# COMP47650 – DEEP LEARNING

## Introduction

SLIDE DECK 1

Guénolé Silvestre

School of Computer Science
**University College Dublin** – Ireland

January 20, 2025

# Overview

- Objectives:
  - Understand the scope of DL in modern ML and how to use these techniques effectively
  - Understand how to build DL systems (in tensorflow/pytorch) and solve some problems in machine vision or NLP
- Pre-requisites:
  - Linear Algebra
  - Multivariate Calculus
  - Foundations on Statistics and Probability
  - Fundamentals in Machine Learning
  - Proficiency in Python (homework and tutorials)
- Evaluation:
  - 20%: engagement & participation (attendance, homework, … )
  - 80%: individual DL project

# Organisation

**Weekly Schedule**

- 2× classes: 09:00–10:50 Monday (QUI-014)
- 1× class/tutorial: 13:00–13:50 Wednesday (H2.38-SCH)

**People**

- Week 1–12: Guénolé Silvestre
- TA: Connor O'Sullivan
- Demonstrators: Daniel Atputharuban, Cheng Xu, Duc-Anh Nguyen and Jiaming Xu

**Mics**

- Moodle enrolment key: `comp47650enrol2025`
- Tutorials: jupyter notebooks, tensorflow.keras, pytorch
- Textbooks: *Deep Learning* (MIT Press, Eds. Goodfellow, Bengio & Courville) and *The Science of Deep Learning* (Cambridge University Press, Drori)



# Syllabus

**Part 1: NN Fundamentals** (G. Silvestre)

- W 01 Module overview and Introduction to DL
- W 02 Neural Networks, Perceptron, Loss functions, Backprop
- W 03 Computational Graphs and Backprop, Optimisation (SGD, RMSProp, Adam), Non-linearities, MLP – (self-study, Bank Holiday)
- W 04 Vectorisation, Data normalisation, Hyperparameter tuning, Learning rate decay, Bias/Variance, Overfitting

# Syllabus

◆ Part 2: CNN (G. Silvestre)

- W 05 Introduction to Image Understanding, Classification, Softmax, Regularisation, Dropout, Batch Norm, Introduction to DL libraries
- W 06 Convolution, Stride and Padding, Pooling, CNN, AlexNet, VGG
- W 07 Case Studies: GoogLeNet (inception module), ResNet (residual module), Misc. Topologies and Image Applications
- W 08 CNN generalisation to 1D and 3D. Transfer learning.



# Syllabus

◆ Part 3: RNN, Sequence Models, Unsupervised Learning (G. Silvestre)

- W 09 Introduction to Language Modelling, Dimensional Embeddings (word2vec), RNN (LSTM & GRU), Backprop through Time
- W 10 Sequence Models. Applications to Machine Translation, Sentiment Analysis, Language modelling, Attention models with Transformers (if time permits)
- W 11 Unsupervised Learning: AutoEncoders, Variational AEs, Generative Adversarial Networks, (and if time permits Diffusion Models)
- W 12 Easter Week

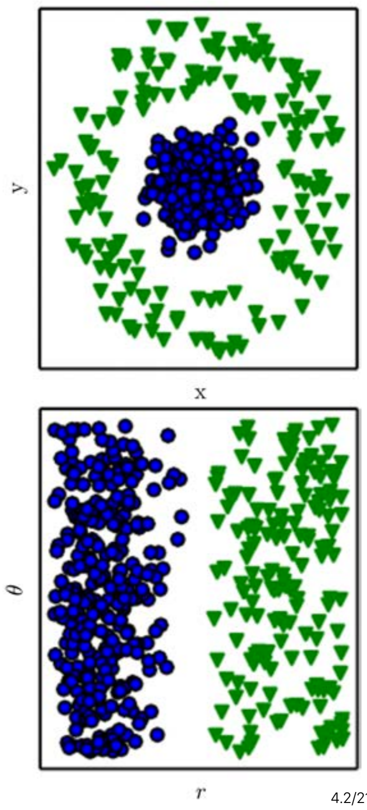
# Deep Learning

- Models: graphs of conceptual hierarchy
- Computer learns complicated concepts, building out of simpler ones
  $\rightarrow$ deep graphs with many layers
- Machines have now matched human ability in most vision or speech tasks, recently in many NLP and cognitive tasks
- Knowledge-base models
  - first AI attempts
  - hard-coding in formal languages
  - struggle to devise rules of sufficient complexity
  - suggests AI system need to acquire own knowledge from raw data
  - → emergence ML
- ML examples:
  - logistic regression
  - naive Bayes

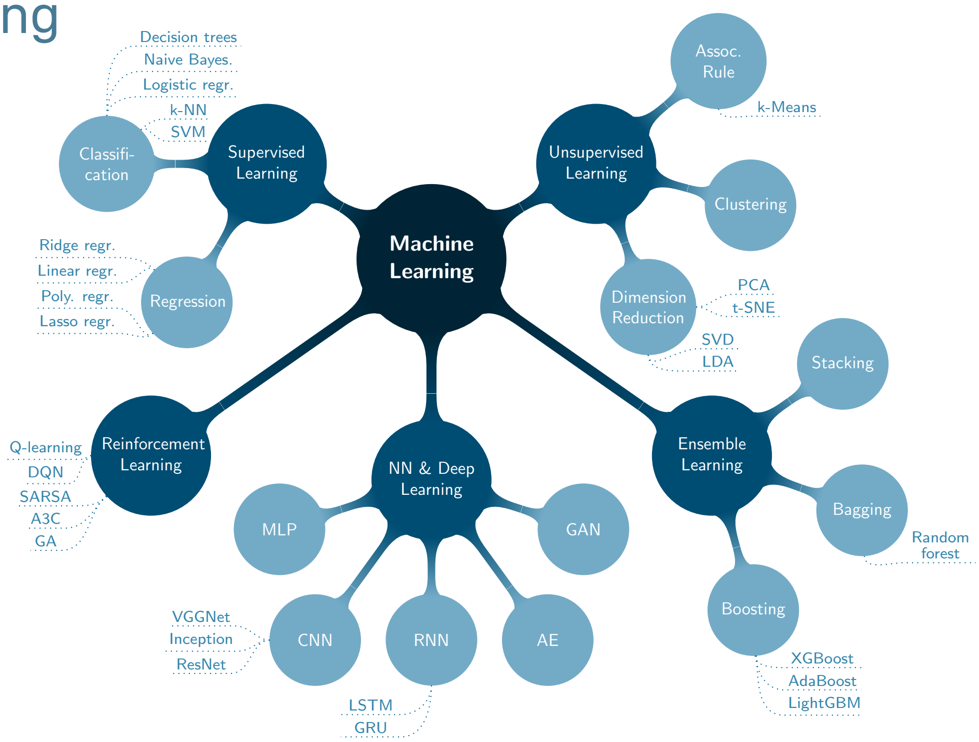


# Deep Learning

- Performance depends heavily on well crafted data representations and input features, a human task requiring tremendous human expertise
- Data input $\equiv$ features
- Models such logistic regression learn how each feature correlates with output
- Often difficult to know what are the best features to extract
- ML reduces to weights optimisation in order to make the best prediction
- Idea: use ML to discover representations along with mapping to output
  $\rightarrow$ representation learning



# Machine Learning



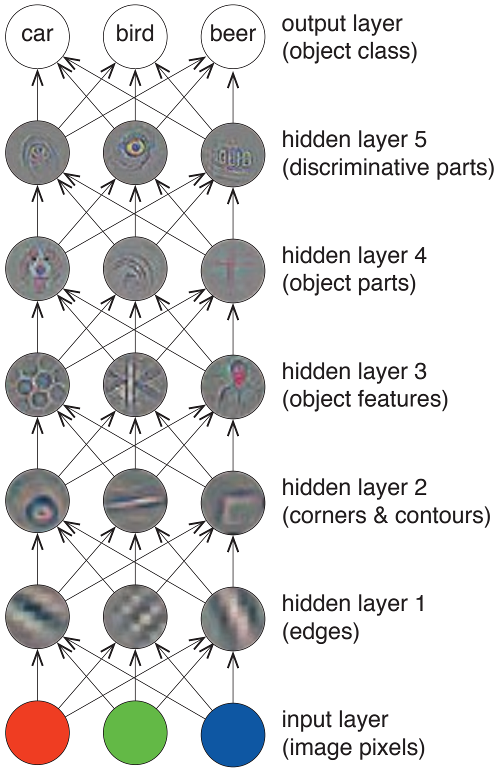



# Representation Learning

- Representation learning attempts to automatically learn good features
  - Better performance than hand-crafted representations
  - Better adaptation to new tasks
  - Faster than manual modelling
- Example: AutoEncoder
  - encoder function converts data input to new representation
  - decoder function converts back to original representation
  - trained to learn new representation with nice properties
- Goal: separate factors of variation that explain observed data
- Factor of variations often affects all observed data → difficult to disentangle
- High-level abstract features are hard to extract, requires almost human-level understanding of data

## Deep Learning

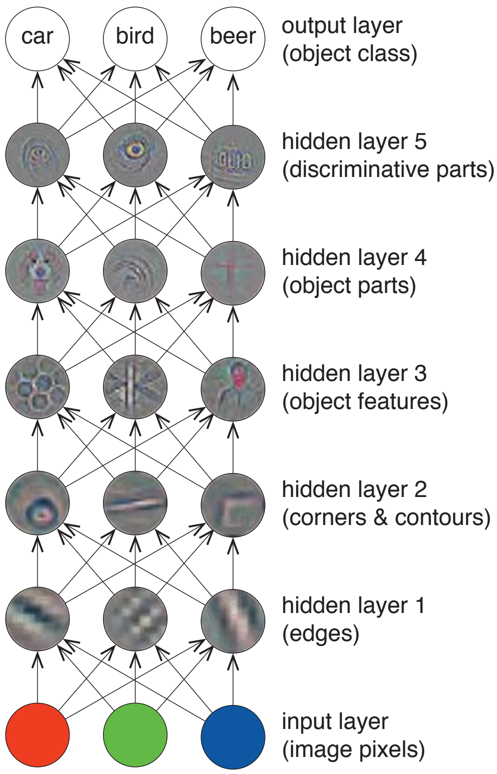
- Objective: solve central problem in representation learning
- Learned representations expressed in terms of simpler representations
- Build complex concepts out of simpler ones
- Example: MLP (deep feedforward net)
  - MLP $\equiv$ non-linear map by composing simpler functions
  - each function application provides new input representation for next layer
  - network depth can also be understood as sequential instructions of computer program





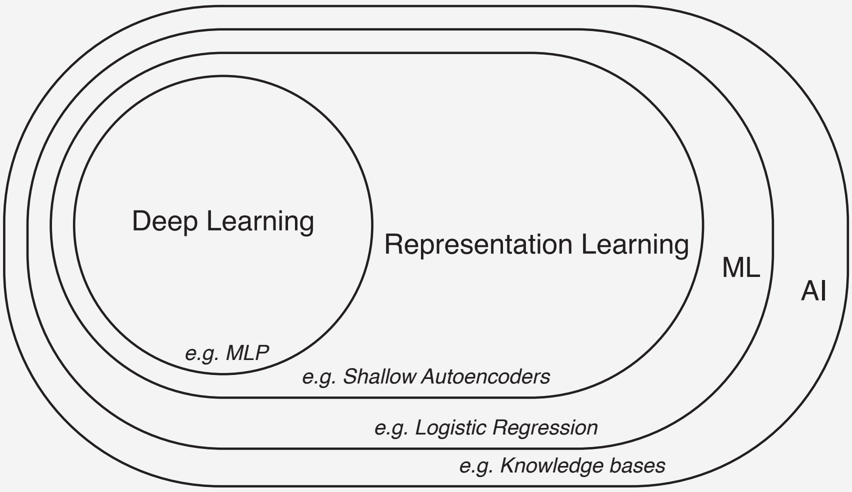

## Deep Learning

- Network depth:
  - longest path in computational graph, number of instructions
  - longest path of modelling graph, relating concepts
- DL = study of models that involves greater amount of composition of learned functions (concepts) than ML
- Great power and flexibility to represent complex world in nested hierarchy of concepts
- Representation abstraction from simpler abstracted representations



# Deep Learning, ML and AI

Venn diagram: DL $\subset$ ML $\subset$ AI




- DL finds applications in many disciplines: computer vision, speech and audio processing, NLP, bioinformatics, search engines, advertising, finance, chemistry, physics...



# Disruptive Technology

- 2010's mark the beginning of DL disruptive intrusion into ML
- DL starts to outperform most traditional ML techniques
- Why? It follows a decade of
  - online (big) data $\rightarrow$ plenty of training information
  - gaming emergence $\rightarrow$ powerful GPUs and fast multicore CPU
- Resurgence of research in NN
  - new models, new ideas, new algorithm
  - better and more flexible learning of intermediate representations
  - emergence of methods for using context knowledge and transfer between tasks
  - end-to-end effective learning
  - advances in training techniques (regularisation, dropout) leading to better optimisation

## Disruptive Technology

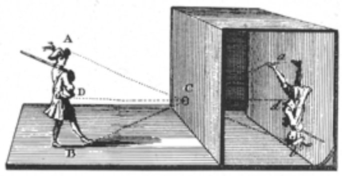
- Speech Recognition (2010) and Computer Vision (2012)
- Initial DL achievement on large dataset occurred in speech recognition
  - → Context dependent pre-trained deep NN for large vocabulary speech recognition (Dahl *et al.*, 2010)
  - – 30% performance improvement over state-of-the-art traditional feature models!
- First focus of DL research was computer vision:
  - → Breakthrough paper: ImageNet Classification with Deep Convolutional Neural Networks (Krizhevsky *et al.*, 2012)
  - – 37% improvement over current state-of-the-art on image classification
- → 10 years later: near human performance reached in vision, speech, and now NLP!

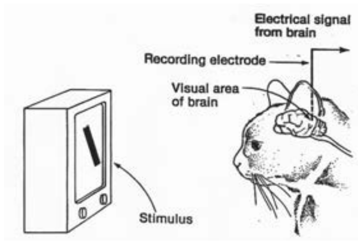


## Evolution, Vision and Neurons

- 540 millions years ago:
  - – live mostly in the ocean, crude, primitive, dark with no vision
  - – feeding on floating nearby organisms
- Cambrian explosion: big bang of evolution
  - – within 10 million years, number of species explodes. Why?
  - – Light switch theory (A. Parker), possibly combined with rise of $O_2$ levels
  - → Subphylum Trilobita (sort of horseshoe crab): first animal with colour vision
  - – life becomes proactive, predatory
  - → start of an arms race that accelerated evolution
- Visual processing is largest sensory system, involves nearly 50% of neurons in humans cortex

# Historical Remarks



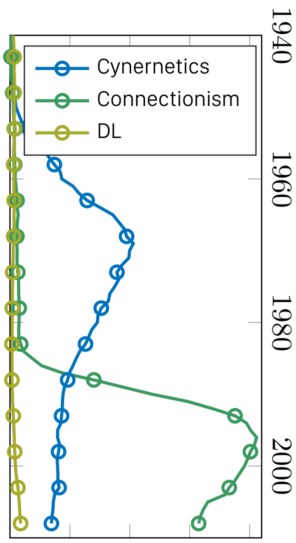



- Mechanical Vision: Cameras
  - Renaissance $\rightarrow$ camera obscura
  - Pinhole camera similar to early eyes
  - 2022: ubiquitous imaging sensors, 4+ in smartphone
- Study of Visual System, Hubel & Wisel, 1956: What are the visual processing mechanisms in mammals?
  - Wire electrodes to primary visual cortex (back of cat's brain), observe response to stimuli
  - Simple cells: oriented edges when they move in certain directions
  - Complex cells: orientation and movement
  - Hypercomplex cells: movement with an end point

→ Visual system builds complexity from with simple structures
$\rightarrow$ Influential work in human and computer vision



# Historical Remarks




- 1940–1960: DL $\equiv$ Cybernetics
  - Inspired by biological learning theories, neuroscience
  - Early brain models were linear
  - 50's: perceptron
- 1980–1995: DL $\equiv$ Connectionism or ANNs
  - Back-propagation
  - MLP, CNN
- 2006–date: DL rebranding
  - Large deep NN
  - Recent neuroscience studies suggest DL algorithms can solve many different tasks

# Historical Remarks: Cybernetics

- Earliest predecessors of modern DL: linear models motivated by neuroscience

  $f(\mathbf{x}, \mathbf{w}) = x_1 w_1 + \ldots + x_n w_n$
- McCulloch-Pitts neuron [†]: $f(\mathbf{x}, \mathbf{w})$ recognize 2 categories of input. $\mathbf{w}$ tuned by human operator
- Perceptron [‡]: threshold function applied to $f(\mathbf{x}, \mathbf{w})$. First model to learn $\mathbf{w}$ given examples from categories
- Training algorithm to adjust $\mathbf{w}$ (of the adaptive linear element, ADALINE): special case of stochastic gradient descent
- Linear models are widely used in ML (albeit trained differently)
- Limitations: cannot learn XOR function [§] $\rightarrow$ NN loss of popularity

[†] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.
[‡] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408.
[§] Minsky, M. L. and Papert, S. A. (1969). *Perceptrons*. MIT Press, Cambridge.



# Historical Remarks: Connectionism

- Resurgence of NN research and new architectures inspired by neuroscience albeit different (in particular learning algorithm)
- Neurons not as rigid guide: modern rectified linear units compute very different functions
- Introduction of non-linear activation function (original Cognitron)
- Neocognitron [†]: (structure of mammalian visual system) $\rightarrow$ basis for modern convolutional network
- Connectionism [‡]: large number of units can achieve intelligent behaviour when networked together (cognition models grounded in neural implementations)

[†] Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36, 193–202.
[‡] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In *Parallel Distributed Processing*, vol. 1, ch 8, pages 318–362. MIT Press.

# Historical Remarks: Connectionism

- 90s: important advances in modelling sequences with neural networks $\rightarrow$ LSTM [†]
- Main accomplishments:
  1. Distributed representations [‡]: each system input represented by many features, each feature involved in representation of many inputs.
  2. Successful use of back-propagation to train deep neural networks with internal representations [§] [¶]
- Hype with unrealistically ambitious claims albeit impressive results on specific tasks $\rightarrow$ research focus switched to ML with kernel machines and graphical models

[†] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, **9**(8), 173–1780

[‡] Hinton, G. E. and Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In *Parallel Distributed Processing*, vol. 1, ch 7, pages 282–317. MIT Press

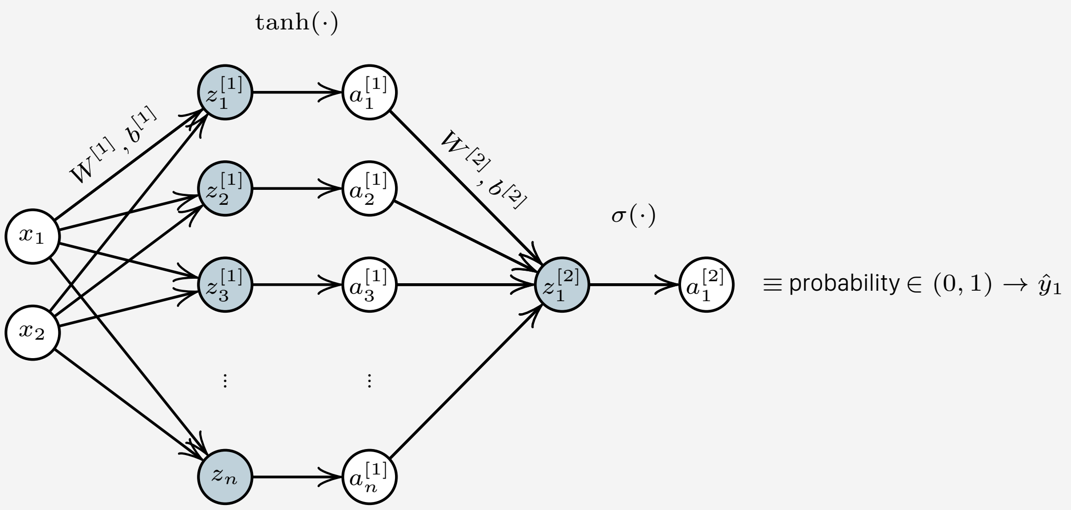
[§] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In *Parallel Distributed Processing*, vol. 1, ch 8, pages 318–362. MIT Press.

[¶] LeCun, Y. (1985). Une procédure d'apprentissage pour Réseau à seuil assymétrique. In *Cognitiva 85: A la Frontière de l'Intelligence Artificielle, des Sciences de la Connaissance et des Neurosciences*, pages 599–604, Paris.



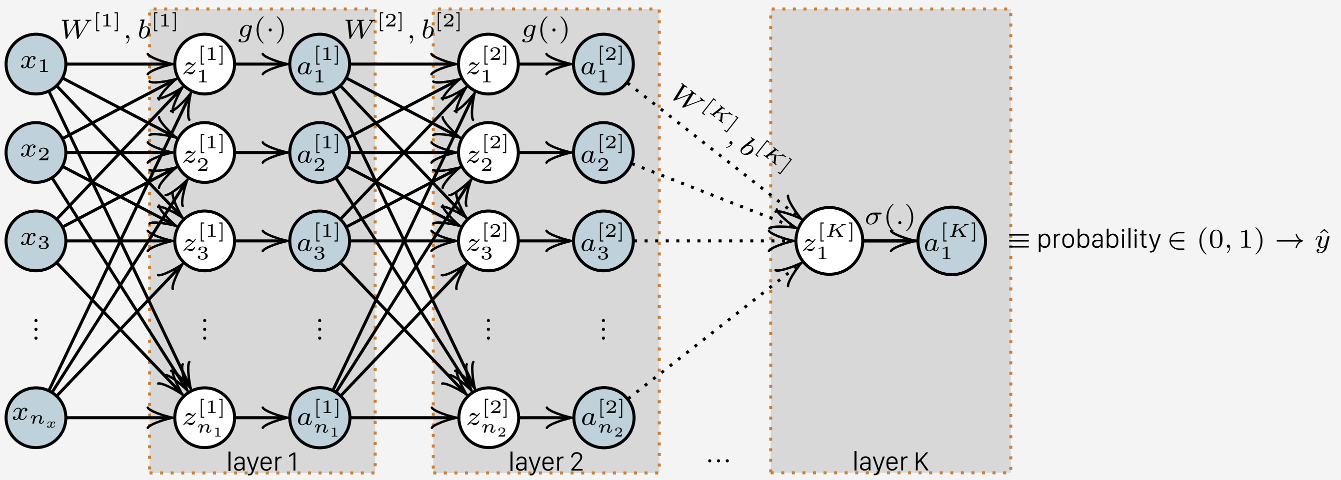
# Historical Remarks: Connectionism

## Single layer feedforward neural network



# Historical Remarks: Connectionism

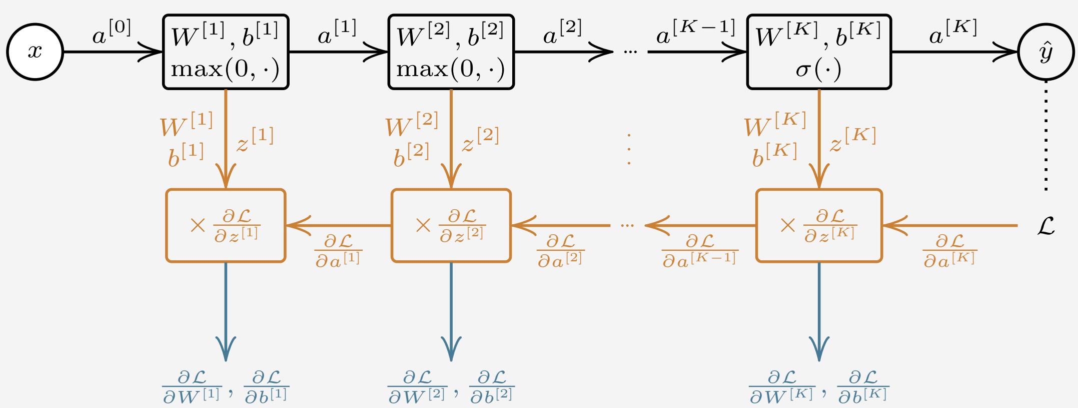
## Deep feedforward neural network






# Historical Remarks: Connectionism

## Backprop



# Historical Remarks: Deep Learning

- 2006: NN research revival with deep belief network [†] efficiently trained using using greedy layer-wise pre-training technique
- → It is possible to train deeper neural networks with focus on theoretical importance of depth [‡]
- Emergence of large datasets and increase of model size
- Deep neural networks start to outperform AI systems based on other ML technologies
- Deep neural networks start to outperform techniques based on hand-designed functionality

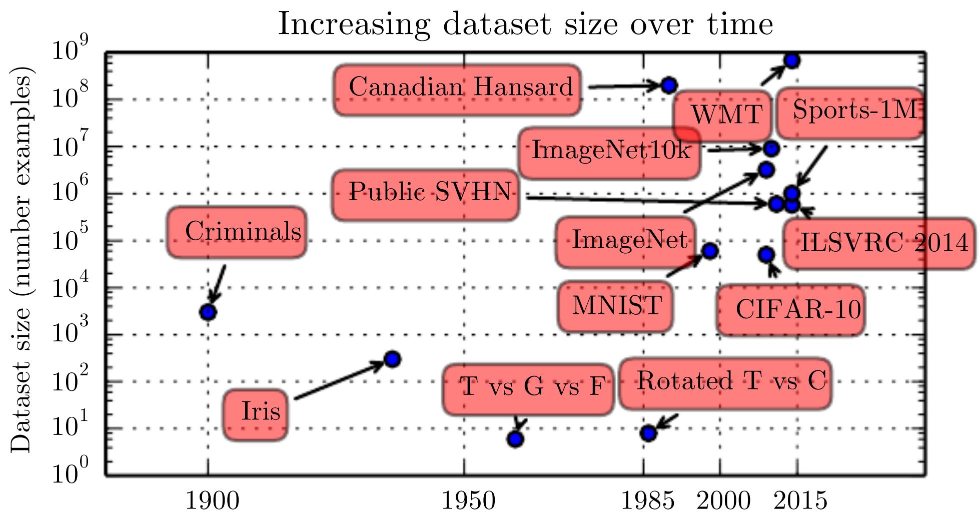
[†] Hinton, G. E., Osindero, S., and Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18, 1527–1554

[‡] Bengio, Y. and LeCun, Y. (2007). Scaling learning algorithms towards AI. In *Large Scale Kernel Machines*.



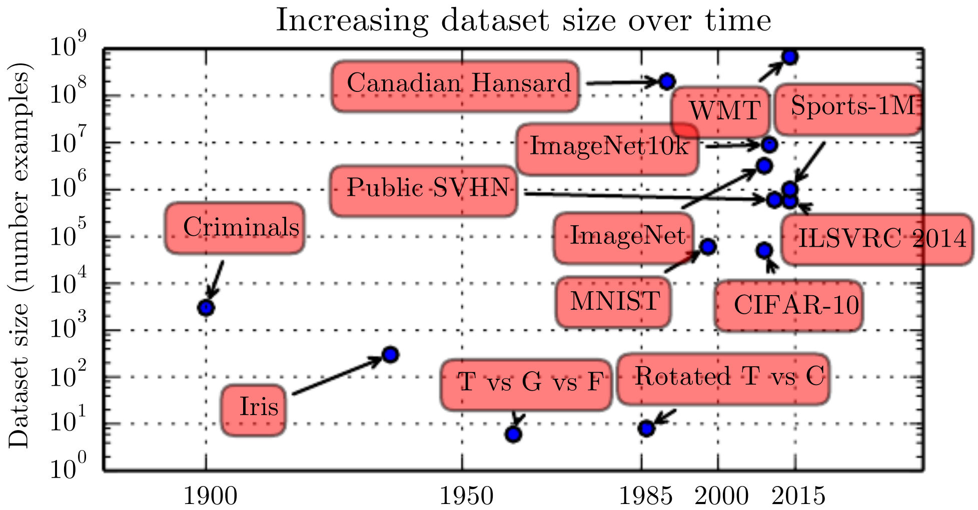
# Emergence of Large Datasets

- Training shallow networks on small datasets is somewhat of an art
- DL algorithms reaching super-human performance are similar to models in the 80s, main difference is dataset size



# Emergence of Large Datasets

- Rule of thumb for DL supervised algorithms:
  - Acceptable performance with 5k labelled examples per category
  - Super human performance with 10k examples per category



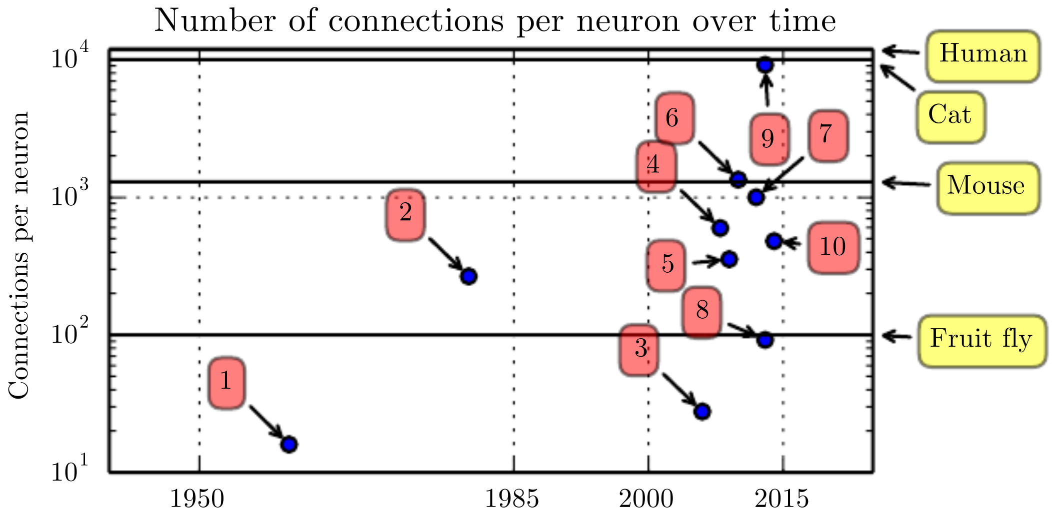



# Increase in Model Sizes

- 2010's: computational resources to train much larger datasets
- Connectionism insight:
  - animals become intelligent when many neurons work together
  - individual neuron is not very useful
- Biological neurons are not particularly densely connected
- Number of neuron units quite small until recently
- Today's networks smaller than frog's nervous systems
- With faster computing (GPU, specific HW), larger memory, larger datasets $\rightarrow$ models with human brain capacity in $\sim$2050s

# Increase in Model Sizes

- Number of neuron connection comparable to mammal since 80's
- Design consideration



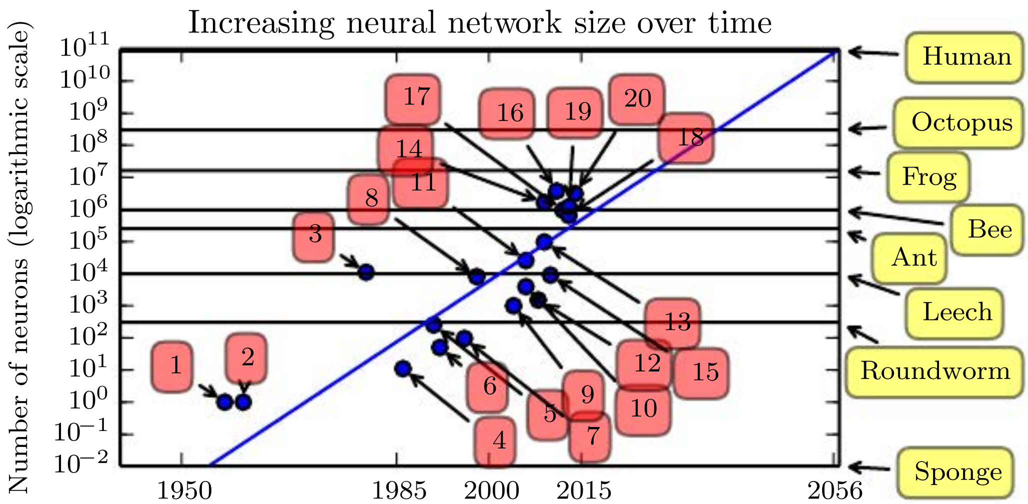

§

§1-perceptron, 2-Neocognitron, 3-GPU CNN, 4-Deep Boltzmann machine, 5-Unsupervised CNN, 6-GPU MLP, 7-Distributed encoder, 8-AlexNet, 9-HPC unsupervised CNN, 10-GoogLeNet



# Increase in Model Sizes

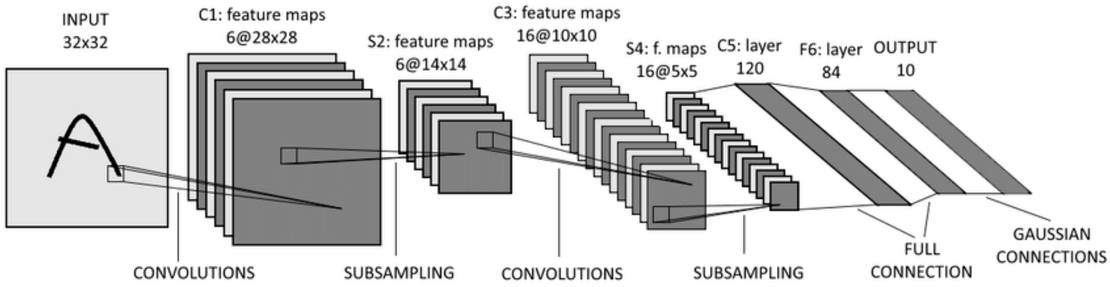
- Since MLP introduction, network size $\times 2$ every 2.4 years

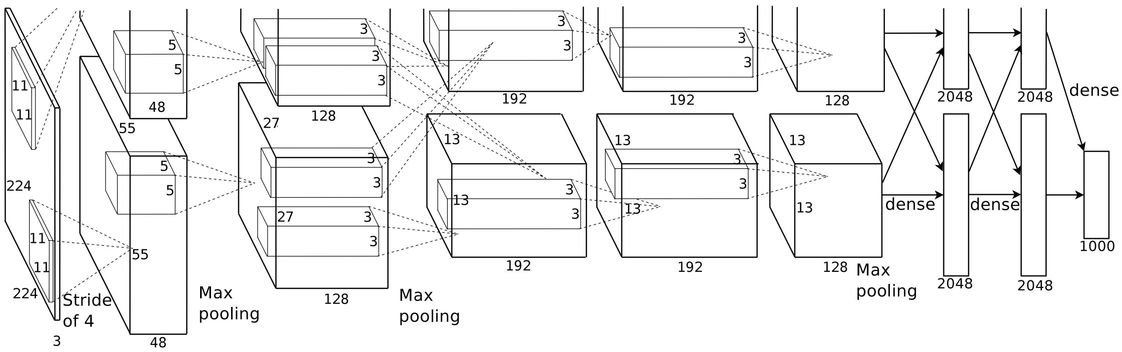



§

§1-perceptron, 2-Adaline, 3-Neocognitron, 4-Early backprop, 5-Speech RNN, 6-MLP for speech, 7- Mean field sigmoid belief network, 8-LeNet5, 9- Echo state network, 10-Deep Belief network, 11-GPU CNN, 12-Deep Boltzmann machine, 13-GPU DBN, 14-Unsupervised CNN, 15-GPU MLP, 16-OMP1 network, 17-Distributed Autoencoder, 18-AlexNet, 19-HPC unsupervised CNN, 20-GoogLeNet

# Convolutional Neural Network

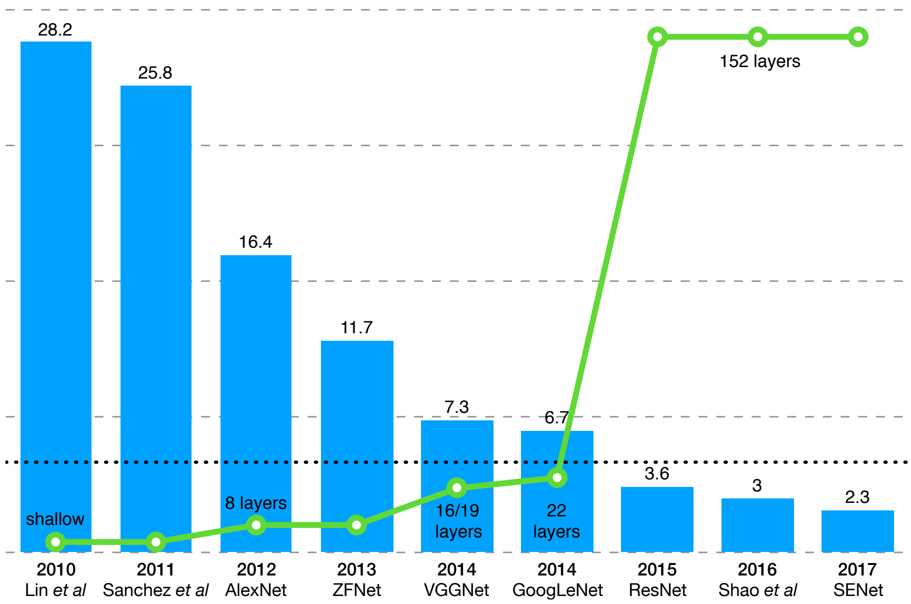




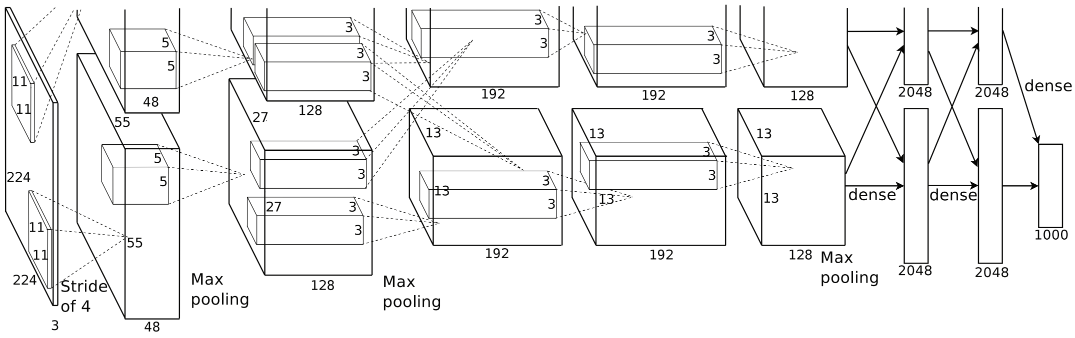

# ImageNet Competition

- 1,431,167 images with 1,000 object categories



# ImageNet Competition

- 1,431,167 images with 1,000 object categories
- AlexNet: 5 CONV + 3 FC layers

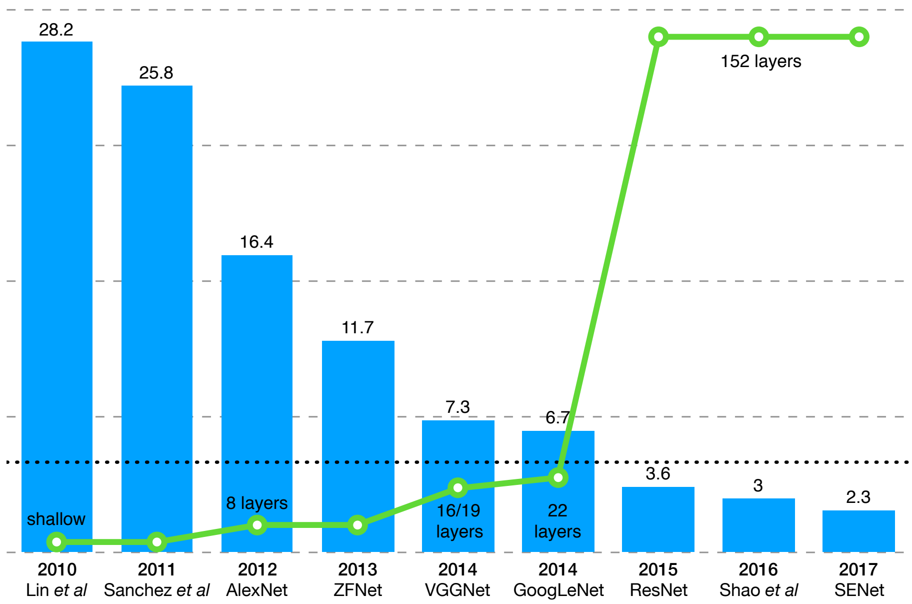






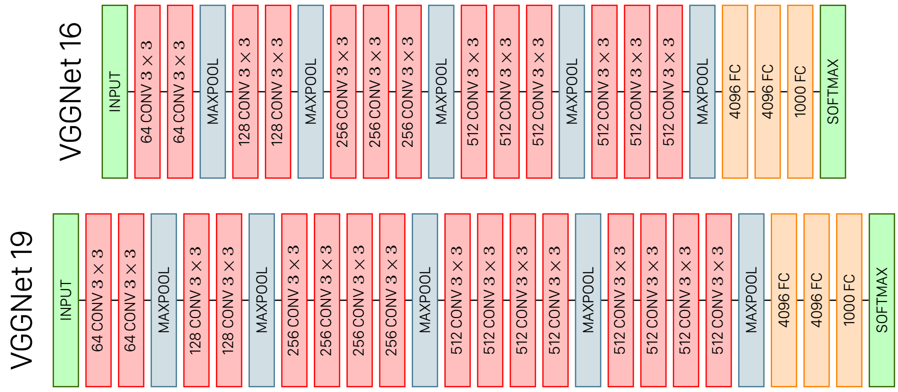
# ImageNet Competition

- 1,431,167 images with 1,000 object categories



# ImageNet Competition

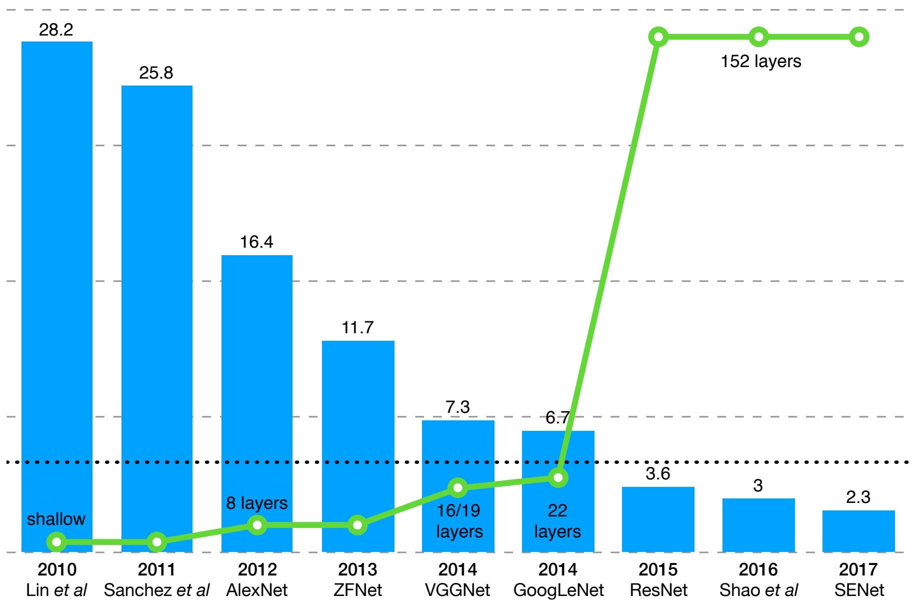
- 1,431,167 images with 1,000 object categories
- VGGNet: 16/19 CONV layers





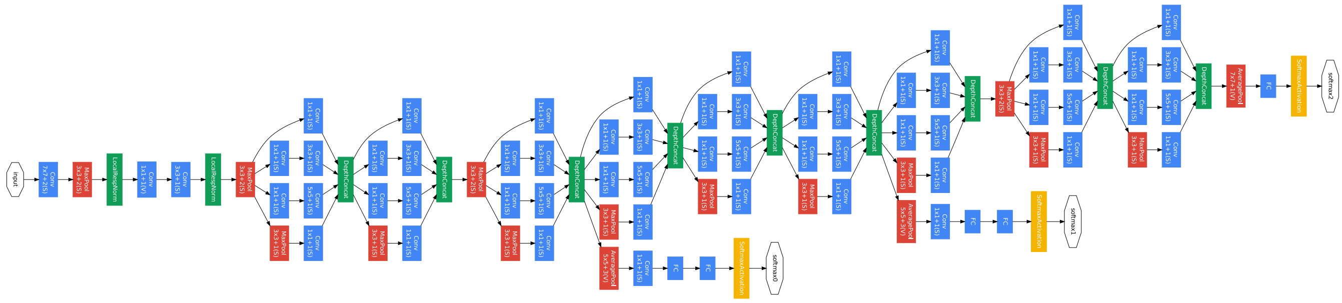

# ImageNet Competition

- 1,431,167 images with 1,000 object categories



# ImageNet Competition

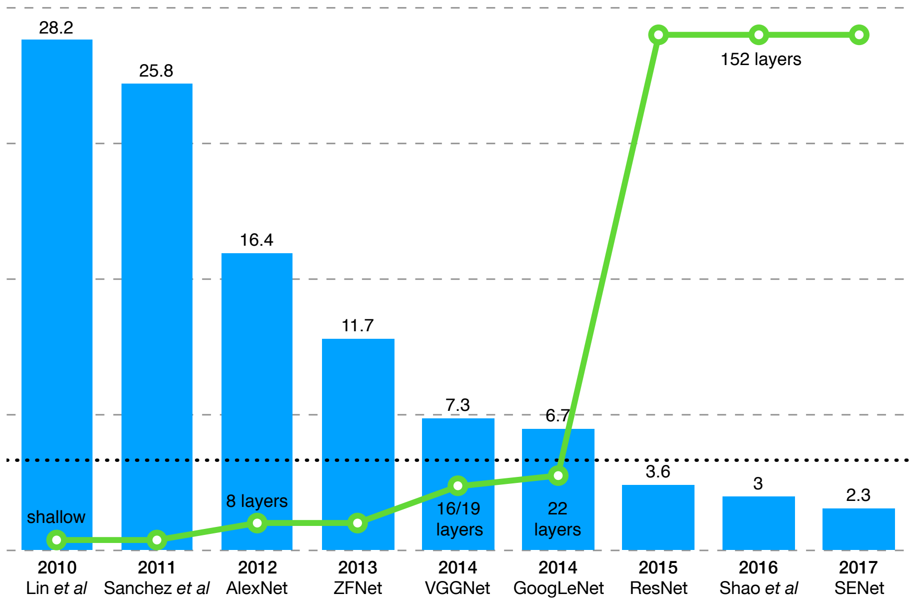
- 1,431,167 images with 1,000 object categories
- GooLeNet: 22 inception CONV layers





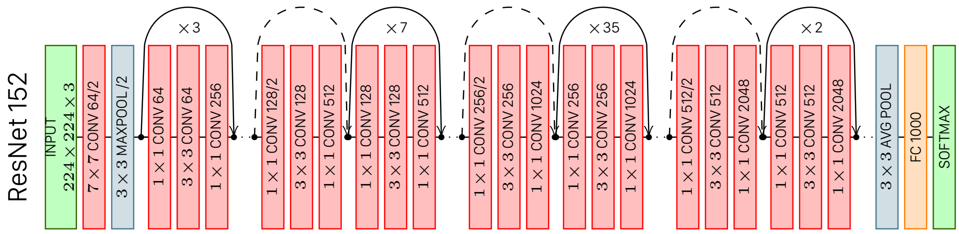
# ImageNet Competition

- 1,431,167 images with 1,000 object categories



# ImageNet Competition

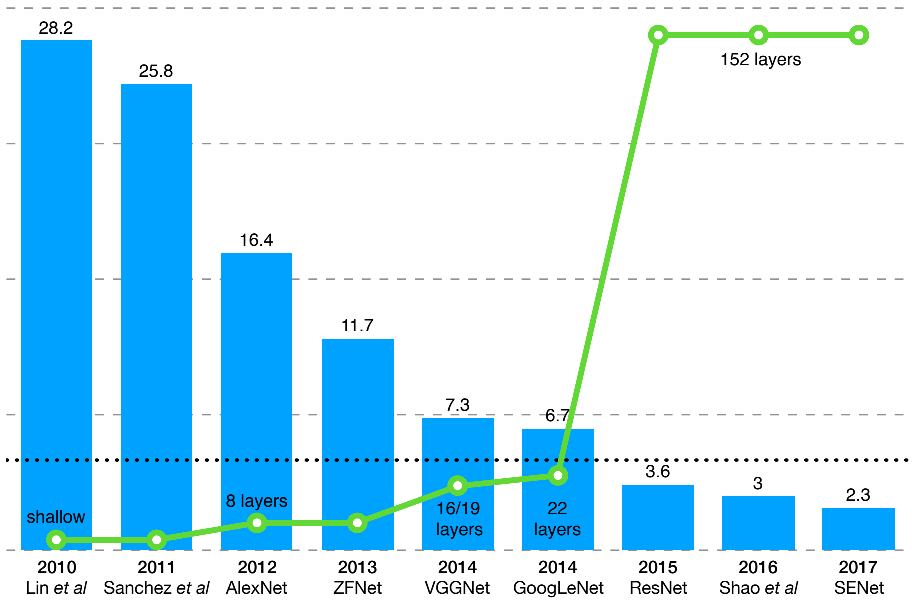
- 1,431,167 images with 1,000 object categories
- ResNet: 152 residual CONV layers



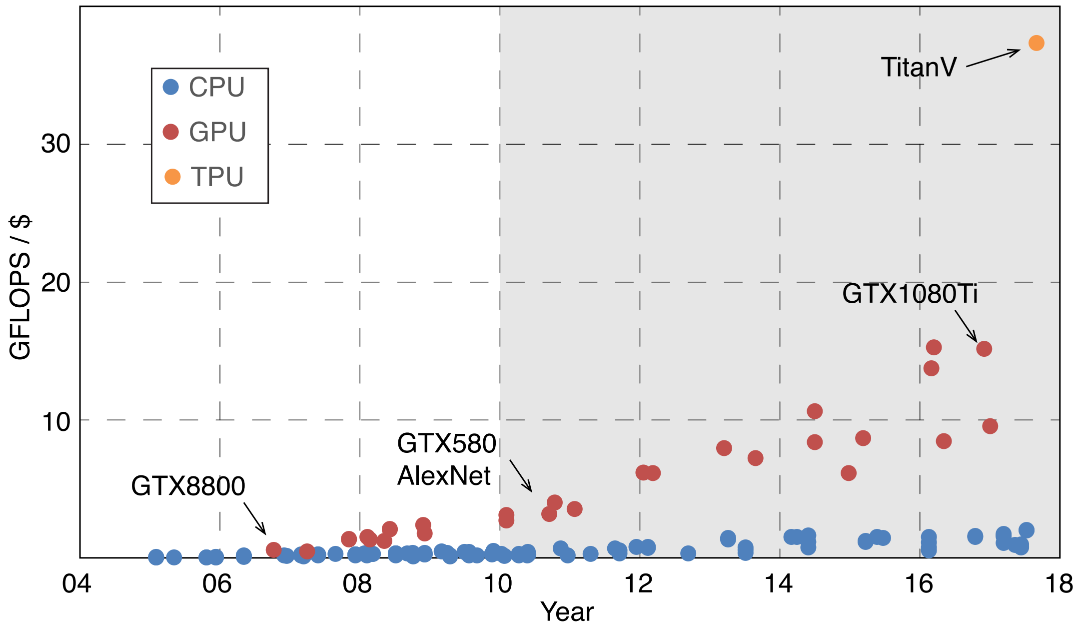



# ImageNet Competition

- 1,431,167 images with 1,000 object categories



# Computing Cost






# Deep Learning Impact

- Since 80s, DL keeps providing better accuracy with success to broader sets of applications
- Vision:
  - Early deep models: object recognition on small cropped images
  - Modern object recognition possible on high-resolution images for 1k+ object categories
  - Dramatic moment with CNN breakthrough in ILSVRC
  - super-human traffic sign classification, object tracking, pedestrian detection
    $\rightarrow$ autonomous car almost a reality

## Deep Learning Impact

- Since 80s, DL keeps providing better accuracy with success to broader sets of applications
- Audio:
  - After stagnation in 90's, error rates $\div 2$ in speech recognition using DL
  - High quality audio synthesis e.g. wavenet



## Deep Learning Impact

- Since 80s, DL keeps providing better accuracy with success to broader sets of applications
- NLP and AI:
  - Automatic Image captioning, combining CNN and RNN
  - Sequence-to-sequence modelling using LSTM $\rightarrow$ revolutionising machine translation
  - AlphaZero (Deepmind): improved significantly reinforcement learning state-of-the-art (go, chess, protein folding)
  - Transformer: explicit attention with mind blowing performance on many tasks (OpenAI GPT, Dall-E and CLIP, Google ImageGen and MUSE, Deepmind AlphaFold)

## Deep Learning Impact

- Since 80s, DL keeps providing better accuracy with success to broader sets of applications
- Highly profitable applications with tremendous industrial interests
- Availability of advanced open-source frameworks: Tensorflow, PyTorch. Caffe, MXNet, ...
- DL expected to impact other sciences providing powerful tools to analyse, process and model vast amount of data e.g. medical drugs, quest for subatomic particles, analysis of 3D map of brain, ...



## Objectives

- Build deep neural network
- Setting up datasets (train, validation and test)
- Learn how to train it to improve performance
- Use regularisation effectively
- Exposure to various optimisation algorithms (SGD, momentum, adam, rmsprop, ...)
- Implement MLP and CNN from scratch
- Apply DL frameworks in a few case studies and project