

# Introduction to Deep Learning

## Chapter 1: Introduction

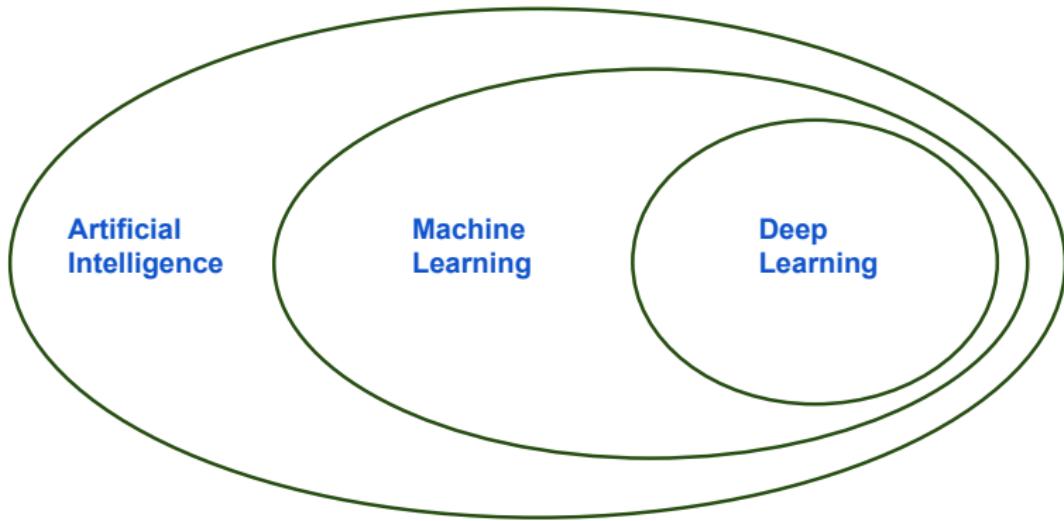
**Bernd Bischl**

Department of Statistics – LMU Munich

WS 2021/2022



# WHAT IS DEEP LEARNING



- Deep learning is the use of artificial neural networks to construct models on large amounts of (unstructured) data.

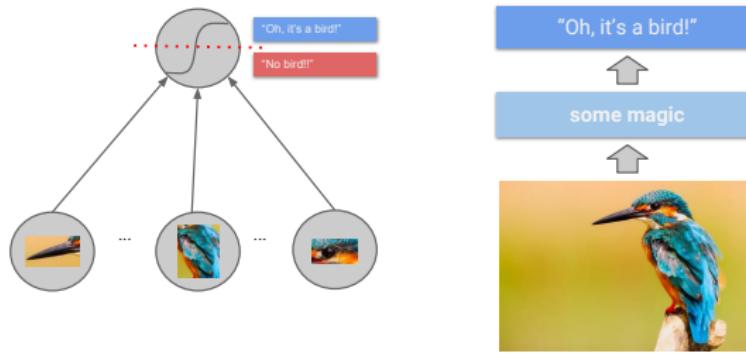
# DEEP LEARNING AND NEURAL NETWORKS

- Deep learning and neural networks are mostly equivalent.
- Deep learning itself is not *new*:
  - Neural networks have been around since the 70s
  - *Deep* neural networks, i.e., networks with multiple hidden layers, are not much younger.
- Why everybody is talking about deep learning now:
  - ❶ Specialized, powerful hardware allows training of huge neural networks to push the state-of-the-art on difficult problems.
  - ❷ Large amount of data is available.
  - ❸ Special network architectures for image/text data.
  - ❹ Better optimization and regularization strategies.

# IMAGE CLASSIFICATION WITH NEURAL NETWORKS

*“Machine learning algorithms, inspired by the brain, based on learning multiple levels of representation/abstraction.”*

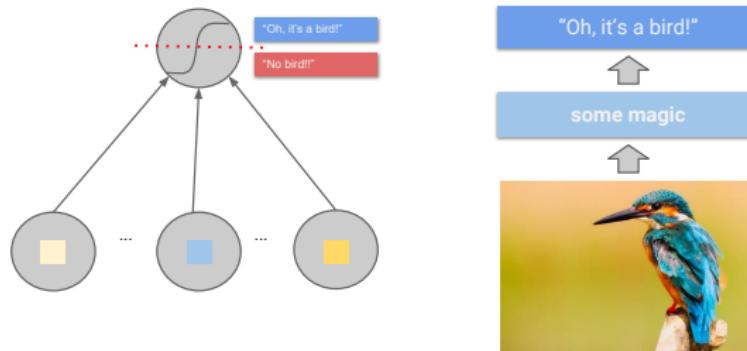
Y. Bengio



# IMAGE CLASSIFICATION WITH NEURAL NETWORKS

*“Machine learning algorithms, inspired by the brain, based on learning multiple levels of representation/abstraction.”*

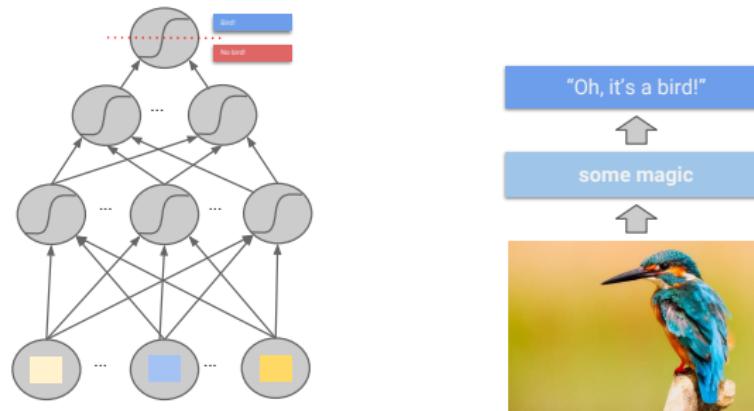
Y. Bengio



# IMAGE CLASSIFICATION WITH NEURAL NETWORKS

*“Machine learning algorithms, inspired by the brain, based on learning multiple levels of representation/abstraction.”*

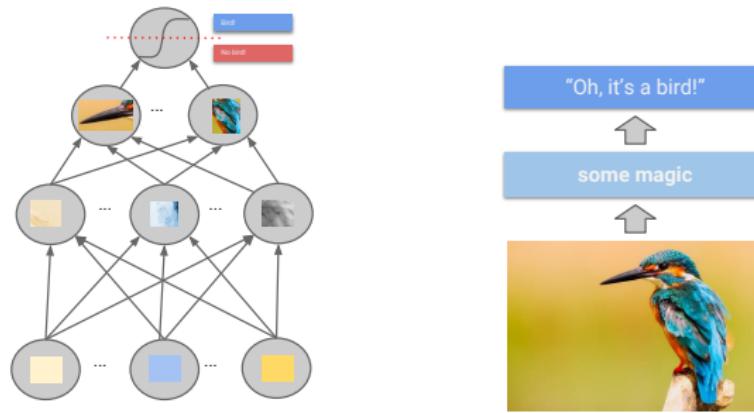
Y. Bengio



# IMAGE CLASSIFICATION WITH NEURAL NETWORKS

*“Machine learning algorithms, inspired by the brain, based on learning multiple levels of representation/abstraction.”*

Y. Bengio



# POSSIBLE USE-CASES

**Deep learning can be extremely valuable if the data has these properties:**

- It is high dimensional.
- Each single feature itself is not very informative but only a combination of them might be.
- There is a large amount of training data.

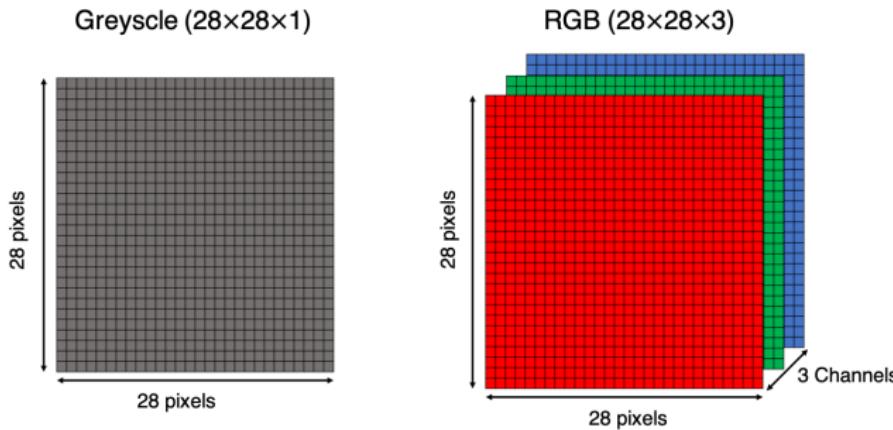
**This implies that for tabular data, deep learning is almost never the correct model choice.**

- Models like random forests or gradient boosting will outperform deep learning most of the time.
- One exception is data with categorical features with many levels.

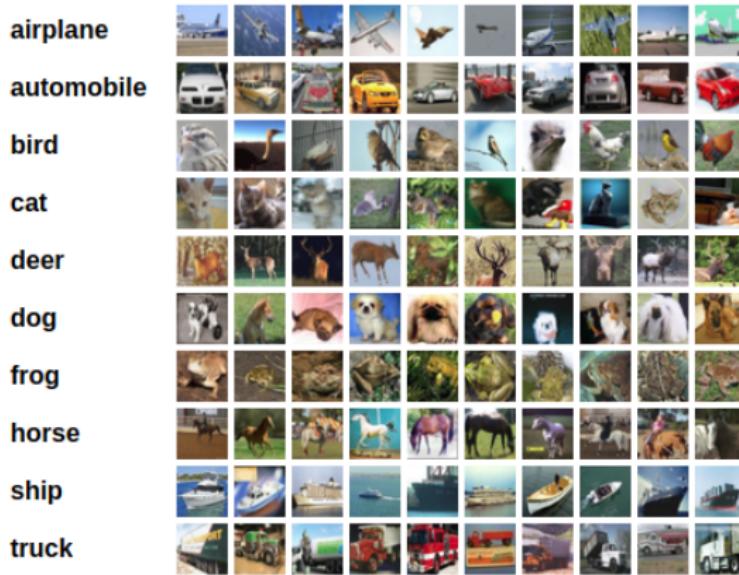
# POSSIBLE USE-CASE: IMAGES

- **High Dimensional:** A color image with  $255 \times 255$  (3 Colors) pixels already has 195075 features.
- **Informative:** A single pixel is not meaningful in itself.
- **Training Data:** Depending on applications huge amounts of data are available.

Architecture: **Convolutional Neural Networks (CNN)**



# POSSIBLE USE-CASE: IMAGES

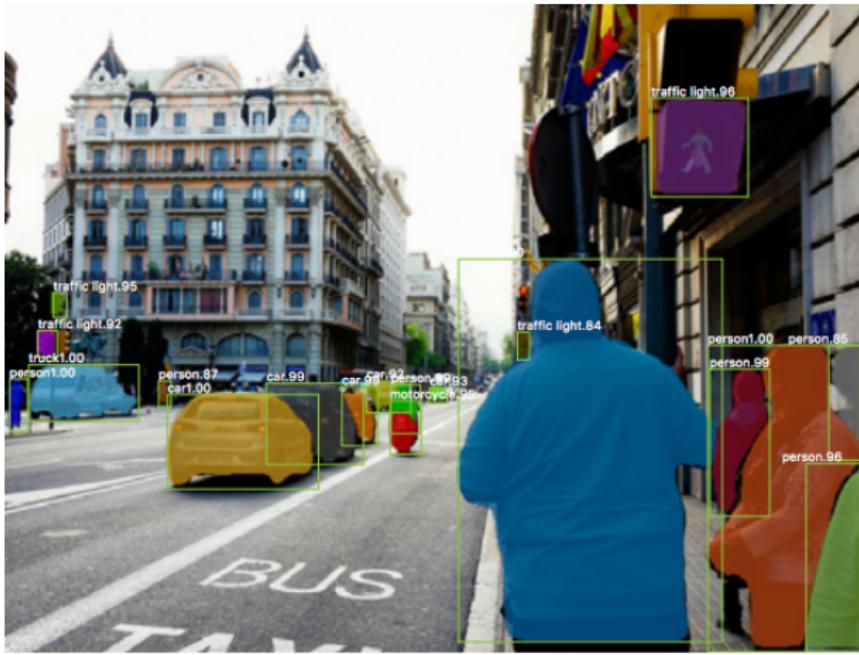


Credit: Alex Krizhevsky (2009)

**Image classification** tries to predict a single label for each image.

CIFAR-10 is a well-known dataset used for image classification. It consists of 60,000 32x32 color images containing one of 10 object classes, with 6000 images per class.

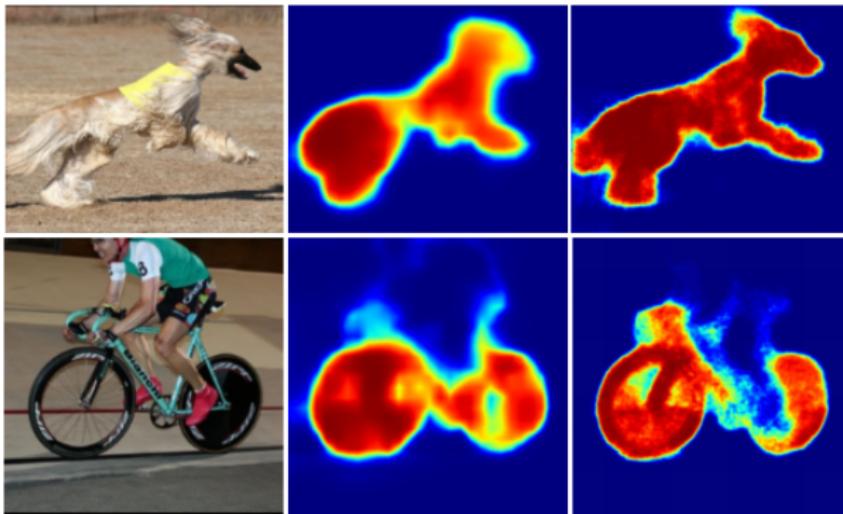
# POSSIBLE USE-CASE: IMAGES



Credit: Kaiming He (2017)

**Object Detection** Mask R-CNN is a general framework for instance segmentation, which efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance.

# POSSIBLE USE-CASE: IMAGES



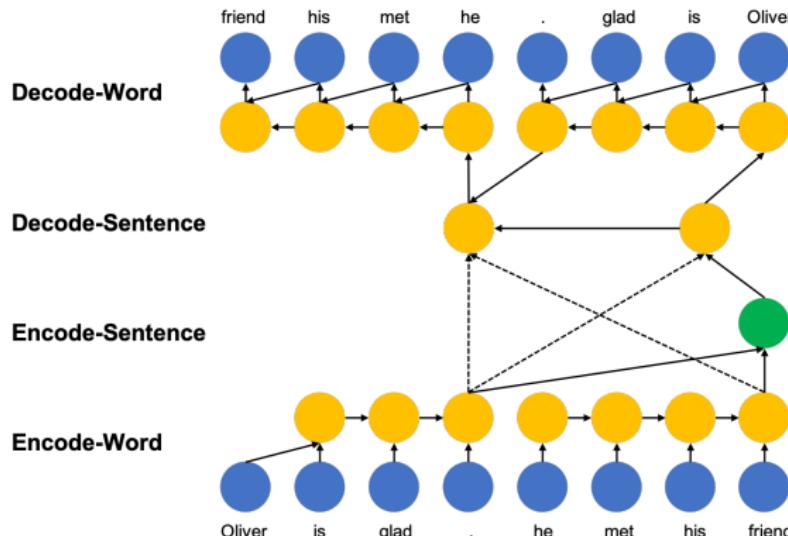
Credit: Hyeonwoo Noh (2015)

**Image segmentation** partitions the image into (multiple) segments.

# POSSIBLE USE-CASE: TEXT

- **High Dimensional:** Each word can be a single feature ( 300000 words in the German language).
- **Informative:** A single word does not provide much context.
- **Training Data:** Huge amounts of text data available.

Architecture: Recurrent Neural Networks (RNN)



# POSSIBLE USE-CASE: TEXT CLASSIFICATION



**Positive**



**Neutral**



**Negative**

Great job! Your customer support is fantastic.

Not bad, but it should be improved in the future.

The worst customer service I have ever seen.

**Sentiment Analysis** is the application of natural language processing to systematically identify the emotional and subjective information in texts.

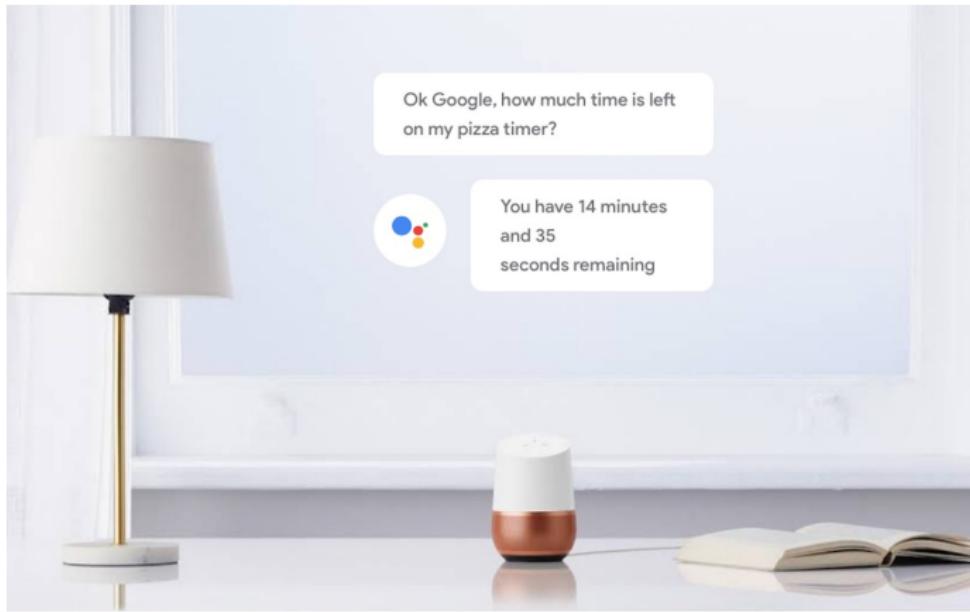
# POSSIBLE USE-CASE: TEXT

The image displays two separate instances of the Google Translate web interface. The top instance shows a translation from English ('He loves to eat') to German ('Er liebt es zu essen'). The bottom instance shows a translation from Norwegian ('Butikken er stengt') to English ('The store is closed'). Both interfaces include language selection dropdowns, microphone and speaker icons for audio, and a copy/paste icon.

From Language	To Language	Text	Edit
English – detected	German	He loves to eat	Er liebt es zu essen
Norwegian	English	Butikken er stengt	The store is closed

**Machine Translation** (e.g. google translate) Neural machine translation exploits neural networks to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.

# APPLICATIONS OF DEEP LEARNING: SPEECH



**Speech Recognition and Generation** (e.g. google assistant) Neural network extracts features from audio data for downstream tasks, e.g., to classify emotions in speech.