Introduction to Deep Learning

Computer Vision

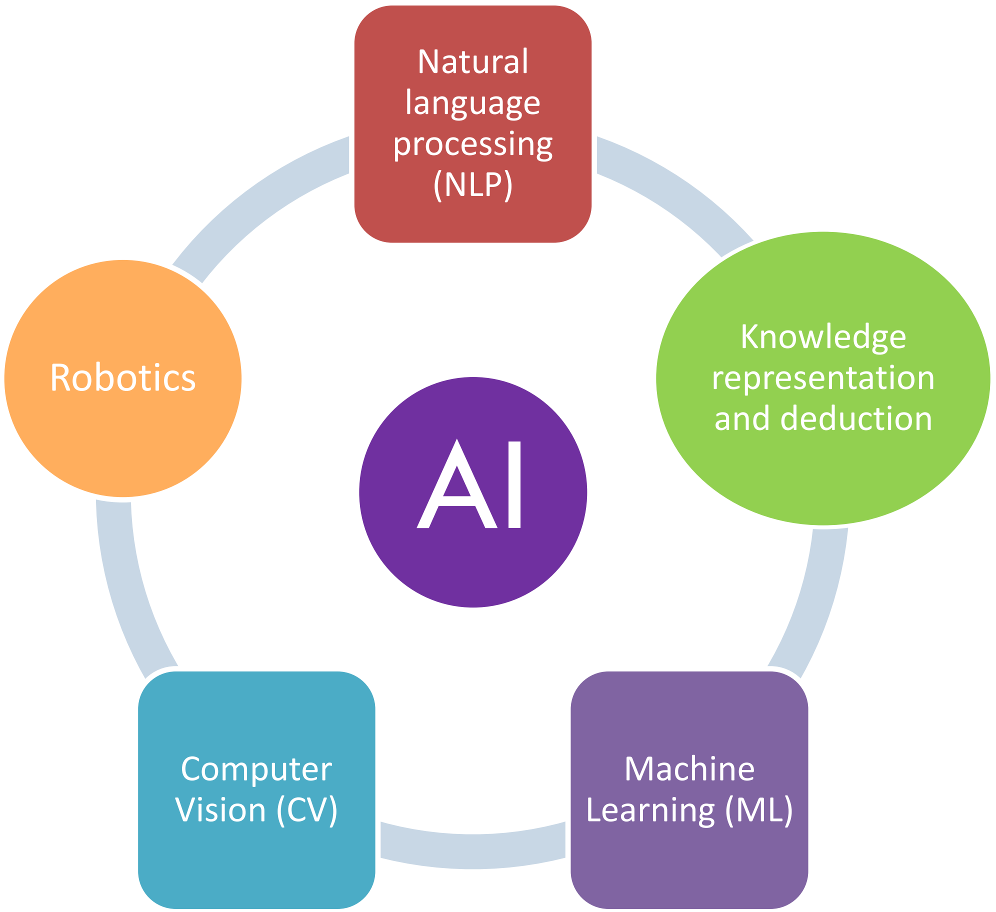
Aziz M. Qaroush

Birzeit University

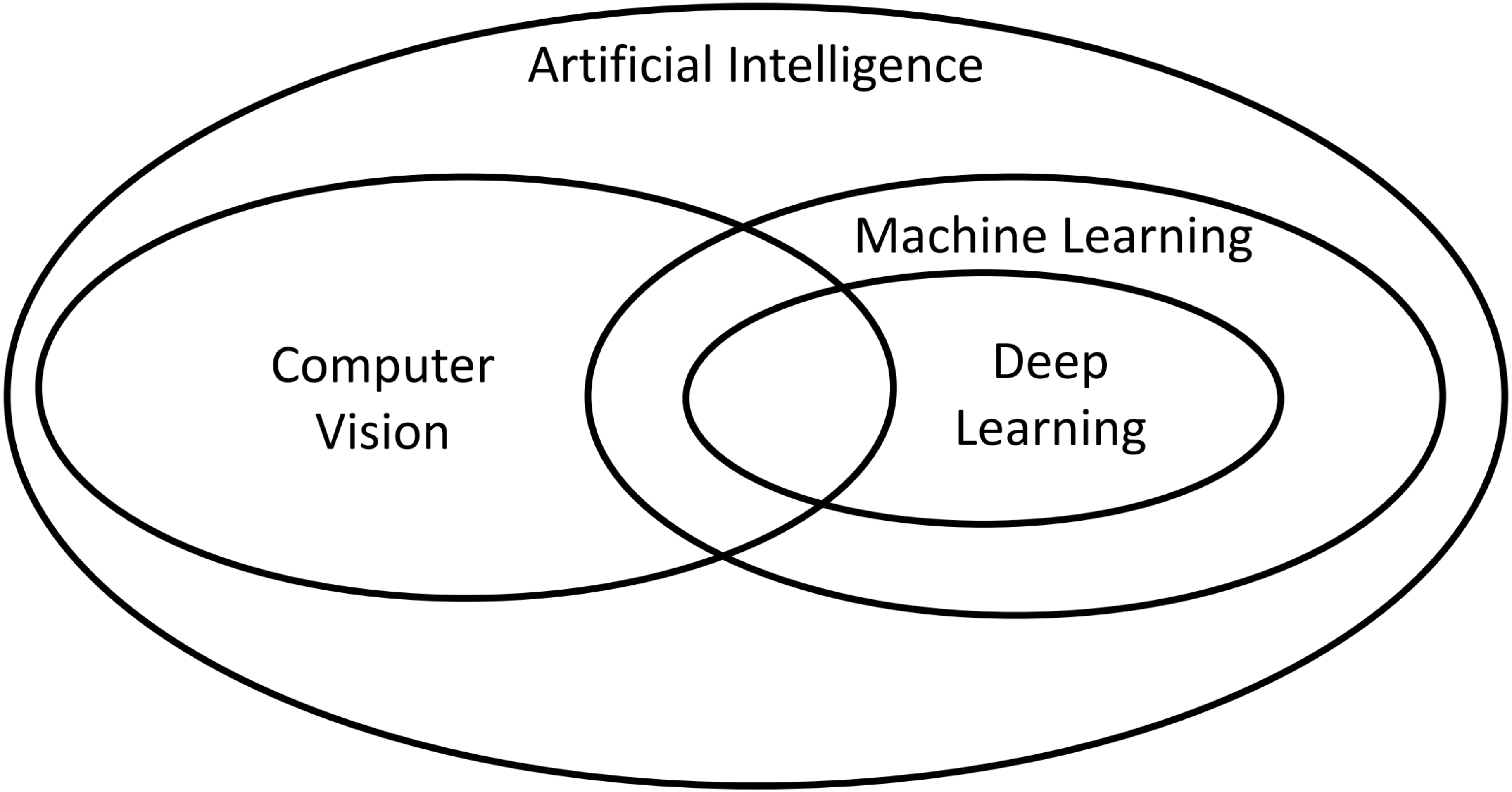
Outline

- Motivation
- Machine Learning Basics
- Machine Learning History
- Performance of Deep Learning Networks
- Applications of Deep Learning in Computer Vision

Artificial Intelligence Areas

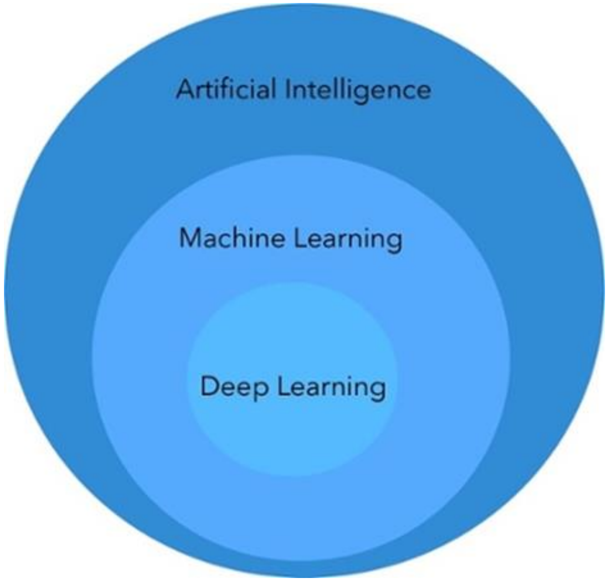


Computer Vision and AI



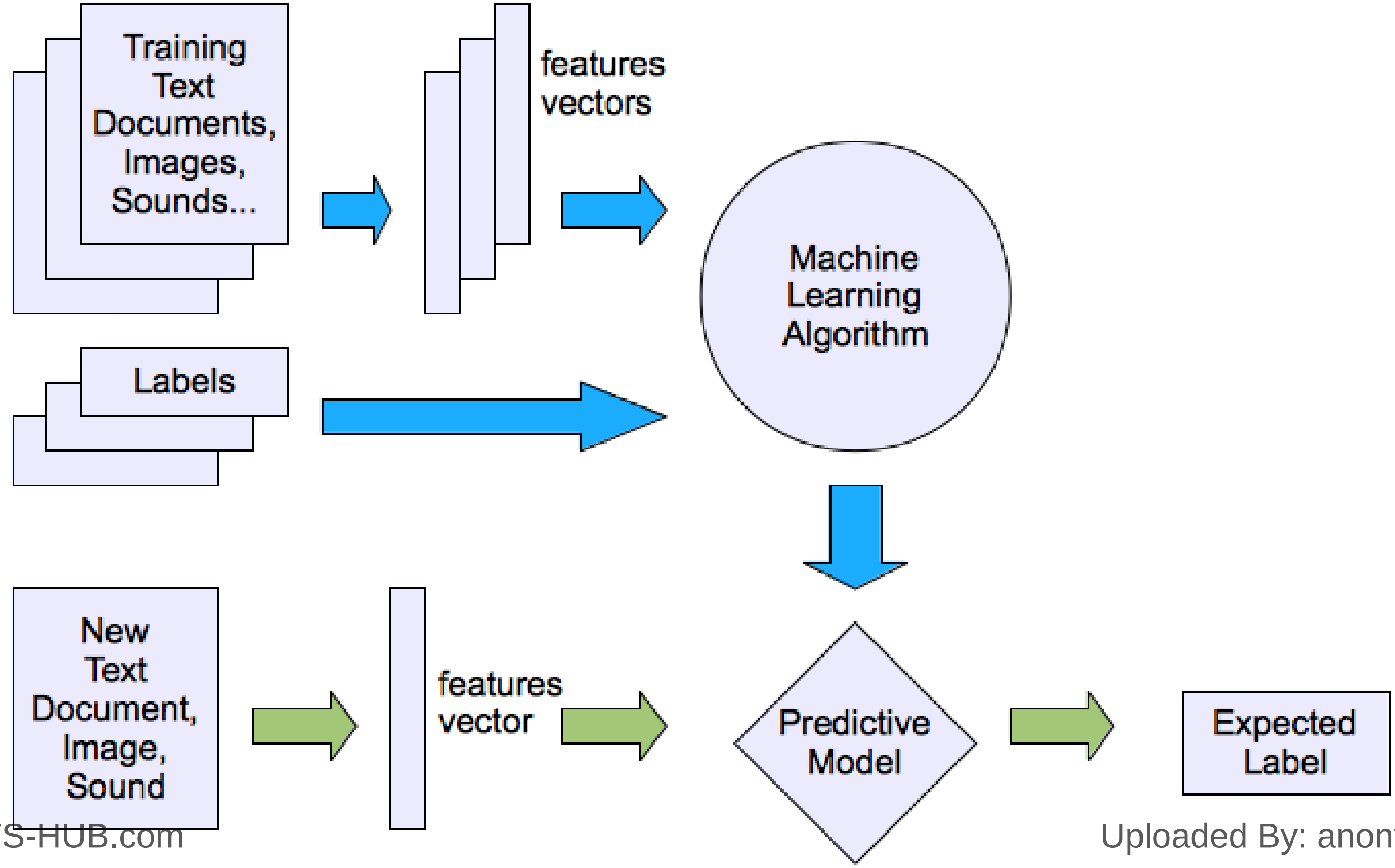
Machine Learning Basics

- Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed
- It is a branch of artificial intelligence, concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data.
- Most machine learning methods work well because of human-designed representations and input features ML becomes just optimizing weights to best make a final prediction



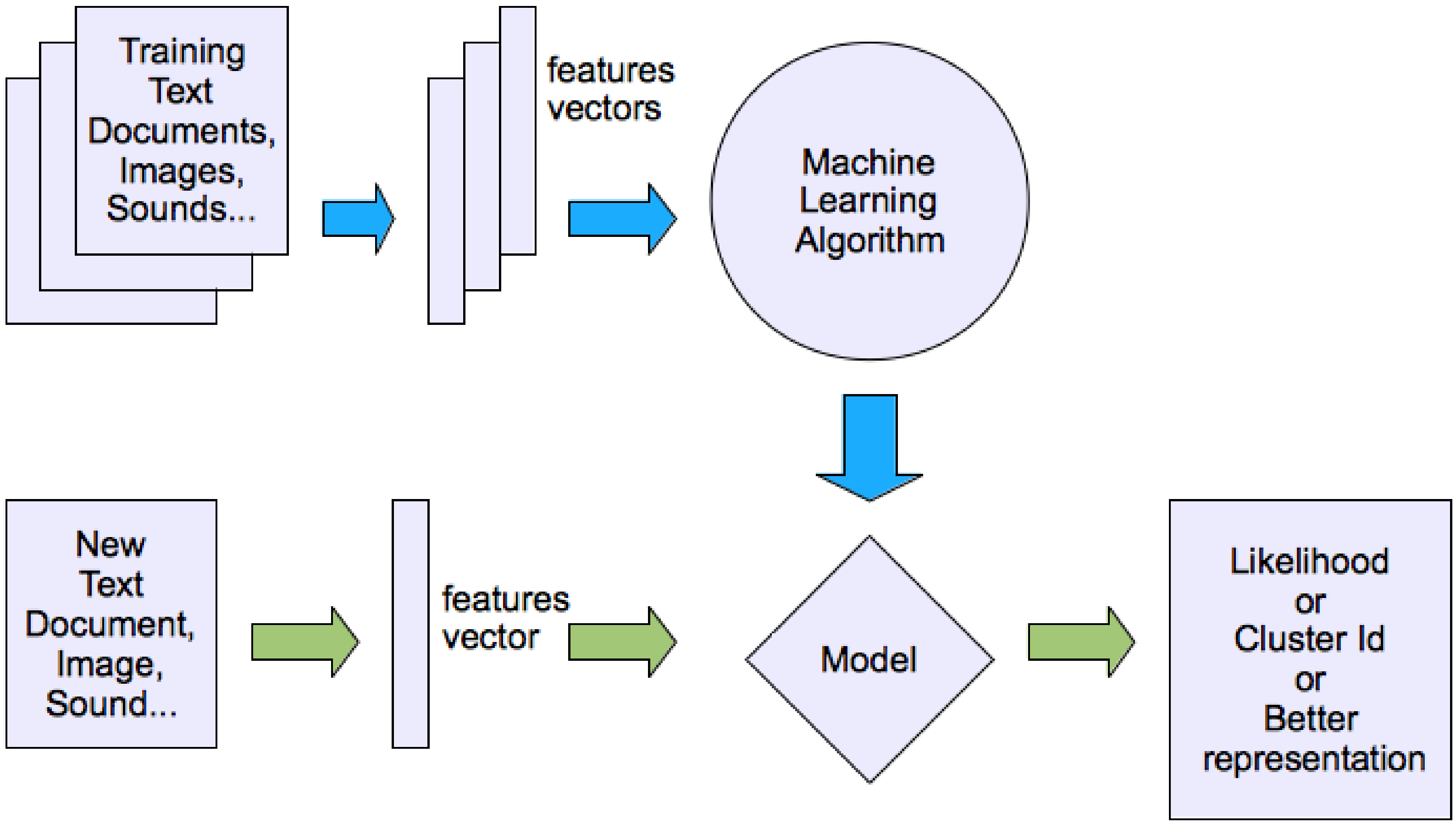
Types of Learning

- **Supervised**: Learning with a labeled training set
  - Example: email classification with already labeled emails



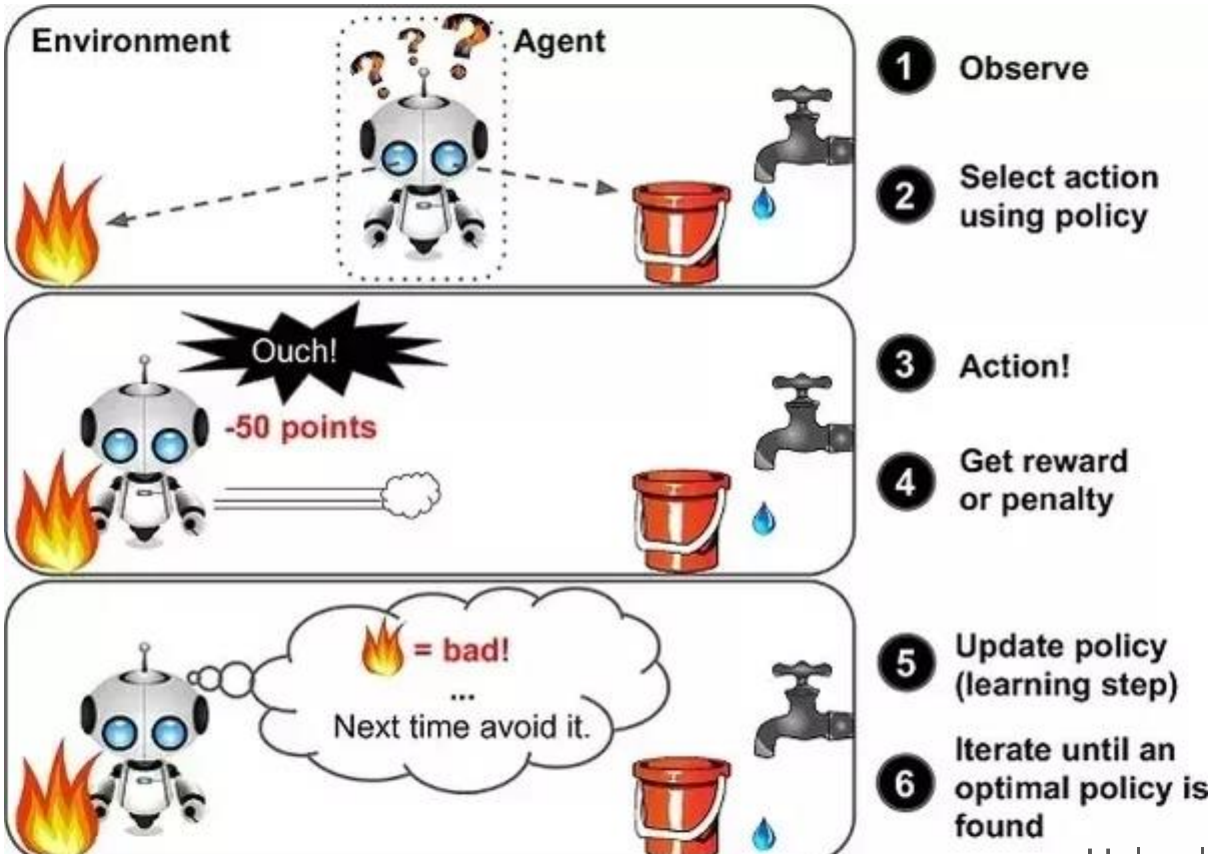
Types of Learning

- **Unsupervised**: Discover patterns in unlabeled data
  - Example: cluster similar documents based on text



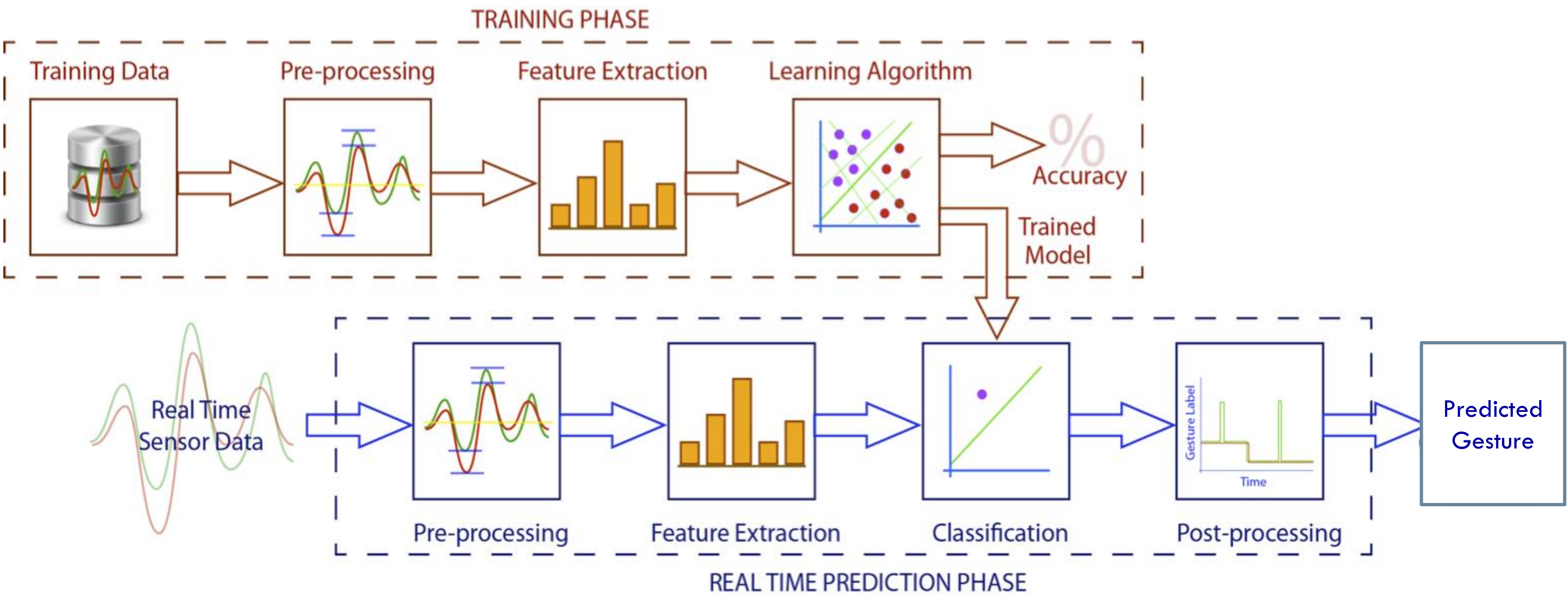
Types of Learning

- **Reinforcement learning**: learn to act based on feedback/reward
  - Example: learn to play Go, reward: win or lose



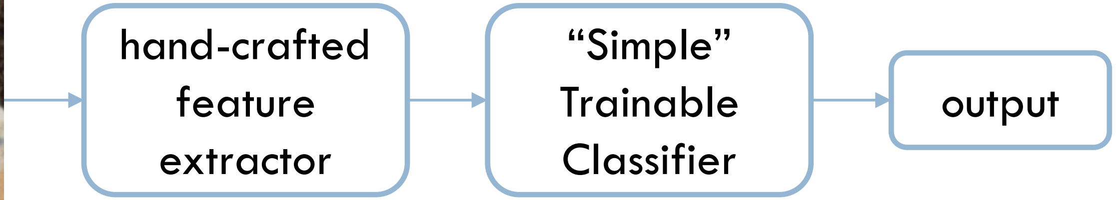
Machine Learning Cycle

Gesture Prediction – An Example



Classical ML Approach

- Traditional pattern recognition models use hand-crafted features and relatively simple trainable classifier.



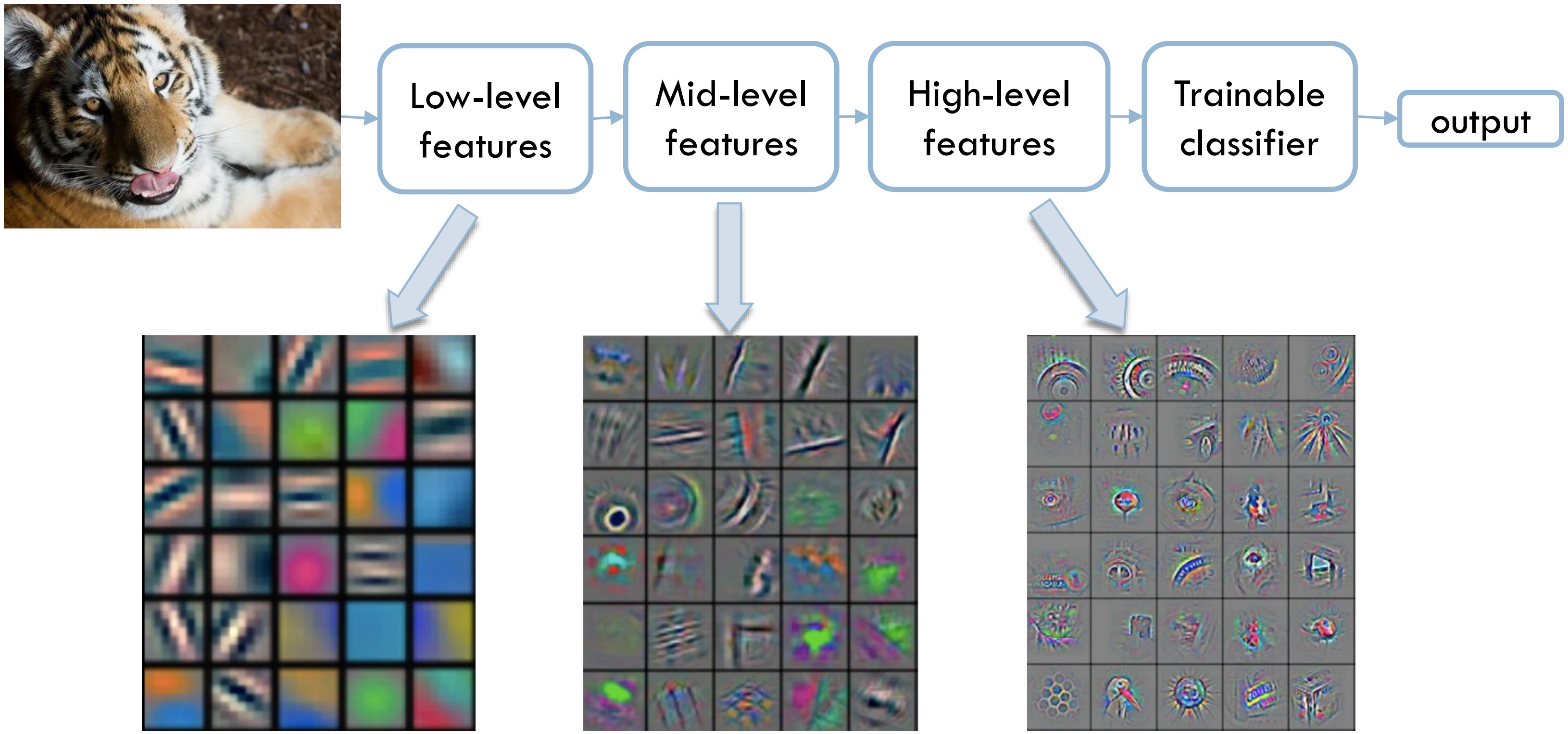
- This approach has the following limitations:
  - It is very tedious and costly to develop hand-crafted features
  - The hand-crafted features are usually highly dependents on one application, and cannot be transferred easily to other applications
- How to know what features/representation is best?
  - Difficult to know!
  - Why not learn the representation as well in addition to the mapping from representation to output
  - Learned representations many times result in much better performance

Deep Learning Approach

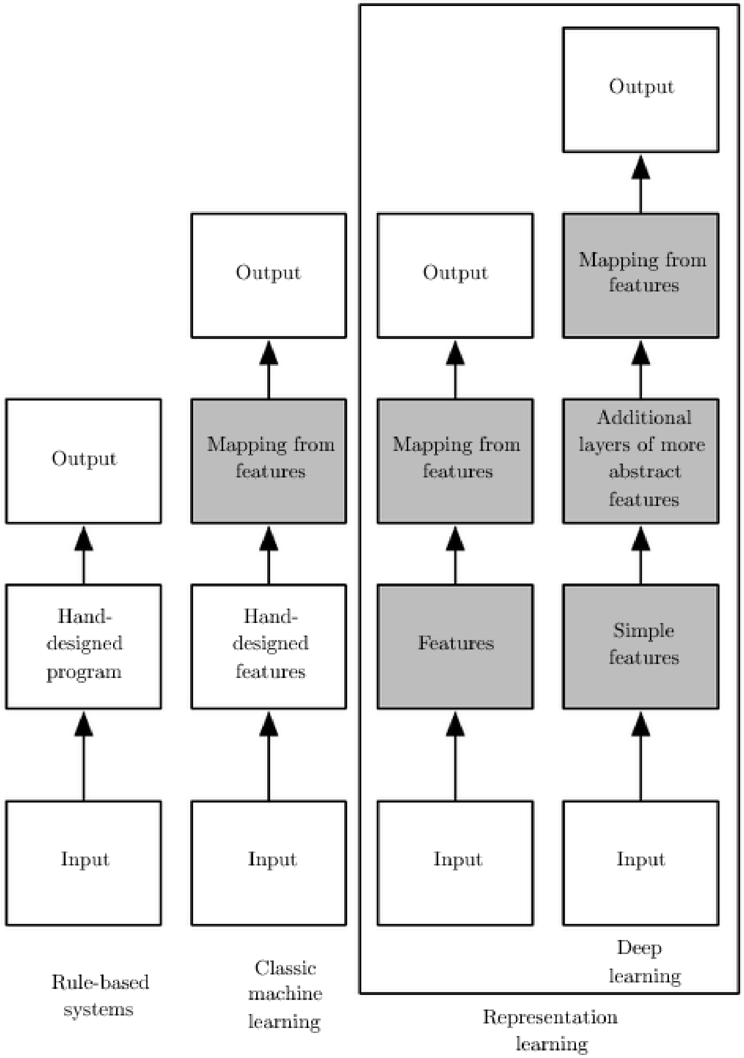
- **Deep Learning** is "a subset of **Machine Learning that achieves great power and flexibility by learning to represent the world as nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones".**
- Practically, Deep Learning' means using a neural network with several layers of nodes between input and output.
- The series of layers between input & output do feature identification and processing in a series of stages, just as our brains seem to.

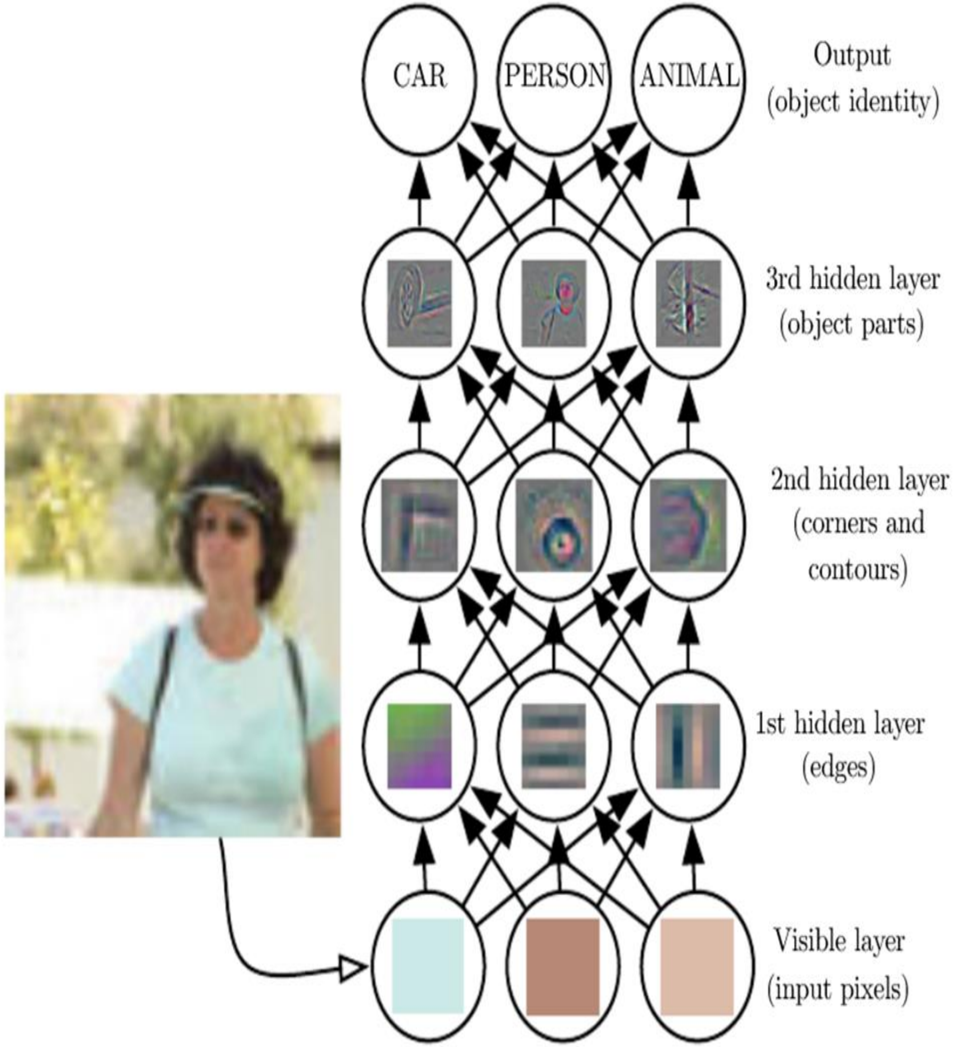
Deep Learning Approach

- Deep learning (a.k.a. representation learning) seeks to learn rich hierarchical representations (i.e. features) automatically through multiple stage of feature learning process.

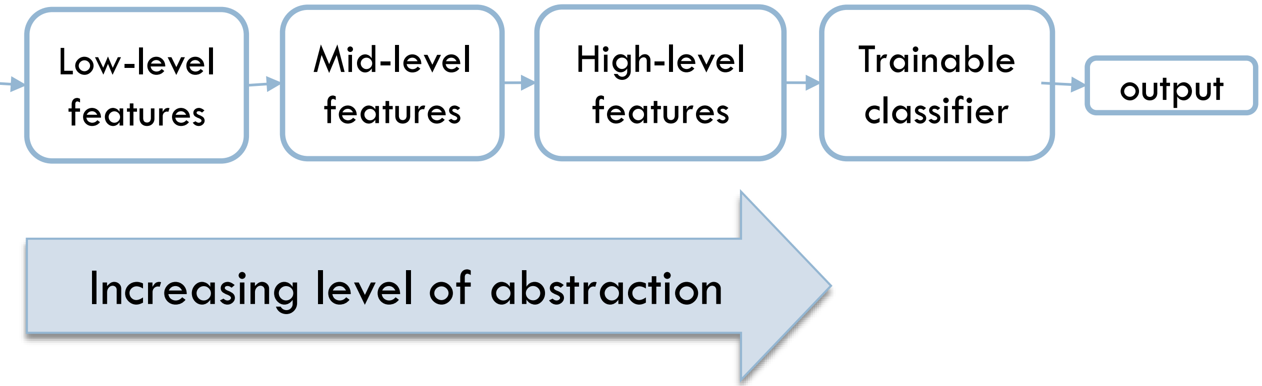


Learning Hierarchical Representations





Learning Hierarchical Representations

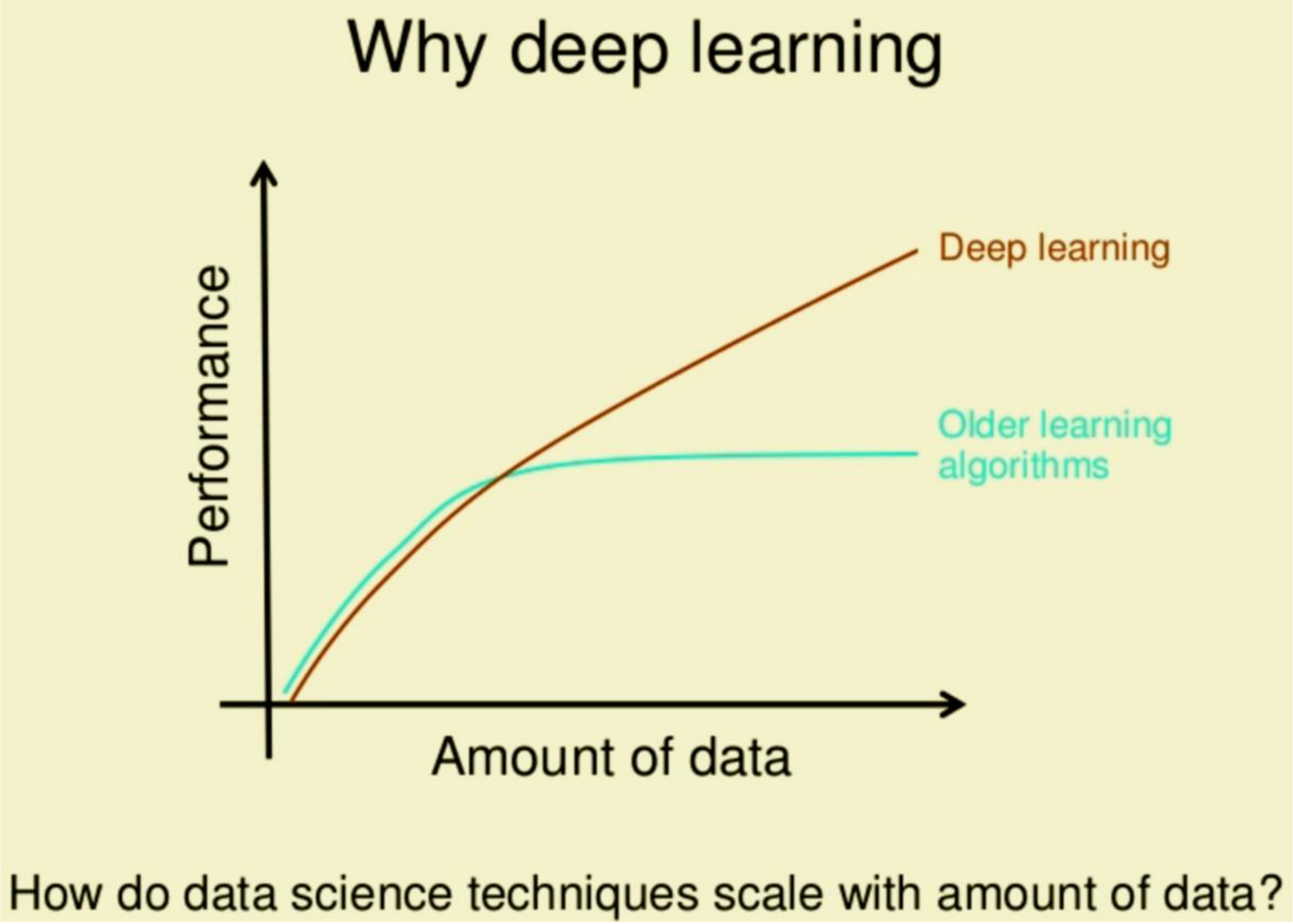


- Hierarchy of representations with increasing level of abstraction. Each stage is a kind of trainable nonlinear feature transform
- Image
  - Pixel → edge → texton → motif → part → object
- Text
  - Character → word → word group → clause → sentence → story

So, Why is DL Useful?

- Handling Complex Data: DL excels at processing and understanding complex, high-dimensional data such as images, audio, text, and video.
- Automated Feature Learning: DL models are capable of automatically learning hierarchical features from data.
  - Manually designed features are often over-specified, incomplete and take a long time to design and validate
  - Learned Features are easy to adapt, fast to learn
- Deep learning provides a very flexible, (almost?) universal, learnable framework for representing world, visual and linguistic information.
- **Adaptability:** DL models can be adapted for various tasks through transfer learning.
- Can learn both unsupervised and supervised
- Effective end-to-end joint system learning

DL Vs. Classical ML in Term of Performance



What made this possible?

- Big Data
  - Larger data sets
  - Easier collection and storage
- Hardware Computing Power
  - Graphics Processing Units (GPUs), Google TPU, Intel Nervana, Movidius HW accelerators

- Open source software
  - Improved techniques, new models, Toolboxes
- Five decades of research in machine learning
- Resources and efforts from large corporations
- Better media coverage and success cases

When to use Deep Learning or not over others?

- **Deep Learning** out perform other techniques if the **data size is large**. But with small data size, traditional **Machine Learning** algorithms are preferable.
- **Deep Learning** techniques need to have **high end infrastructure** to train in reasonable time.
- When there is **lack of domain understanding for feature introspection**, **Deep Learning** techniques outshines others as you have to worry less about feature engineering.
- **Deep Learning** really shines when it comes to **complex problems such as image classification, natural language processing, and speech recognition**.

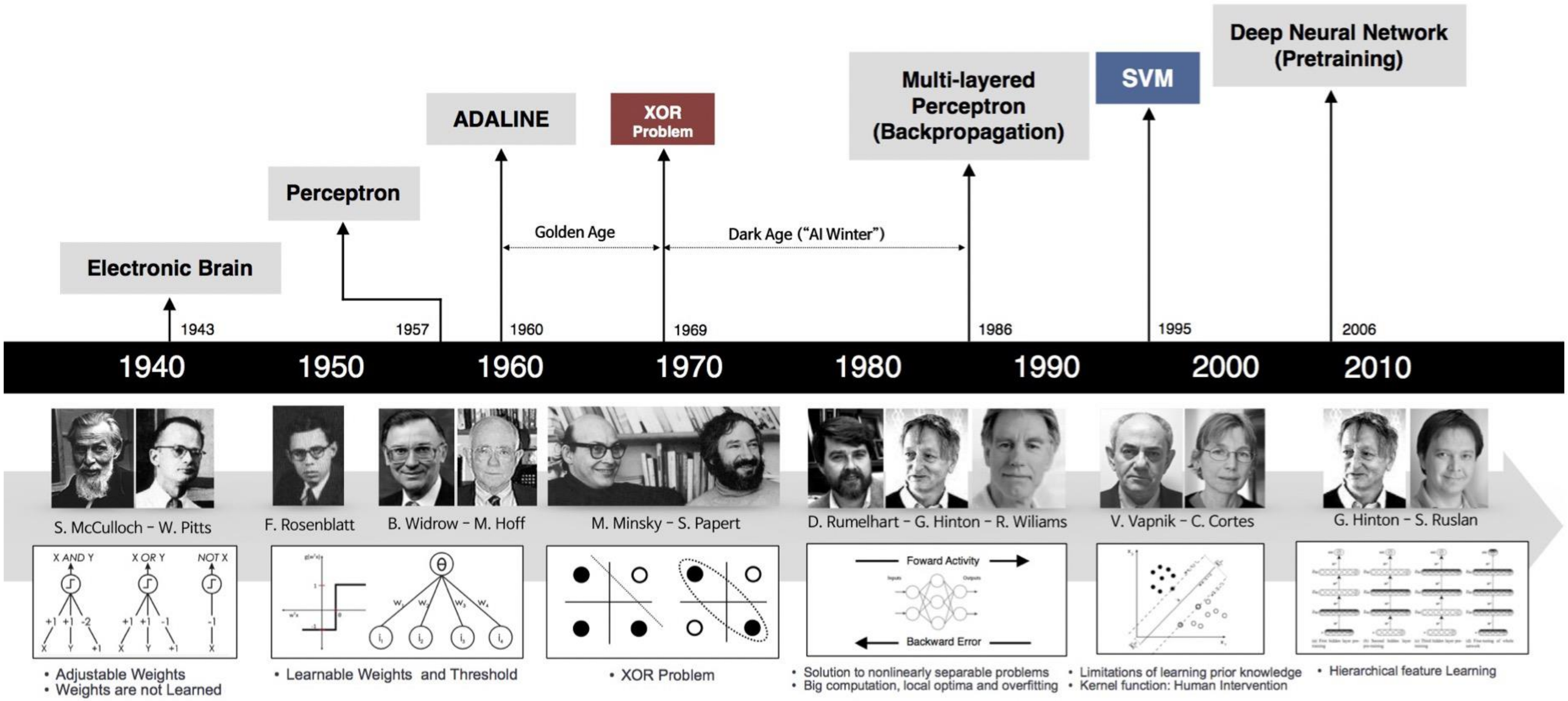
ML History

- 1805–Least squares (Adrien-Marie Legendre)
- 1812–Bayes Theorem (named after Thomas Bayes)
- 1913–Markov Chains (Andrey Markov)
- Probabilistic models–one of the earliest forms of machine learning, and it’s still widely used to this day, e.g., Naïve Bayes, HMMs.
- Early neural networks were proposed in 1950s but lacked efficient training methods.
- KNN was conceived in 1967—the beginning of basic pattern recognition.
- Rabiner’s (AT&T Bell Labs) work on HMMs in 1980s made them the model of choice for sequential data.
- Backpropagation algorithm for network training discovered in 1980s
- The first successful practical application of neural nets came in 1989 from Bell Labs (Yann LeCun), LeNet for US postal services.

ML History

- SVM was developed by Vladimir Vapnik and Corinna Cortes in the early 1990s at Bell Labs and published in 1995.
- Decisions trees began to receive significant research interest in the 2000s, and by 2010 they were often preferred to kernel methods.
- Kaggle competitions started in 2010. Gradient boosting and Random forest (Tin Kam Ho in 1995) were the top two models most of the time.
- Neural network winner–until 2012.
- The Canadian Institute for Advanced Research (CIFAR) and Neural Computation and Adaptive Perception (NCAP) research initiative:
  - The groups of Geoffrey Hinton (University of Toronto),
  - Yoshua Bengio (University of Montreal),
  - Yann LeCun (New York University and IDSIA in Switzerland) were still active in the area.
  - Others

ML History



https://www.slideshare.net/deview/251-implementing-deep-learning-using-cu-dnn/4

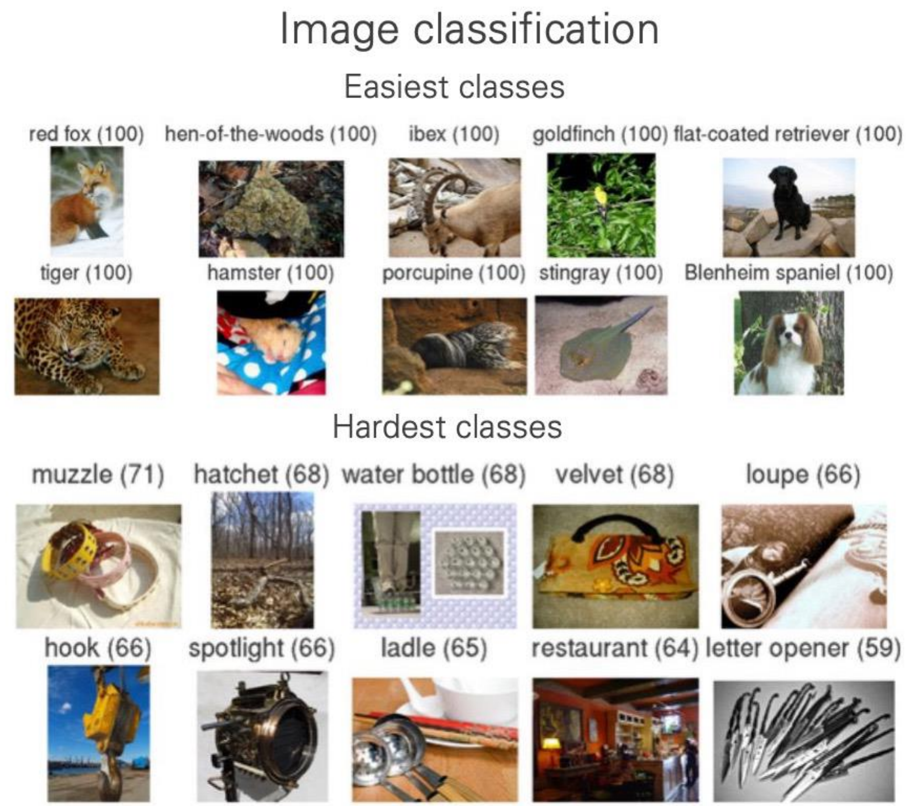
ML History

- ImageNet competitions:
  - 2011, accuracy around 74-75%
  - 2012, system from Alex Krizhevsky (supervised by Geoffrey Hinton) reached an accuracy of 83.6% (AlexNet)
  - 2015 onwards ImageNet considered a ***'solved' problem***.

- Kaggle is currently dominated by two approaches: Boosting machines (for structured data) and deep learning (for perceptual problems)

The 2012 Revolution

- IMAGENET Large Scale Visual Recognition Challenge (ILSVRC)
  - **1.2M** training images with **1K** categories
  - Measure top-5 classification error

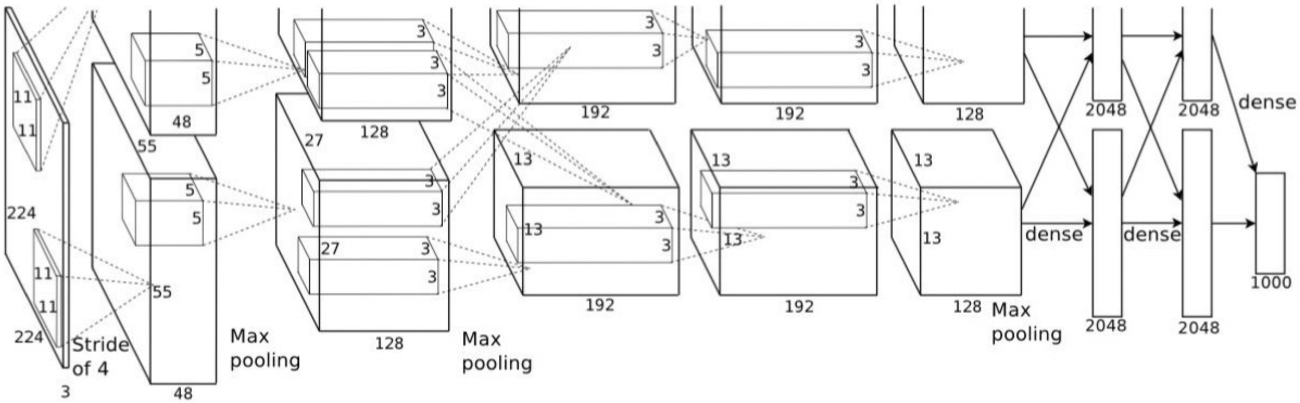


J. Deng, Wei Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei , "**ImageNet: A Large-Scale Hierarchical Image Database**", CVPR 2009.
O. Russakovsky et al., "**ImageNet Large Scale Visual Recognition Challenge**", Int. J. Comput. Vis.,, Vol. 115, Issue 3, pp 211-252, 2015.

The 2012 Revolution

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

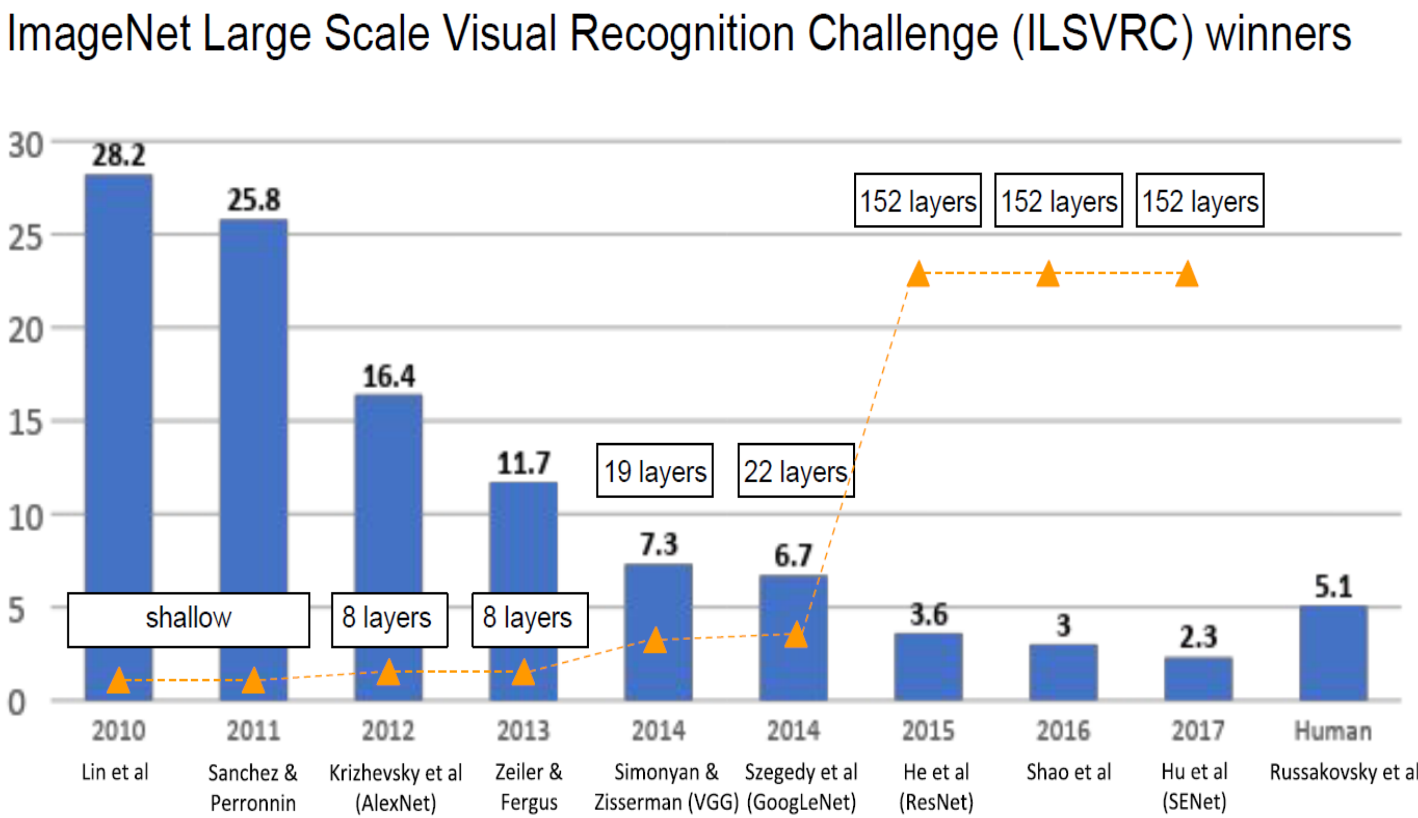
CNN based, non-CNN based

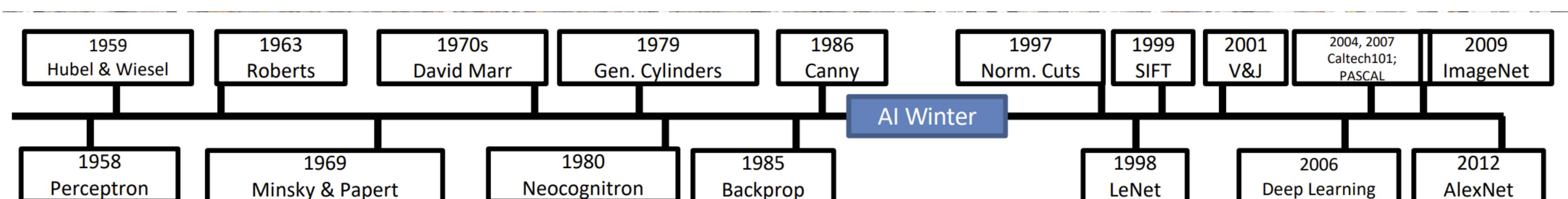


- The success of AlexNet, a deep convolutional network
  - 7 hidden layers (not counting some max pooling layers)
  - 60M parameters
- Combined several tricks
  - ReLU activation function, data augmentation, dropout

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

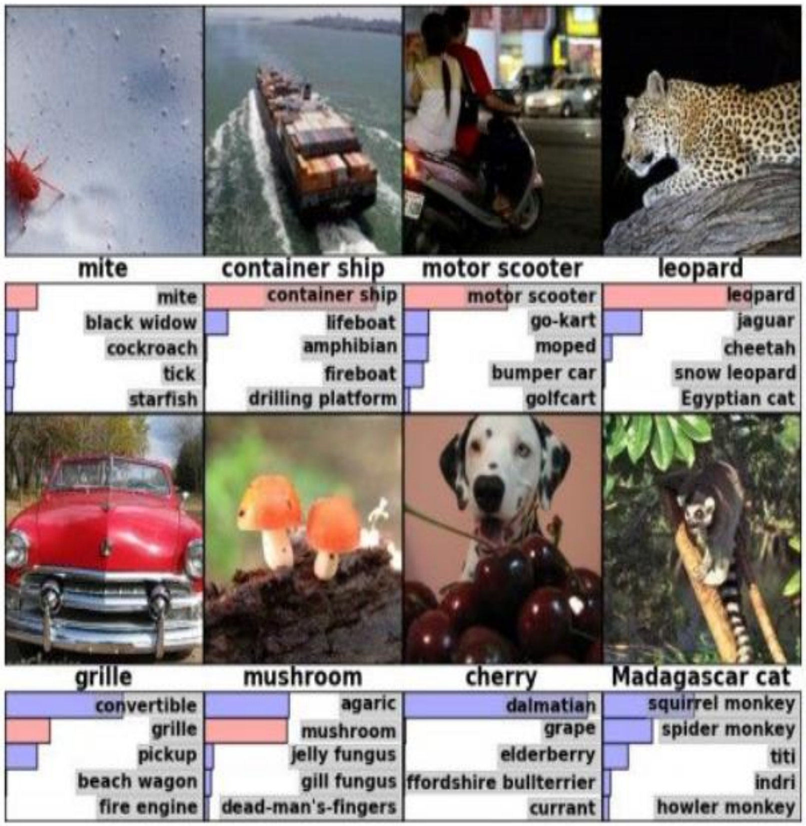
Performance of CNN Networks



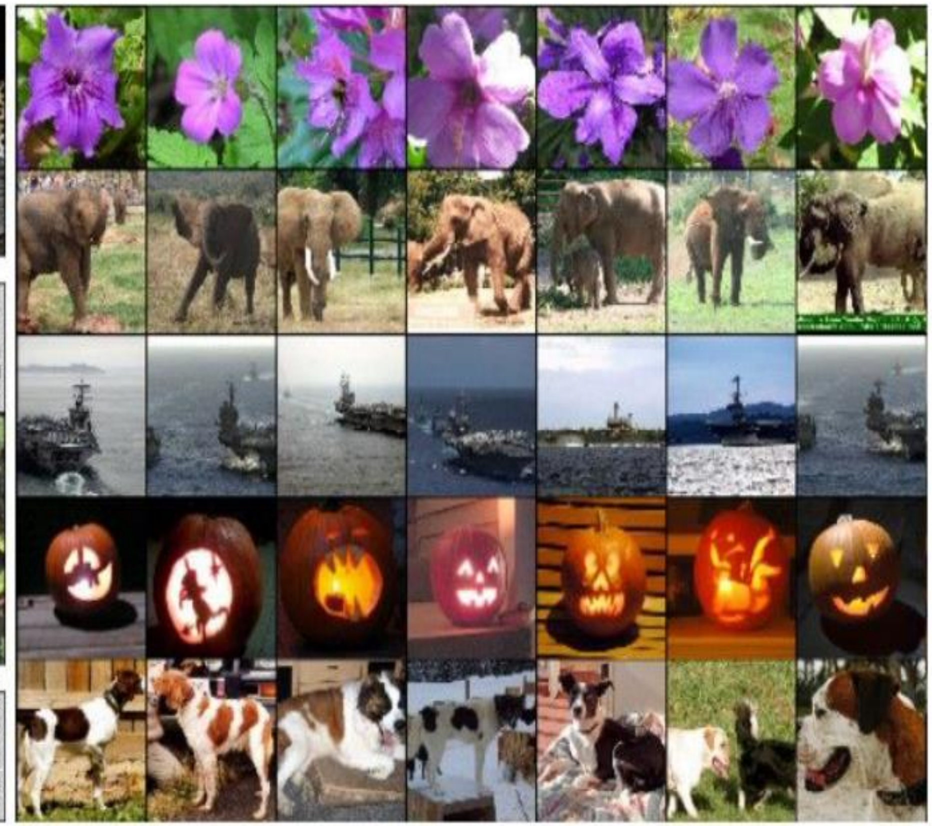


2012 to Present: Deep Learning is Everywhere

Classification

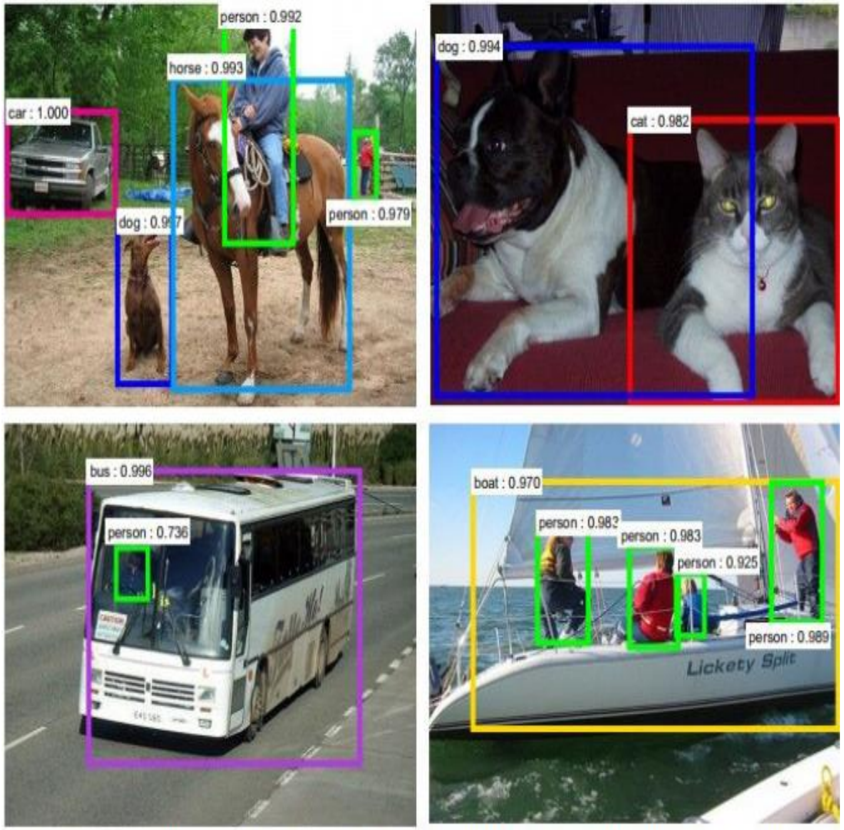


Retrieval



2012 to Present: Deep Learning is Everywhere

Detection



Figures copyright Shaoqing Ren, Kaiming He, Ross Girschick, Jian Sun, 2015. Reproduced with permission.

[Faster R-CNN: Ren, He, Girshick, Sun 2015]

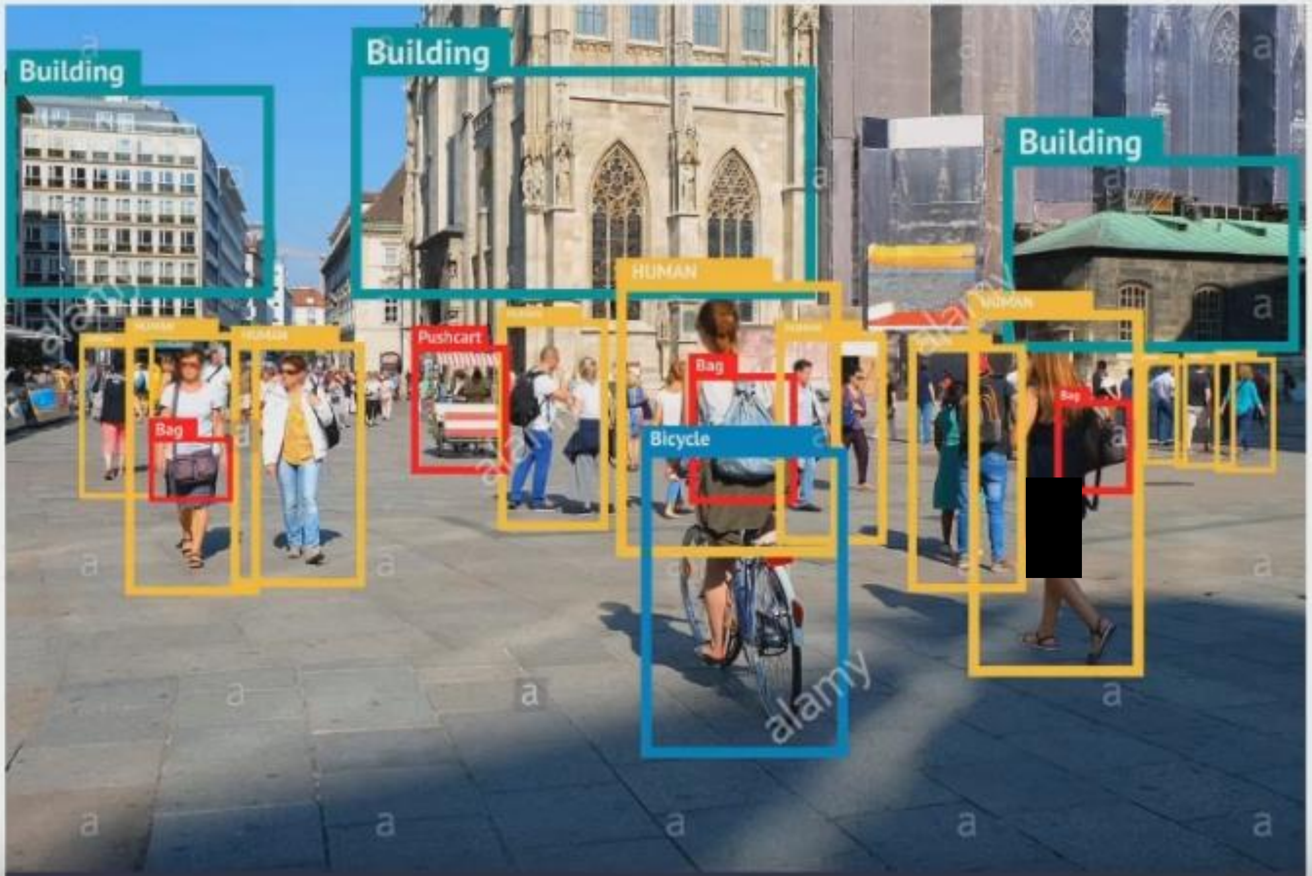
Segmentation

Figures copyright Clement Farabet, 2012. Reproduced with permission.

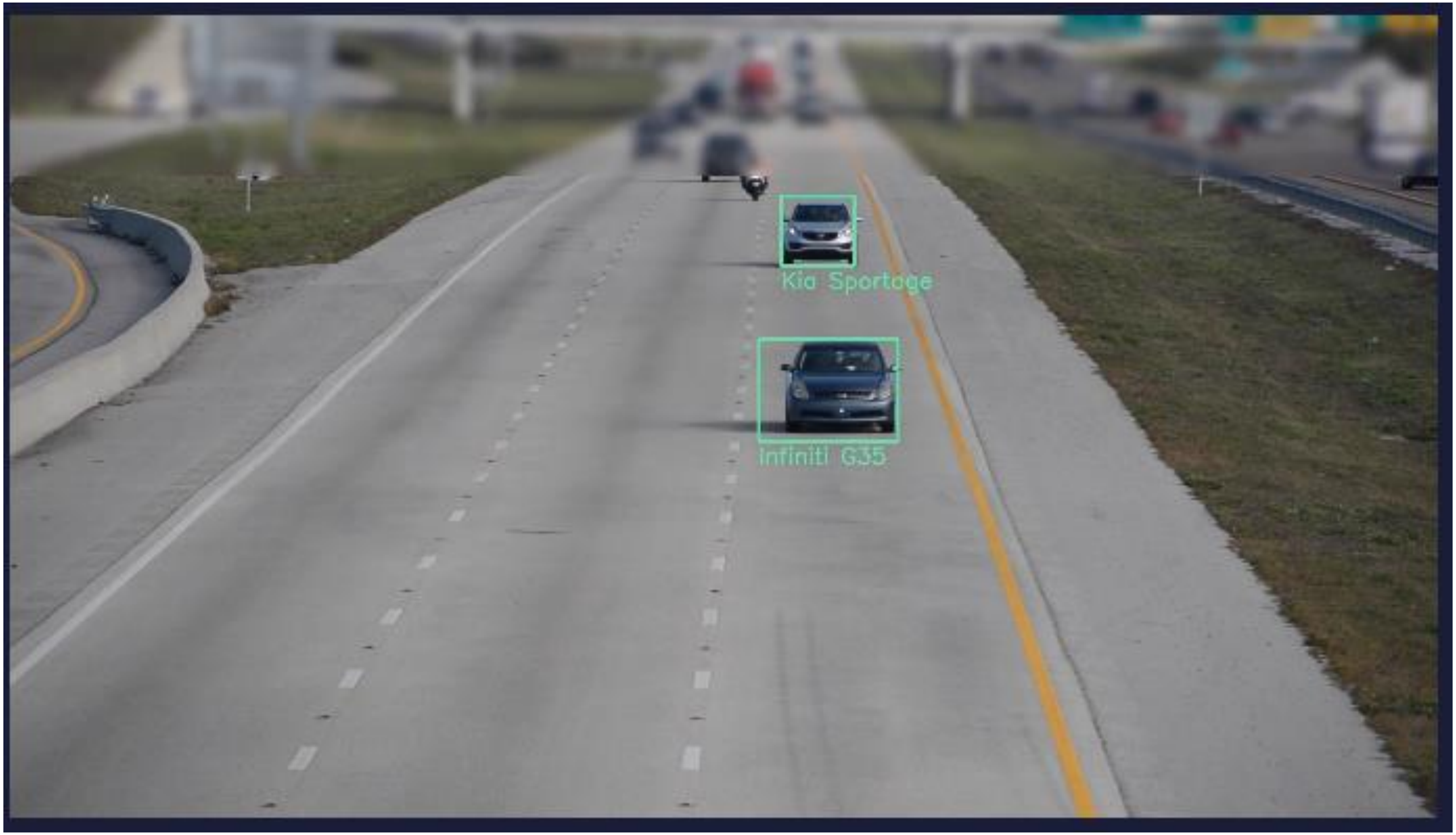
[Farabet et al., 2012]

2012 to Present: Deep Learning is Everywhere

Image segmentation and recognition



2012 to Present: Deep Learning is Everywhere



2012 to Present: Deep Learning is Everywhere

Image Captioning

[Vinyals et al., 2015]
[Karpathy and Fei-Fei, 2015]

No errors

A white teddy bear sitting in the grass

A man riding a wave on top of a surfboard

Minor errors

A man in a baseball uniform throwing a ball

A cat sitting on a suitcase on the floor

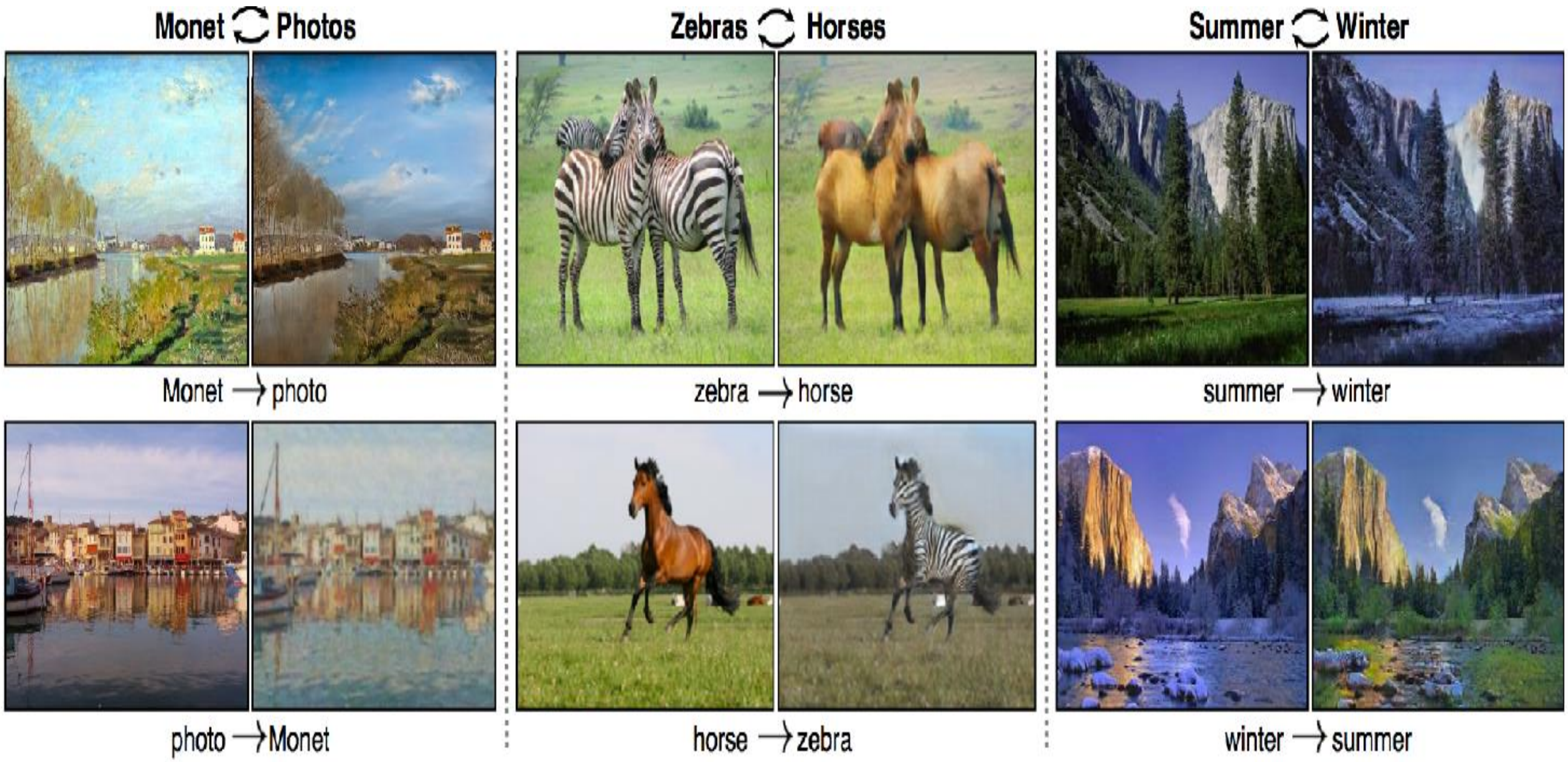
2012 to Present: Deep Learning is Everywhere

- Deep Art: Combining Content and Style from Different Images
- Coarse-scale Content from one image, Fine-scale Style from another image
- Dynamic Capacity Networks (DCNs) learn sophisticated multi-scale representations

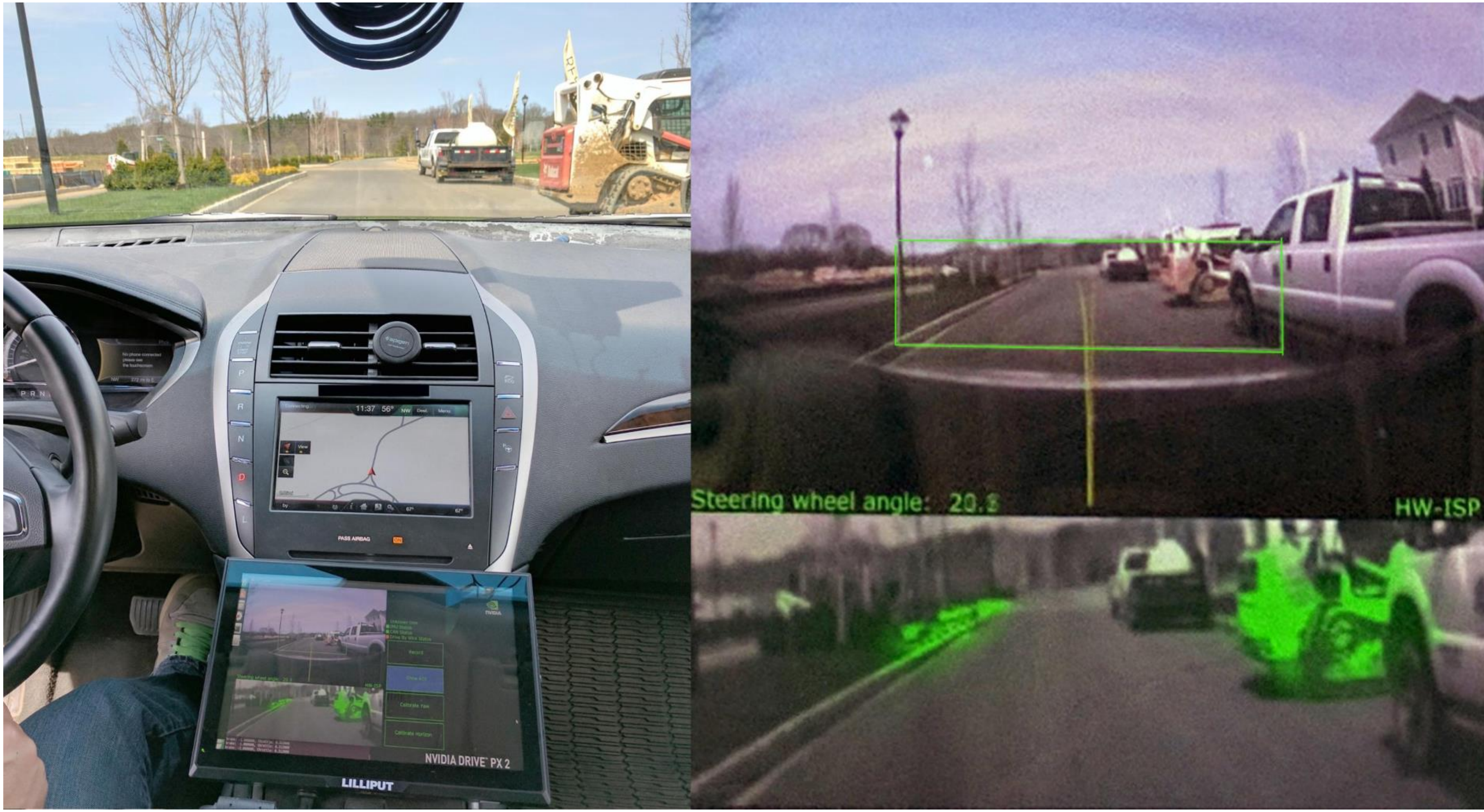
L. Gatys, A. Ecker, M. Bethge
A Neural Algorithm of Artistic Style (ArXiV 2015: eprint arXiv:1508.06576)

2012 to Present: Deep Learning is Everywhere

Generative Adversarial Nets (GANs) for Natural Image *Translation*



2012 to Present: Deep Learning is Everywhere



Deep learning crucial for the global success of automotive autonomy – Automotive World, 6/26/2018

Most Recent

Input: An armchair in the shape of an avocado

Most Recent


(a) AttnGAN
(b) MirrorGAN
(c) DM-GAN
(d) ARRPNGAN
(e) Ground Truth
Text Input
This is a dark blue bird with a dark purple head and a long tail.
This bird has wings that are grey and black and has a white belly.
A small bird with a yellow belly, a long tail, and a short pointy beak.
This bird is brown and white and has a very short beak.

Acknowledgement

- The material in these slides are based on:
  - Digital Image Processing: Rafael C. Gonzalez, and Richard
  - Forsythe and Ponce: Computer Vision: A Modern Approach
  - Rick Szeliski's book: Computer Vision: Algorithms and Applications
  - cs131@ Stanford University
  - cs131n@ Stanford University
  - CS198-126@ University of California, Berkely
  - CAP5415@ University of Central Florida
  - CSW182 @ University of California, Berkely
  - 11-785@ Carnegie Mellon University
  - CSCI1430@ Brown University
  - Computer Vision@ Bonn University
  - ICS 505@ KFUPM
  - Digital Image Processing@ University of Jordan

