- Compositions
  - Deep Learning
  - Grammar models
  - …

$$f(x) = g_1(g_2(\dots(g_n(x))\dots))$$



Building A Complicated Function

Library of simple functions

$cos(x)$

x^n

$sin(x)$

$log(x)$

e^x

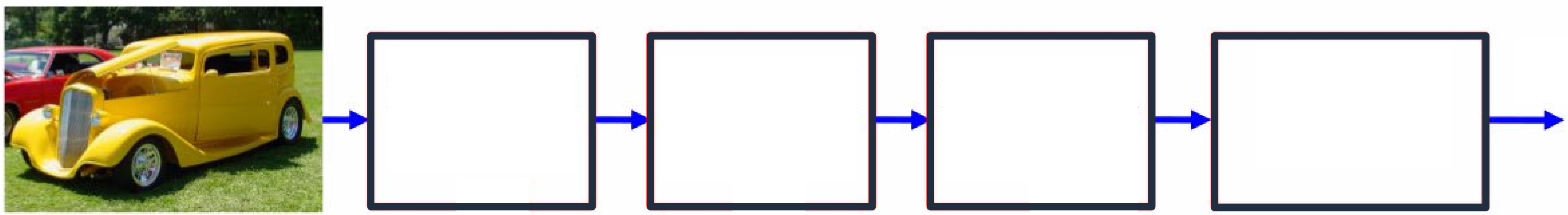
Compose

- Compositions
  - Deep Learning
  - Grammar models
  - …

$$f(x) = log\left(cos\left(exp\left(sin^3(x)\right)\right)\right)$$

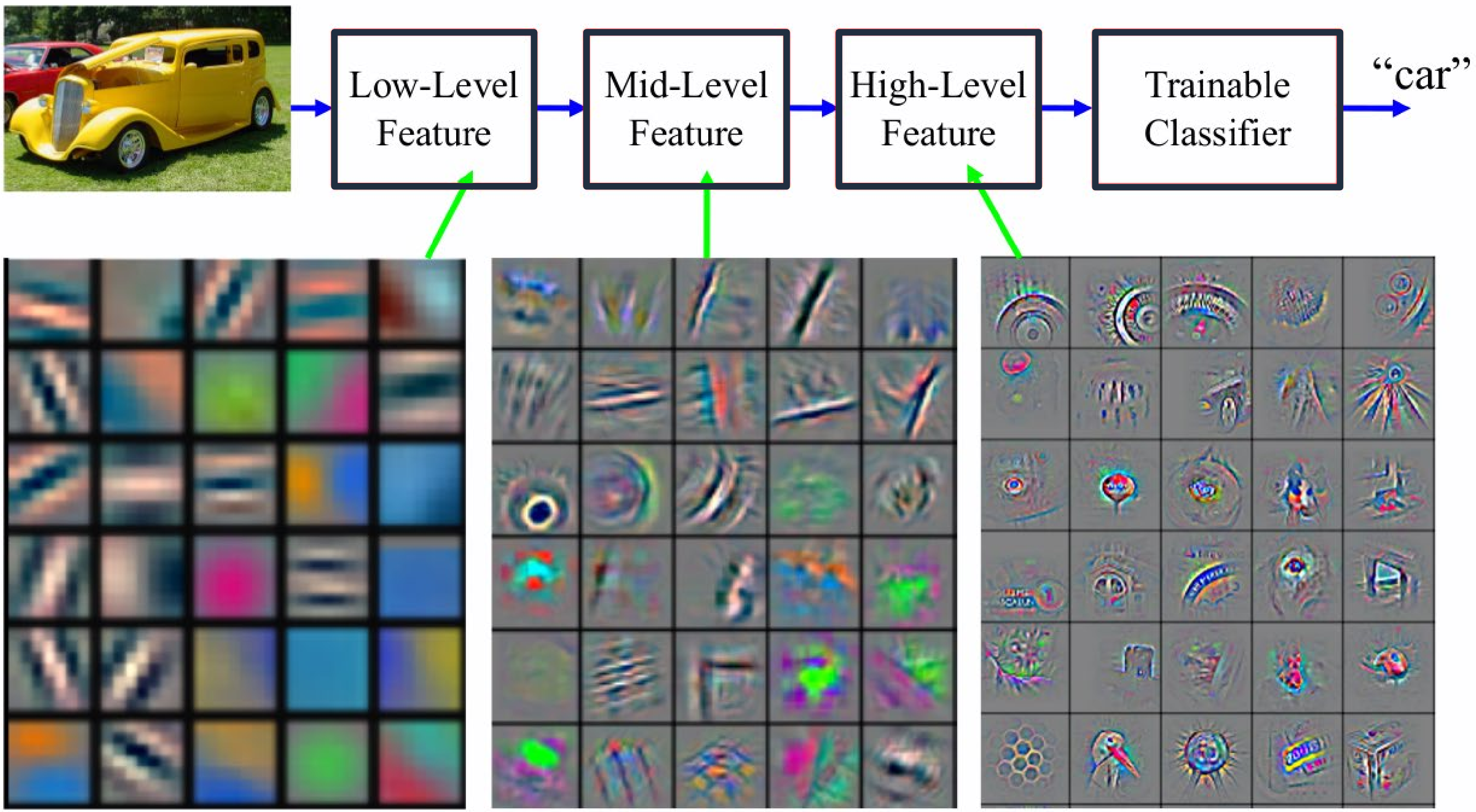
$$sin(x) \rightarrow x^3 \rightarrow e^x \rightarrow cos(x) \rightarrow log(x)$$

Hierarchical Composition



➢ Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Hierarchical Composition



➢ Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

The Driving Ideas of Deep Learning

- Representation Learning using deep neural networks

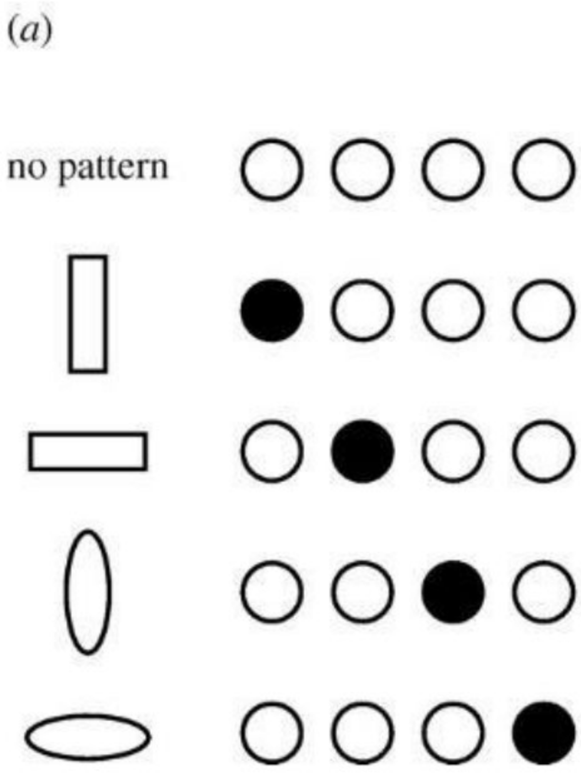
- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- **Distributed Representations**
  - **No single neuron “encodes” everything**
  - **Groups of neurons work together**

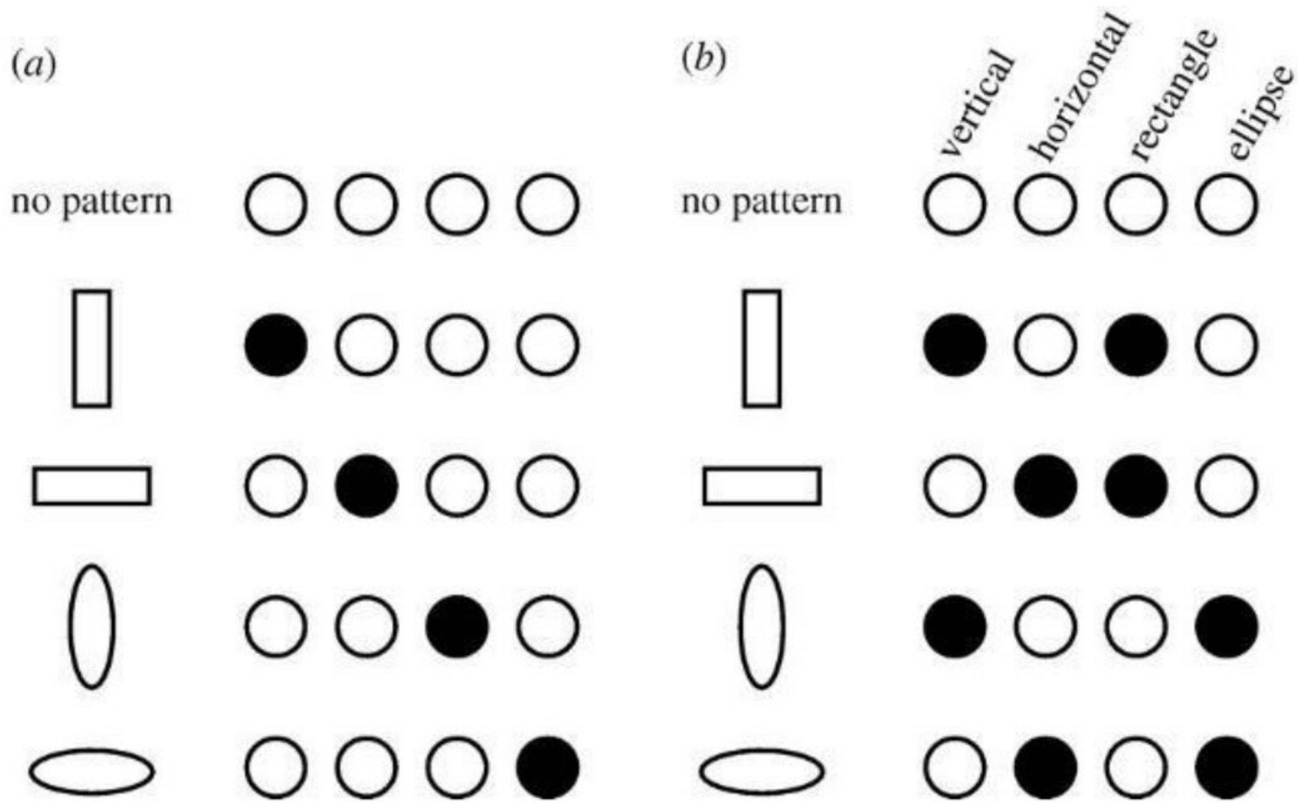
Distributed Representations

- Local

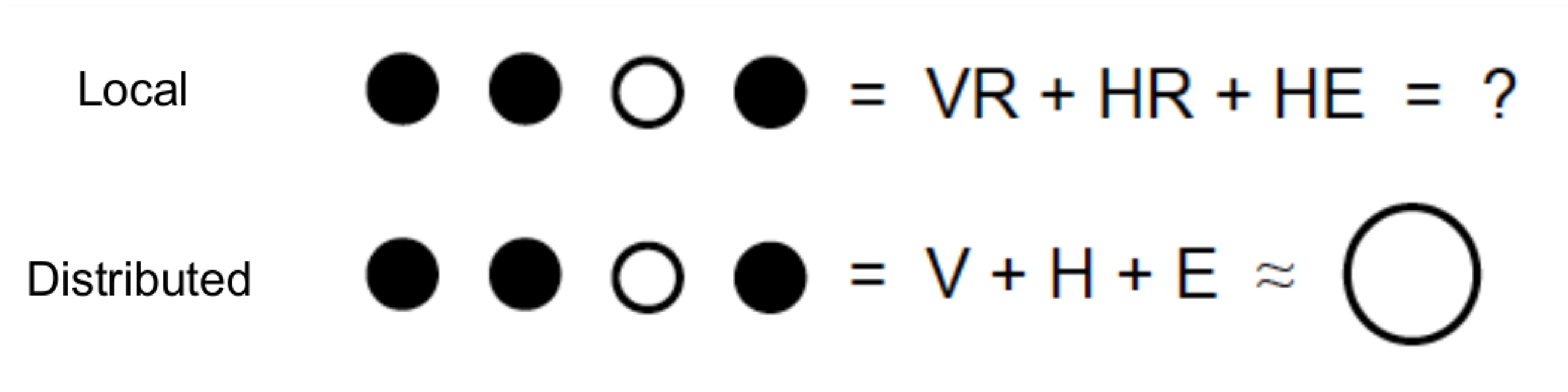


Distributed Representations

- Local vs. Distributed
- Can we interpret each dimension?

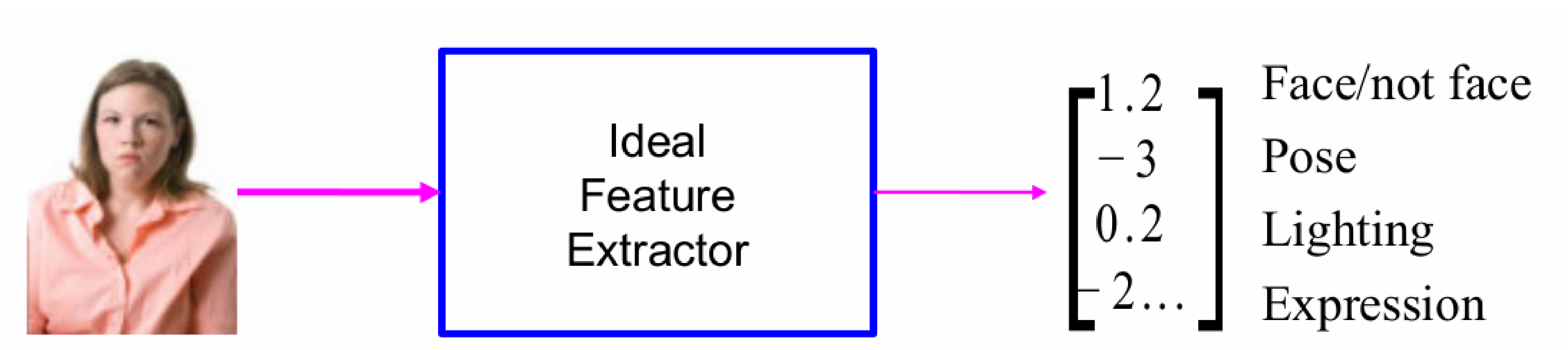


Power of Distributed Representations!

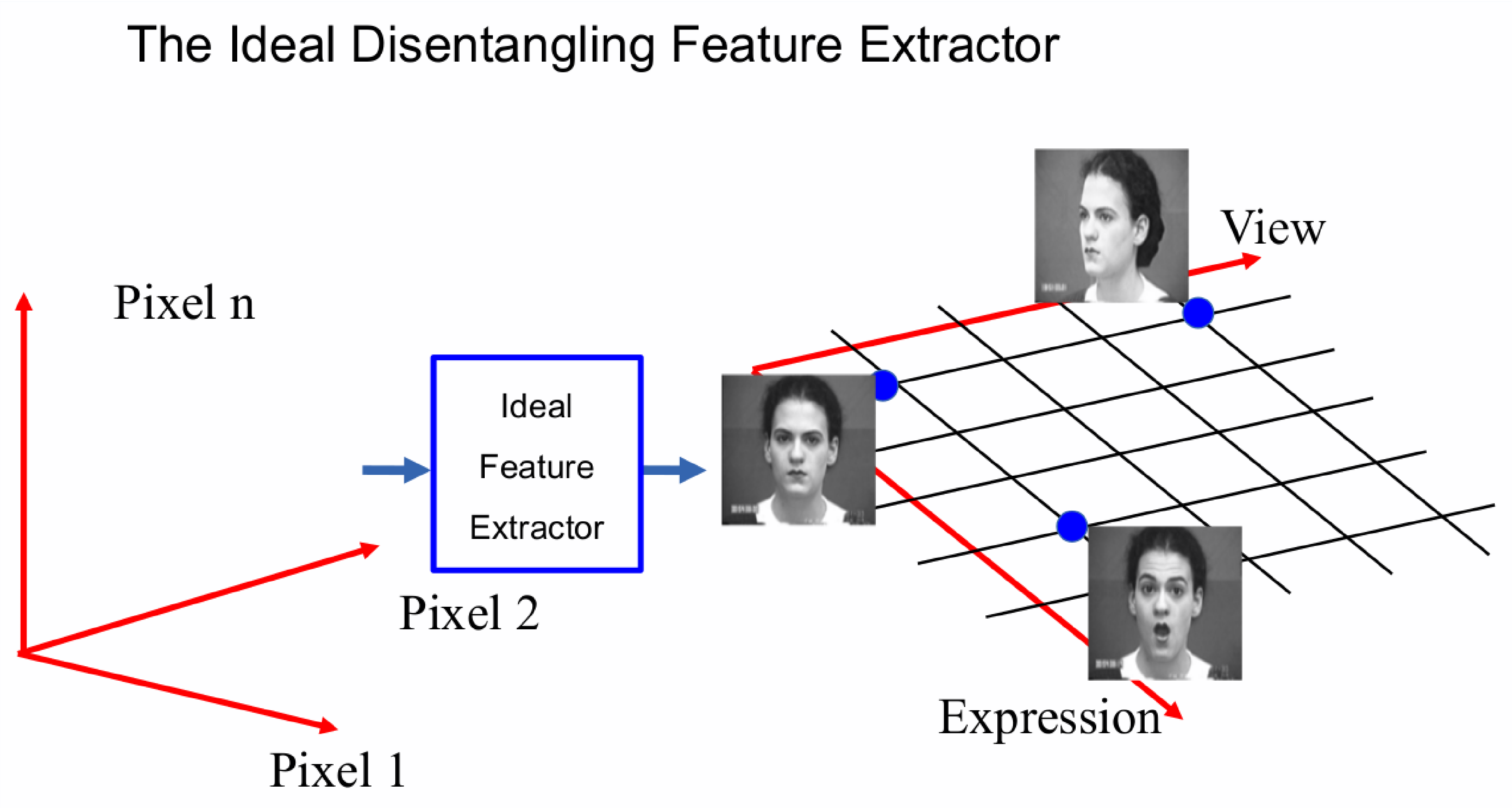


Power of Distributed Representations!

- Example: all face images of a person
  - 1000x1000 pixels = 1,000,000 dimensions
  - But the face has 3 cartesian coordinates and 3 Euler angles
  - And humans have less than about 50 muscles in the face
  - Hence the manifold of face images for a person has <56 dimensions
- The perfect representations of a face image:
  - Its coordinates on the face manifold
  - Its coordinates away from the manifold



Power of Distributed Representations!



Representation Matters