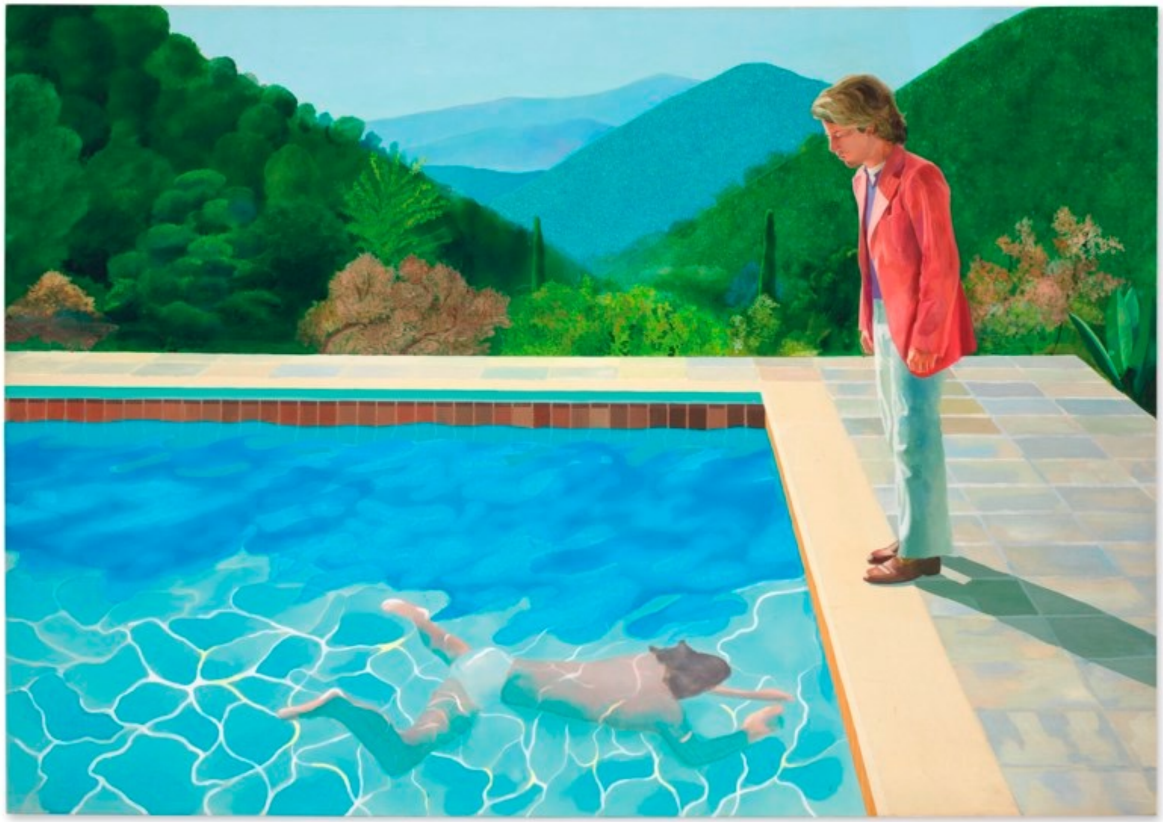
CS 444: Deep Learning for Computer Vision



D. Hockney, Pool with two figures, 1972

https://slazebni.cs.illinois.edu/spring24/

Overview

- Logistics
- Motivation: The statistical learning viewpoint
- A taxonomy of learning problems
- Topics to be covered in class

How can we build an agent to…

Play chess?

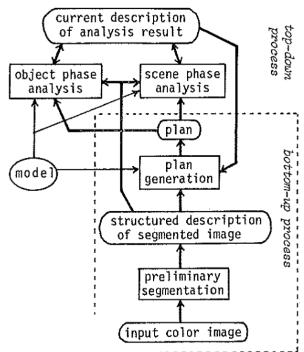
Translate between languages?

Recognize object categories?

Fly a drone?

How can we build an agent to achieve expertise?

- Good old-fashioned AI (GOFAI) answer:
  Program expertise into the agent



low of data in the analysis process.

›lock as a set of
; it must satisfy

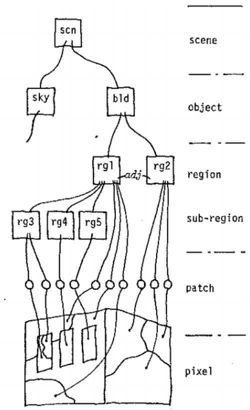


Figure 3. Structure of description built by the system.

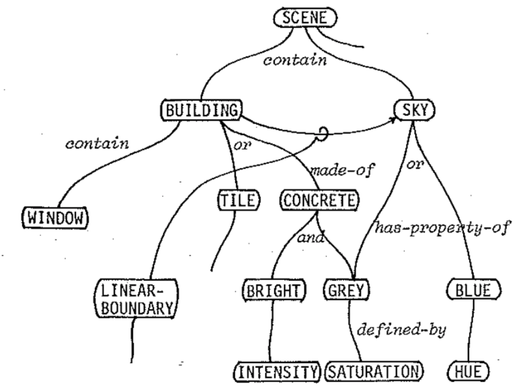


Figure 2. Semantic network for knowledge representation.

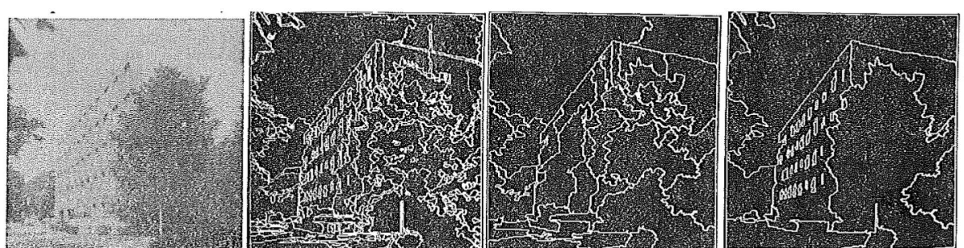


Figure 5-a. Digitized color scene.

5-b. Result of preliminary segmentation.

5-c. Plan image.

5-d. Result of semantic segmentation.

Y. Ohta, T. Kanade and T. Sakai. An Analysis System for Scenes Containing objects with Substructures. Proc. of the Fourth International Joint Conference on Pattern Recognition, pp. 752-754, 1978

How can we build an agent to achieve expertise?

- Good old-fashioned AI (GOFAI) answer:
  Program expertise into the agent

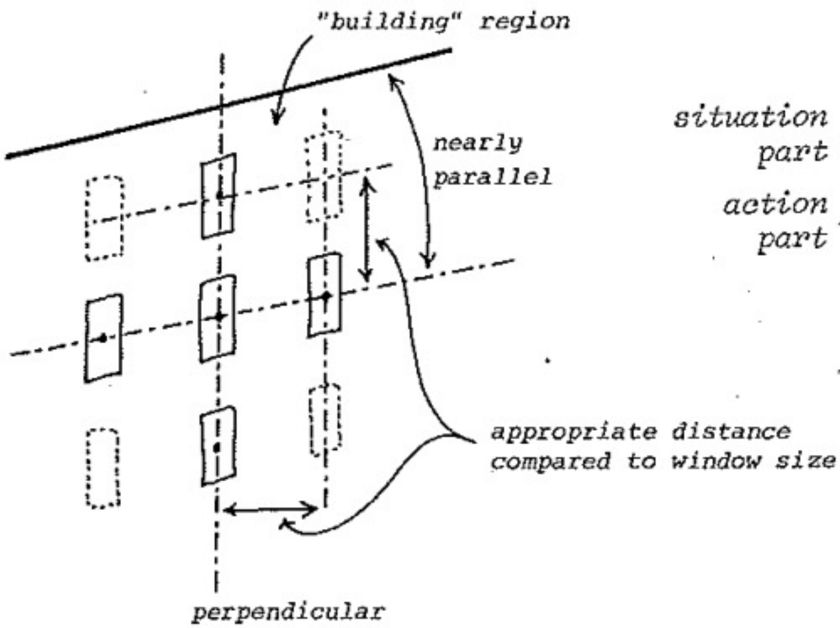


Figure 4-a. "Building" region and "windows".

situation part

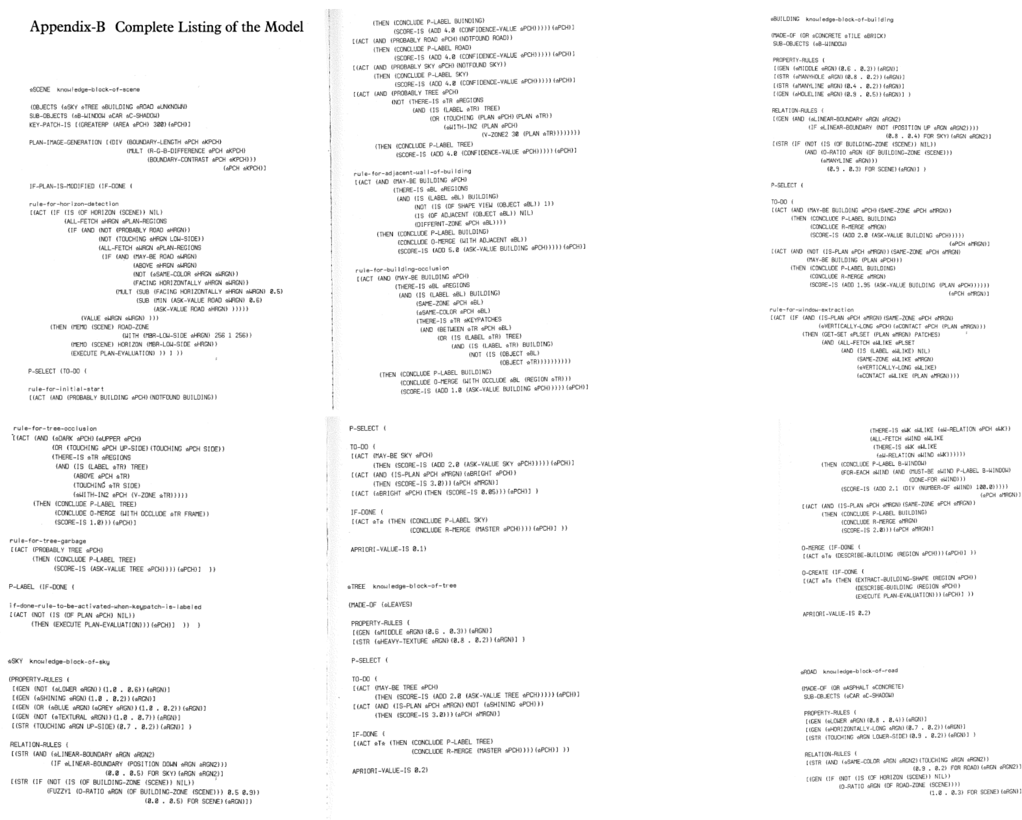
action part

```
(ACT                  "indicates following is production"
(IF (*WINDOW-LIKE *RGN)
   (THEN (GET-SET *PLSET (PLAN *MRGN) PATCHES)
                      "assign to *PLSET all patches which belong to the plan of *MRGN"
        (AND (ALL-FETCH *WLIKE *PLSET
                  (AND (IS (LABEL *WLIKE) NIL)(*WINDOW-LIKE *WLIKE)))
              (ALL-FETCH *WIND *WLIKE
                  (THERE-IS *WK *WLIKE (*W-RELATION *WIND *WK)))
                   "search all patches that have at least one partner patch
                    with which *W-RELATION holds"
              (THERE-IS *WK *WLIKE (IS *RGN *WK))))))
(THEN (CONCLUDE P-LABEL B-WINDOW)   "*RGN is 'window'"
      (FOR-EACH *WIND (MUST-BE *WIND P-LABEL B-WINDOW))
                      "all patches in *WIND must be 'window' if *RGN is 'window'"
      (DONE-FOR *WIND)   "no need to examine the *WIND patches any more
                          under this context --- for controler"
      (PRIORITY (ADD 2100 (NUMBER *WIND))) ))
                      "premium is the number of patches in *WIND"
```

Figure 4-b. The production for analyzing "windows".

How can we build an agent to achieve expertise?

- Good old-fashioned AI (GOFAI) answer:
  Program expertise into the agent



How can we build an agent to achieve expertise?

- Good old-fashioned AI (GOFAI) answer:
  Program expertise into the agent
  - Never worked (in general)

How can we build an agent to achieve expertise?

- Good old-fashioned AI (GOFAI) answer:
  Program expertise into the agent
  - Never worked (in general)
  - Though not without exceptions…

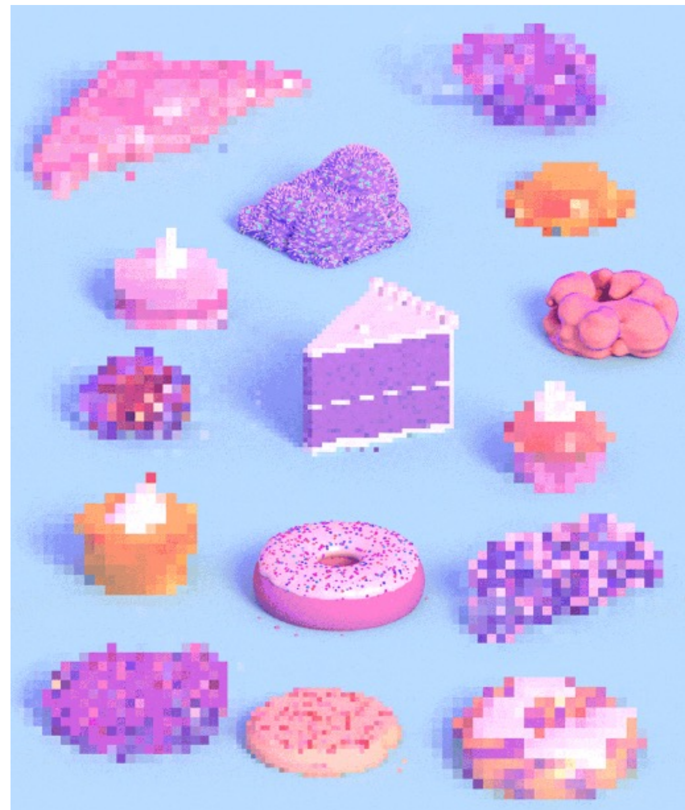
ANNALS OF TECHNOLOGY

THE PASTRY A.I. THAT LEARNED TO FIGHT CANCER

In Japan, a system designed to distinguish croissants from bear claws has turned out to be capable of a whole lot more.

By James Somers

March 18, 2021



https://www.newyorker.com/tech/annals-of-technology/the-pastry-ai-that-learned-to-fight-cancer

How can we build an agent to achieve expertise?

- Good old-fashioned AI (GOFAI) answer:
  Program expertise into the agent
  - Never worked (in general)
  - Though not without exceptions…

ANNALS OF TECHNOLOGY

THE PASTRY A.I. THAT LEARNED TO FIGHT CANCER

In Japan, a system designed to distinguish croissants from bear claws has turned out to be capable of a whole lot more.

By James Somers

March 18, 2021

not well fed.) But this was all under carefully controlled conditions. In a real bakery, the lighting changes constantly, and BRAIN's software had to work no matter the season or the time of day. Items would often be placed on the device haphazardly: two pastries that touched looked like one big pastry. A subsystem was developed to handle this scenario. Another subsystem, called "Magnet," was made to address the opposite problem of a pastry that had been accidentally ripped apart.

be used elsewhere. Today, solving the pastry problem without deep learning would seem impossible; it's a wonder that, in 2007, when neural networks weren't a viable option, Kambe even took it on. The system that he and his team managed to build over the following fifteen years must surely be one of the more sophisticated achievements in "classical" computer vision—a fact obscured, perhaps, by its origin in baked goods.

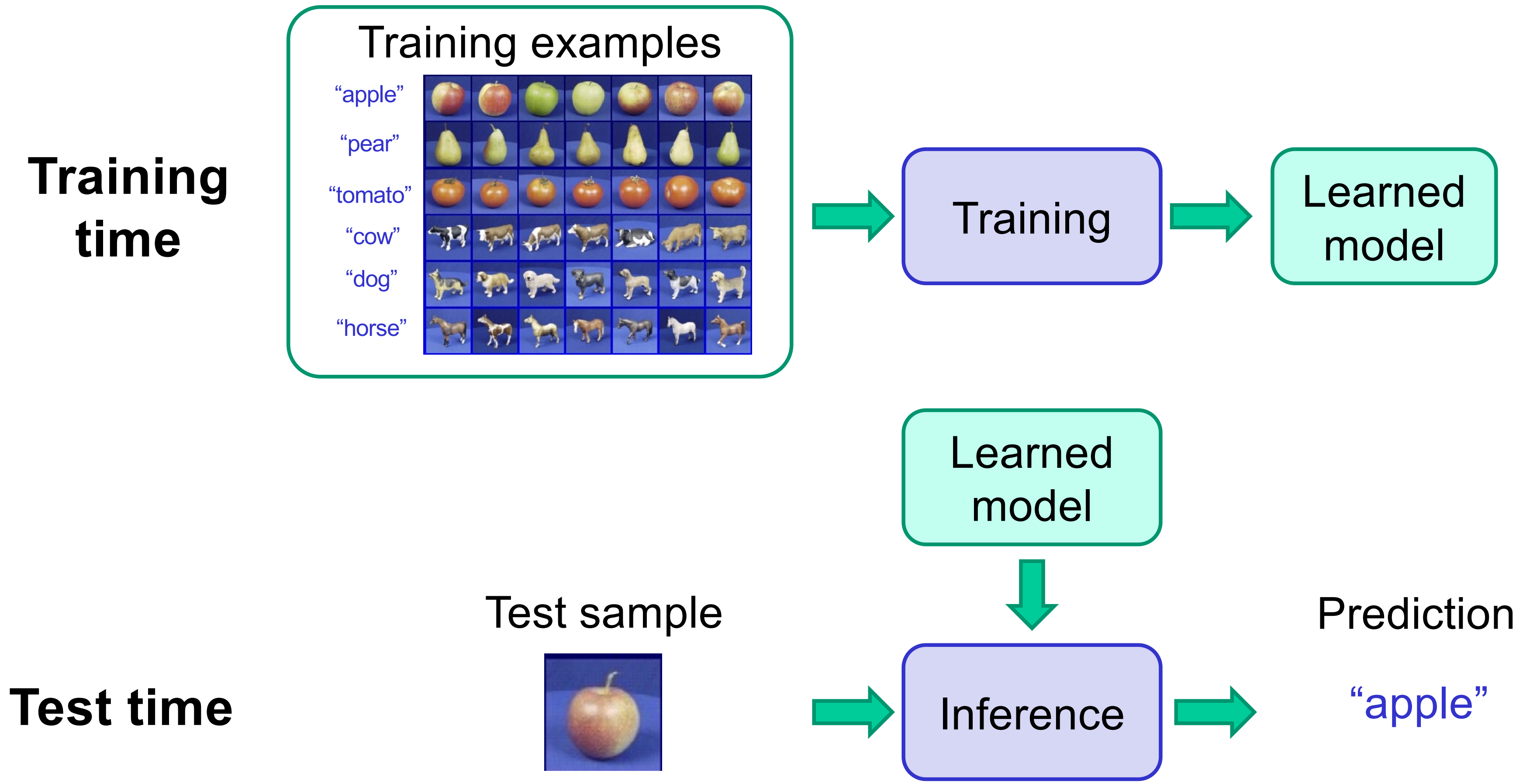
https://www.newyorker.com/tech/annals-of-technology/the-pastry-ai-that-learned-to-fight-cancer

How can we build an agent to achieve expertise?

- Good old-fashioned AI (GOFAI) answer:
  Program expertise into the agent

- Modern answer: Program into the agent the ability to **improve performance** based on **experience**
  - **Performance** needs to be quantified using some score or metric (loss, reward, etc.)
  - **Experience** comes from *training data* or *demonstrations*
  - **Improvement** results from the *learning algorithm*
  - Leap of faith: agent that can achieve good performance on training data will *generalize* to never-before-seen inputs

The basic statistical learning framework



Overview

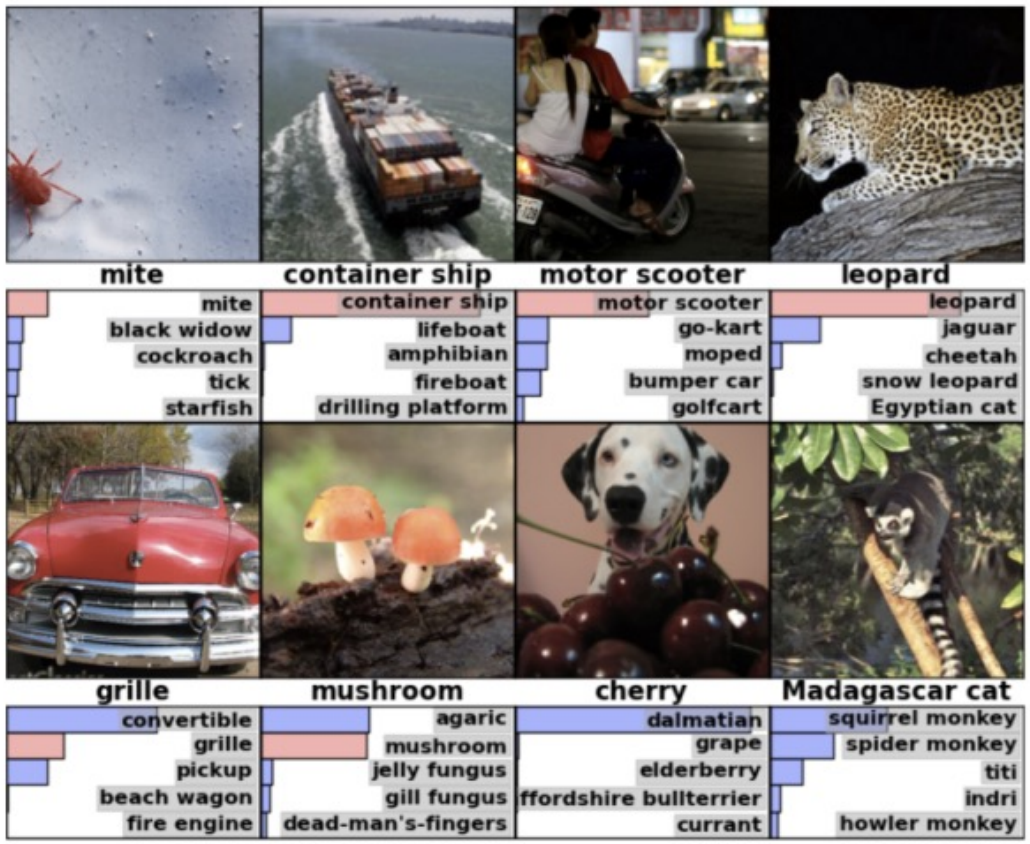
- Logistics
- Motivation: The statistical learning viewpoint
- A taxonomy of learning problems

Taxonomy of learning problems

- **Type of output**
  - Classification
  - Regression
  - Structured prediction
  - Dense prediction
  - Multi-modal prediction
- **Type of supervision**
  - Fully supervised
  - Unsupervised
  - Self-supervised or predictive learning
- **Training regime**
  - Batch offline learning
  - Online/continual learning
  - Active learning
  - Reinforcement learning

Type of output: Classification

ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)



Type of output: Regression

Age estimation

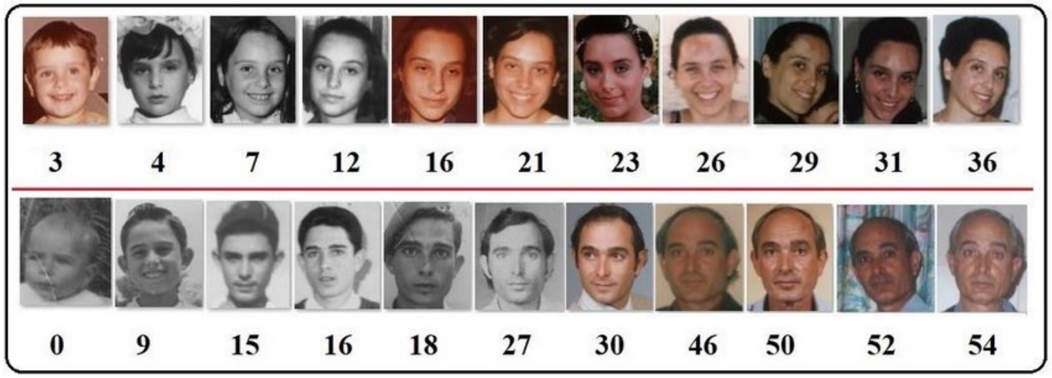
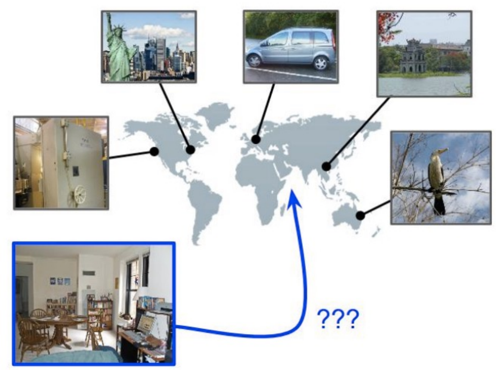


Figure source

Date prediction

Vittayakorn et al. (2017)

Location prediction



Vo et al. (2017)

Type of output: Dense prediction

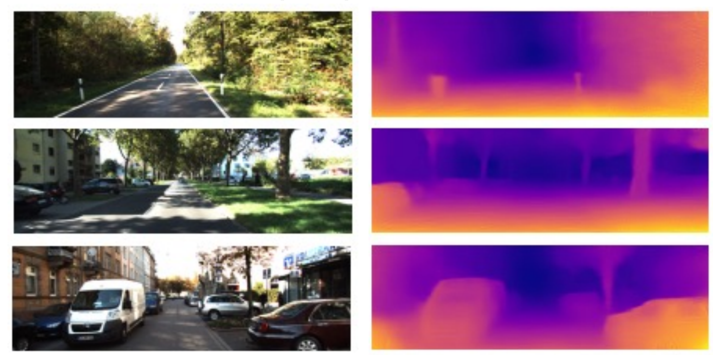
Image colorization

Zhang et al. (2016)

Semantic segmentation

Long et al. (2016)

Depth prediction



Wang et al. (2017)

Type of output: Structured prediction

Object detection

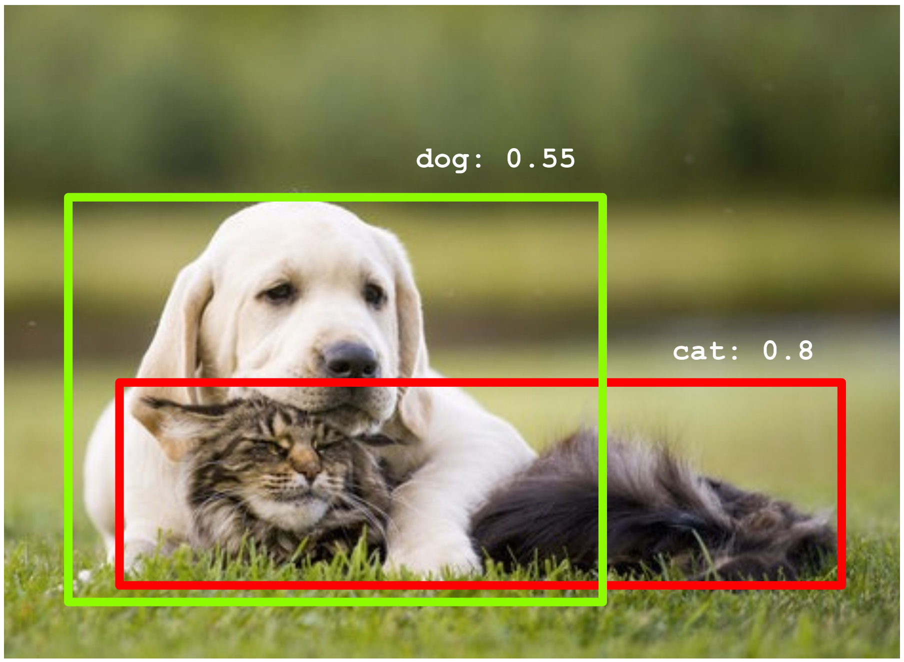
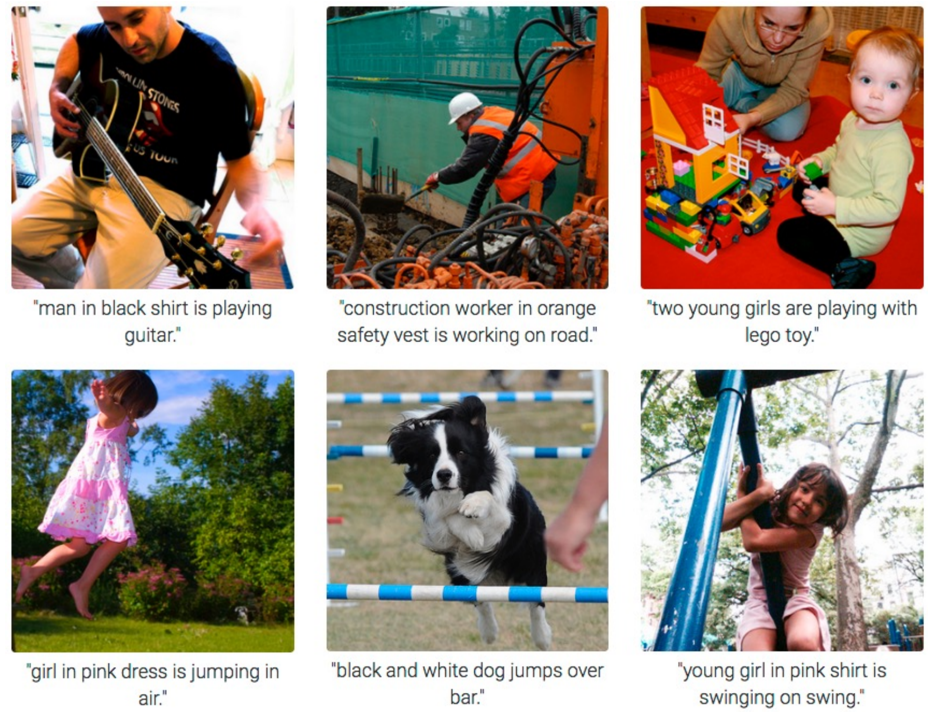


Image description



Karpathy & Fei-Fei (2015)

Dense + structured prediction

Object detection + instance segmentation

Keypoint detection

K. He, G. Gkioxari, P. Dollar, and R. Girshick, [Mask R-CNN](), ICCV 2017

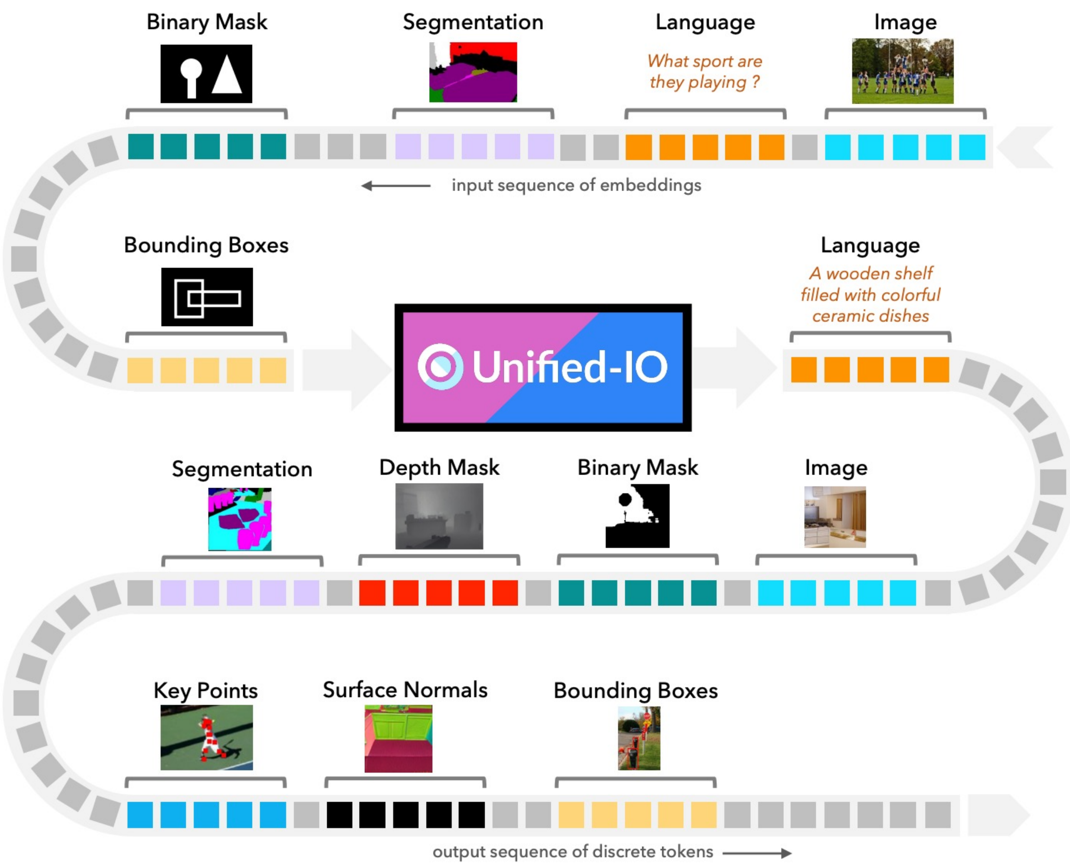
Overview

- Logistics
- Motivation: The statistical learning viewpoint
- A taxonomy of learning problems
- Topics to be covered in class

Taxonomy of learning problems

- **Type of output**
  - Classification
  - Regression
  - Structured prediction
  - Dense prediction
  - Multi-modal prediction
- **Type of supervision**
  - Fully supervised
  - Unsupervised
  - Self-supervised or predictive learning
- **Training regime**
  - Batch offline learning
  - Online/continual learning
  - Active learning
  - Reinforcement learning

Multi-modal prediction

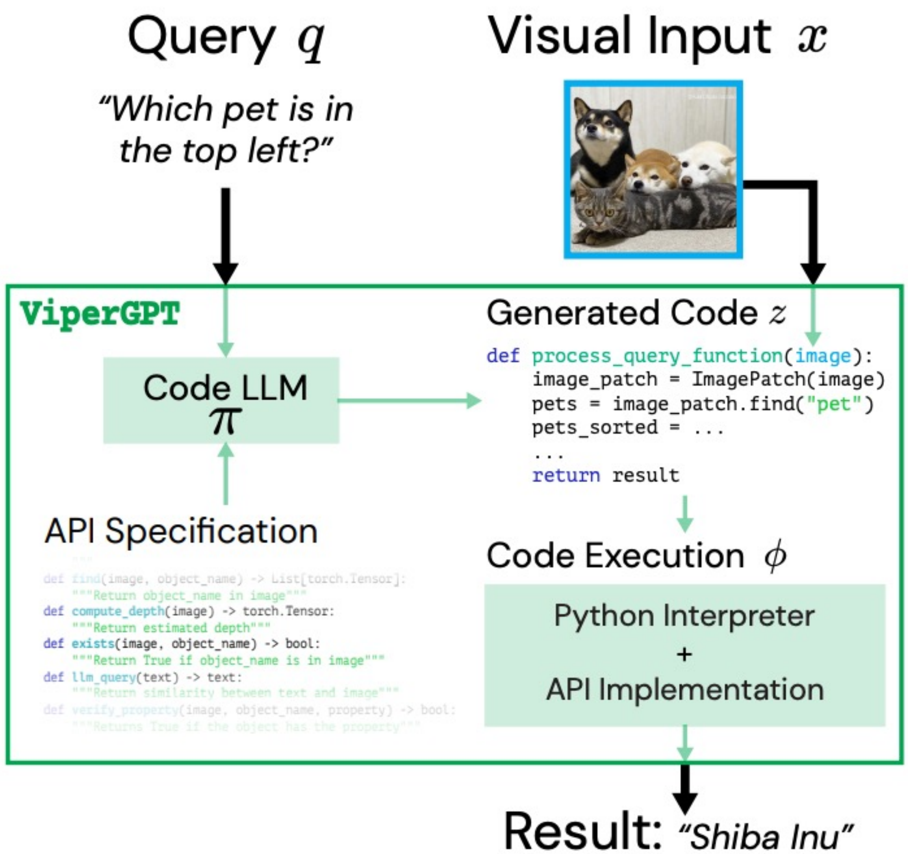


Tasks

Image Classification
Object Detection
Semantic Segmentation
Depth Estimation
Surface Normal Estimation
Segment-based Image Generation
Image Inpainting
Pose Estimation
Relationship Detection
Image Captioning
Visual QA
Referring Expressions
Situation Recognition
Text-based Image Generation
Visual Commonsense
Classification in context
Region Captioning
GLUE Benchmark tasks
Reading comprehension
Natural Language Inference
Grounded Commonsense Inference

J. Lu et al. Unified-IO: A Unified Model for Vision, Language, and Multi-Modal Tasks. ICLR 2023

Multi-modal prediction



S. Menon et al. [ViperGPT: Visual Inference via Python Execution for Reasoning.] ICCV 2023

Multi-modal prediction

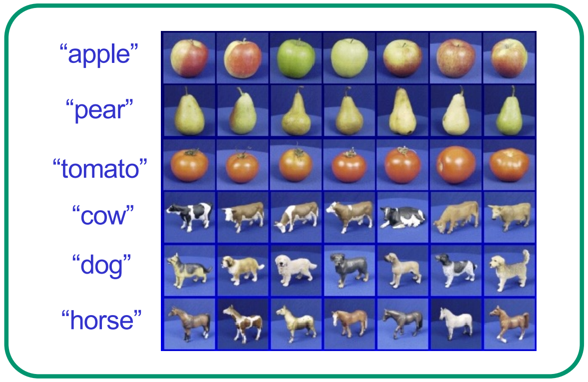
S. Menon et al. [ViperGPT: Visual Inference via Python Execution for Reasoning.] ICCV 2023

Taxonomy of learning problems

- **Type of output**
  - Classification
  - Regression
  - Structured prediction
  - Dense prediction
  - Multi-modal prediction
- **Type of supervision**
  - Fully supervised
  - Unsupervised
  - Self-supervised or predictive learning

Type of supervision

- Traditional (over-simplified) dichotomy



Supervised learning: clean, complete training labels for the task of interest

Unsupervised learning: no labels

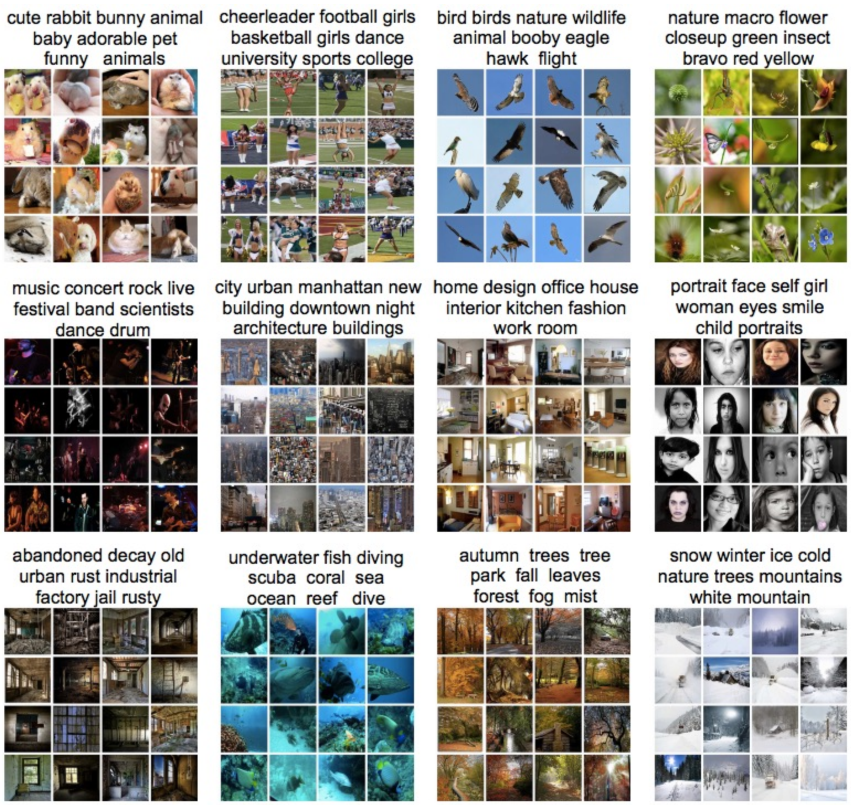
Unsupervised learning

- Given: large collection of unlabeled data
- Goal: ???

Image source

Unsupervised learning

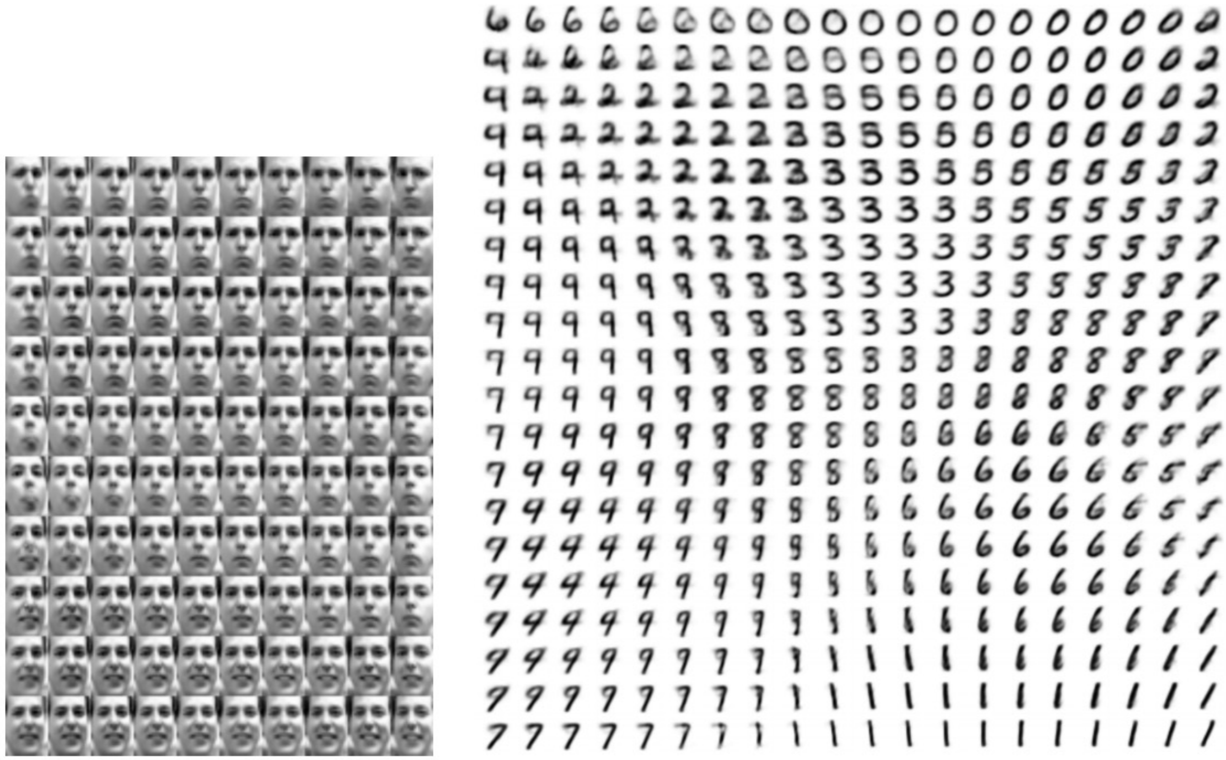
- **Clustering**
  - Discover groups of “similar” data points



Y. Gong, Q. Ke, M. Isard, and S. Lazebnik. A Multi-View Embedding Space for Modeling Internet Images, Tags, and Their Semantics. IJCV 2014

Unsupervised learning

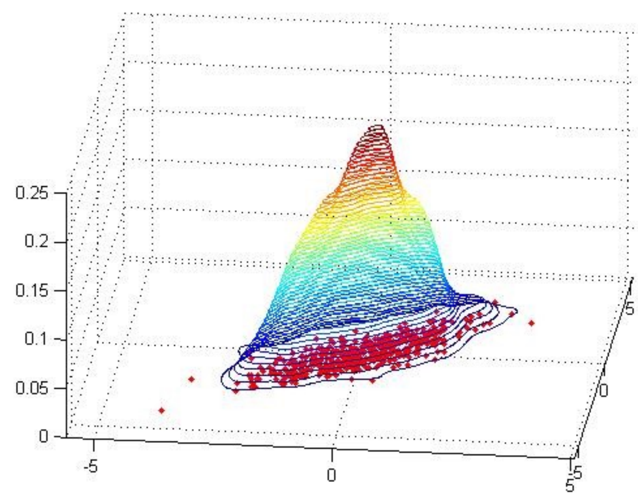
- **Dimensionality reduction, manifold learning**
  - Discover a lower-dimensional surface on which the data lives



D. Kingma and M. Welling, [Auto-Encoding Variational Bayes](), ICLR 2014

Unsupervised learning

- Learning the data distribution
  - **Density estimation**: Find a function that approximates the probability density of the data (i.e., value of the function is high for “typical” points and low for “atypical” points)
  - An extremely hard problem for high-dimensional data…



Unsupervised learning

- Learning the data distribution
  - **Learning to sample:** Produce samples from a data distribution that mimics the training set

Generative adversarial networks (GANs)

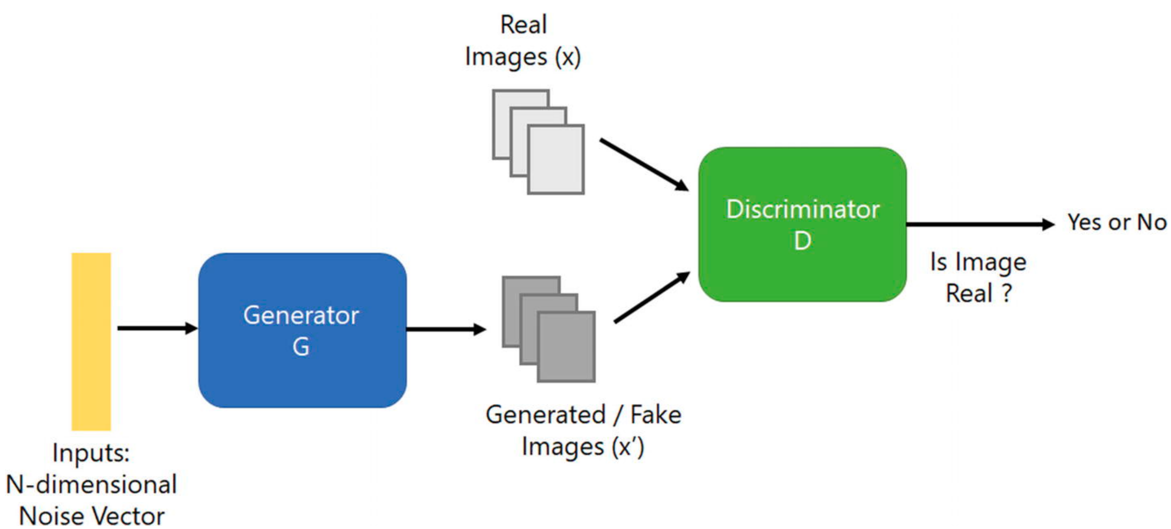
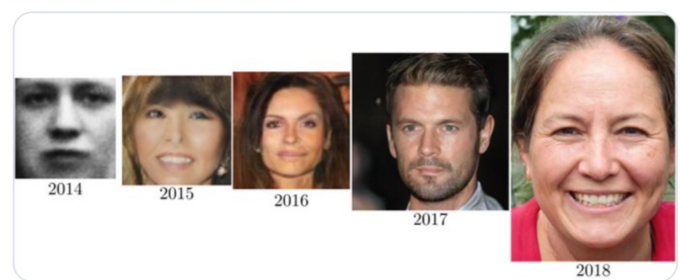


Image source

Ian Goodfellow
@goodfellow_ian

4.5 years of GAN progress on face generation.
arxiv.org/abs/1406.2661 arxiv.org/abs/1511.06434
arxiv.org/abs/1606.07536 arxiv.org/abs/1710.10196
arxiv.org/abs/1812.04948

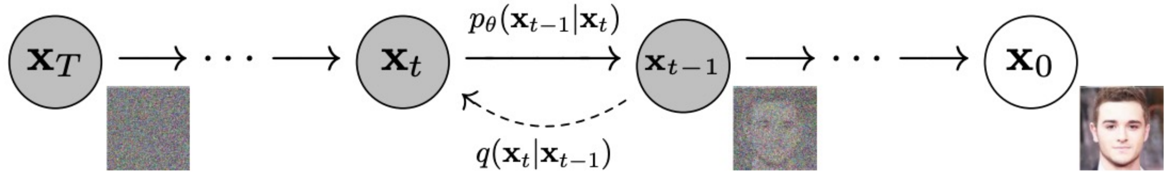


6:40 PM · Jan 14, 2019

Unsupervised learning

- Learning the data distribution
  - **Learning to sample:** Produce samples from a data distribution that mimics the training set

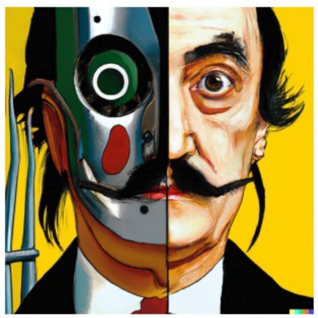
[Denoising diffusion probabilistic models](#) (DDPMs)



Unsupervised learning

- Learning the data distribution
  - **Learning to sample:** Produce samples from a data distribution that mimics the training set

Denoising diffusion probabilistic models (DDPMs)



vibrant portrait painting of Salvador Dalí with a robotic half face

a shiba inu wearing a beret and black turtleneck

a close up of a handpalm with leaves growing from it

an espresso machine that makes coffee from human souls, artstation

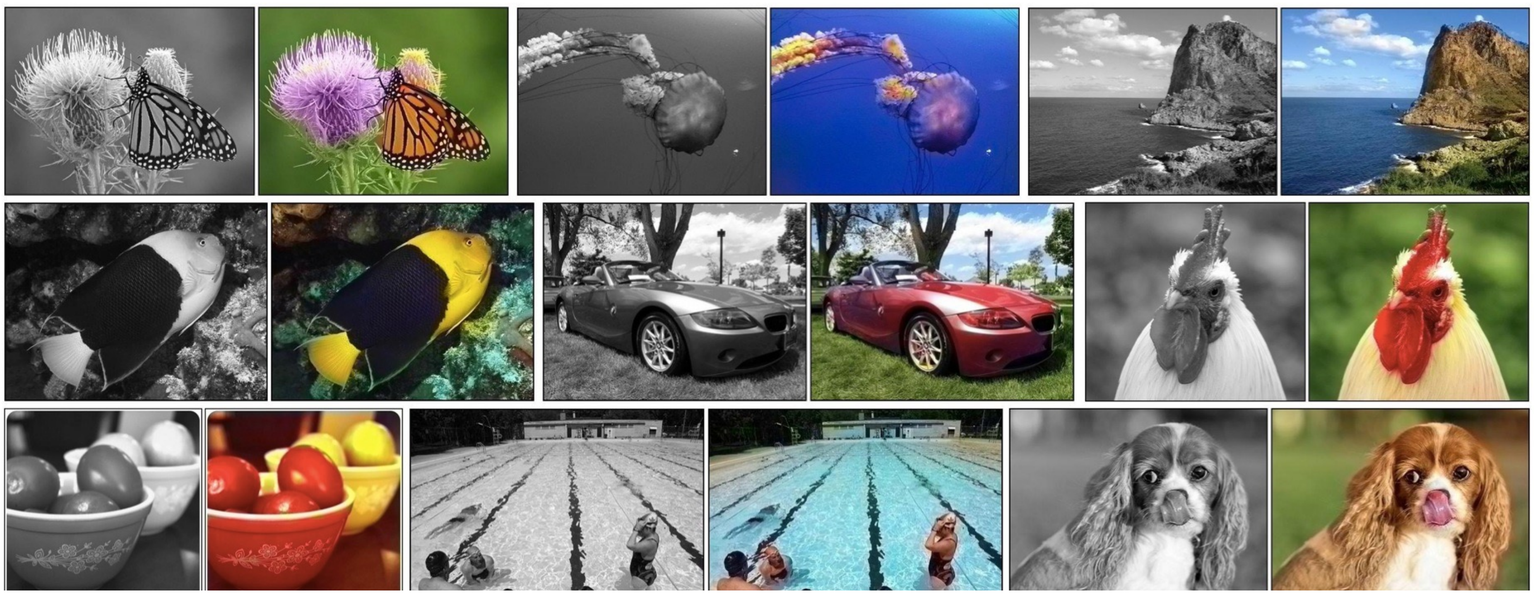
panda mad scientist mixing sparkling chemicals, artstation

a corgi's head depicted as an explosion of a nebula

Source: DALL-E 2

Self-supervised or predictive learning

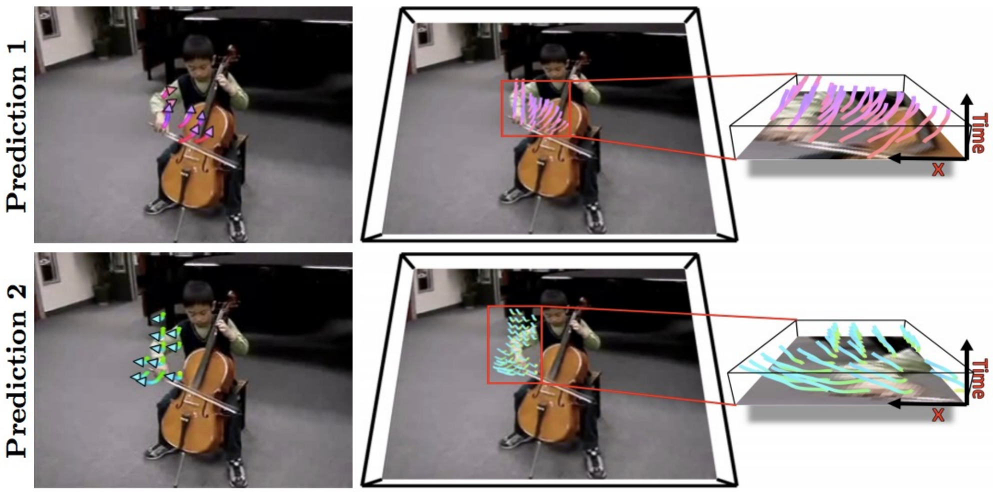
- Use part of the data to predict other parts of the data
  - Example: Image colorization

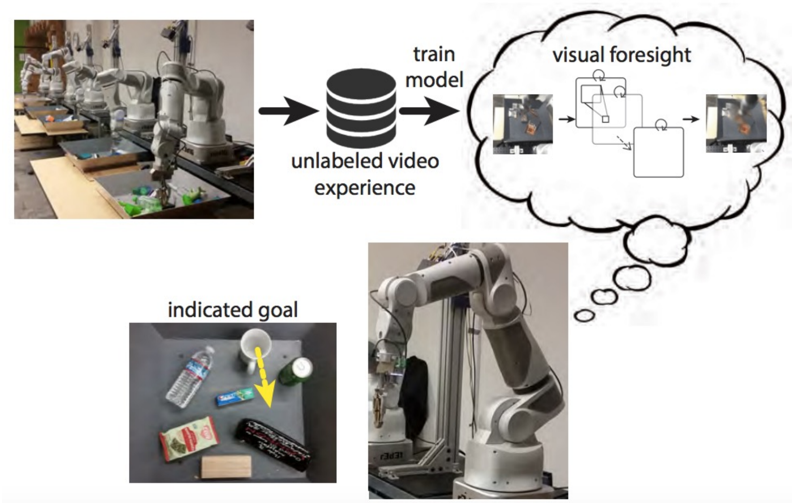


R. Zhang et al., Colorful Image Colorization, ECCV 2016

Self-supervised or predictive learning

- Use part of the data to predict other parts of the data
  - Example: Future prediction





J. Walker et al. [An Uncertain Future: Forecasting from Static Images Using Variational Autoencoders](). ECCV 2016

C. Finn and S. Levine. [Deep Visual Foresight for Planning Robot Motion](). ICRA 2017. [YouTube video]()

Self-supervised or predictive learning

- Use part of the data to predict other parts of the data
  - Example: Grasp prediction

L. Pinto and A. Gupta. [Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours.]() ICRA 2016

Self-supervised or predictive learning

- Use part of the data to predict other parts of the data
  - Example: Next/masked word prediction

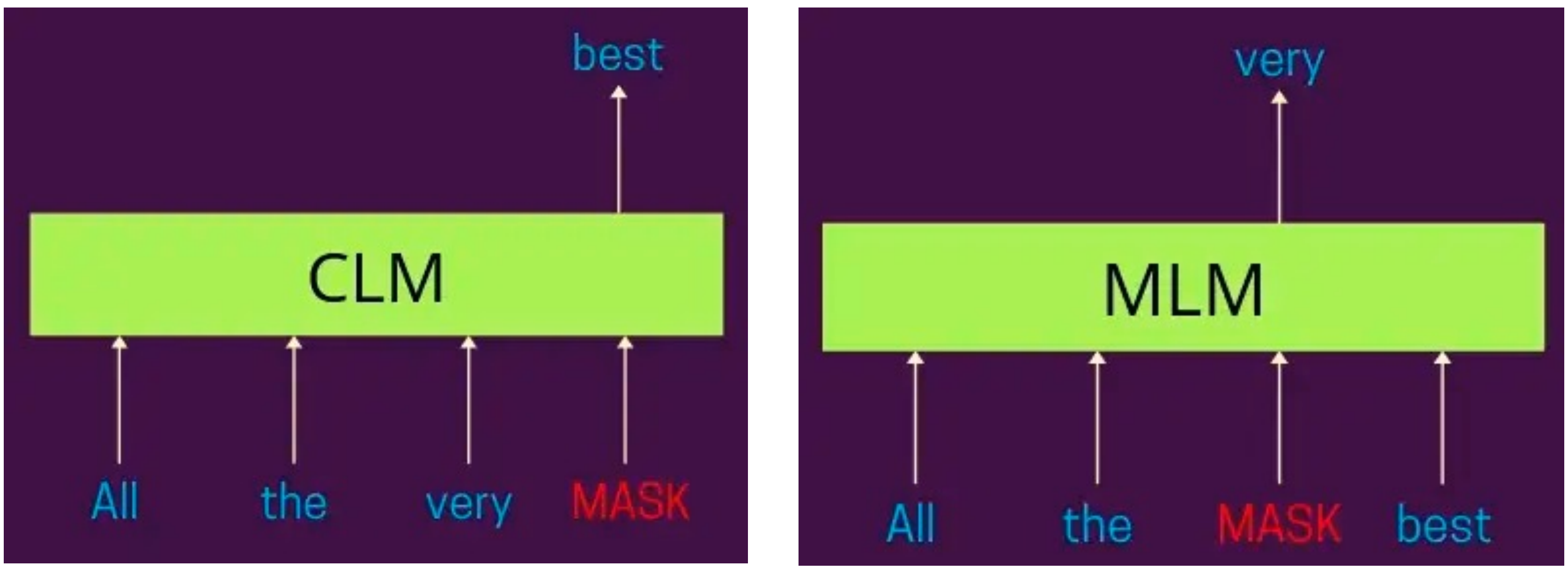
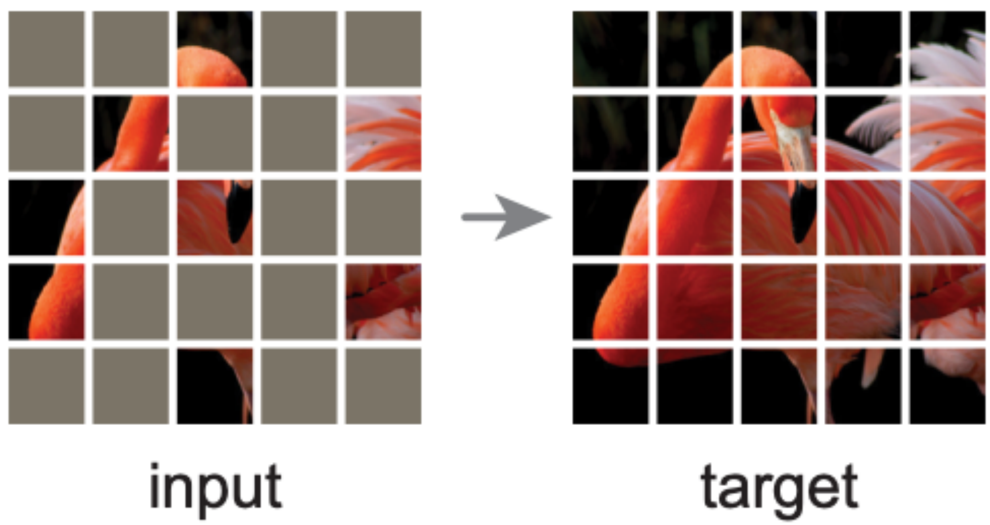


Figure source

Self-supervised or predictive learning

- Use part of the data to predict other parts of the data
  - Example: Masked patch prediction



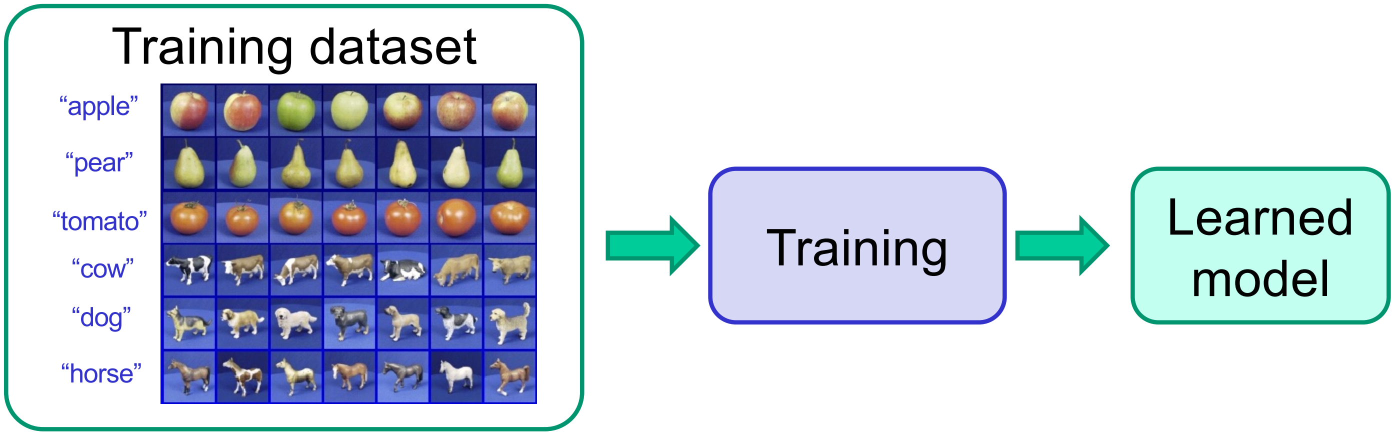
K. He et al. [Masked autoencoders are scalable vision learners](). CVPR 2022

Taxonomy of learning problems

- **Type of output**
  - Classification
  - Regression
  - Structured prediction
  - Dense prediction
  - Multi-modal prediction
- **Type of supervision**
  - Fully supervised
  - Unsupervised
  - Self-supervised or predictive learning
- **Training regime**
  - Batch offline learning
  - Online/continual learning
  - Active learning
  - Reinforcement learning

Training regime

Offline learning

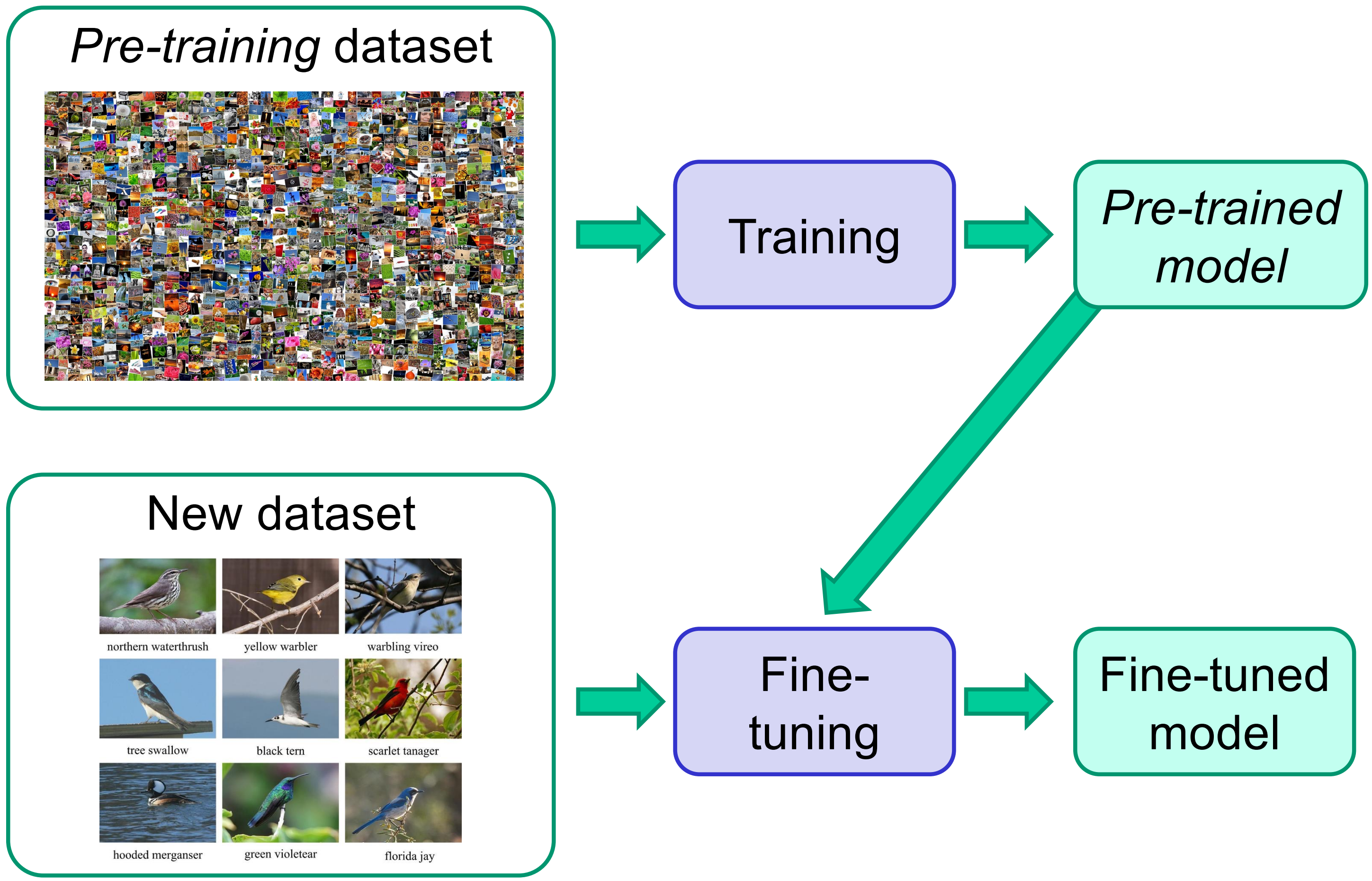


Process the entire training set (typically in multiple passes)

Challenges: static dataset, high storage, memory requirements

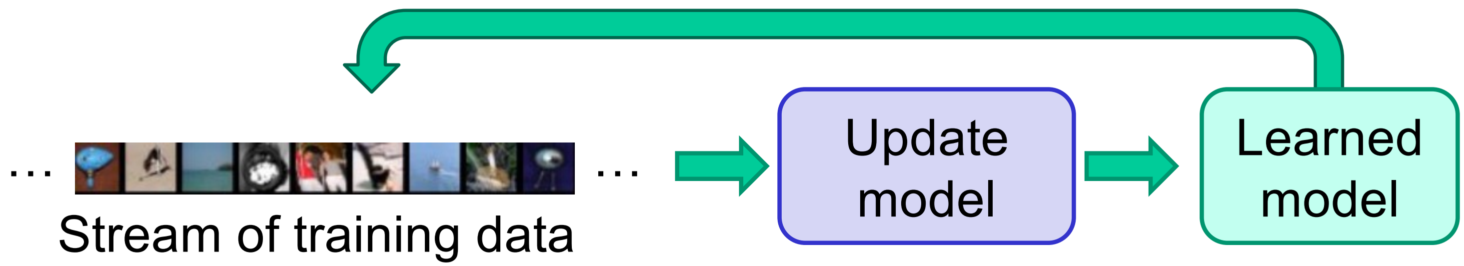
Training regime

Transfer learning



Training regime

Online learning, continual learning

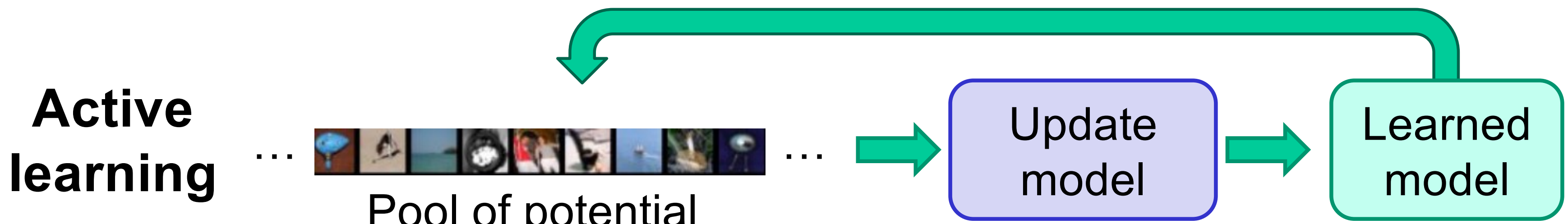


Update model based on continuous stream of data, do not revisit samples

Challenges: changes in the data distribution, *catastrophic forgetting*

Training regime

Model asks for next training example, annotation

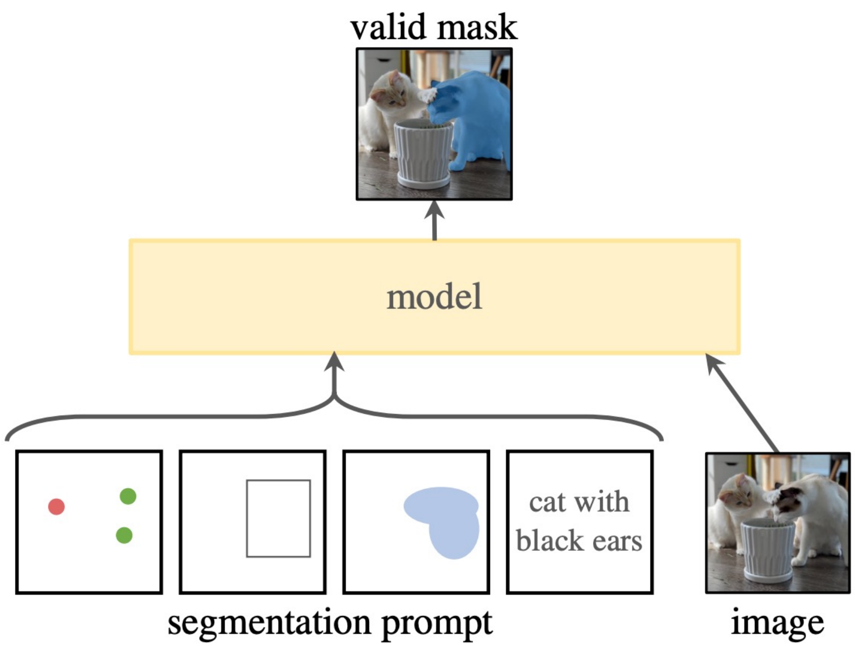


Pool of potential training examples

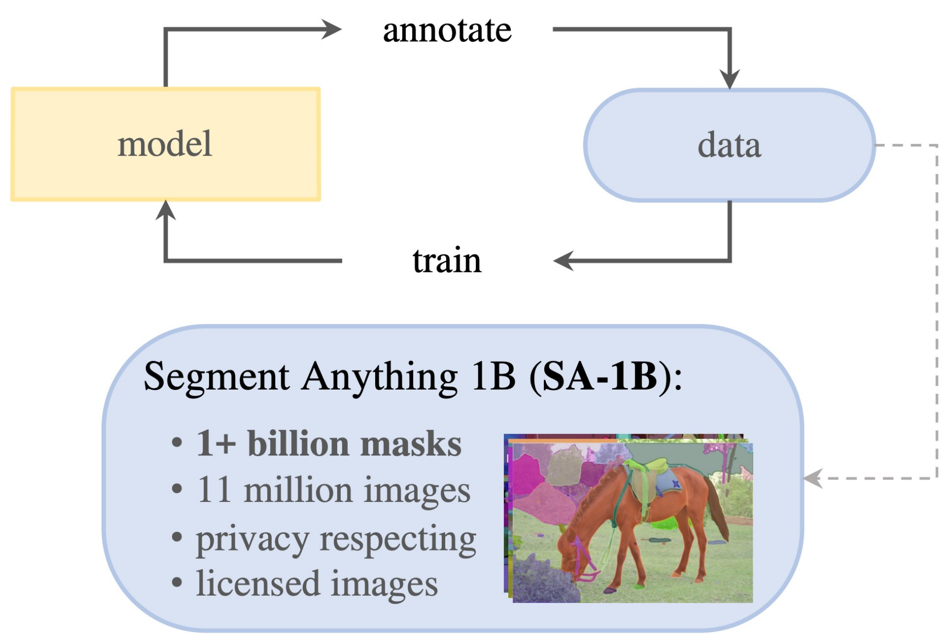
Challenges: scalability, availability of annotators, difficulty of evaluation

Today's trend: *Data engines*

Task: Promptable segmentation



Data engine

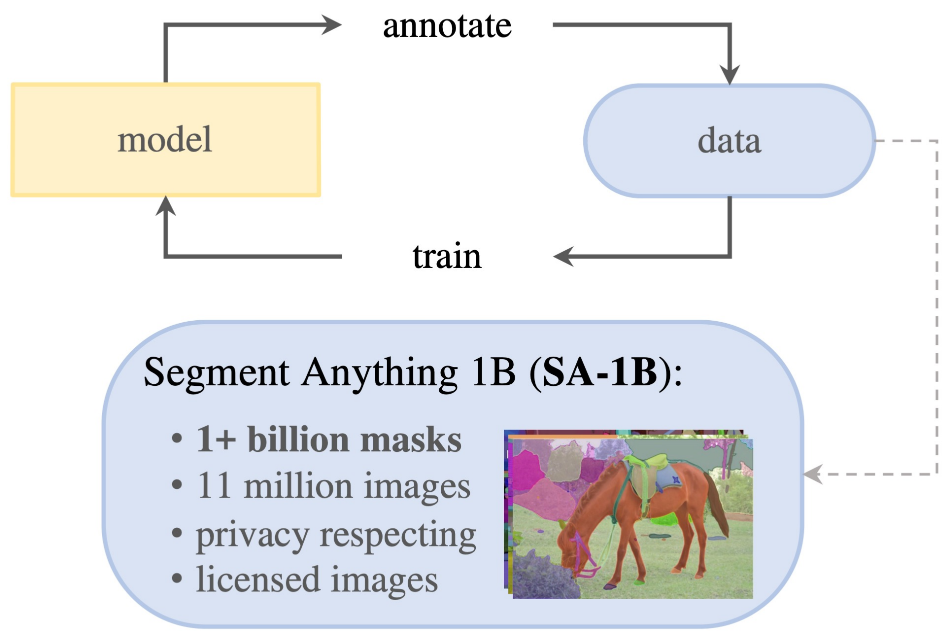


A. Kirilllov et al. [Segment anything](). ICCV 2023

Today’s trend: *Data engines*

- Data engine steps:
  1. **Pre-training** using public datasets
  2. **Assisted manual stage**: interactive segmentation with SAM assisting annotators
  3. **Semi-automatic stage**: SAM generates confident masks, annotators add masks to improve diversity
  4. **Fully automatic stage**: SAM generates ~100 masks per image starting with a grid of points

Data engine



A. Kirilllov et al. [Segment anything](). ICCV 2023

Reinforcement learning

- Learning for an agent that can affect the world through actions

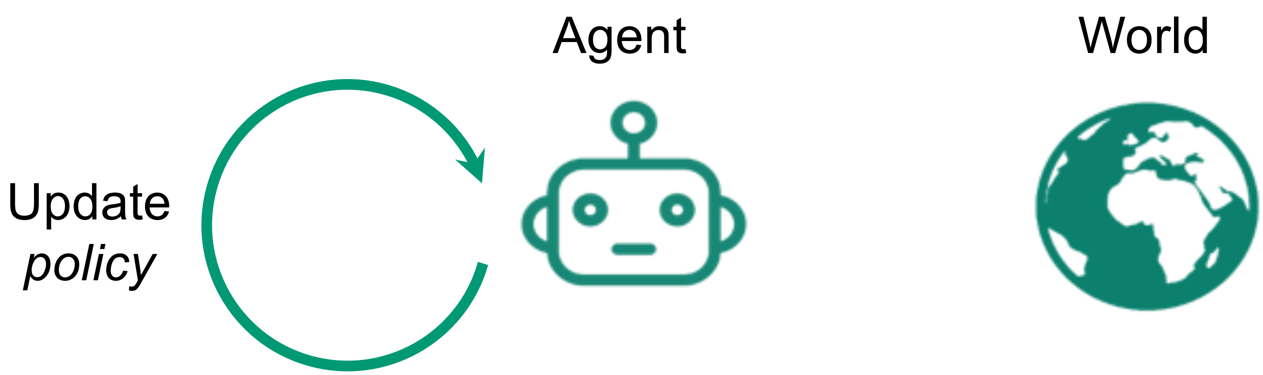
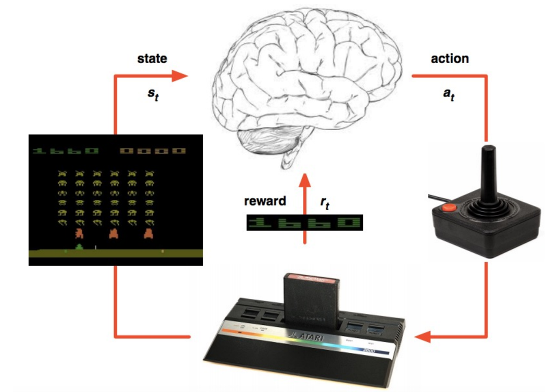


Figure source

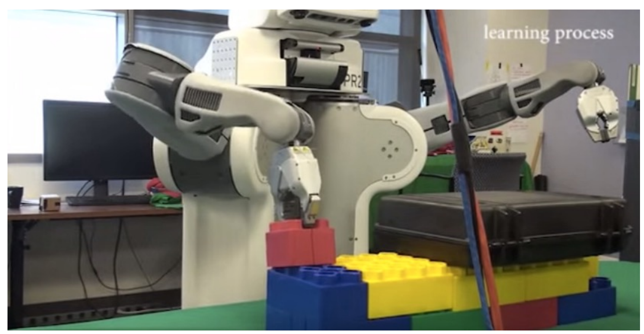
Reinforcement learning: Examples

[DeepMind's AlphaGo]

[DeepMind's Atari system]



[Sensorimotor learning]



Sensorimotor learning

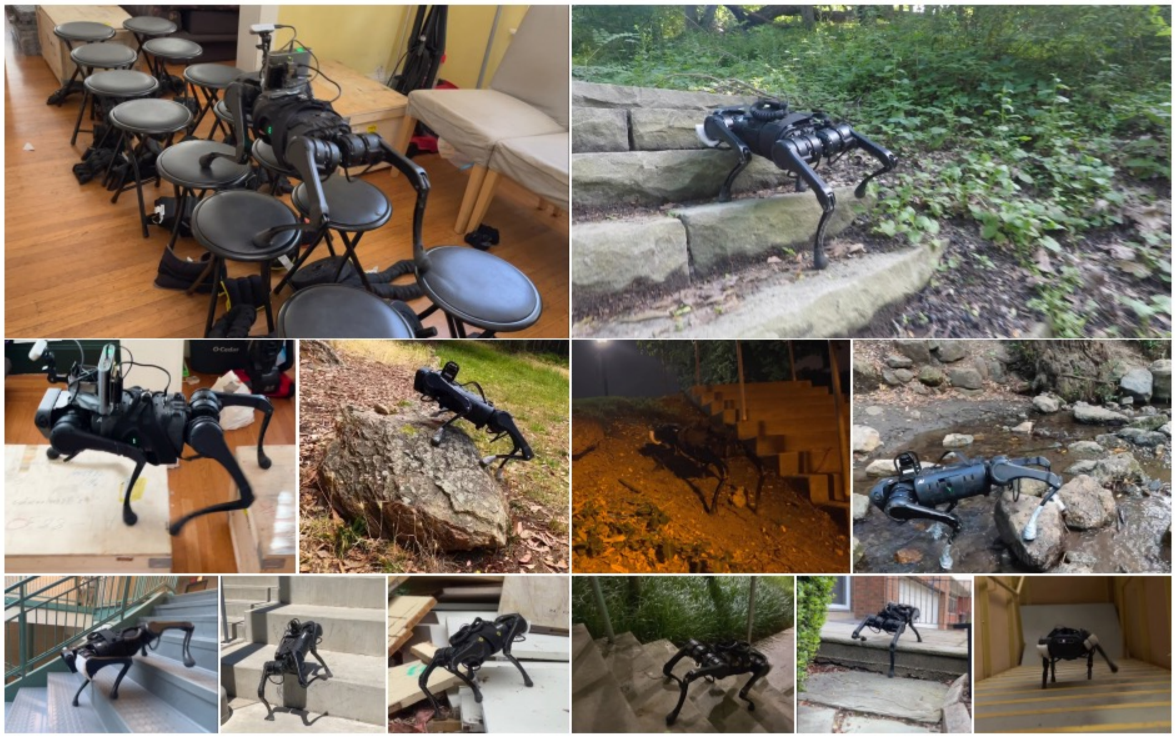


Figure 1: Our robot can traverse a variety of challenging terrain in indoor and outdoor environments, urban and natural settings during day and night using a single front-facing depth camera. The robot can traverse curbs, stairs and moderately rocky terrain. Despite being much smaller than other commonly used legged robots, it is able to climb stairs and curbs of a similar height. Videos at https://vision-locomotion.github.io

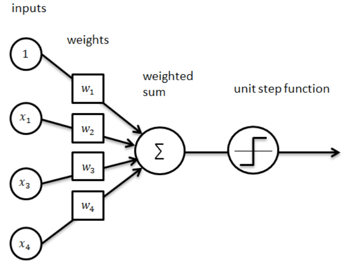
A. Agarwal, A. Kumar, J. Malik, and D. Pathak. Legged Locomotion in Challenging Terrains using Egocentric Vision. CoRL 2022

Overview

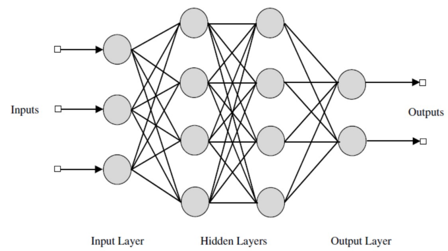
- Logistics
- Motivation: The statistical learning viewpoint
- A taxonomy of learning problems
- Topics to be covered in class

Topics to be covered in class

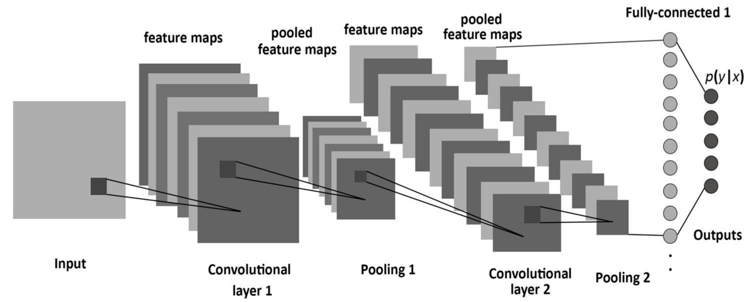
ML basics, linear classifiers



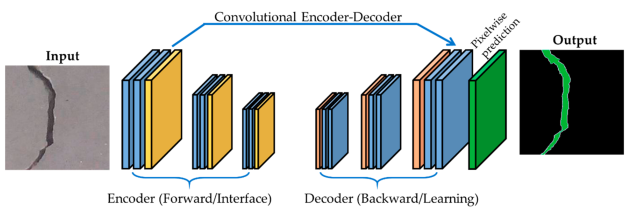
Multilayer neural networks, backpropagation



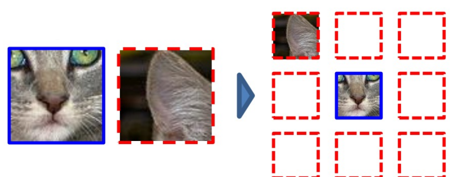
Convolutional networks for classification



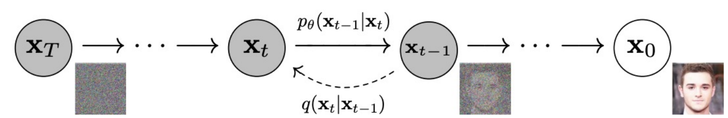
Networks for detection, dense prediction



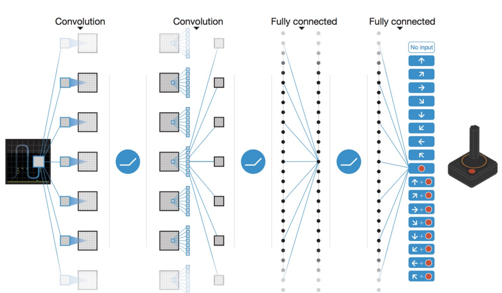
Self-supervised learning



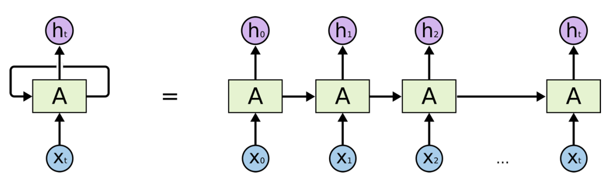
Generative models: GANs, image-to-image translation, diffusion models



Deep reinforcement learning



Models for sequence data



Transformers, large language models, transformers for vision