DSE 3121 DEEP LEARNING [4 0 0 4]

Slide1: Introduction + History of Neural Networks & Deep learning

B.Tech Data Science & Engineering

Rohini R Rao & Abhilash Pai
Department of Data Science and Computer Applications
MIT Manipal

Course Outcomes

CO1: Design, optimize and evaluate shallow and deep neural networks to solve real world problems.

CO2: Design, optimize and evaluate Convolutional Neural Networks for Image Classification.

CO3: Design, optimize and evaluate Recurrent Neural Networks for sequence modelling tasks like Time series prediction and NLP.

CO4: Design and evaluate Autoencoders and Generative models for representational learning.

References

1. Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning*, MIT Press 2016.
2. Charu C. Aggarwal, Neural Networks and Deep Learning, Springer 2018.
3. Course Notes – Neural Network and Deep Learning, Dr. Andrew NG
4. Course Notes- Deep Learning IIT Madras, Prof Dr. Mithesh Khapra
5. Course Notes- Deep Learning IIT Ropar, Prof Dr. Sudharshan Iyengar
6. Aurelien Geron, "Hands-On Machine Learning with Scikit-Learn , Keras & Tensorflow, OReilly Publications

Credits

The contents for this slide set was adapted from the following sources:

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

2. Aurelien Geron, “Hands-On Machine Learning with Scikit-Learn , Keras & Tensorflow, OReilly Publications

3. Course Notes- Deep Learning IIT Madras, Prof Dr. Mithesh Khapra

4. Course Notes- Deep Learning University of Tübingen, Prof. Dr. Andreas Geiger

Introduction to AI

- Evolution of AI
  - Goal: To create machines that think and perceive the world like humans.
  - Initially solved problems that can be described by a list of formal mathematical rules.
    - Ex- Playing Chess

  - Challenge in solving tasks that are easy for people to perform but difficult for people to describe formally (problems that we solve intuitively)
    - Ex- Identifying words, Recognizing people in images

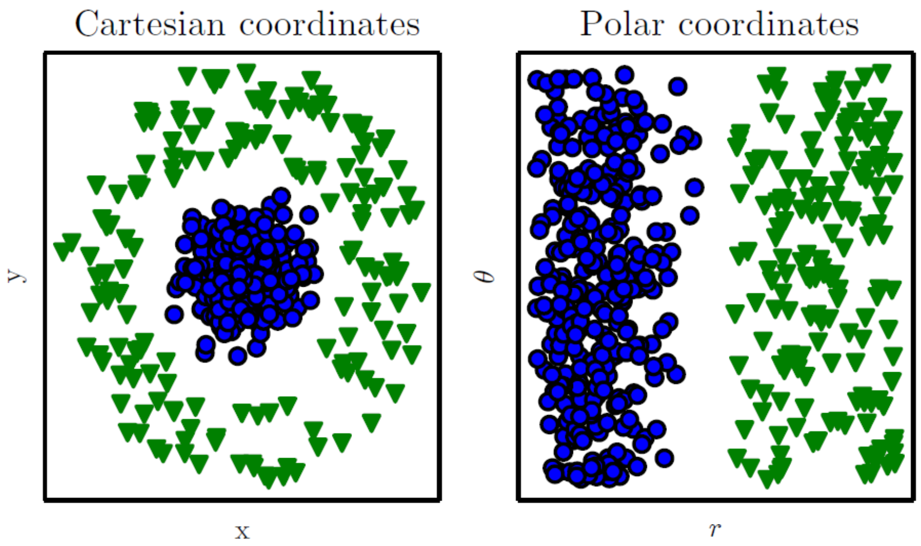
  - How to get the informal knowledge into a computer?
    - **Knowledge base** approach
      - Hard code knowledge in formal languages
      - Computers can reason about statements automatically using logical inference rules
        - Ex- Cyc project
      - However, none of these projects has led to a major success and fail model intricate aspects of the world.

Machine Learning

- AI systems need the ability to acquire their own knowledge, by extracting patterns from raw data. This capability is known as **machine learning.**

- Machine learning algorithms depends heavily on the **representation** (collection of features).

- For many tasks, it is difficult to know what features should be extracted.



Representations matter

Deep Learning

- Solution is to use machine learning to discover not only the mapping from representation to output but also the representation itself.
- This approach is known as **representation learning**.

- Learned representation often produces better performance in comparison to the hand-designed representations.

- They also allow AI systems to rapidly adapt to new tasks, with minimal human intervention.

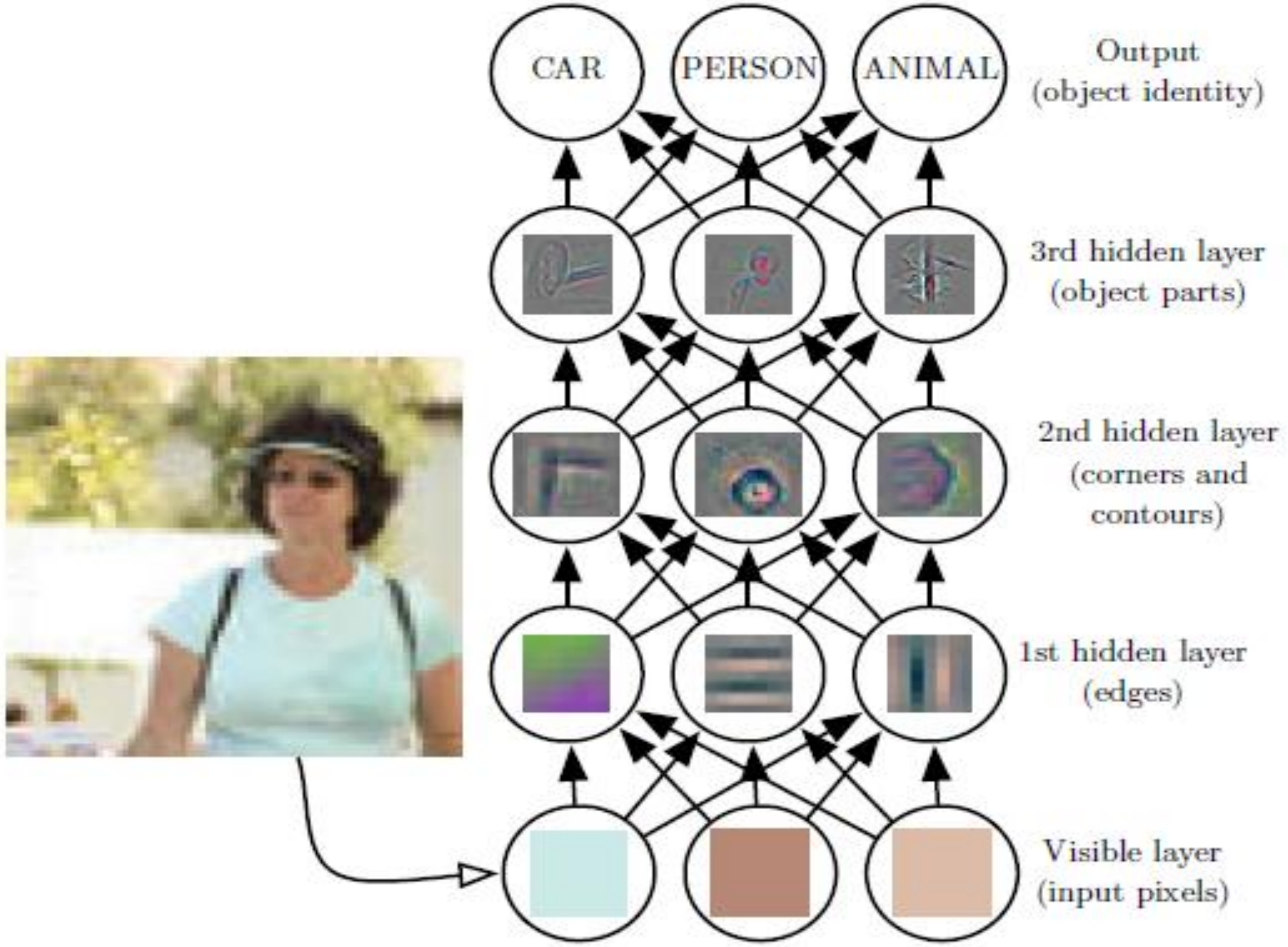
Deep Learning

- While designing algorithms for learning features the goal is usually to separate the **factors of variation.**
  - Unobserved objects or unobserved forces in the physical world that affect observable quantities.
  - They may also exist as constructs in the human mind that provide useful simplifying explanations or inferred causes of the observed data.
    - Ex: In voice data – speaker’s accent, gender, age

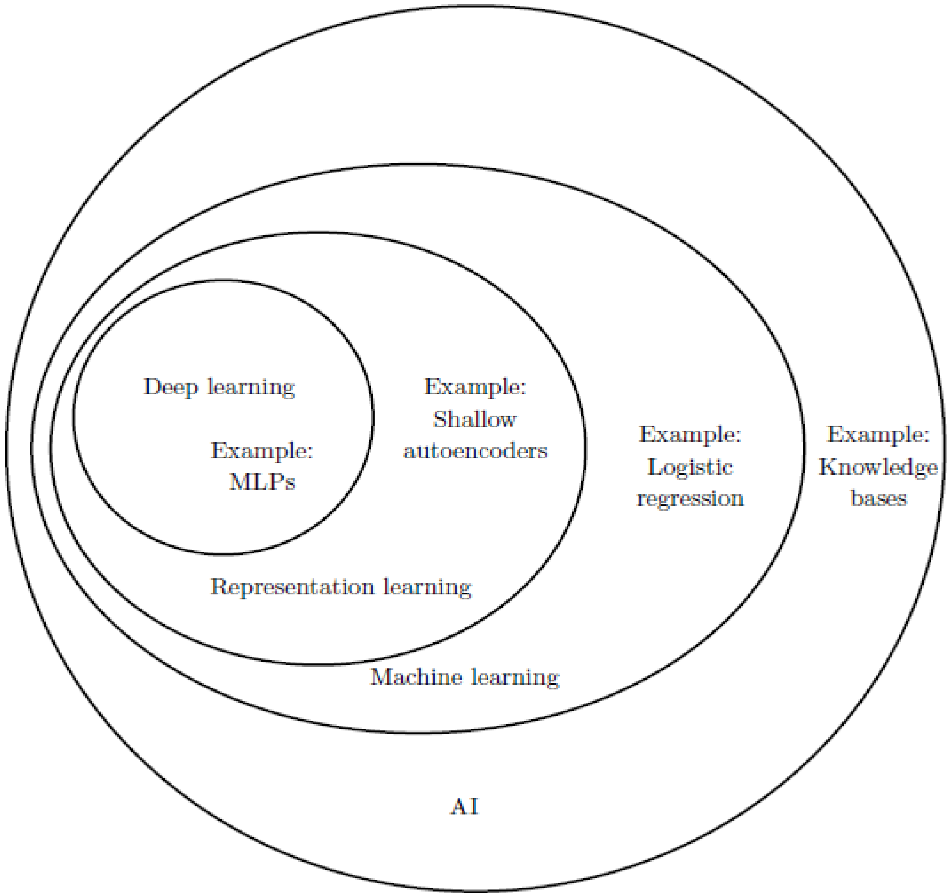
- However, many factors of variation influence every single piece of data we observe.
- Also, it can be very difficult to extract such high-level, abstract features from raw data.

- **Deep learning** solves this central problem in representation learning by introducing representations that are expressed in terms of other, simpler representations.

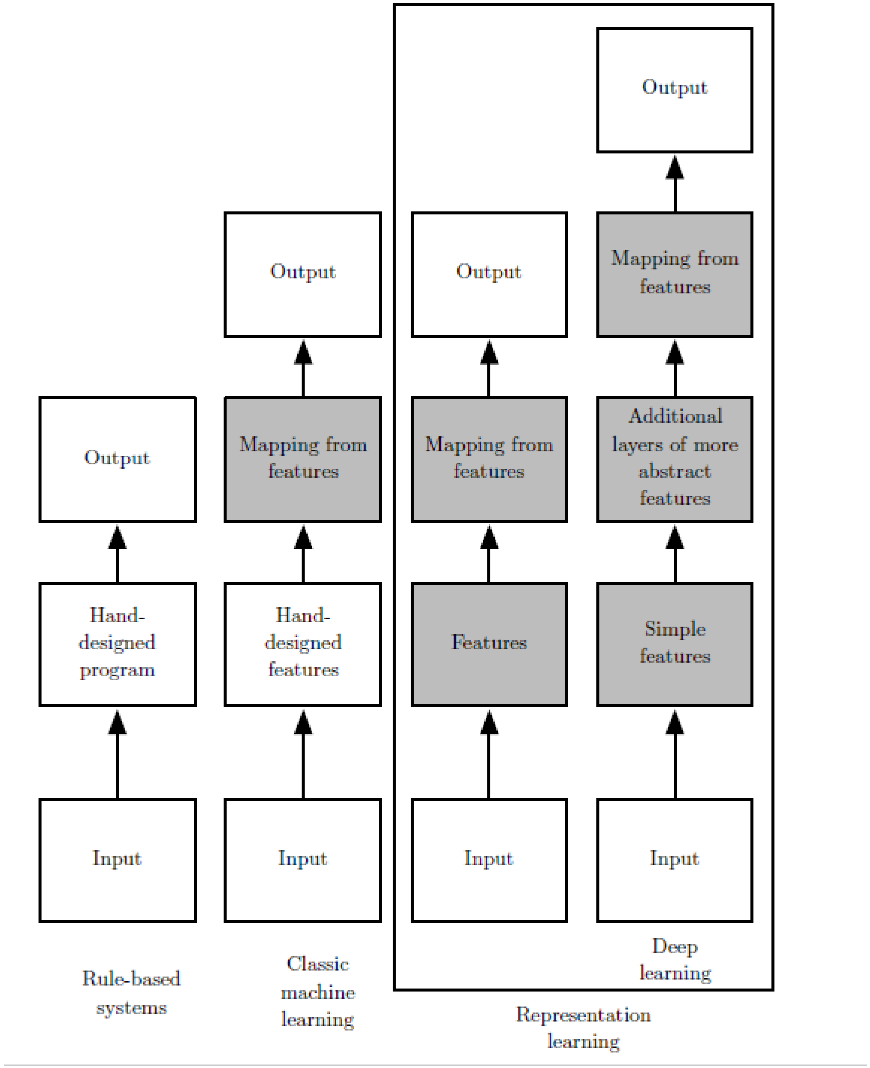
Depth: Repeated Composition



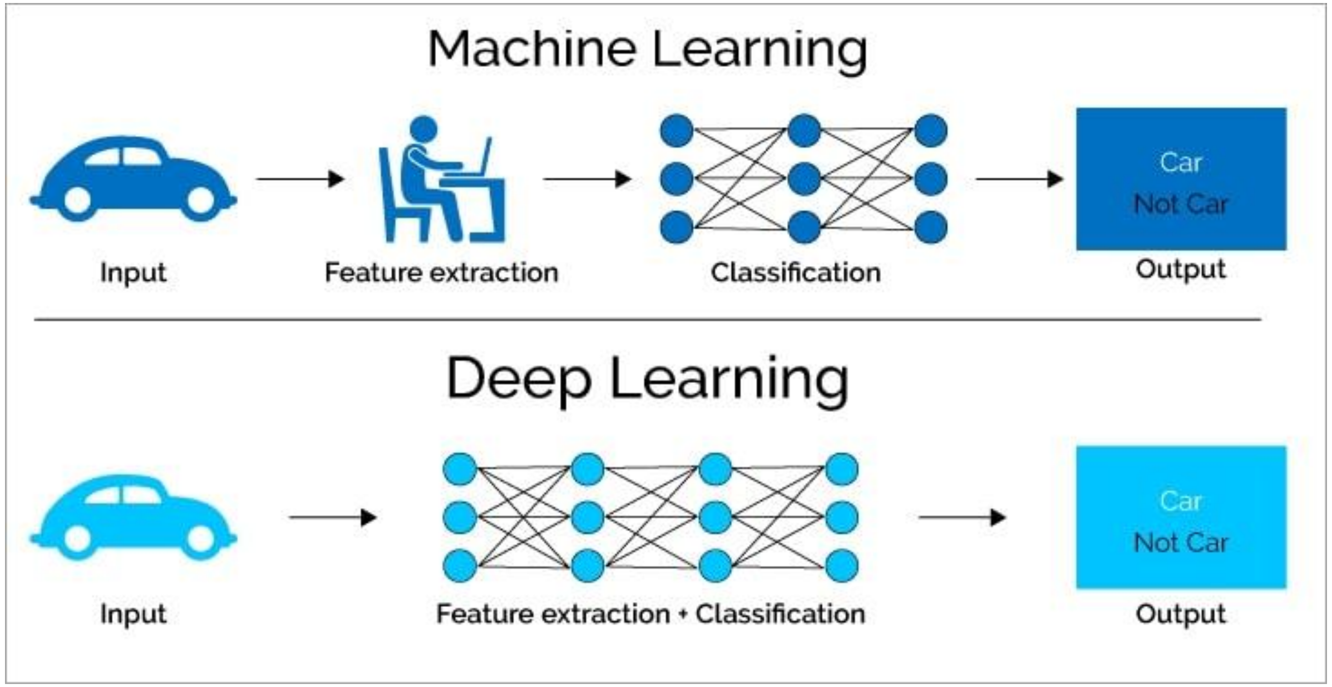
Deep Learning, Machine Learning and AI



Learning Multiple Components



Deep Learning and Machine Learning

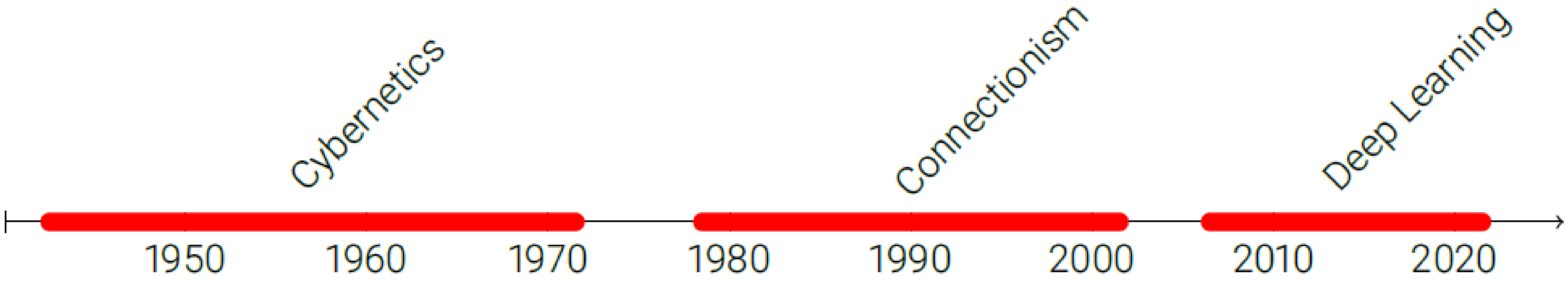


(Source: softwaretestinghelp.com)

A Brief History of Deep Learning

Three waves of development:

- 1940-1970: “Cybernetics” (Golden Age)
  - Simple computational models of biological learning, simple learning rules
- 1980-2000: “Connectionism” (Dark Age)
  - Intelligent behavior through large number of simple units, Backpropagation
- 2006-now: “Deep Learning” (Revolution Age)
  - Deeper networks, larger datasets, more computation, state-of-the-art in many areas



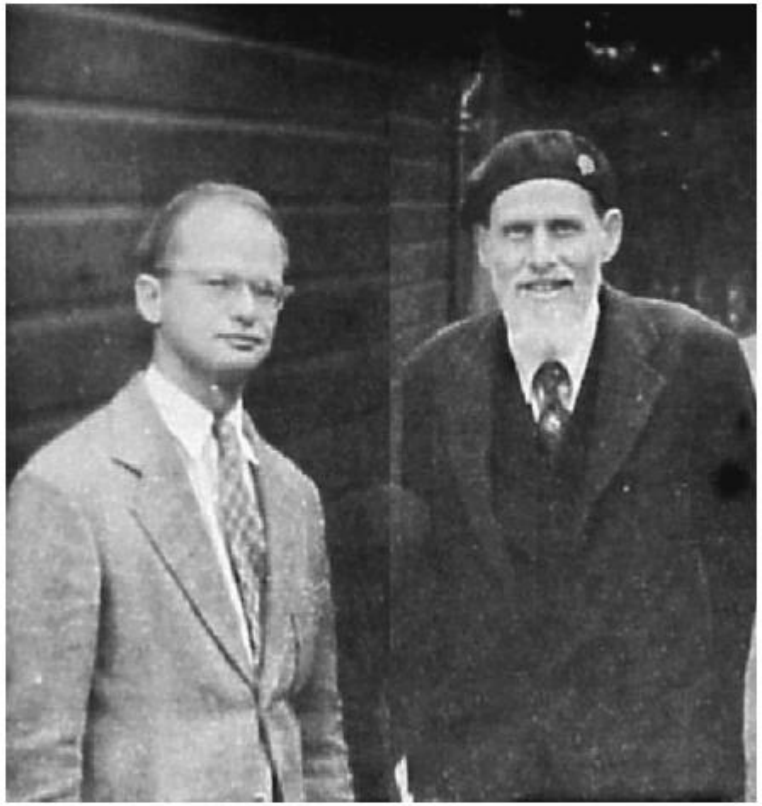
A Brief History of Deep Learning

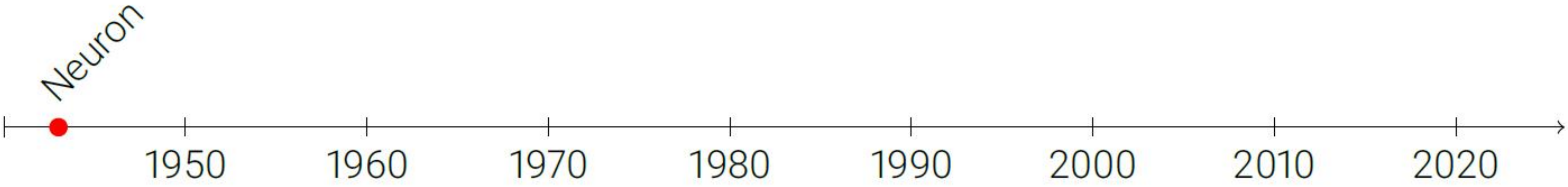
1943: McCullock and Pitts

- Early model for neural activation
- Linear threshold neuron (binary):

$$
\begin{aligned}
y = f(g(\mathbf{x})) &= 1 \quad if \quad g(\mathbf{x}) \geq \theta \\
&= 0 \quad if \quad g(\mathbf{x}) < \theta
\end{aligned}
$$

- More powerful than AND/OR gates
- But no procedure to learn weights



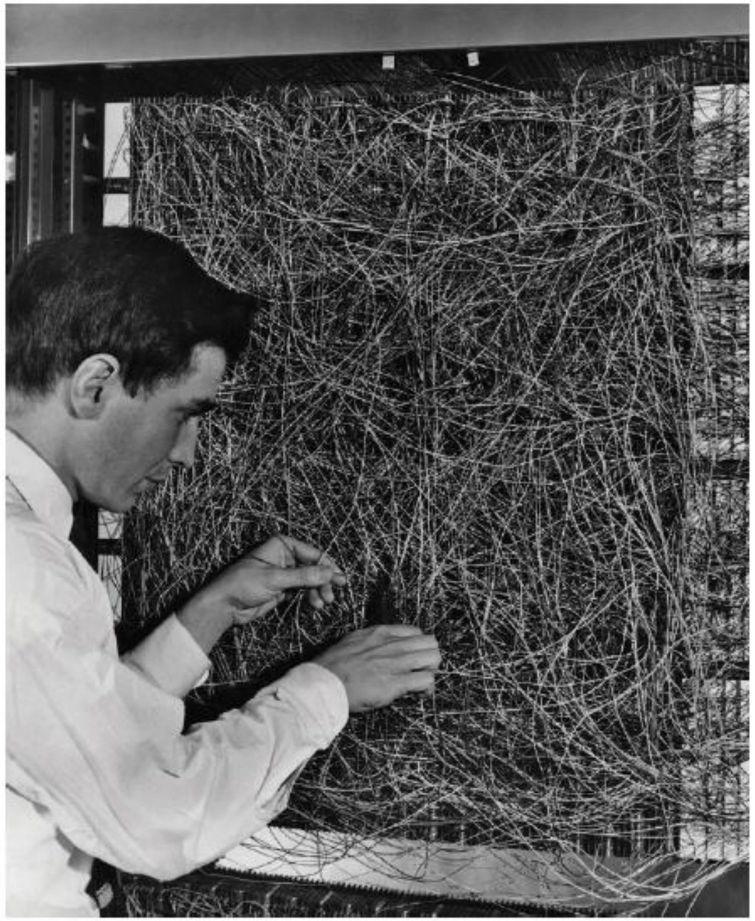


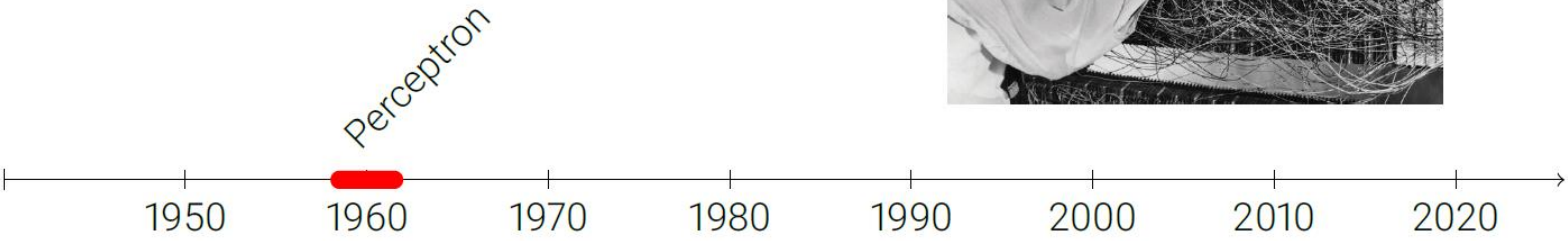
McCulloch and Pitts: A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 1943.

A Brief History of Deep Learning

1958-1962: Rosenblatt's Perceptron

- First algorithm and implementation to train single linear threshold neuron
- Optimization of perceptron criterion:
- Novikoff proved convergence



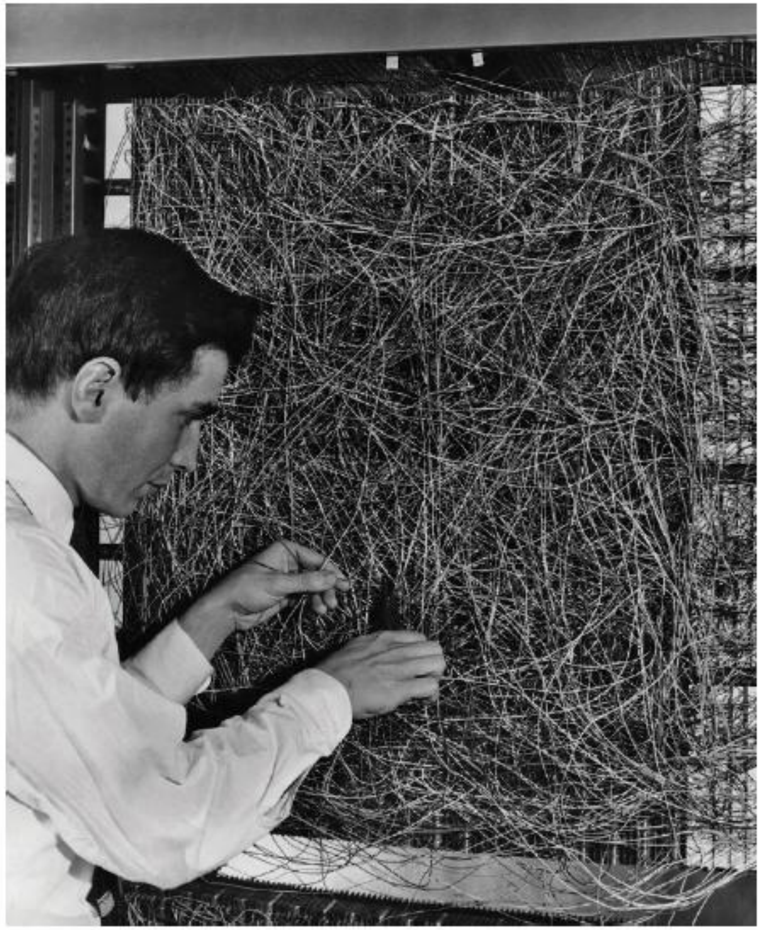


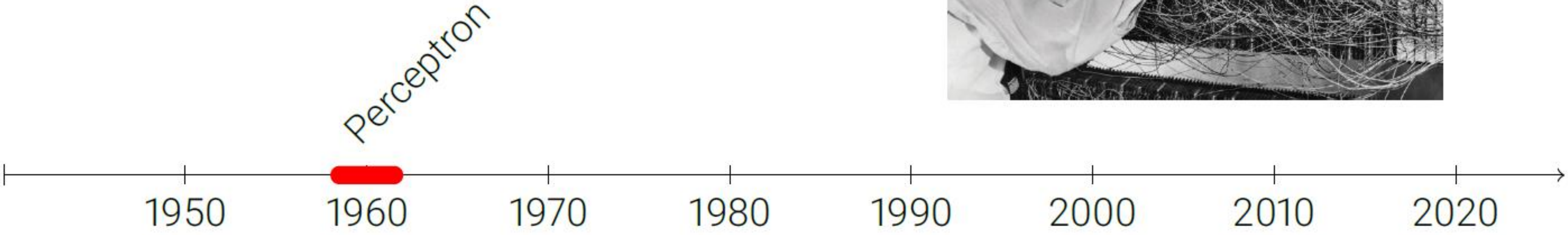
Rosenblatt: The perceptron - a probabilistic model for information storage and organization in the brain. Psychological Review, 1958.

A Brief History of Deep Learning

1958-1962: Rosenblatt's Perceptron

- First algorithm and implementation to train single linear threshold neuron
- Overhyped: Rosenblatt claimed that the perceptron will lead to computers that walk, talk, see, write, reproduce and are conscious of their existence



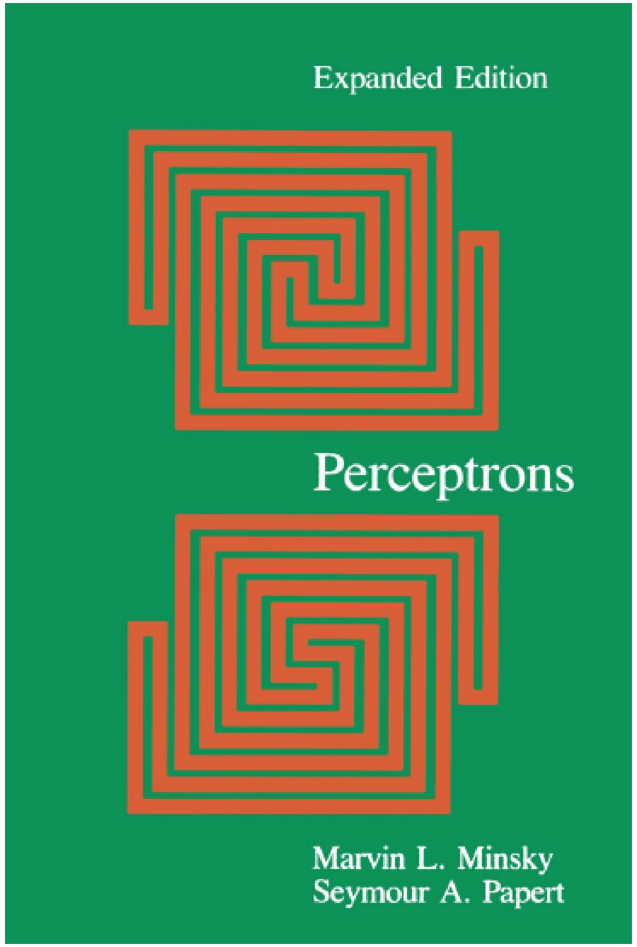


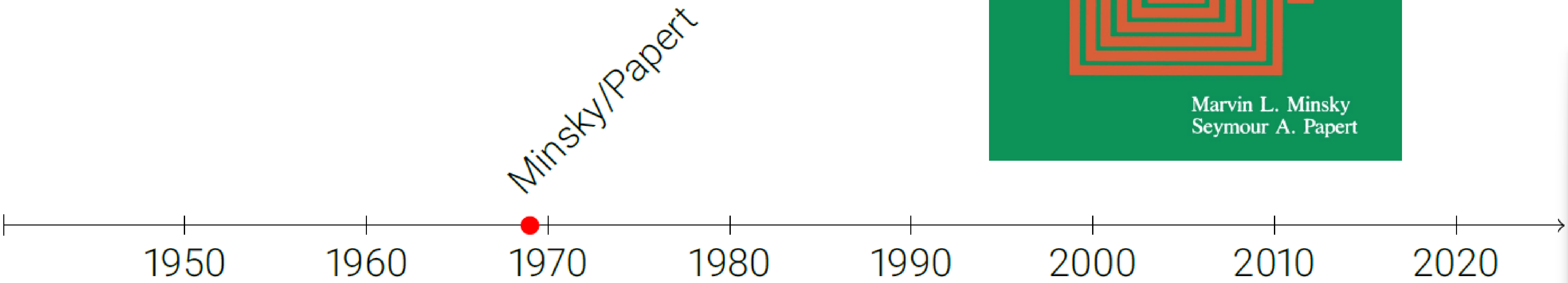
Rosenblatt: The perceptron - a probabilistic model for information storage and organization in the brain. Psychological Review, 1958.

A Brief History of Deep Learning

1969: Minsky and Papert publish book

- Several discouraging results
- Showed that single-layer perceptrons cannot solve some very simple problems (XOR problem, counting)
- Symbolic AI research dominates 70s



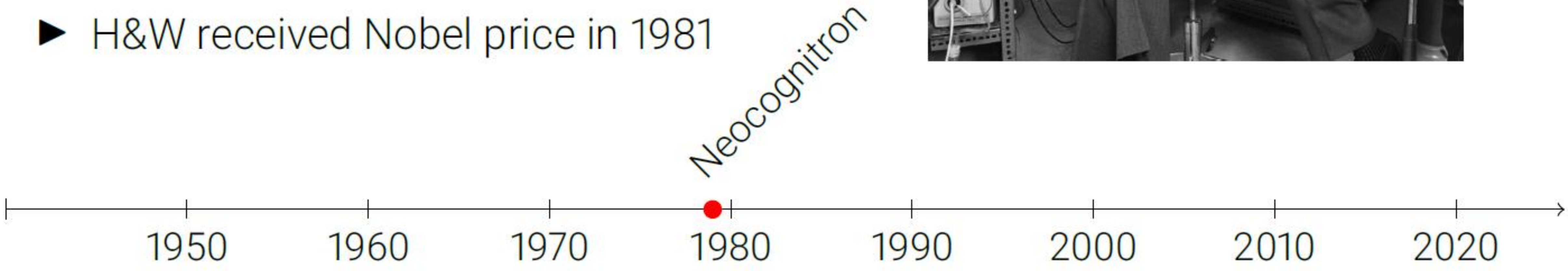


Minsky and Papert: Perceptrons: An introduction to computational geometry. MIT Press, 1969.

A Brief History of Deep Learning

1979: Fukushima's Neocognitron

- Inspired by Hubel and Wiesel experiments in the 1950s
- Study of visual cortex in cats
- Found that cells are sensitive to orientation of edges but insensitive to their position (simple vs. complex)
- H&W received Nobel price in 1981

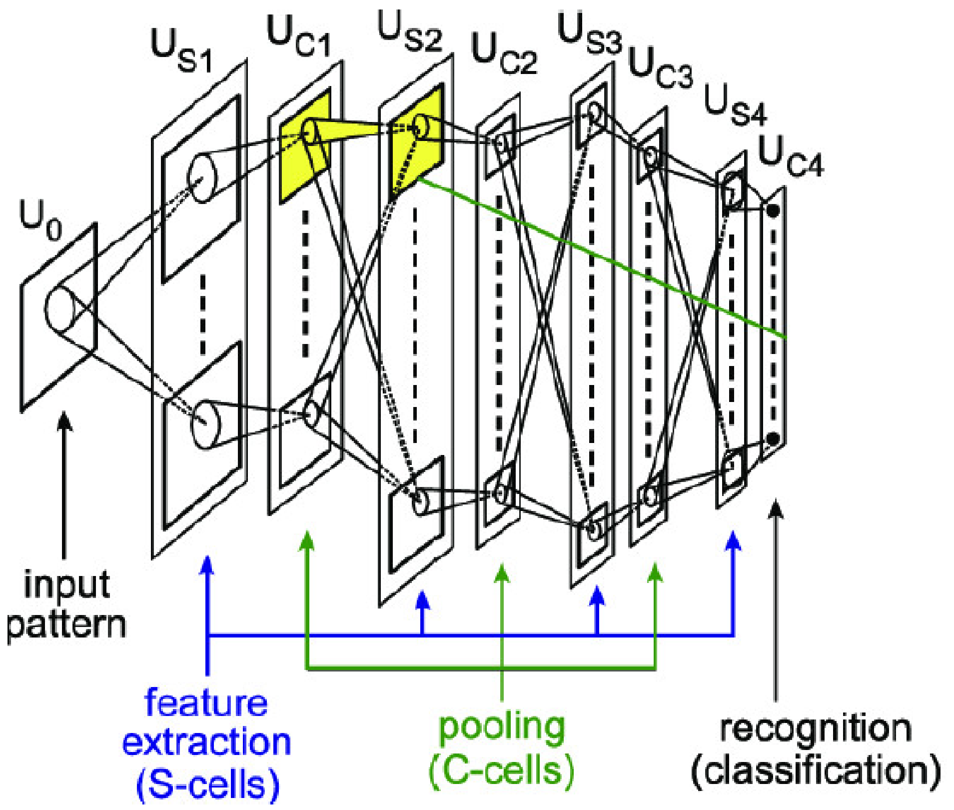


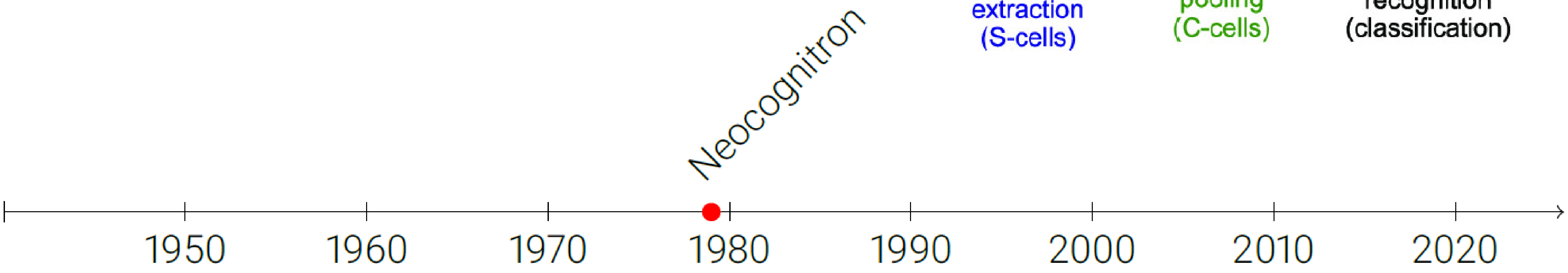
Fukushima: Neural network model for a mechanism of pattern recognition unaffected by shift in position. IECE (in Japanese), 1979.

A Brief History of Deep Learning

1979: Fukushima's Neocognitron

- Multi-layer processing
  to create intelligent behavior
- Simple (S) and complex (C) cells
  implement convolution and pooling
- Reinforcement based learning
- Inspiration for modern ConvNets



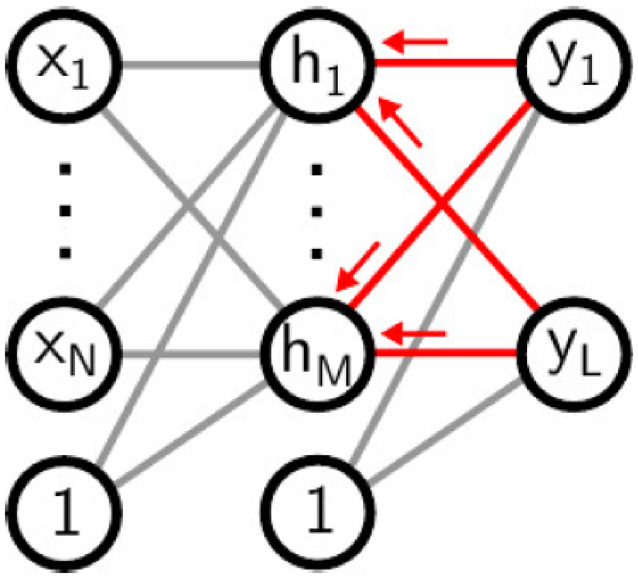


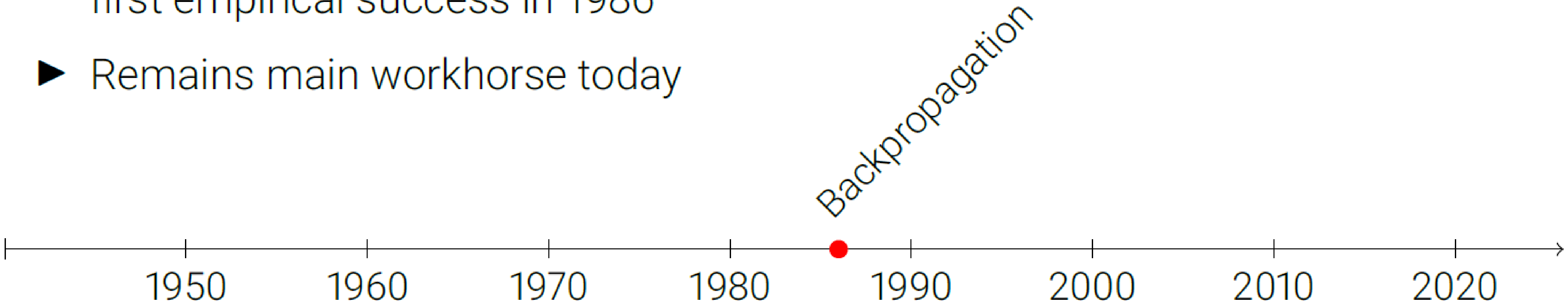
Fukushima: Neural network model for a mechanism of pattern recognition unaffected by shift in position. IECE (in Japanese), 1979.

A Brief History of Deep Learning

1986: Backpropagation Algorithm

- Efficient calculation of gradients in a deep network wrt. network weights
- Enables application of gradient based learning to deep networks
- Known since 1961, but first empirical success in 1986
- Remains main workhorse today





Rumelhart, Hinton and Williams: Learning representations by back-propagating errors. Nature, 1986.

A Brief History of Deep Learning

1997: Long Short-Term Memory

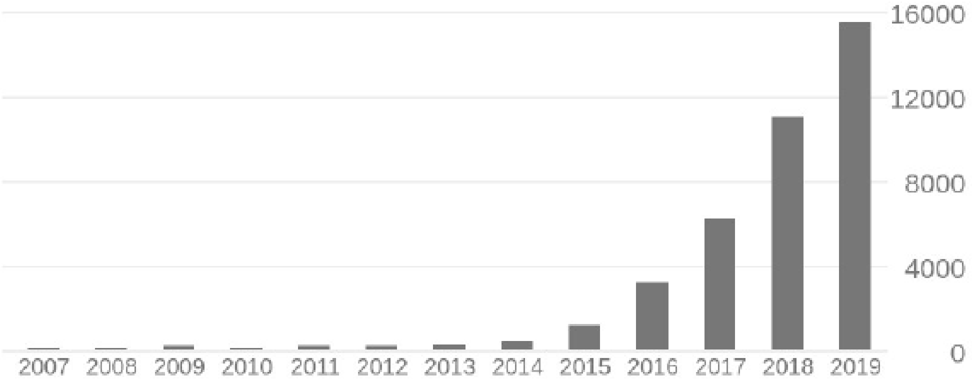
- In 1991, Hochreiter demonstrated the problem of vanishing/exploding gradients in his Diploma Thesis
- Led to development of long-short term memory for sequence modeling
- Uses feedback and forget/keep gate

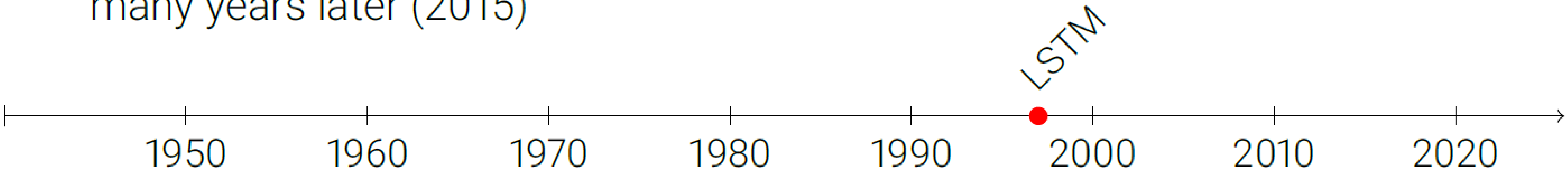
Hochreiter, Schmidhuber: Long short-term memory. Neural Computation, 1997.

A Brief History of Deep Learning

1997: Long Short-Term Memory

- In 1991, Hochreiter demonstrated the problem of vanishing/exploding gradients in his Diploma Thesis
- Led to development of long-short term memory for sequence modeling
- Uses feedback and forget/keep gate
- Revolutionized NLP (e.g. at Google) many years later (2015)



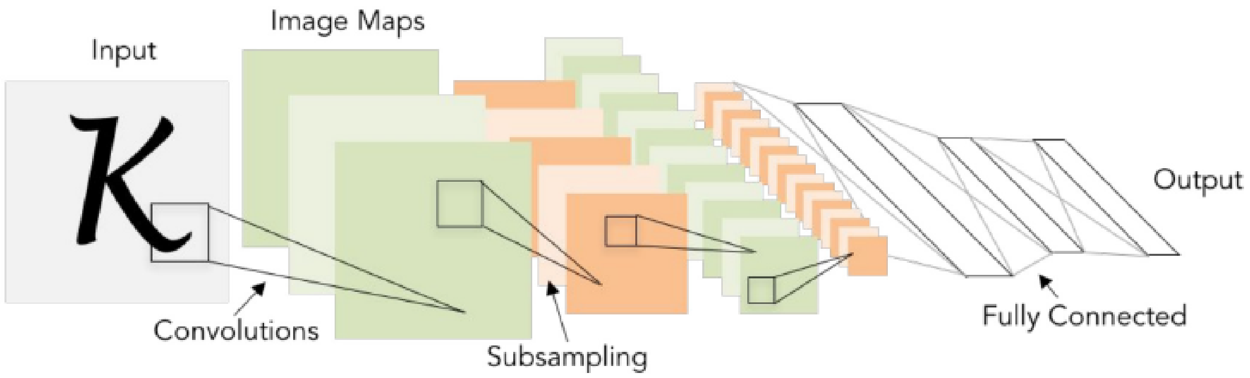


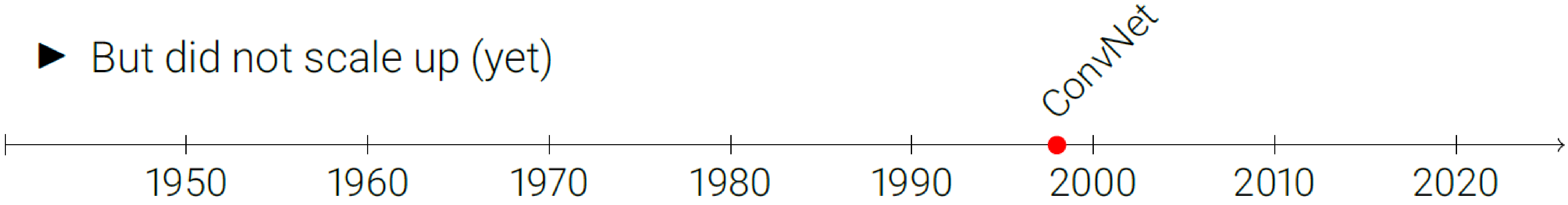
Hochreiter, Schmidhuber: Long short-term memory. Neural Computation, 1997.

A Brief History of Deep Learning

1998: Convolutional Neural Networks

- Similar to Neocognitron, but trained end-to-end using backpropagation
- Implements spatial invariance via convolutions and max-pooling
- Weight sharing reduces parameters
- Tanh/Softmax activations
- Good results on MNIST
- But did not scale up (yet)





LeCun, Bottou, Bengio, Haffner: Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998.

A Brief History of Deep Learning

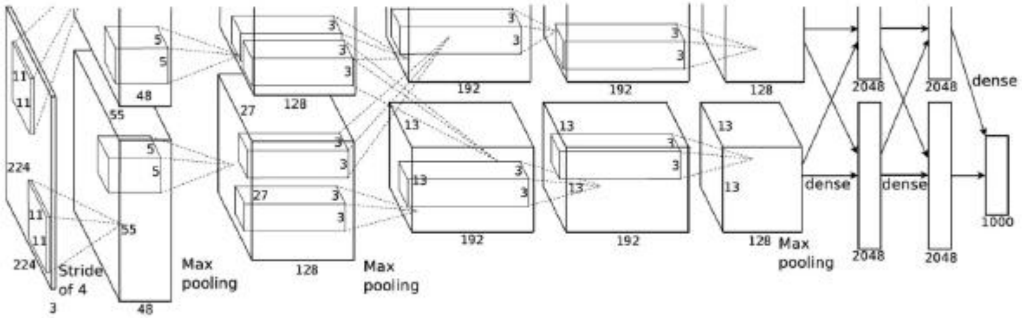
2009-2012: ImageNet and AlexNet

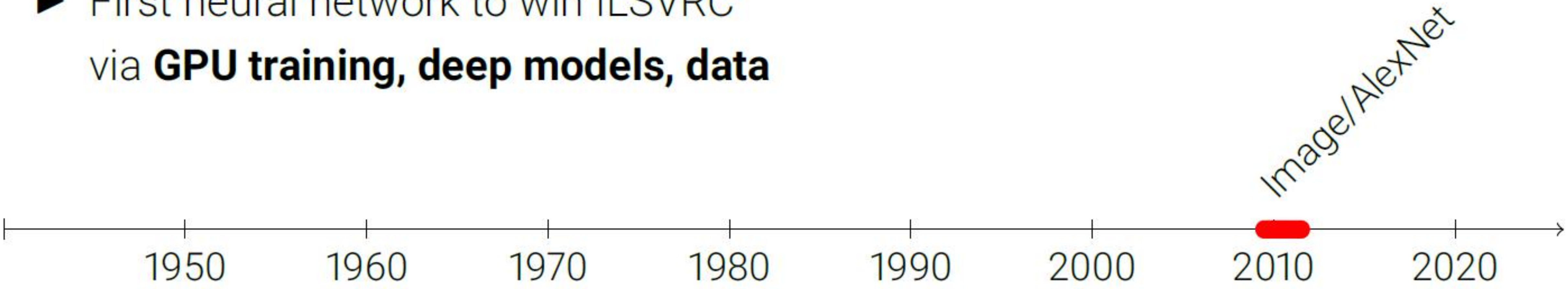
ImageNet

- Recognition benchmark (ILSVRC)
- 10 million annotated images
- 1000 categories

AlexNet

- First neural network to win ILSVRC via **GPU training, deep models, data**





Krizhevsky, Sutskever, Hinton. ImageNet classification with deep convolutional neural networks. NIPS, 2012.

A Brief History of Deep Learning

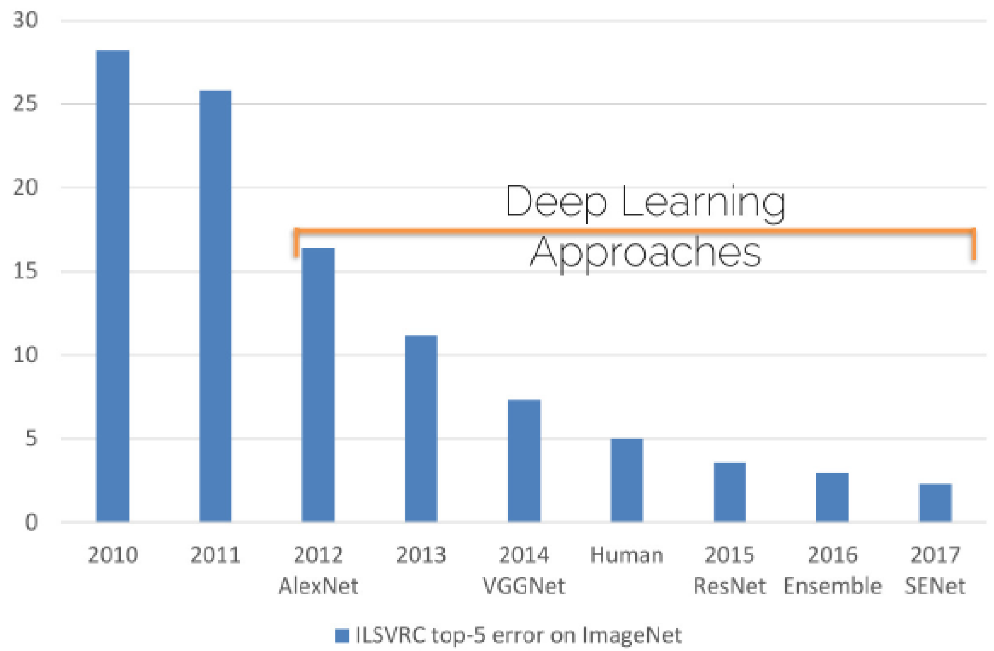
2009-2012: ImageNet and AlexNet

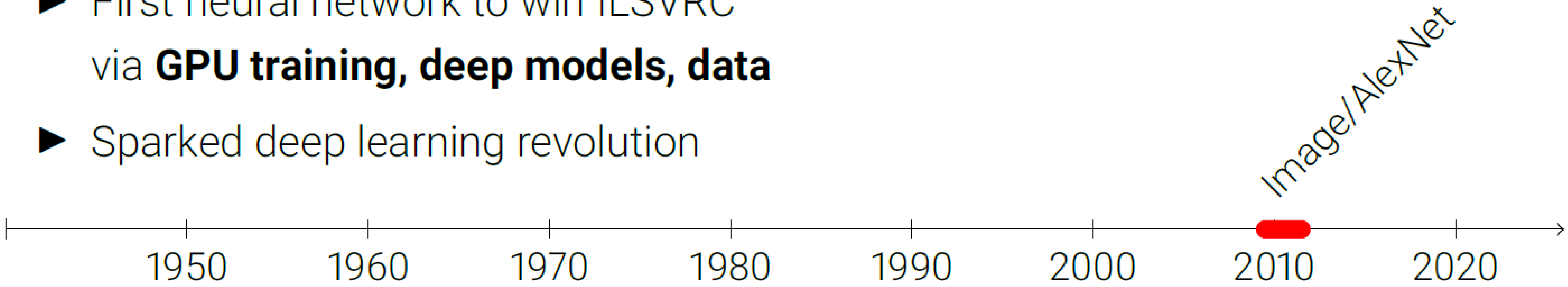
ImageNet

- Recognition benchmark (ILSVRC)
- 10 million annotated images
- 1000 categories

AlexNet

- First neural network to win ILSVRC via **GPU training, deep models, data**
- Sparked deep learning revolution



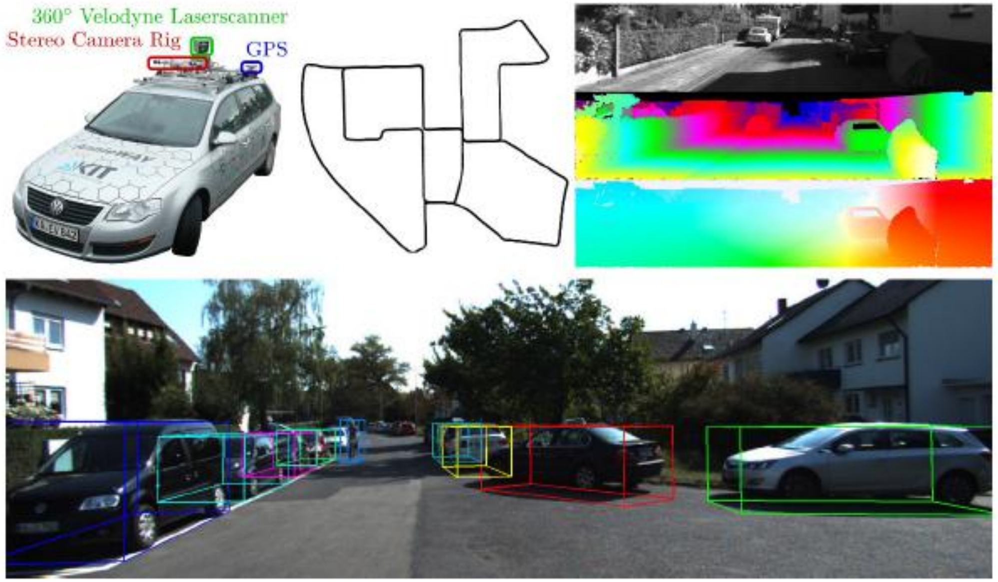


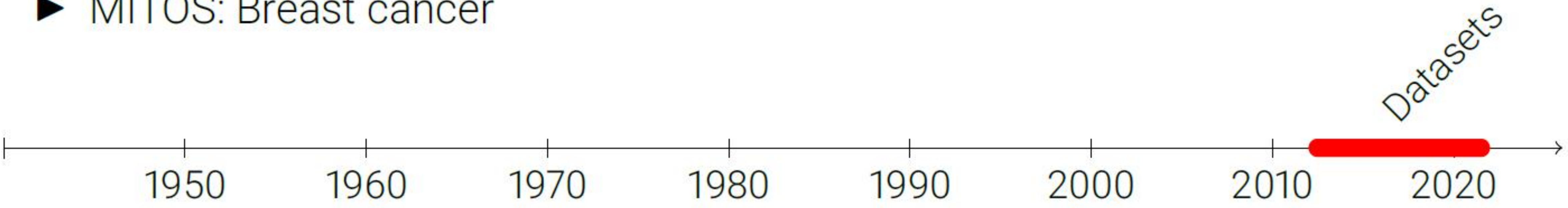
Krizhevsky, Sutskever, Hinton. ImageNet classification with deep convolutional neural networks. NIPS, 2012.

A Brief History of Deep Learning

2012-now: Golden Age of Datasets

- KITTI, Cityscapes: Self-driving
- PASCAL, MS COCO: Recognition
- ShapeNet, ScanNet: 3D DL
- GLUE: Language understanding
- Visual Genome: Vision/Language
- VisualQA: Question Answering
- MITOS: Breast cancer



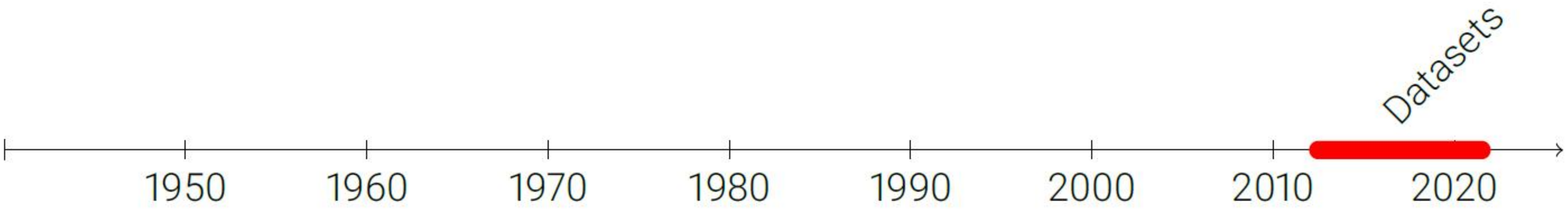


Geiger, Lenz and Urtasun. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. CVPR, 2012.

A Brief History of Deep Learning

2012-now: Synthetic Data

- Annotating real data is expensive
- Led to surge of synthetic datasets
- Creating 3D assets is also costly

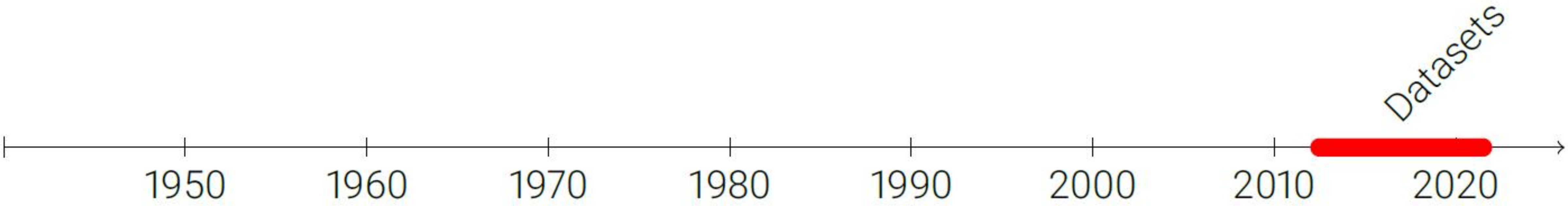


Dosovitskiy et al.: FlowNet: Learning Optical Flow with Convolutional Networks. ICCV, 2015.

A Brief History of Deep Learning

2012-now: Synthetic Data

- Annotating real data is expensive
- Led to surge of synthetic datasets
- Creating 3D assets is also costly
- But even very simple 3D datasets proved tremendously useful for pre-training (e.g., in optical flow)

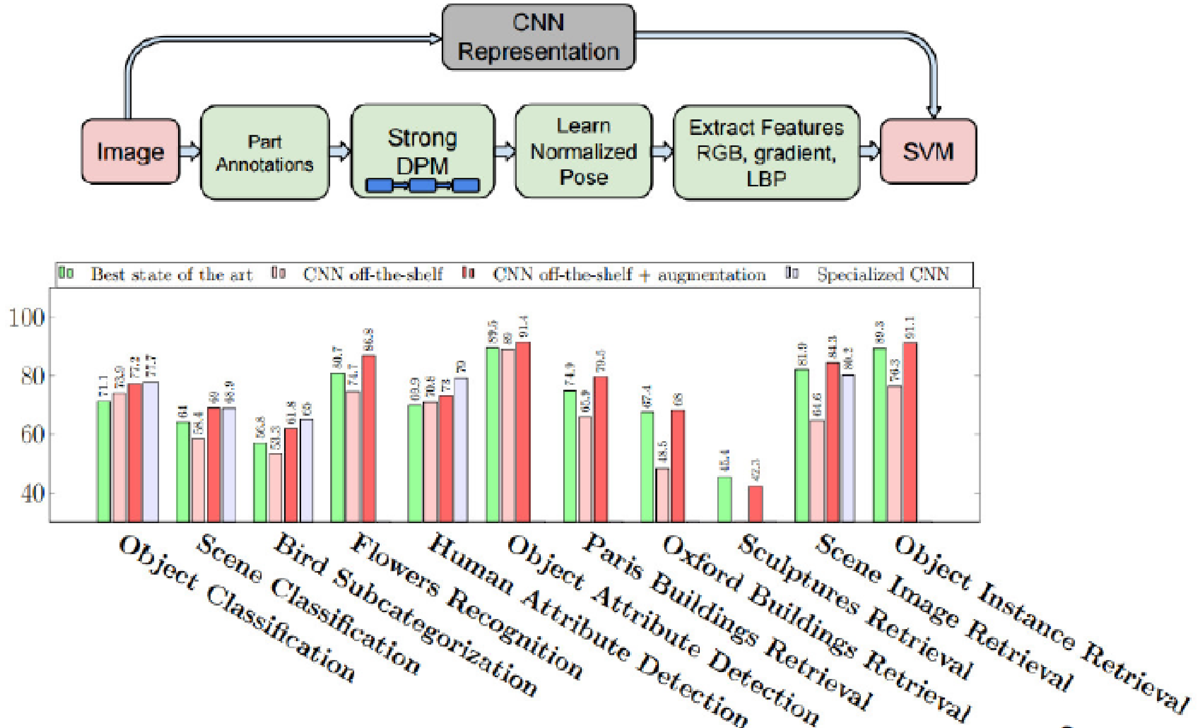


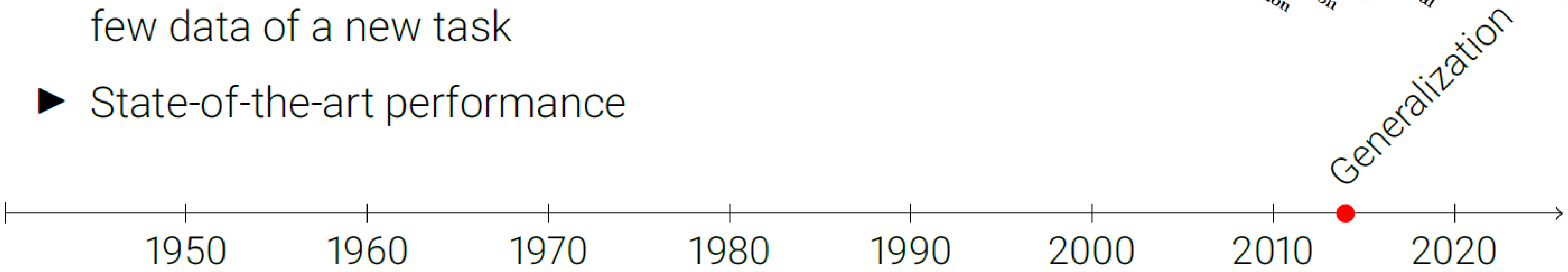
Dosovitskiy et al.: FlowNet: Learning Optical Flow with Convolutional Networks. ICCV, 2015.

A Brief History of Deep Learning

2014: Generalization

- Empirical demonstration that deep representations generalize well despite large number of parameters
- Pre-train CNN on large amounts of data on generic task (e.g., ImageNet)
- Fine-tune (re-train) only last layers on few data of a new task
- State-of-the-art performance



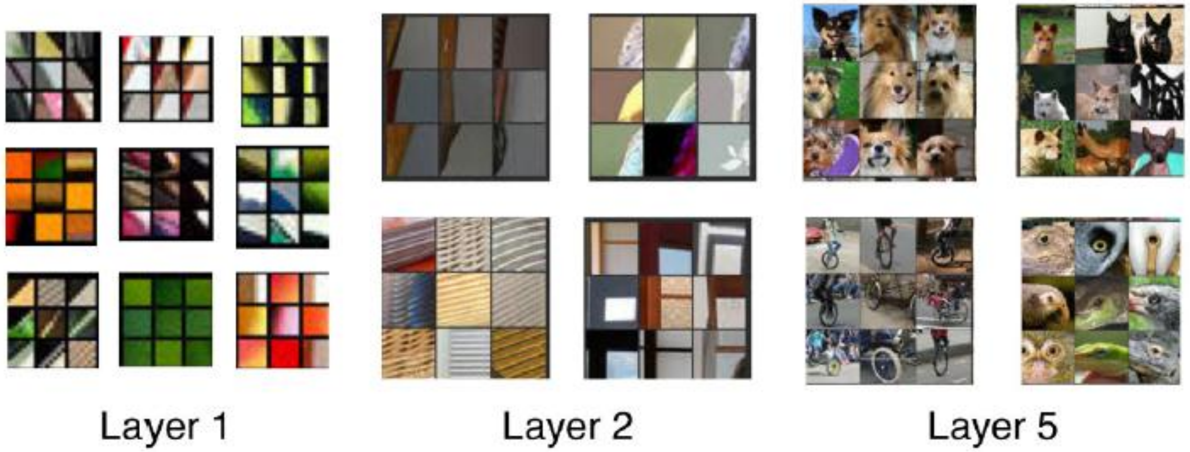


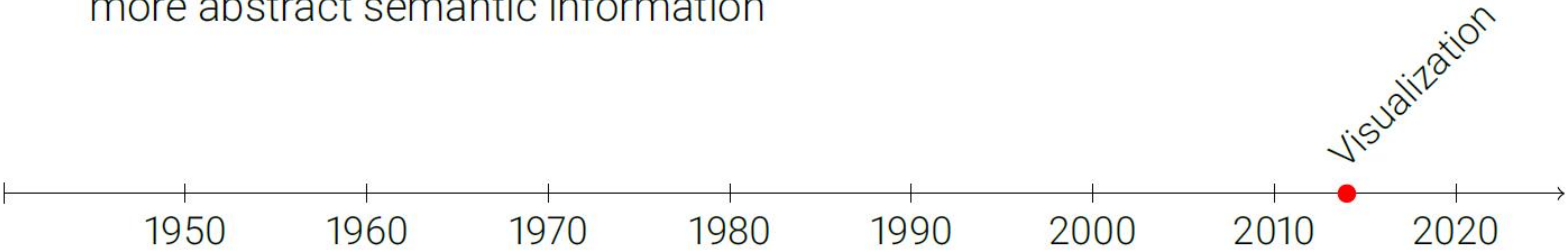
Razavian, Azizpour, Sullivan, Carlsson: CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. CVPR Workshops, 2014.

A Brief History of Deep Learning

2014: Visualization

- Goal: provide insights into what the network (black box) has learned
- Visualized image regions that most strongly activate various neurons at different layers of the network
- Found that higher levels capture more abstract semantic information





Zeiler and Fergus: CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. CVPR Workshops, 2014.

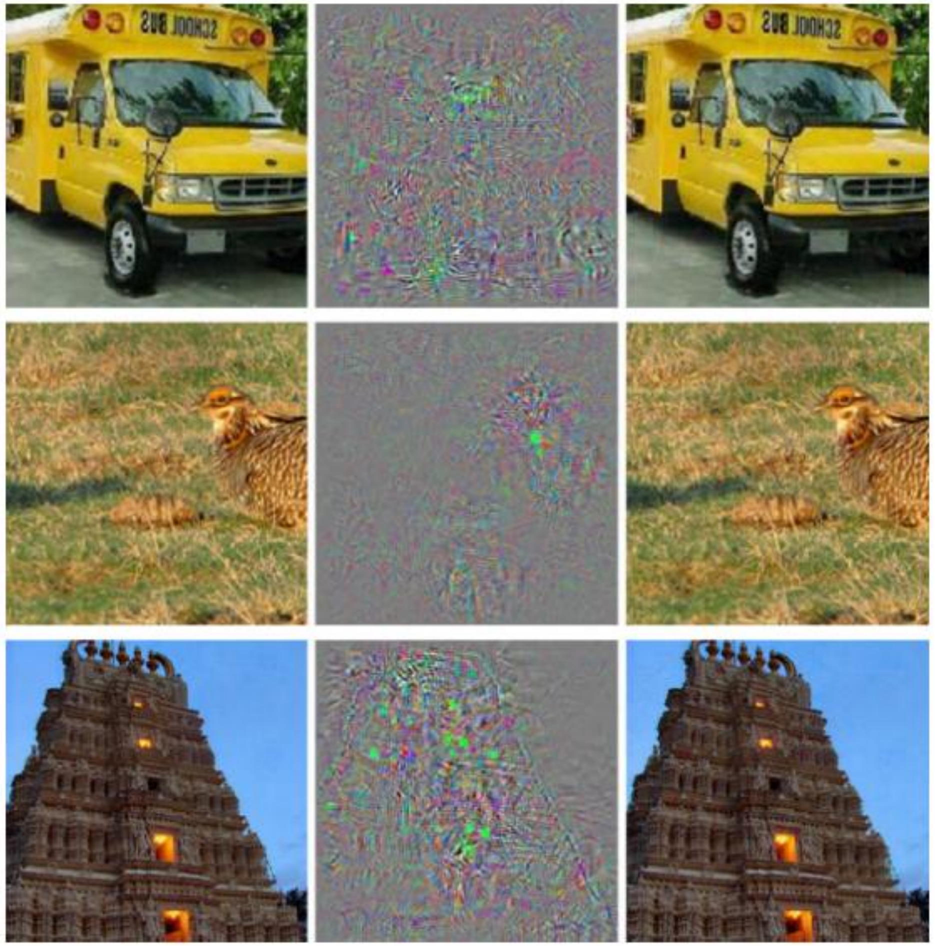
A Brief History of Deep Learning

2014: Adversarial Examples

- Accurate image classifiers can be fooled by imperceptible changes
- Adversarial example:

$$x + \underset{\Delta x}{\operatorname{argmin}} \{ \|\Delta x\|_2 : f(x + \Delta x) \neq f(x) \}$$

- All images classified as “ostrich”



Szegedy et al.: Intriguing properties of neural networks. ICLR, 2014.

A Brief History of Deep Learning

2014: Domination of Deep Learning

- Machine translation (Seq2Seq)

Type	Sentence
Our model	Ulrich UNK , membre du conseil d' administration du constructeur automobile Audi , affirme qu' il s' agit d' une pratique courante depuis des années pour que les téléphones portables puissent être collectés avant les réunions du conseil d' administration afin qu' ils ne soient pas utilisés comme appareils d' écoute à distance .
Truth	Ulrich Hackenberg , membre du conseil d' administration du constructeur automobile Audi , déclare que la collecte des téléphones portables avant les réunions du conseil , afin qu' ils ne puissent pas être utilisés comme appareils d' écoute à distance , est une pratique courante depuis des années .
Our model	" Les téléphones cellulaires , qui sont vraiment une question , non seulement parce qu' ils pourraient potentiellement causer des interférences avec les appareils de navigation , mais nous savons , selon la FCC , qu' ils pourraient interférer avec les tours de téléphone cellulaire lorsqu' ils sont dans l' air " , dit UNK .
Truth	" Les téléphones portables sont véritablement un problème , non seulement parce qu' ils pourraient éventuellement créer des interférences avec les instruments de navigation , mais parce que nous savons , d' après la FCC , qu' ils pourraient perturber les antennes-relais de téléphonie mobile s' ils sont utilisés à bord " , a déclaré Rosenker .
Our model	Avec la crémation , il y a un " sentiment de violence contre le corps d' un être cher " , qui sera " réduit à une pile de cendres " en très peu de temps au lieu d' un processus de décomposition " qui accompagnera les étapes du deuil " .
Truth	Il y a , avec la crémation , " une violence faite au corps aimé " , qui va être " réduit à un tas de cendres " en très peu de temps , et non après un processus de décomposition , qui " accompagnerait les phases du deuil " .

Sutskever, Vinyals, Quoc: Sequence to Sequence Learning with Neural Networks. NIPS, 2014.

A Brief History of Deep Learning

2014: Domination of Deep Learning

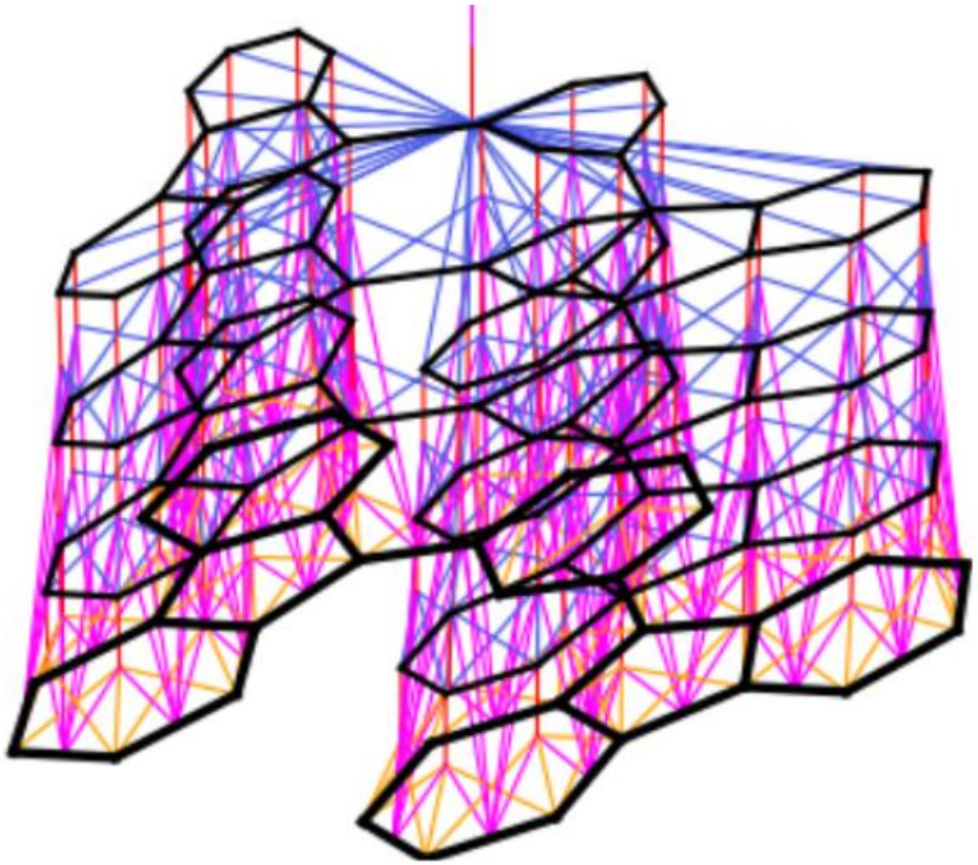
- Machine translation (Seq2Seq)
- Deep generative models (VAEs, GANs) produce compelling images

Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, Bengio: Generative Adversarial Networks. NIPS, 2014.

A Brief History of Deep Learning

2014: Domination of Deep Learning

- Machine translation (Seq2Seq)
- Deep generative models (VAEs, GANs) produce compelling images
- Graph Neural Networks (GNNs) revolutionize the prediction of molecular properties

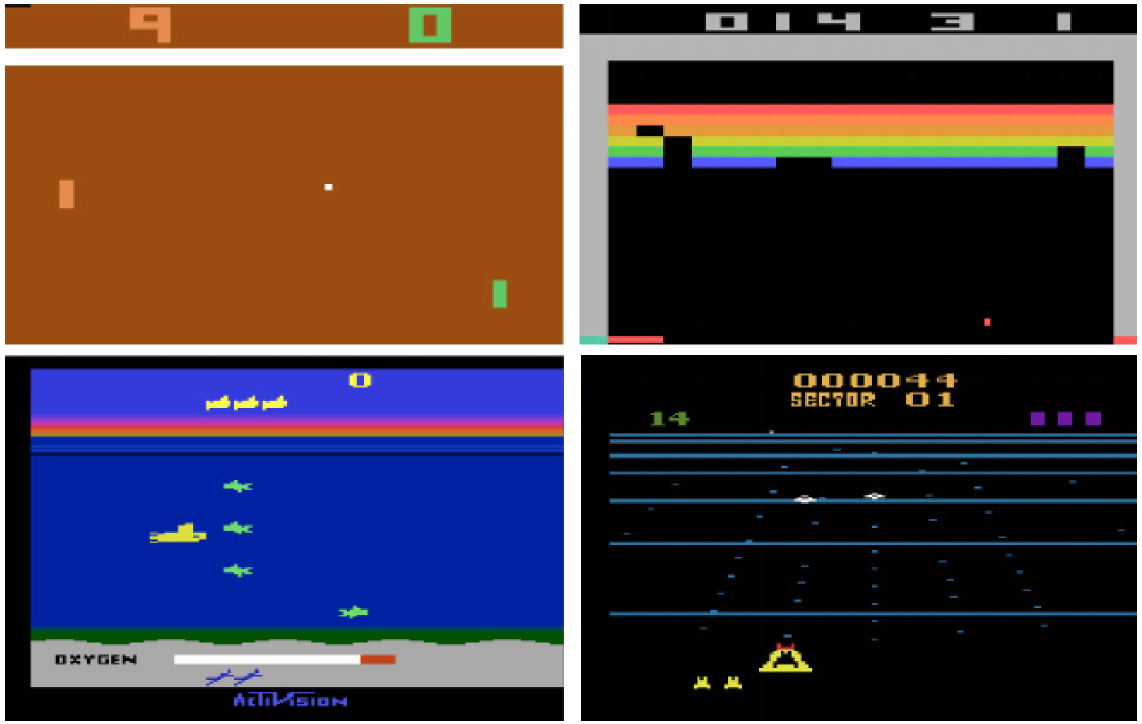


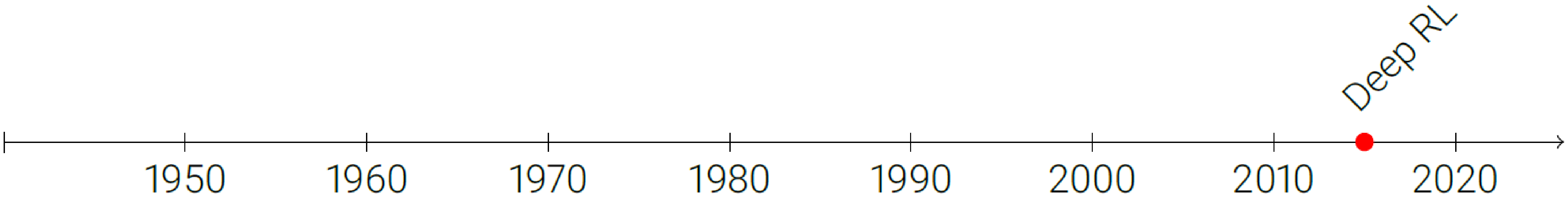
Duvenaud et al.: Convolutional Networks on Graphs for Learning Molecular Fingerprints. NIPS 2015.

A Brief History of Deep Learning

2015: Deep Reinforcement Learning

- Learning a policy (state→action) through random exploration and reward signals (e.g., game score)
- No other supervision
- Success on many Atari games
- But some games remain hard



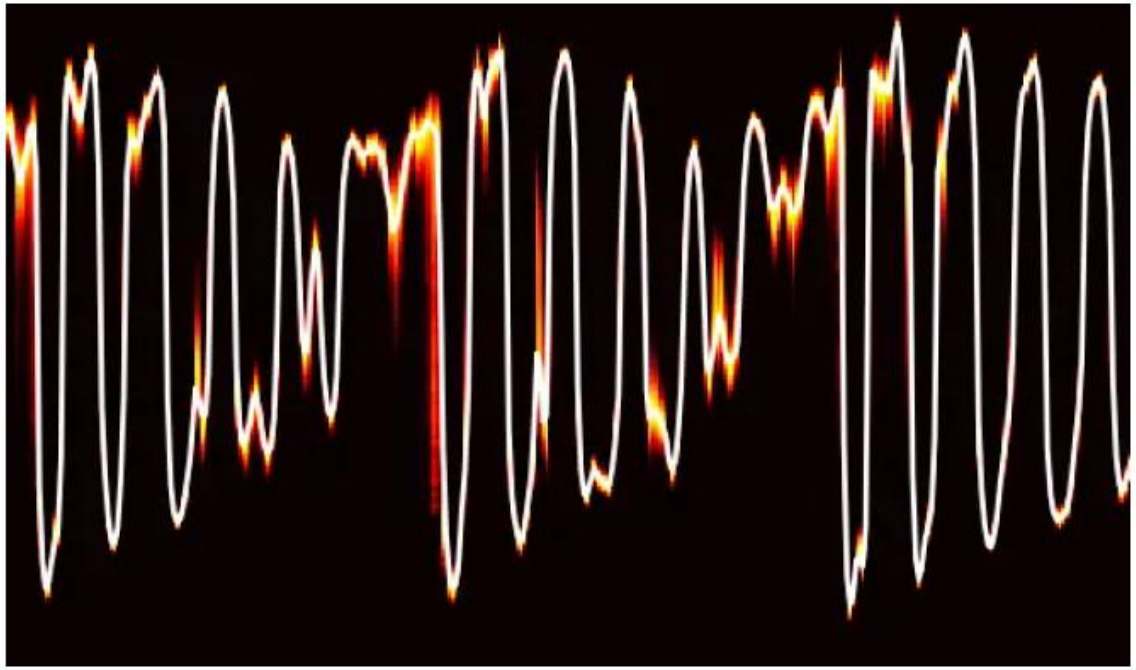


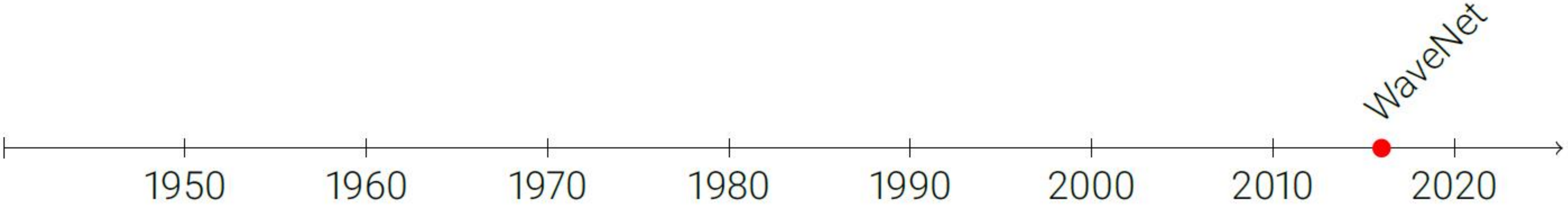
Mnih et al.: Human-level control through deep reinforcement learning. Nature, 2015.

A Brief History of Deep Learning

2016: WaveNet

- Deep generative model of raw audio waveforms
- Generates **speech** which mimics human voice
- Generates **music**



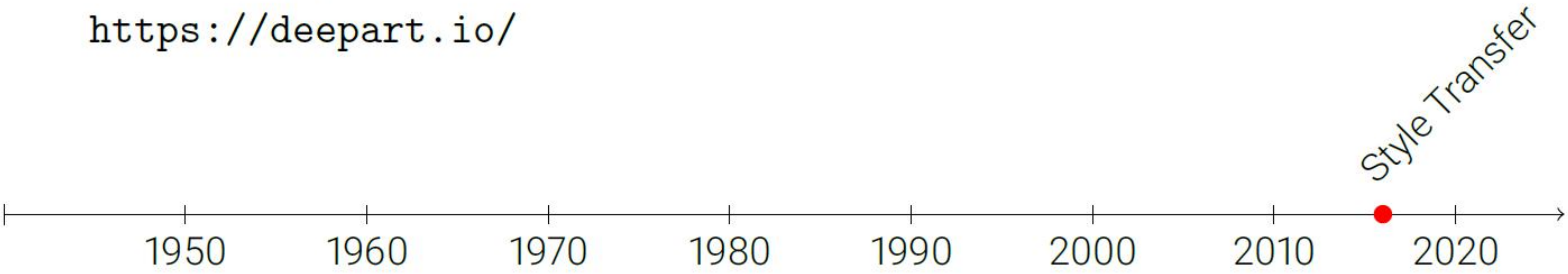


Oord et al.: WaveNet: A Generative Model for Raw Audio. Arxiv, 2016.

A Brief History of Deep Learning

2016: Style Transfer

- Manipulate photograph to adopt style of a another image (painting)
- Uses deep network pre-trained on ImageNet for disentangling content from style
- It is fun! Try yourself: `https://deepart.io/`

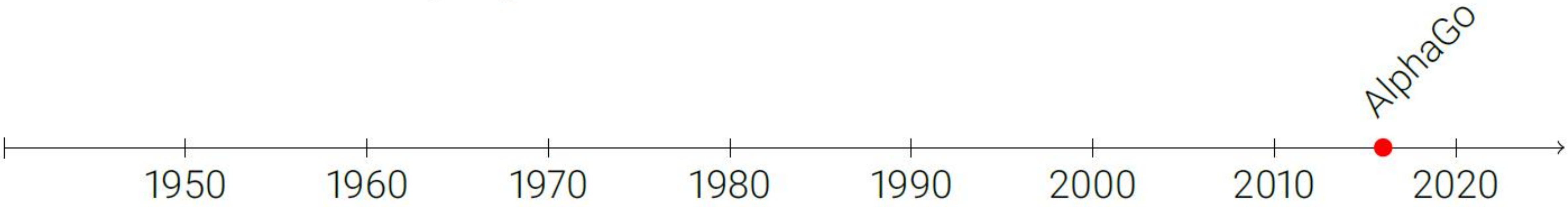


Gatys, Ecker and Bethge: Image Style Transfer Using Convolutional Neural Networks. CVPR, 2016.

A Brief History of Deep Learning

2016: AlphaGo defeats Lee Sedol

- Developed by DeepMind
- Combines deep learning with Monte Carlo tree search
- First computer program to defeat professional player
- AlphaZero (2017) learns via self-play and masters multiple games

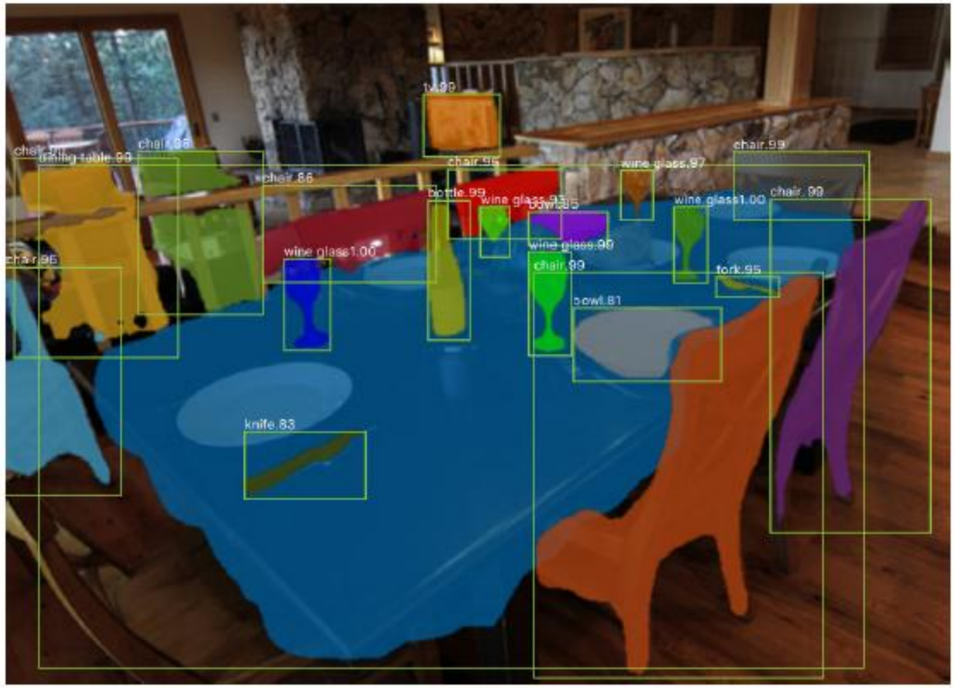


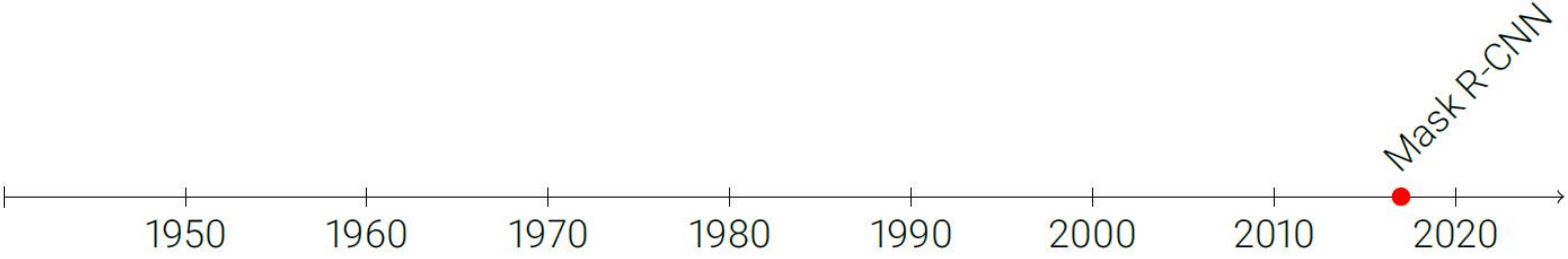
Silver et al.: Mastering the game of Go without human knowledge. Nature, 2017.

A Brief History of Deep Learning

2017: Mask R-CNN

- Deep neural network for joint object detection and instance segmentation
- Outputs "structured object", not only a single number (class label)
- State-of-the-art on MS-COCO



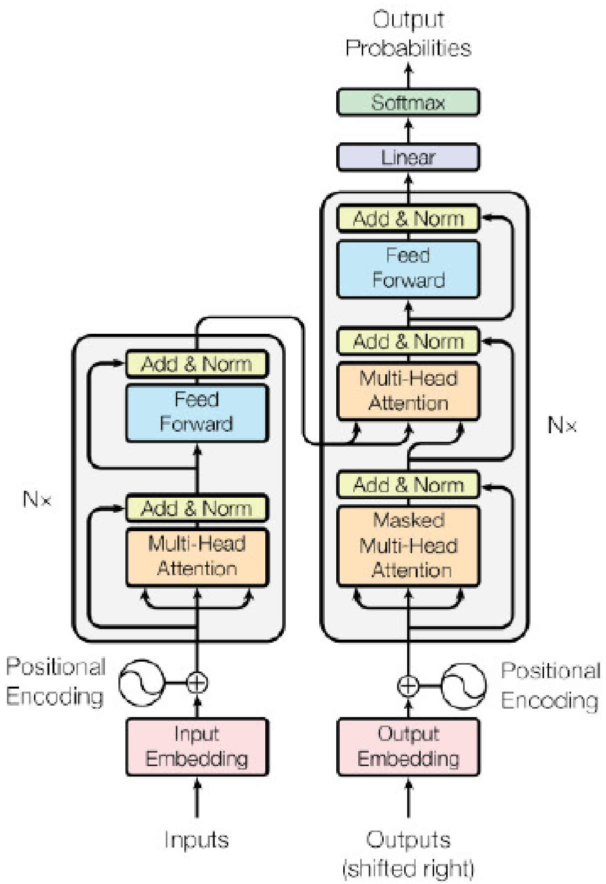


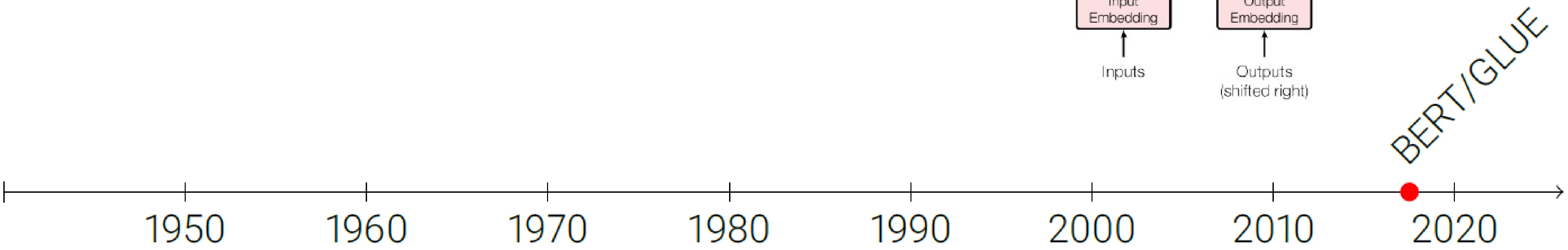
He, Gkioxari, Dollár and Ross Girshick: Mask R-CNN. ICCV, 2017.

A Brief History of Deep Learning

2017-2018: Transformers and BERT

- Transformers: Attention replaces recurrence and convolutions



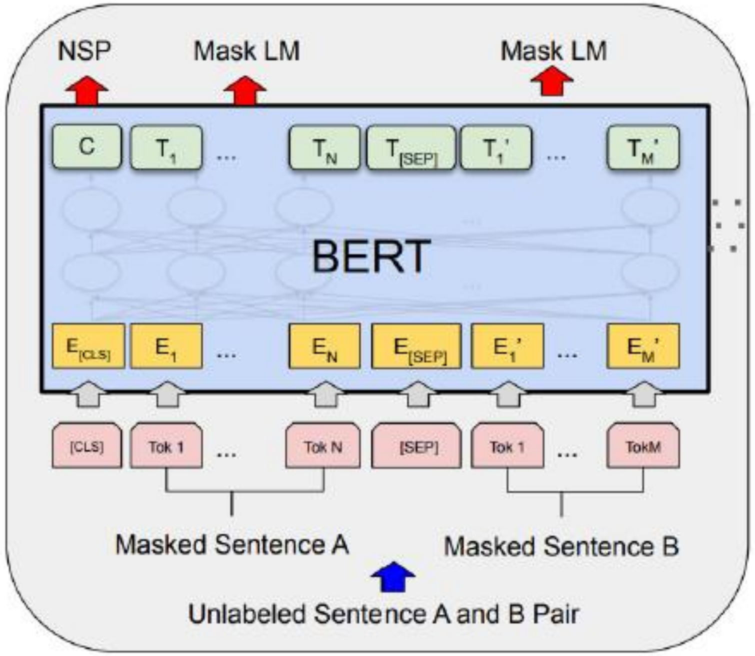


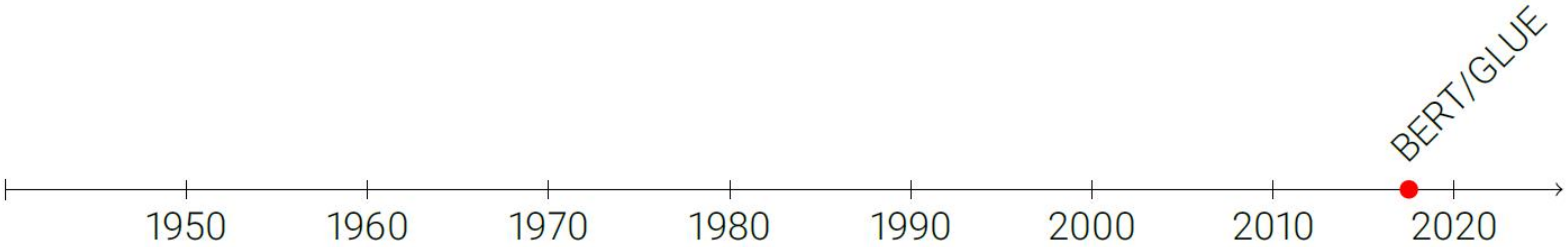
Vaswani et al.: Attention is All you Need. NIPS 2017.

A Brief History of Deep Learning

2017-2018: Transformers and BERT

- Transformers: Attention replaces recurrence and convolutions
- BERT: Pre-training of language models on unlabeled text





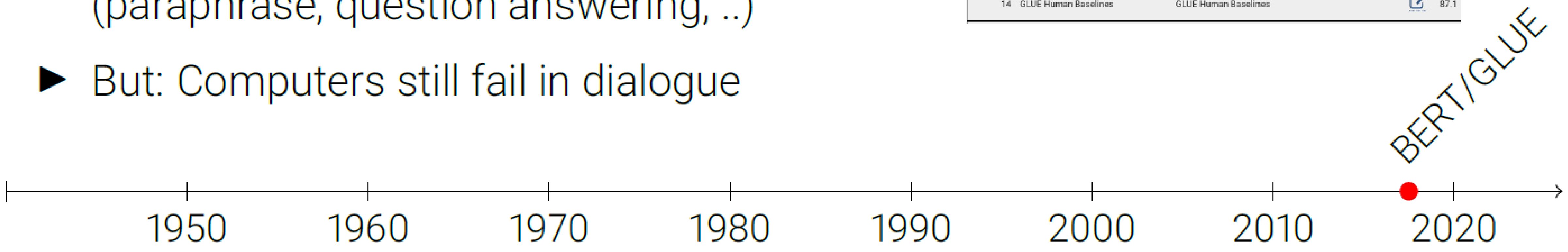
Devlin, Chang, Lee and Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Arxiv, 2018.

A Brief History of Deep Learning

2017-2018: Transformers and BERT

- Transformers: Attention replaces recurrence and convolutions
- BERT: Pre-training of language models on unlabeled text
- GLUE: Superhuman performance on some language understanding tasks (paraphrase, question answering, ..)
- But: Computers still fail in dialogue

Rank	Name	Model	URL	Score
1	HFL IFLYTEK	MacALBERT + DKM		90.7
2	Alibaba DAMO NLP	StructBERT + TAPT		90.6
3	PING-AN Omni-Sinitic	ALBERT + DAAF + NAS		90.6
4	ERNIE Team - Baidu	ERNIE		90.4
5	T5 Team - Google	T5		90.3
6	Microsoft D365 AI & MSR AI & GATECH	MT-DNN-SMART		89.9
7	Zihang Dai	Funnel-Transformer (Ensemble B10-10-10H1024)		89.7
8	ELECTRA Team	ELECTRA-Large + Standard Tricks		89.4
9	Huawei Noah's Ark Lab	NEZHA-Large		89.1
10	Microsoft D365 AI & UMD	FreeLB-RoBERTa (ensemble)		88.4
11	Junjie Yang	HIRE-RoBERTa		88.3
12	Facebook AI	RoBERTa		88.1
13	Microsoft D365 AI & MSR AI	MT-DNN-ensemble		87.6
14	GLUE Human Baselines	GLUE Human Baselines		87.1



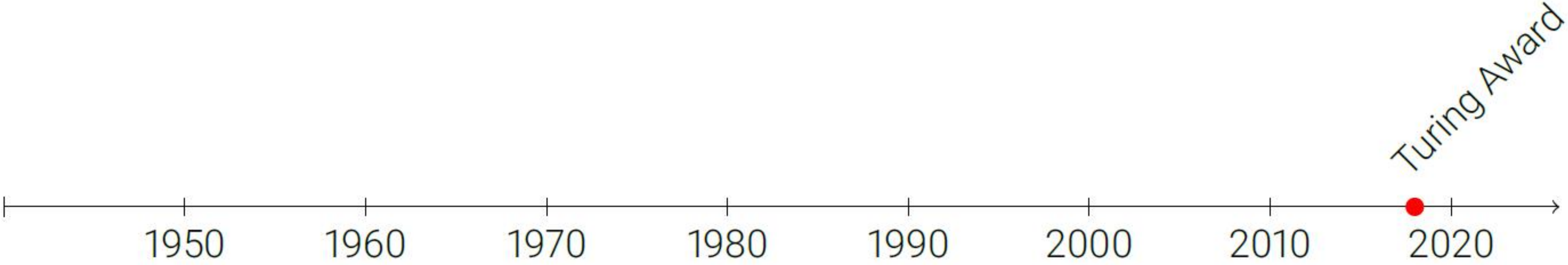
Wang et al.: GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. ICLR, 2019.

A Brief History of Deep Learning

2018: Turing Award

In 2018, the "nobel price of computing" has been awarded to:

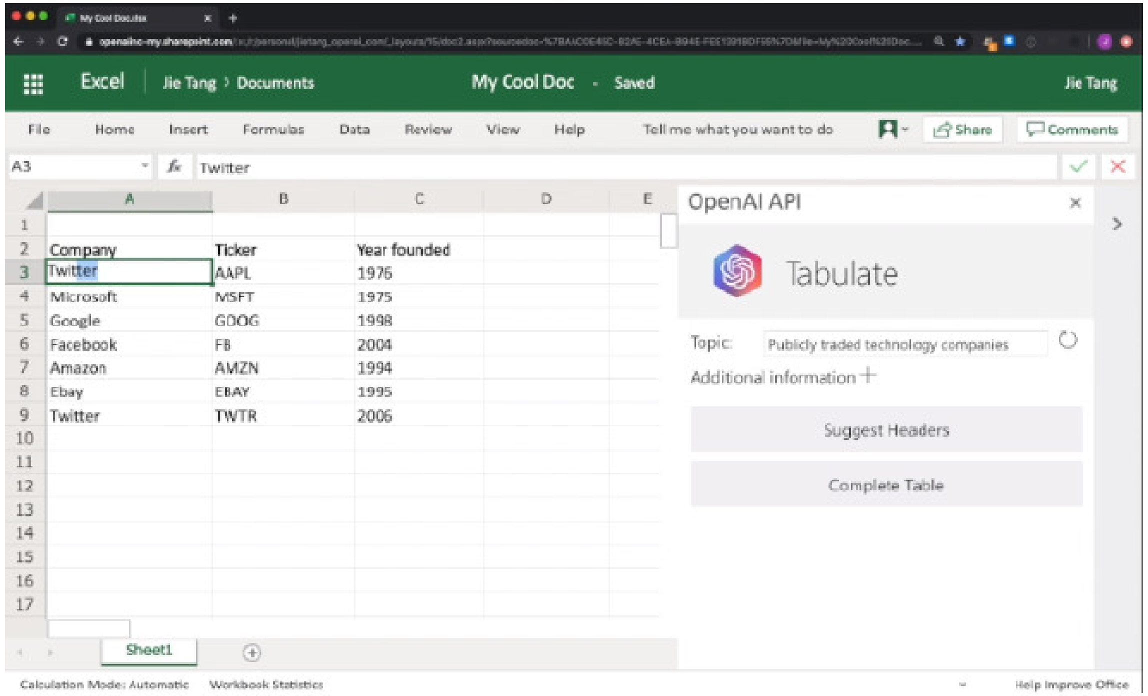
- Yoshua Bengio
- Geoffrey Hinton
- Yann LeCun

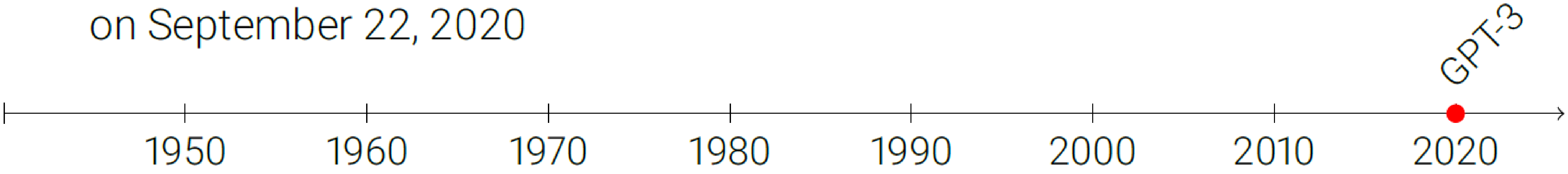


A Brief History of Deep Learning

2020: GPT-3

- Language model by OpenAI
- 175 Billion parameters
- Text-in / text-out interface
- Many use cases: coding, poetry, blogging, news articles, chatbots
- Controversial discussions
- Licensed exclusively to Microsoft on September 22, 2020





Brown et al.: Language Models are Few-Shot Learners. Arxiv, 2020.