Deep Learning for Visual Computing

Motivation, Image Classification

Christopher Pramerdorfer

Computer Vision Lab, TU Wien

Image from reddit, created using midjourney

This Week in AI
GPT-4

Image from OpenAI

Topics

Deep learning

- Motivation
- Primer

Image classification

- Challenges
- Datasets
- Manual approach

Motivation for Deep Learning

Course is called Deep Learning for Visual Computing

- Very generic term (includes computer graphics etc.)

We'll focus on computer vision

- Make computers gain high-level understanding of images
- Goal is human-like understanding

Deep learning has revolutionized this field

- Reason for the current AI hype

Motivation for Deep Learning
Image Classification

"What thing is shown in this image?"

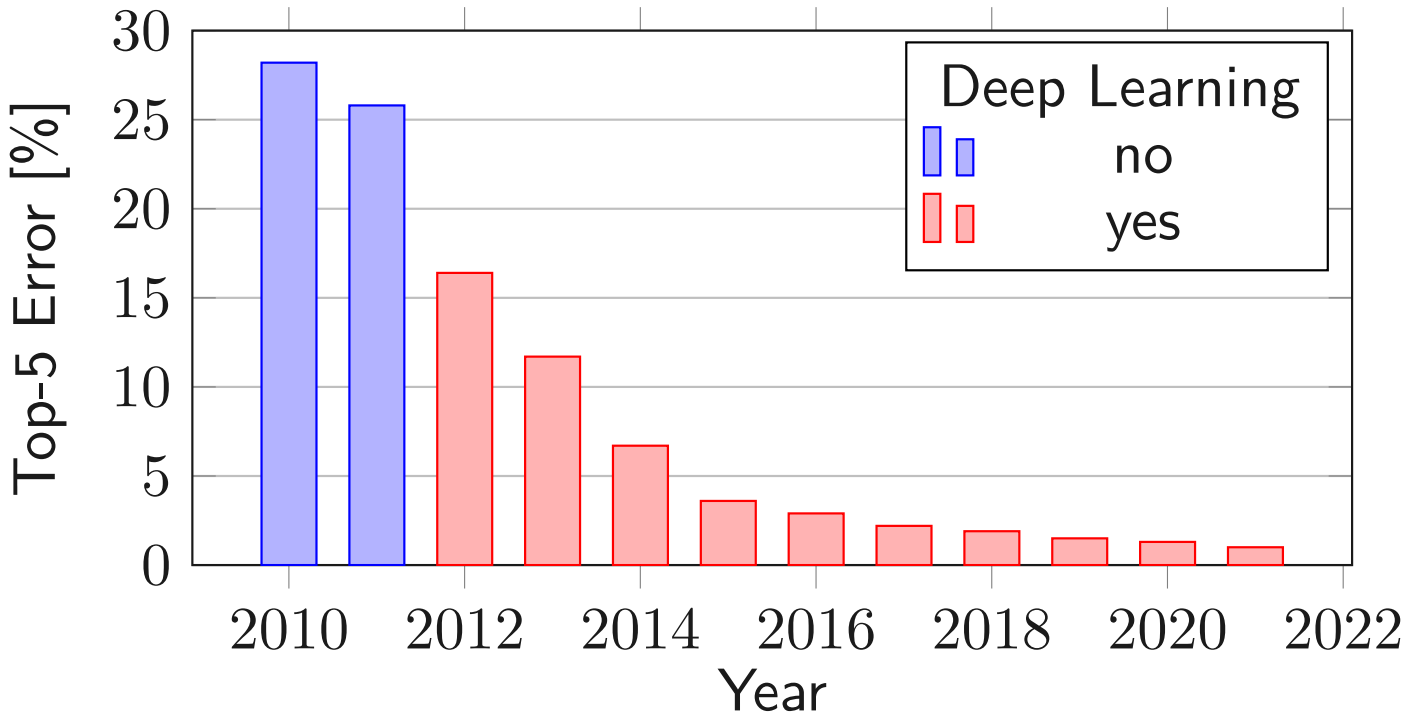
Image from youtube.com

$\Rightarrow$ cat

Motivation for Deep Learning
Image Classification

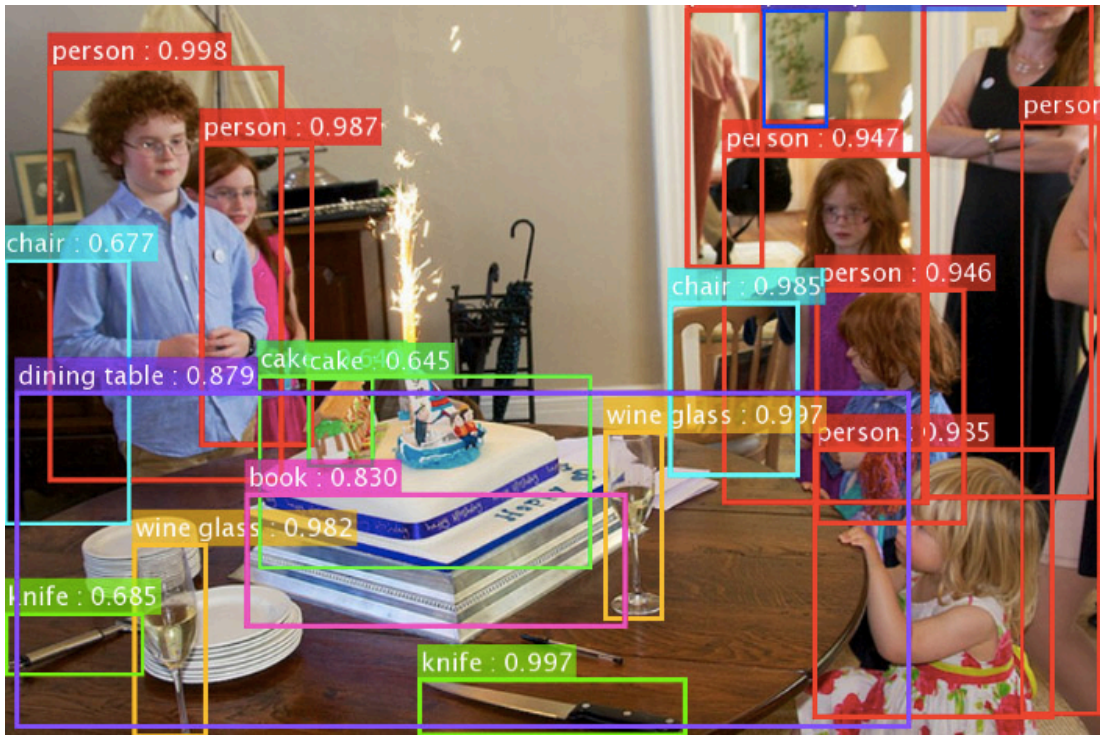
ImageNet benchmark performance over time

- 100k test images of 1000 classes



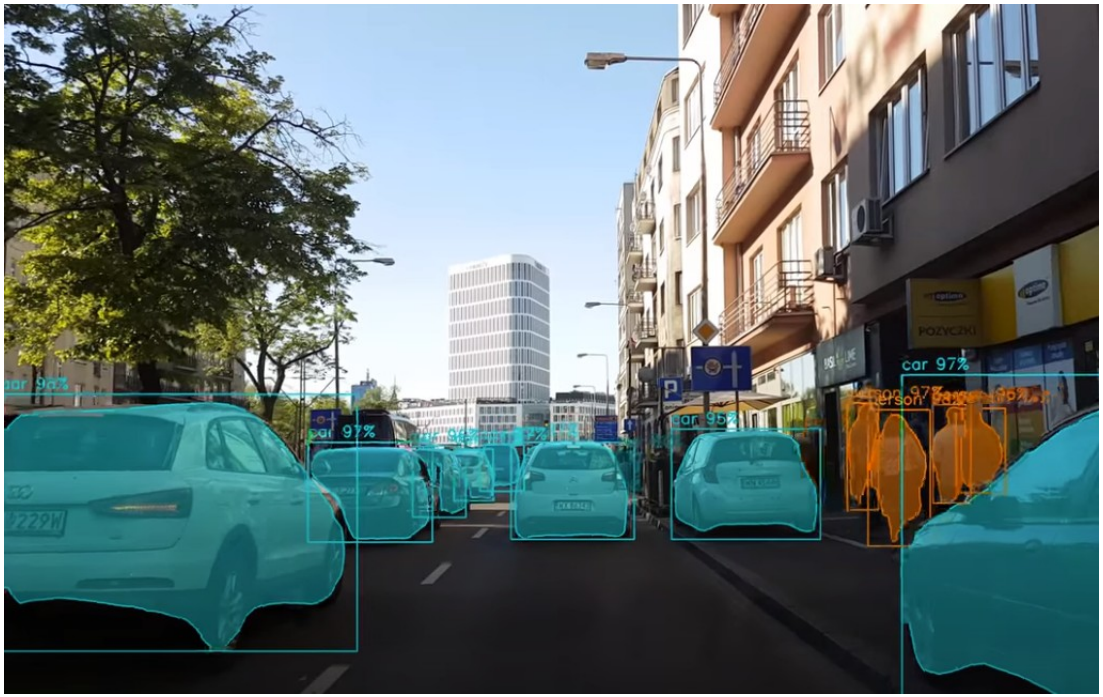
Motivation for Deep Learning
Object Detection

"Detect objects of interest"



Motivation for Deep Learning
Instance Segmentation

"Delineate objects of interest"



link

Motivation for Deep Learning

Pose Estimation

"Estimate people's poses"

link

Motivation for Deep Learning
Image Synthesis

"Be an artist"

link

Motivation for Deep Learning
Image Synthesis

"Be an artist" (really)

midjourney.com

Motivation for Deep Learning
Image Synthesis

"Be a 3D artist"

link

Motivation for Deep Learning
Image Synthesis

"Generate videos that look real"

link

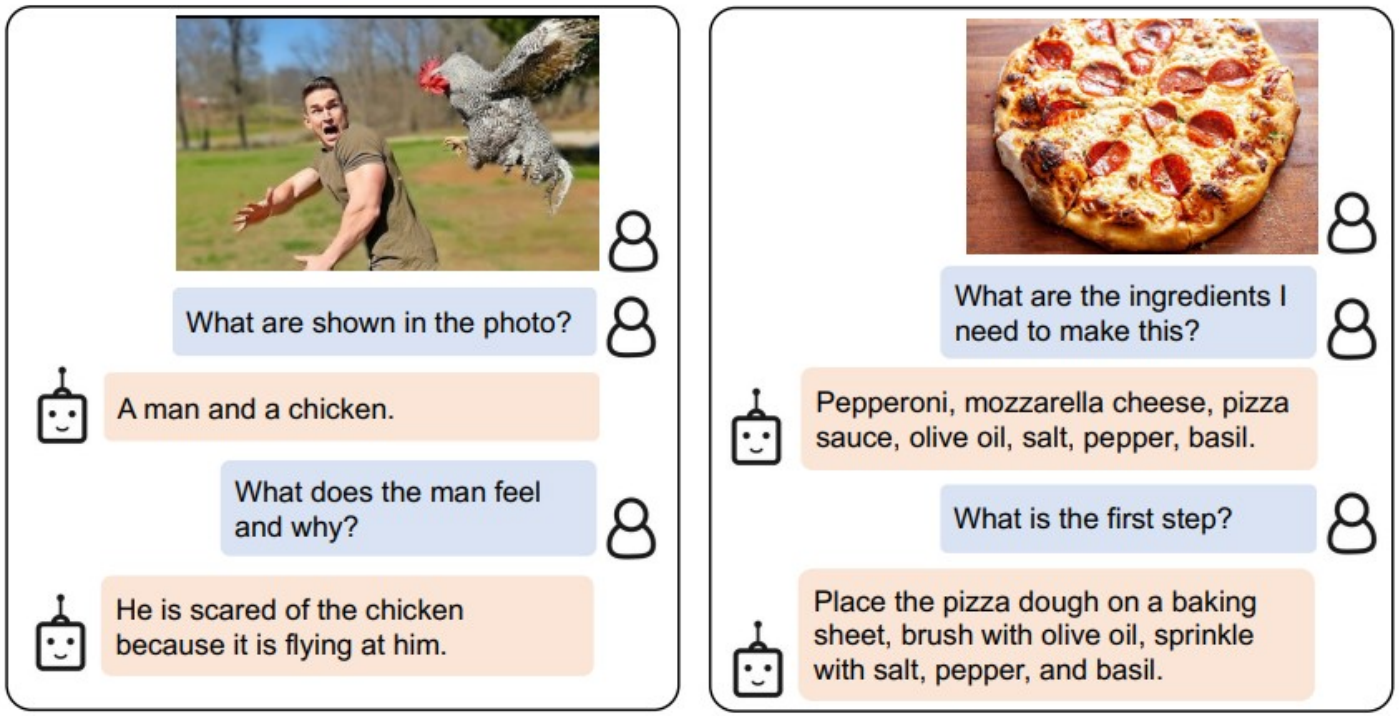
Motivation for Deep Learning
Image Synthesis

"Generate videos that look real" (danger!)

link

Motivation for Deep Learning
Image Understanding

"Explain what's going on in an image"



paper

Motivation for Deep Learning

All these examples are based on deep learning

- Would be impossible otherwise (at this quality)
- We will take a closer look throughout the lecture

Deep learning is state of the art

- In virtually any computer vision task
- In most other fields as well (e.g. speech recognition)

So knowledge of deep learning is essential

blank page

Deep Learning Primer

Deep learning is not magic

- Implemented using neural networks
- With neurons that are adapted for image data
- And arranged in many layers (hence “deep”)

Machine learning concepts still apply

- Parametric models
- Loss functions & iterative optimization
- Overfitting, regularization

Deep Learning Primer

The power of deep learning comes from

- Hierarchical local feature transformations
- That are learned using large amounts of data

Key ingredients needed

- Large datasets
- Lots of processing power

Deep Learning Primer

Learn high-level concepts from lower-level ones

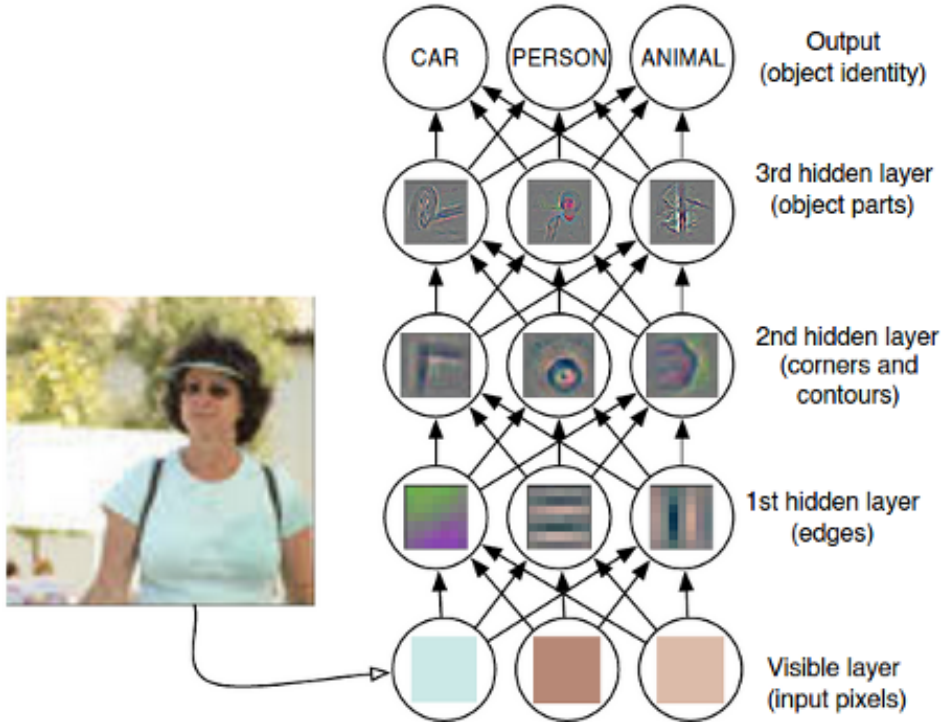


Image from deeplearningbook.org

Deep Learning Primer

Deep learning is $30+$ years old

- But data and processing power were limited until recently

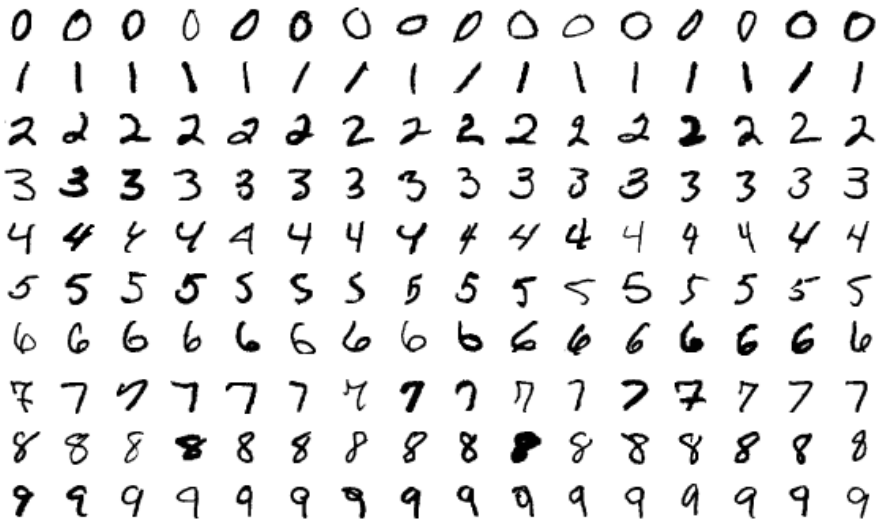
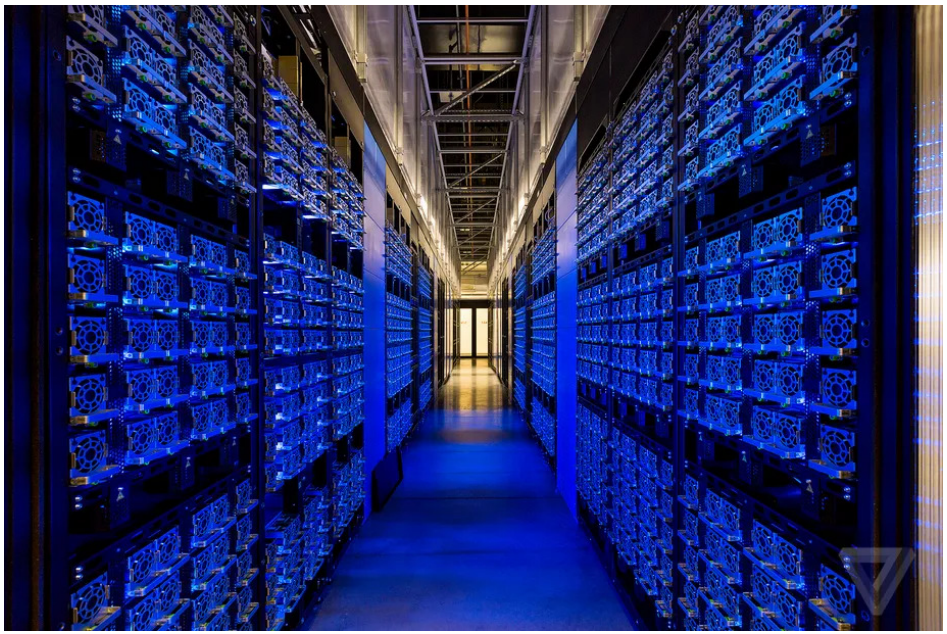


Image from Wikipedia

Deep Learning Primer

Nowadays deep learning is a billion dollar business

- Big tech companies embraced it years ago
- We interact with deep learning daily (e.g. phones, cars)



link

Deep Learning Primer

Deep learning has become accessible

- The software used by Google, Meta etc. is open source
- Runs on your PC and even on your phone
- Cloud services available (e.g. Google Colab)

link

blank page

Image Classification

Fundamental computer vision task

Definition

- Given a set of class labels (e.g. {bird, cat, dog})
- Which class does the given image belong to?

$\Rightarrow$ cat

Image from youtube.com

Image Classification

Image belongs to exactly one class in the set

- Comparatively easy task (but still challenging)
- On many datasets deep learning outperforms humans

Simple problem formulation

- Great for learning deep learning basics
- Why we will stick to classification for now

Image Classification
Challenges – Pose and Viewpoint

Image adapted from warrenphotographic.co.uk

Image Classification
Challenges – Illumination

Image from studioddt.com

Image Classification
Challenges – Deformation

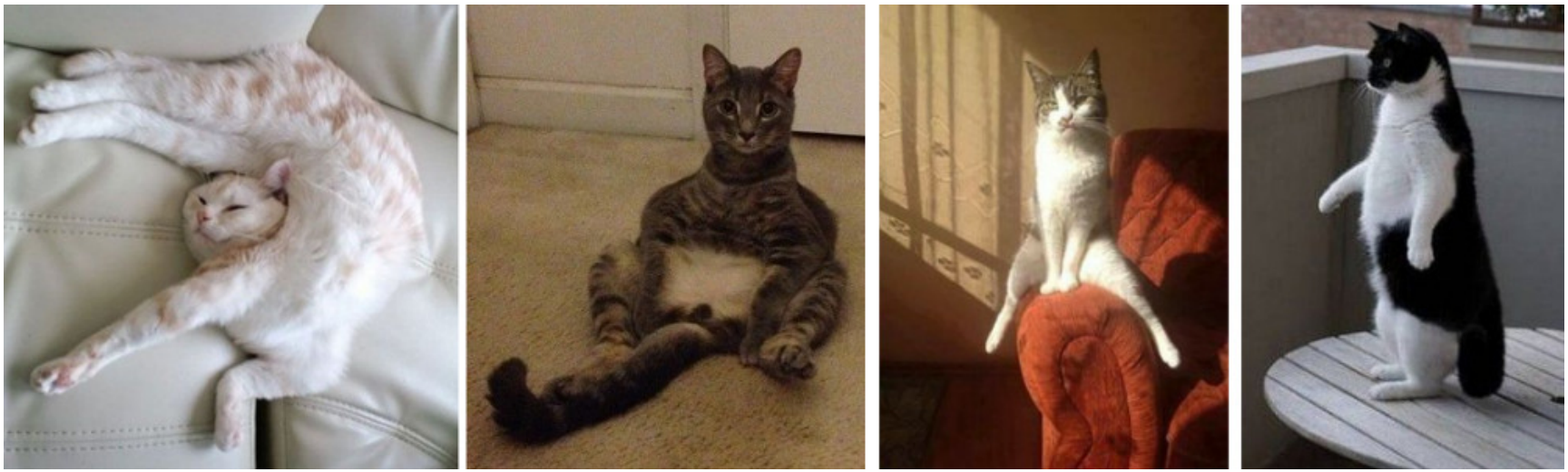


Image from cs231n.github.io

Image Classification
Challenges – Occlusion

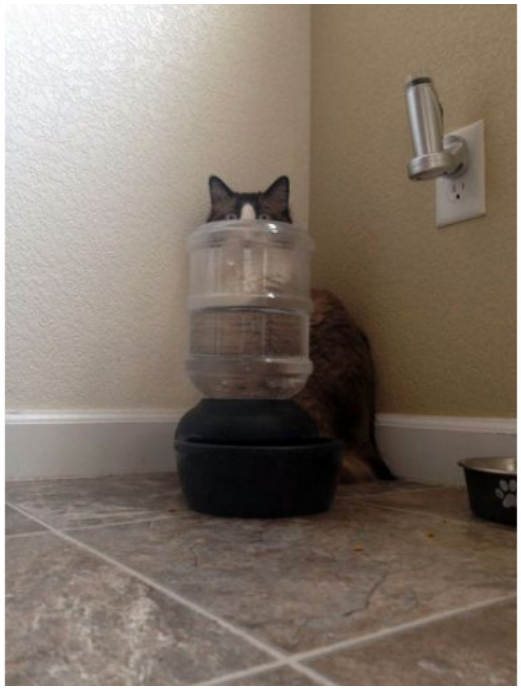




Image from cs231n.github.io

Image Classification
Challenges – Background

Image from cs231n.github.io

Image Classification
Challenges – Intraclass Variation

Image from cs231n.github.io

Image Classification
Datasets

A good classifier must cope with these challenges

- To verify this we need a representative dataset
- Such datasets are usually large

If we employ machine learning we also need training data

- Datasets must be disjoint (so need even more data)
- Deep learning requires lots of data

Image Classification
Datasets

Dataset acquisition takes lots of effort

- Collect many (thousands or more) of images
- Assign class labels to enable automatic training and testing

Data acquisition and processing is central in deep learning

- Often the most time-consuming task
- Usually main bottleneck for performance

Thankfully many public datasets are available

Image Classification
Datasets – CIFAR-10

10 classes, 60k images

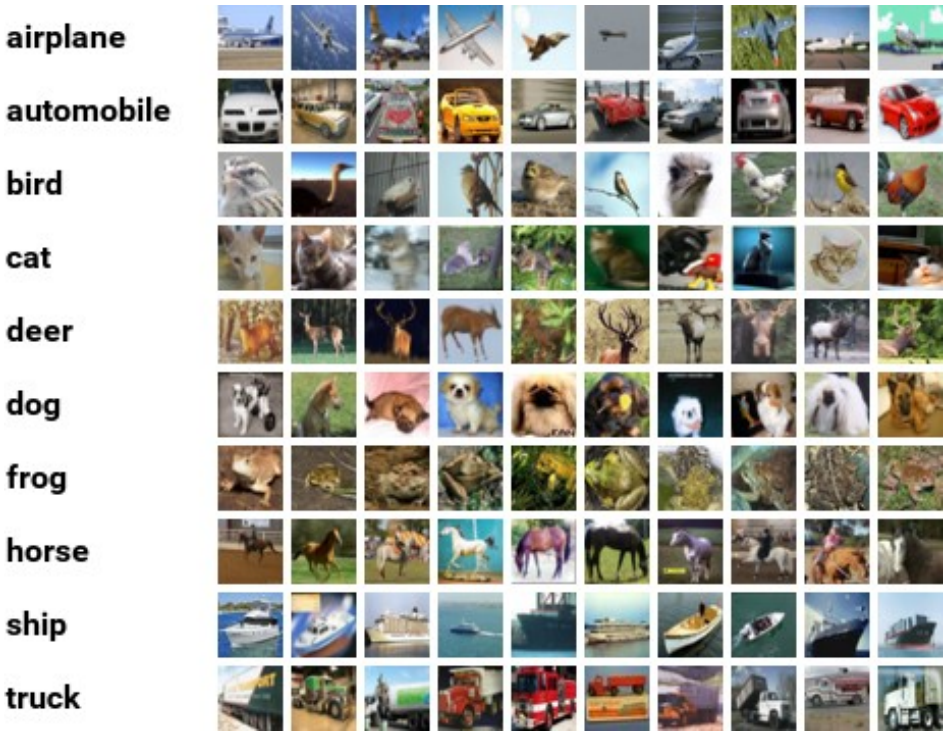
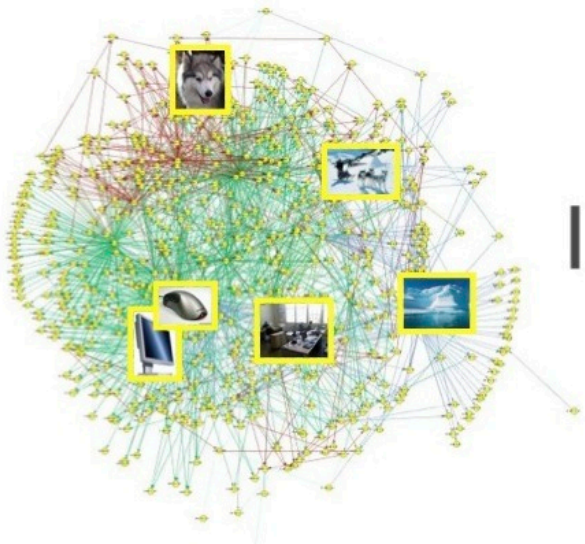


Image from cs.toronto.edu

Image Classification
Datasets – ImageNet

20k classes, 14m images

- One of main drivers for deep learning performance



IMAGENET

Image from umich.edu

Image Classification
Datasets – COCO

300k images, labels for classification, detection, segmentation, ...

Image from cocodataset.org

Image Classification

Let's build an image classifier

- Should support the classes {dog, cat}
- Using the CIFAR-10 dataset

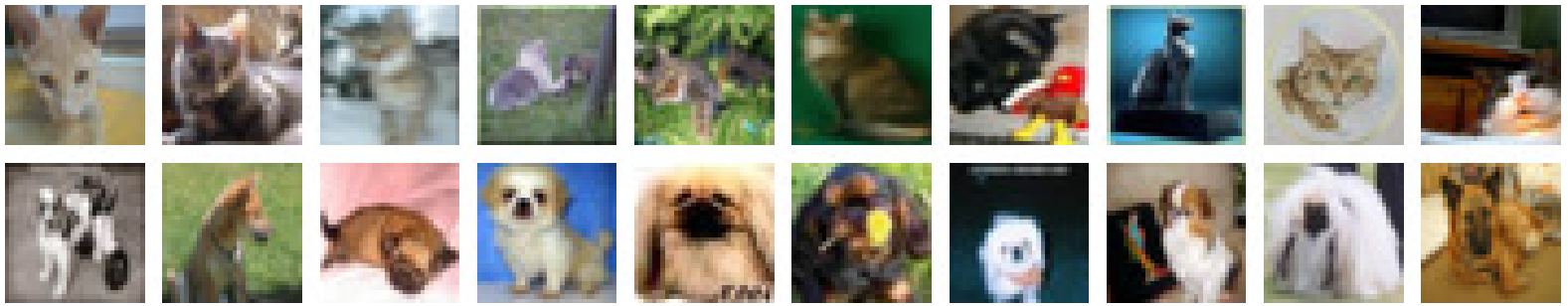


Image from cs.toronto.edu

Image Classification

How can we write an algorithm for this purpose?

Image from cs.toronto.edu

Image Classification

We cannot!

- No obvious unique and reliable features
- Not clear how to represent and use them

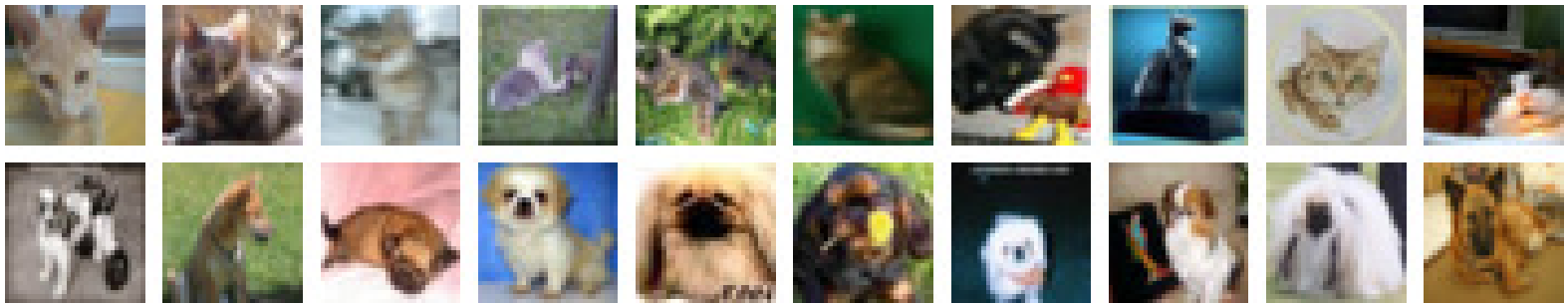


Image from cs.toronto.edu

Image Classification

We humans are incredible image classifiers

But we cannot describe formally how we do so

- Thus the standard `if {} else {}` approach fails

This applies to most vision problems

- Reason we need machine and deep learning