Deep Learning for Data Science
DS 542

https://dl4ds.github.io/fa2024

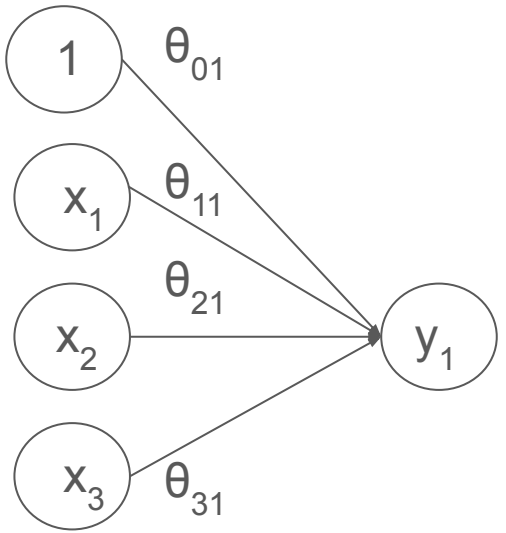
Today

- Deep learning - what/why/now?
- Applications of deep learning
- Course logistics

Perceptrons

“The Perceptron: A Probabilistic Model For Information Storage And Organization in the Brain”
by Rosenblatt (1958)

- First instance of “artificial neural networks”.
- Loosely inspired by our brain…
- Perceptron special case
  - No hidden layer
  - Easy training algorithm
  - Easy to prove limits on expressiveness (parity can not work)



General Neural Networks

- Limits of Perceptrons were misunderstood
- Did not apply to general neural network configurations

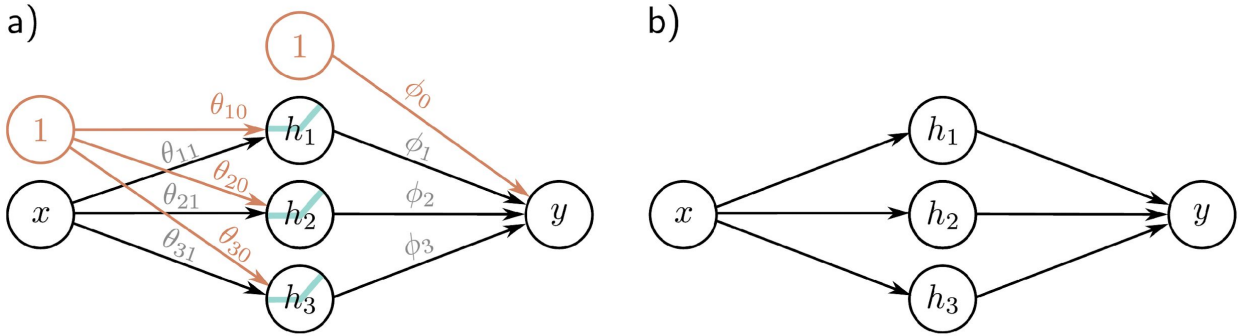


Image source: Understanding Deep Learning

Universal Approximation Theorem for Neural Networks

“Approximation by superpositions of a sigmoidal function”
by Cybenko (1989)

- One hidden layer is enough to approximate any smooth function
  - Can approximate to arbitrary precision
  - Catch is that hidden layer might be really big
  - Does not say how to train such a neural network from data
- Many variations of this theorem covering almost any way we design a neural network
  - Main exception is that polynomial activation functions don’t work.
- Some exponential lower size bounds for low numbers of hidden layers (e.g. parity requires exponential sized layers if # of hidden layers is fixed)

Bigger Neural Networks Organized in Layers

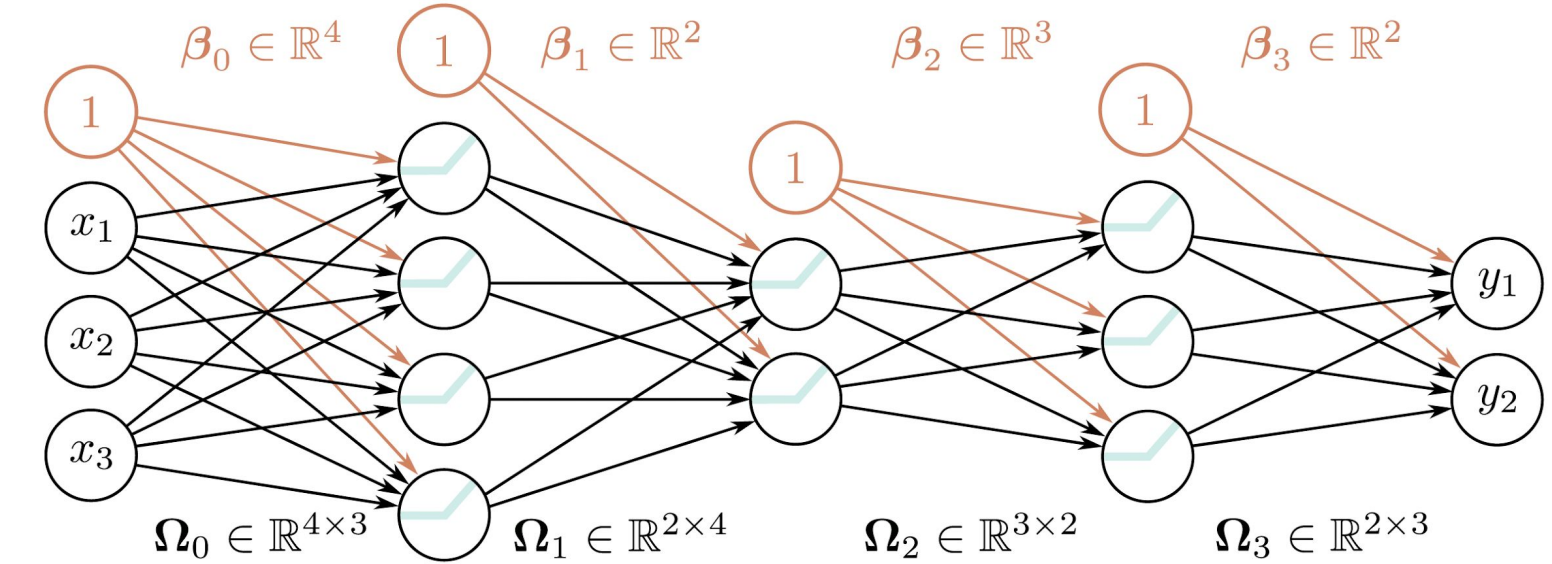


Image source: Understanding Deep Learning

Neural Network Training by Back Propagation

“Learning representations by back-propagating errors”
by Rumelhart, Hinton, and Williams (1986)

TLDR: use calculus to figure out how to update neural network parameters

- Incremental improvement process
- Does not guarantee best possible parameter values will be chosen

What is Deep Learning?

Neural network definition:

- Learning with lots of layers in the neural network
- Historically 5-10 layers was ~~difficult to train~~ deep
- Once we could train deep neural networks consistently, jumped to 100s

There are other deep differentiable computations, but assume neural networks unless a different context is explicit.

Why Deep Learning?

Costs of Universal Approximation

- Theory results
  - Some problems require huge shallow networks
  - Size bounds tend to drop for deeper networks
  - Parity: fixed layers → exponential size, logarithmic layers → linear size

Why Deep Learning Now?

- Deep neural network advantages have been known for decades.
- Why traction now?
  - Historically 5-10 layers was difficult to train
  - Many incremental results in the 1990s and early 2000s
  - Then they were all combined in 2012, and
  - Suddenly deep learning was both viable and better…

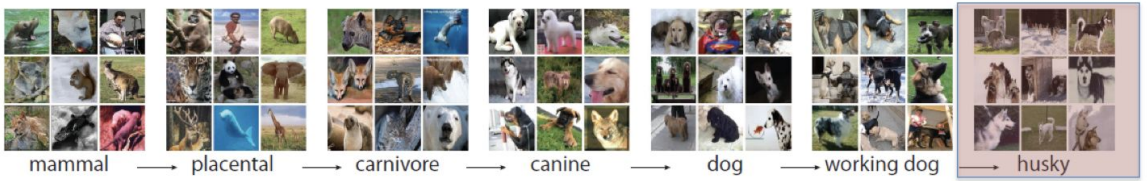
“Large Scale Visual Recognition Challenge 2010” by Berg, Brook, Deng, Li (2010)

aka the ImageNet competition…

- Dataset built in 2009
- First Competition in 2010
- Match images to 1000 classes
- Score on top 5 accuracy

IMAGENET is a knowledge ontology

- Taxonomy



- S: (n) Eskimo dog, **husky** (breed of heavy-coated Arctic sled dog)
  - *direct hypernym / **inherited hypernym** / sister term*
    - S: (n) working dog (any of several breeds of usually large powerful dogs bred to work as draft animals and guard and guide dogs)
      - S: (n) dog, domestic dog, Canis familiaris (a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) *"the dog barked all night"*
        - S: (n) canine, canid (any of various fissiped mammals with nonretractile claws and typically long muzzles)
          - S: (n) carnivore (a terrestrial or aquatic flesh-eating mammal) *"terrestrial carnivores have four or five clawed digits on each limb"*
            - S: (n) placental, placental mammal, eutherian, eutherian mammal (mammals having a placenta; all mammals except monotremes and marsupials)
              - S: (n) mammal, mammalian (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
                - S: (n) vertebrate, craniate (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
                  - S: (n) chordate (any animal of the phylum Chordata having a notochord or spinal column)
                    - S: (n) animal, animate being, beast, brute, creature, fauna (a living organism characterized by voluntary movement)
                      - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
                        - S: (n) living thing, animate thing (a living (or once living) entity)
                          - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
                            - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
                              - S: (n) physical entity (an entity that has physical existence)
                                - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Image source:
https://image-net.org/static_files/files/pascal_ilsvrc.pdf

Then in 2012…

One of these is not like the others…

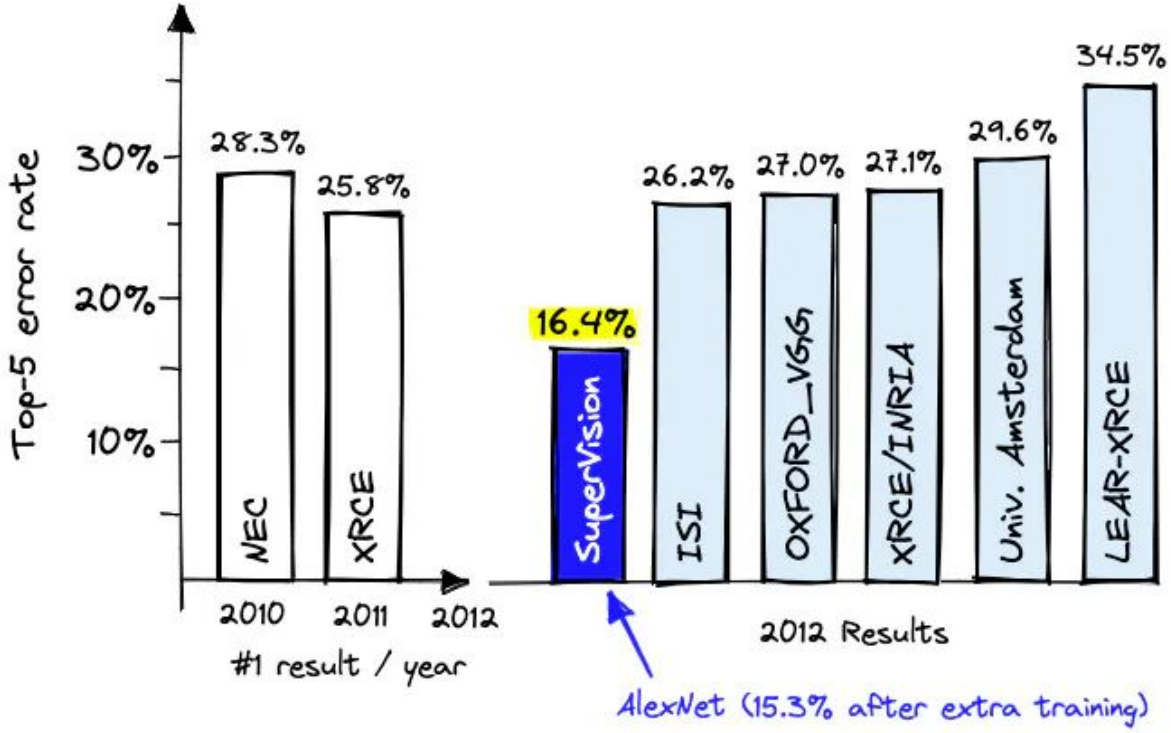


Image source: https://www.pinecone.io/learn/series/image-search/imagenet/

“ImageNet Classification with Deep Convolutional Neural Network” by Krizhevsky, Sutskever, Hinton (2012)

aka AlexNet

- Key changes*
  - Deep neural network
  - Pre-training (15M labeled images)
  - GPU usage (more compute)

Result: “achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry”

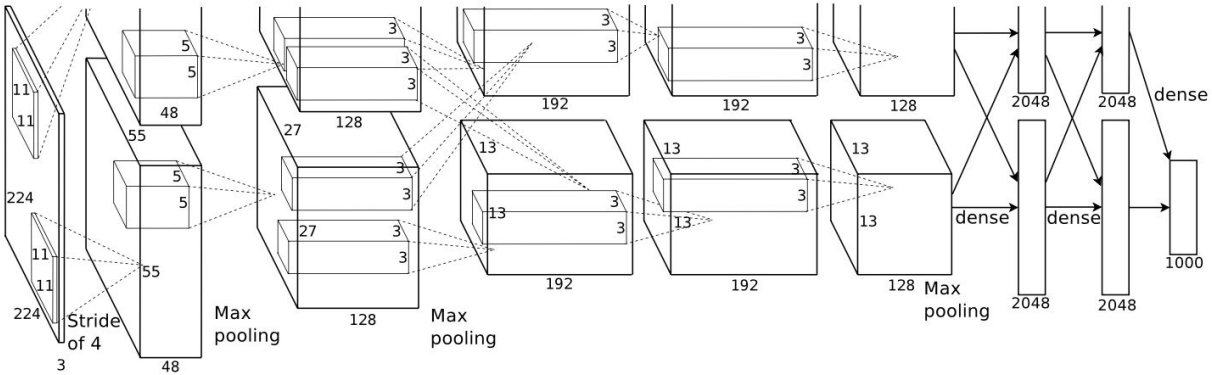
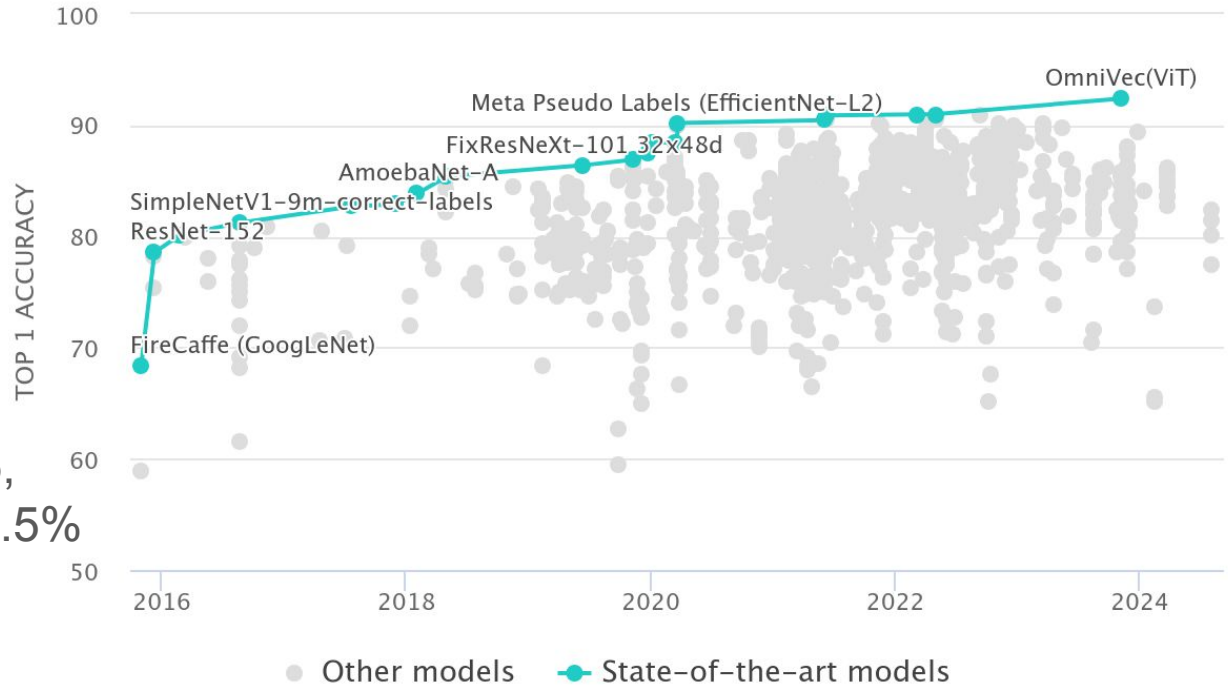


Image source: AlexNet paper

ImageNet Now



Top 1, not top 5, AlexNet had 62.5% top 1 accuracy

Image Source: https://paperswithcode.com/sota/image-classification-on-imagenet

Machine Translation → Large Language Models

Learning representations by back-propagating errors by Rumelhart, Hinton, and Williams (1986)	backpropagation, recursive neural networks (called “iterated” here)
“Sequence to Sequence Learning with Neural Networks” by Sutskever, Vinyals, and Le (2014)	seq2seq
“Neural Machine Translation by Jointly Learning to Align and Translate” by Bahdanau, Cho, and Bengio (2015)	attention
“Attention is All You Need” by Vaswani et al (2017)	transformers
“Language Models are Unsupervised Multitask Learners” by OpenAI (2020)	GPT-2 aka the model too dangerous to release
“Language Models are Few-Shot Learners” by OpenAI (2020)	GPT-3 → ChatGPT → …

“The Bitter Lesson” by Richard Sutton (2019)

TLDR: Leveraging more computation consistently beats clever humans.

- Chess vs Kasparov, light heuristics and very deep search
- Alpha Go vs Lee Sedol
- Hidden Markov models for speech recognition
- Computer vision…

Combination of search and learning is very powerful.

- Specifically avoid humans trying to explain the right answer to the computer.

http://www.incompleteideas.net/IncIdeas/BitterLesson.html

“A Better Lesson” by Rodney Brooks (2019)

- Rebuttal to “The Bitter Lesson”.
- Figuring out the problem structure to be learnable is actually the hard work.
- Infinite cost avoidance would be nice.

https://rodneybrooks.com/a-better-lesson/

Applications of Deep Learning

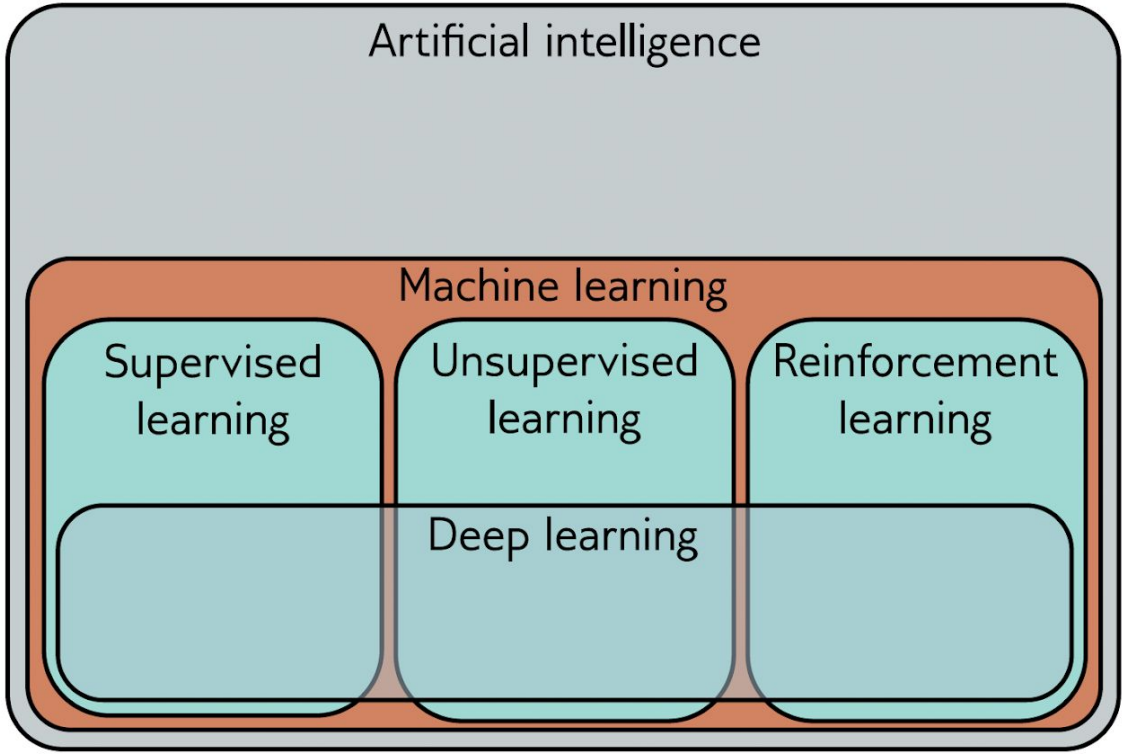


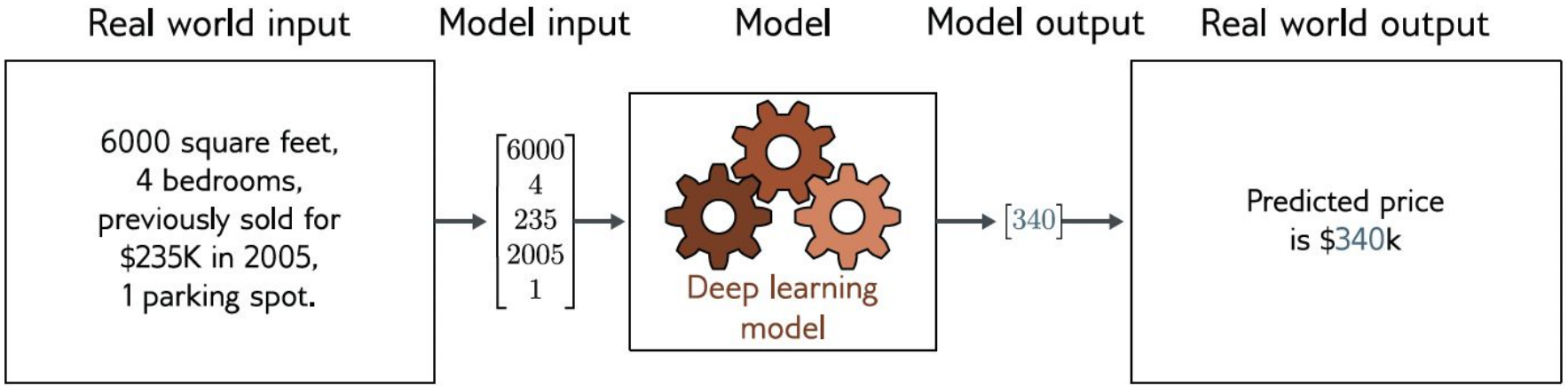
Image source: Understanding Deep Learning

Supervised Learning

Given input/output pairs, train a model that generalizes them.

- What is a model?
- What is training?
- What is generalization?

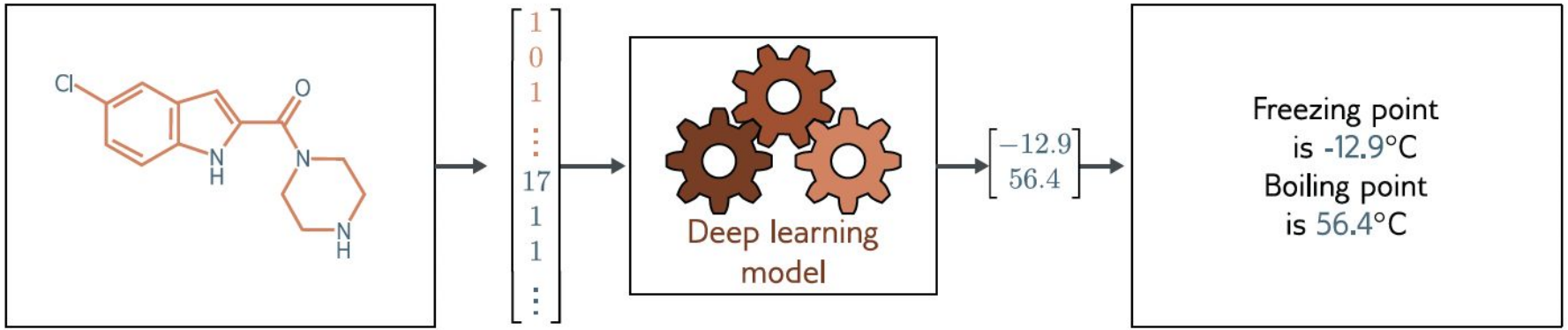
Supervised Learning - Regression



- Inputs and outputs straightforward (for a human) to extract from text.

Image source: Understanding Deep Learning

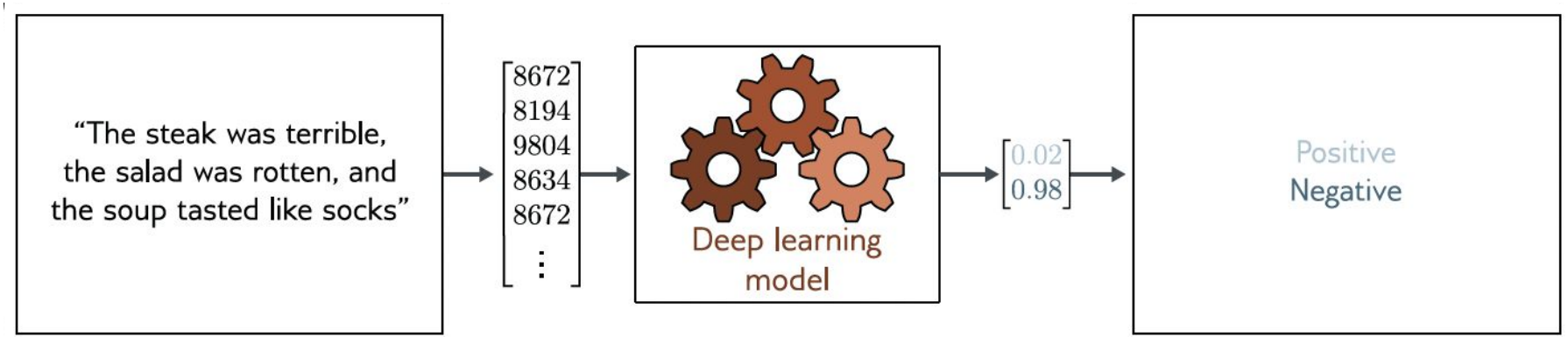
Supervised Learning - Multivariate Regression



- Where do these input numbers come from?
- Is the input size fixed?
- Do we get an advantage from making both predictions together?

Image source: Understanding Deep Learning

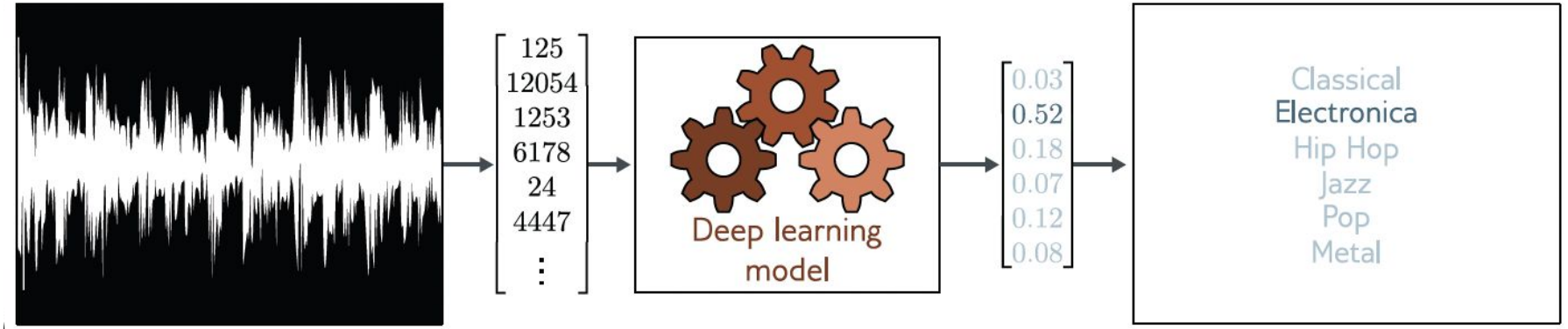
Supervised Learning - Binary Classification



- What are these input numbers?
- Input length definitely looks variable.
- Are those output numbers probabilities?

Image source: Understanding Deep Learning

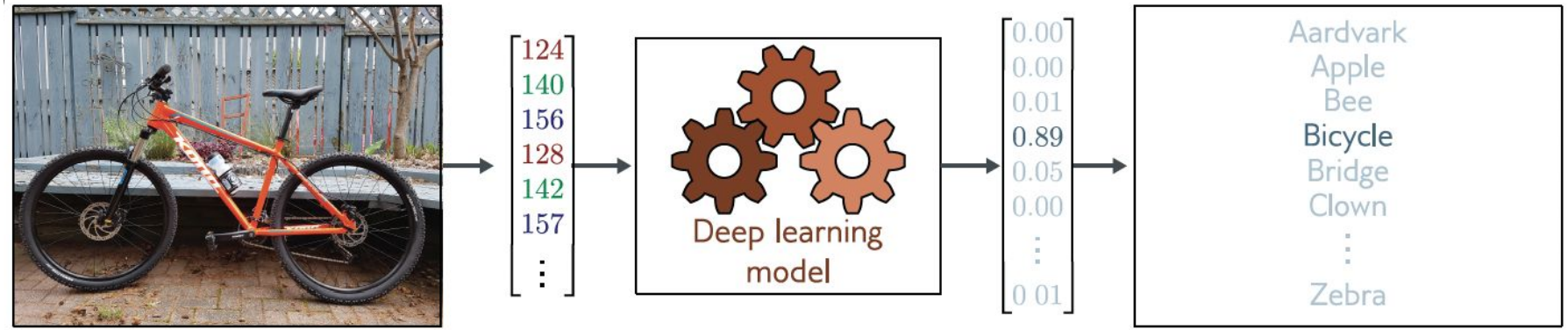
Supervised Learning - Multiclass Classification



- Are those input numbers sound samples? Or spectrograms?
- Are the output numbers probabilities?

Image source: Understanding Deep Learning

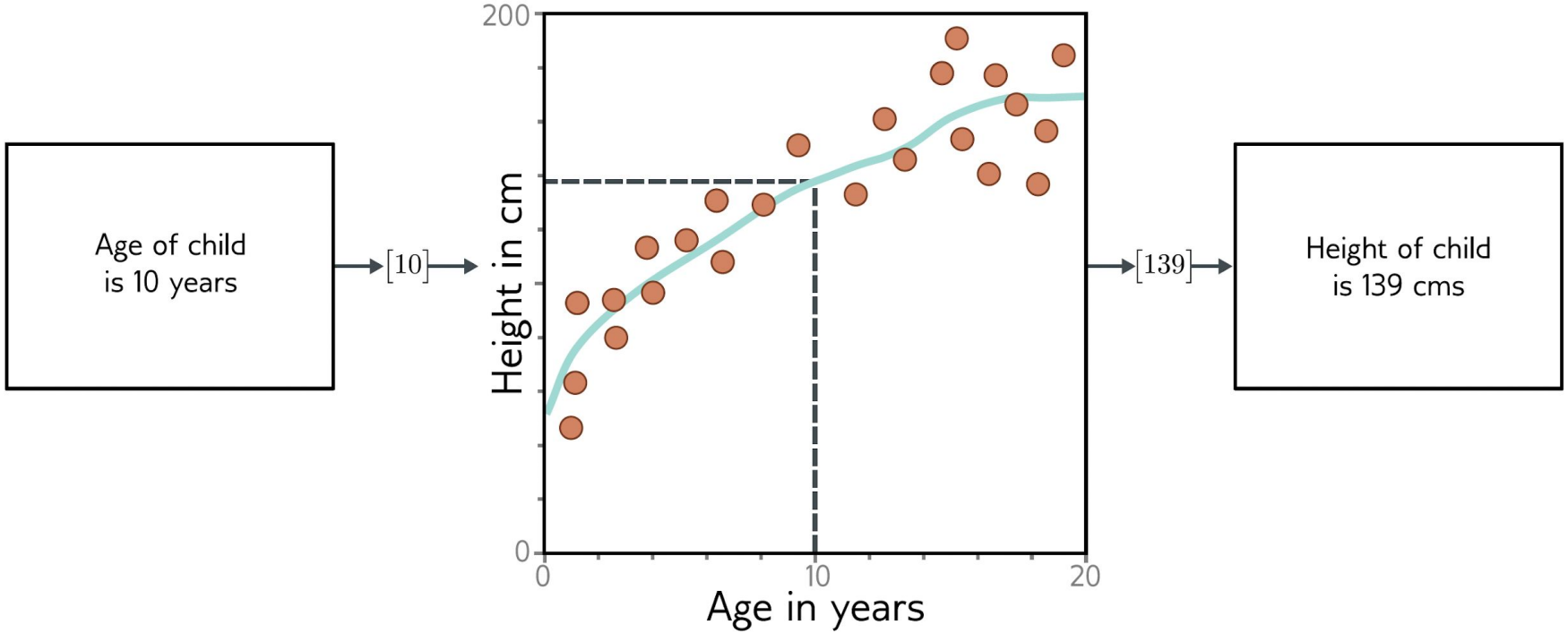
Supervised Learning - Multiclass Classification



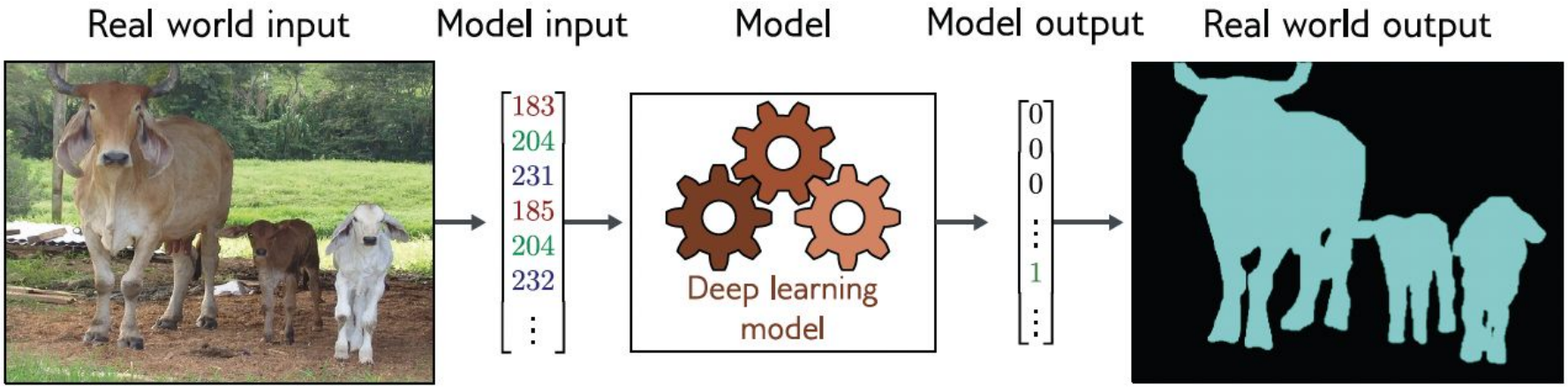
- Inputs are red, green, blue values per pixel?
- Outputs are probabilities?

Image source: Understanding Deep Learning

Models Map Inputs to Outputs Based on Training Data



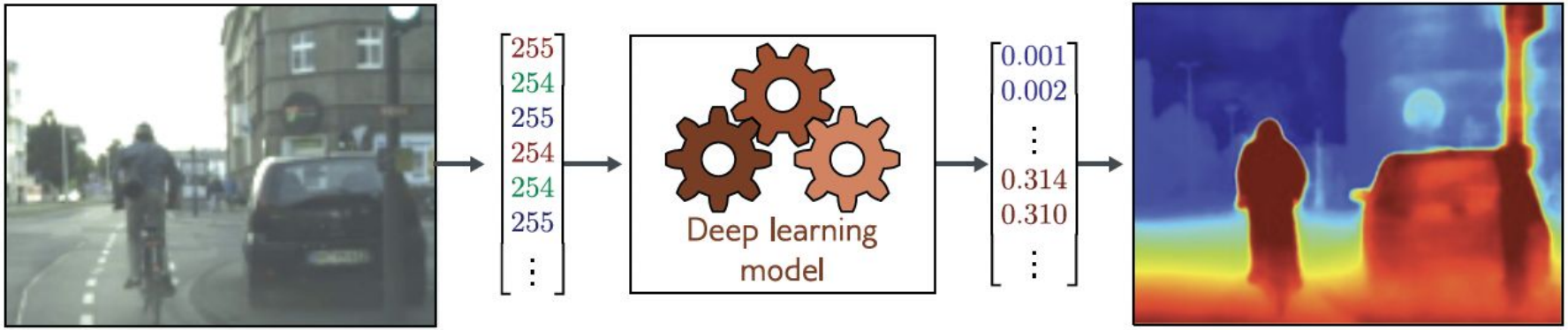
Supervised Learning - Segmentation



- Before deep learning, segmentation was mostly ad-hoc and hand-coded.

Image source: Understanding Deep Learning

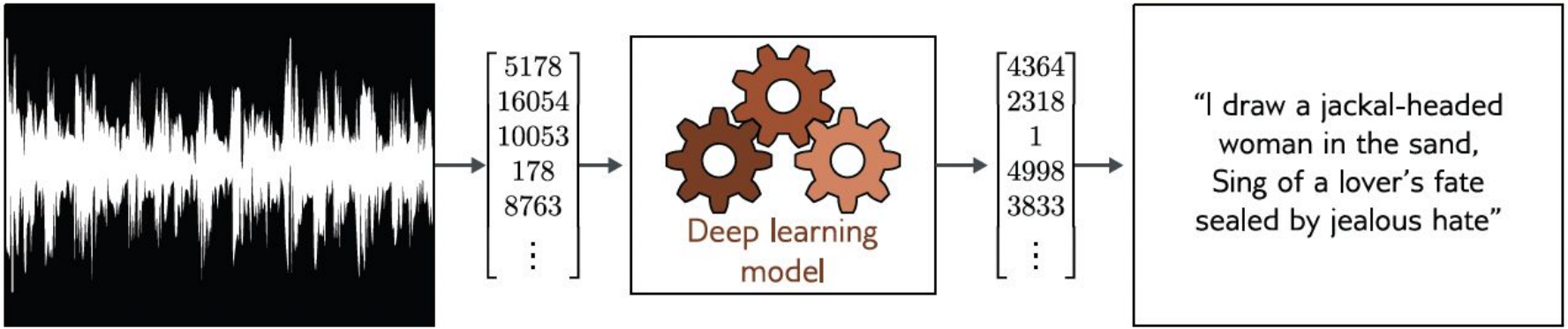
Supervised Learning - Depth Estimation



- Output based on distance
- Previous approaches like “Shape from shading” were fragile.
- But note blurriness, and mistakes.

Image source: Understanding Deep Learning

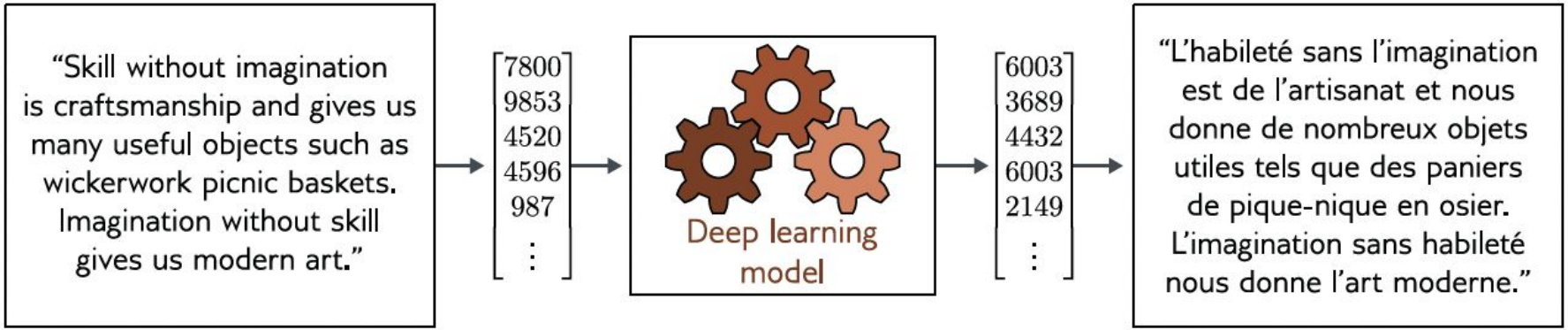
Supervised Learning - Audio Transcription



- Magic output numbers?
- Variable length output?

Image source: Understanding Deep Learning

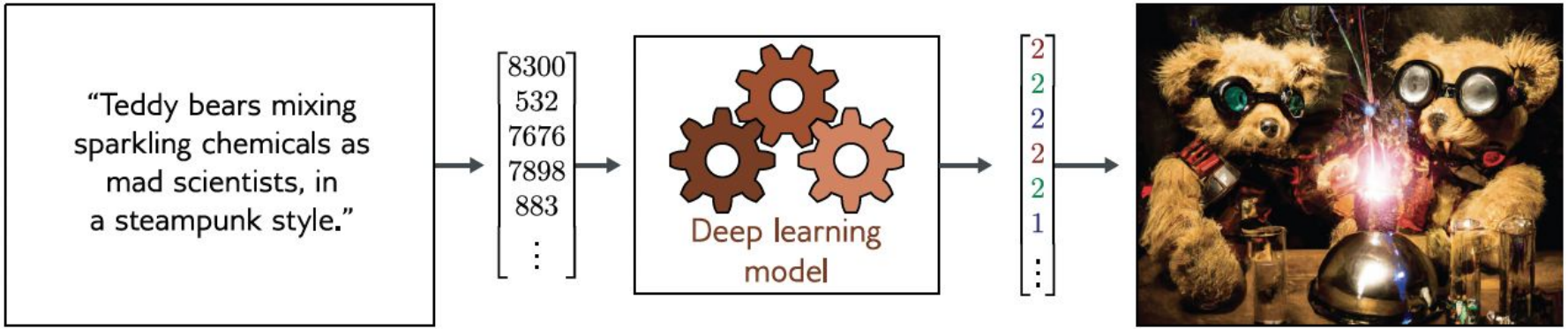
Supervised Learning - Machine Translation



- Variable length input and output.
- Before deep learning, this was mostly implemented with rules and pattern matching.

Image source: Understanding Deep Learning

Supervised Learning - Text to Image Synthesis



- Before… not even imagined?

Image source: Understanding Deep Learning, using OpenAI's DALL-E 2

Unsupervised Learning

Given a bunch of data, learn something about the data…

- Historically, this sounded like clustering.
- Sometimes is about modeling probability distributions.
- Nowadays, usually about generative models.

Unsupervised Learning - Generative Modeling

Given a bunch of data, generate more samples from the same distribution.

- Images are a favorite.
- Sometimes (base) language models are considered unsupervised.
- Boundaries get blurred with conditional generation…

Unsupervised Learning - Image Generation

"Analyzing and Improving the Image Quality of StyleGAN"
by Karras et al (2019)

Image Source: https://thispersondoesnotexist.com/

Unsupervised Learning - Image Inpainting

Image source: Understanding Deep Learning

Unsupervised Learning - Text Generation

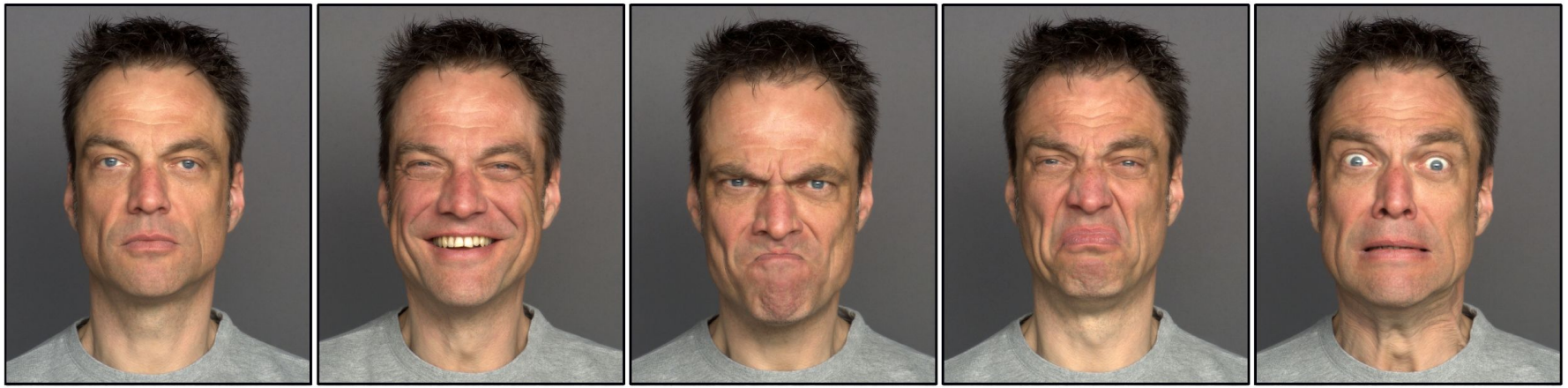
Please write me a one paragraph story starting with "The moon had risen".

ChatGPT said:

The moon had risen over the sleepy village, casting a silvery glow across the cobblestone streets and the rooftops draped in shadows. As the townsfolk slumbered, a lone figure moved silently through the mist, a cloak trailing behind like a whisper. With practiced ease, the figure approached the old, abandoned well at the edge of the forest, where legends spoke of hidden secrets buried deep. Tonight was not a night for legends, however, but for truths; and with a steady hand, the figure reached into the darkness, retrieving an ancient, dust-covered box. As the lid creaked open, the moonlight revealed a trove of forgotten memories, each one a piece of the puzzle that could restore the village's lost history and heal old wounds.

Text by https://chatgpt.com/ with prompt based on example in Understanding Deep Learning

Understanding Learning - Facial Poses



A human face has 42 muscles. Could you generate these faces with 42 inputs?

Image source: Dynamic FACES database (Holland et al. 2019)
via Understanding Deep Learning

Unsupervised Learning - Latent Variables

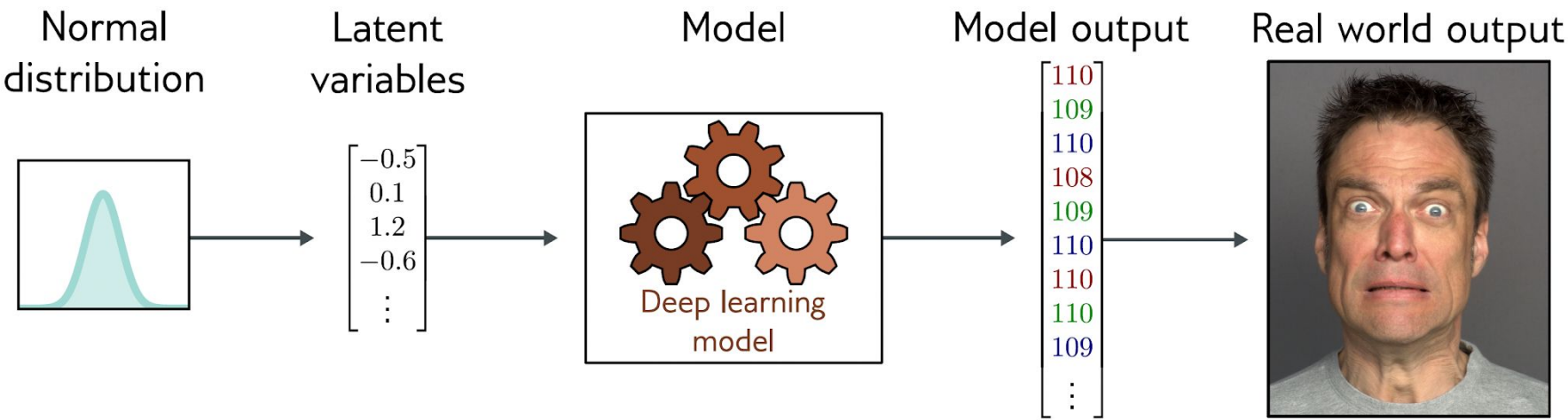


Image source: Understanding Deep Learning

Unsupervised Learning - Image Interpolation

Image source: Understanding Deep Learning

Unsupervised Learning - Text to Image Synthesis

Train text and image models with the same latent space…

Prompt: “A teddy bear on a skateboard in Times Square”

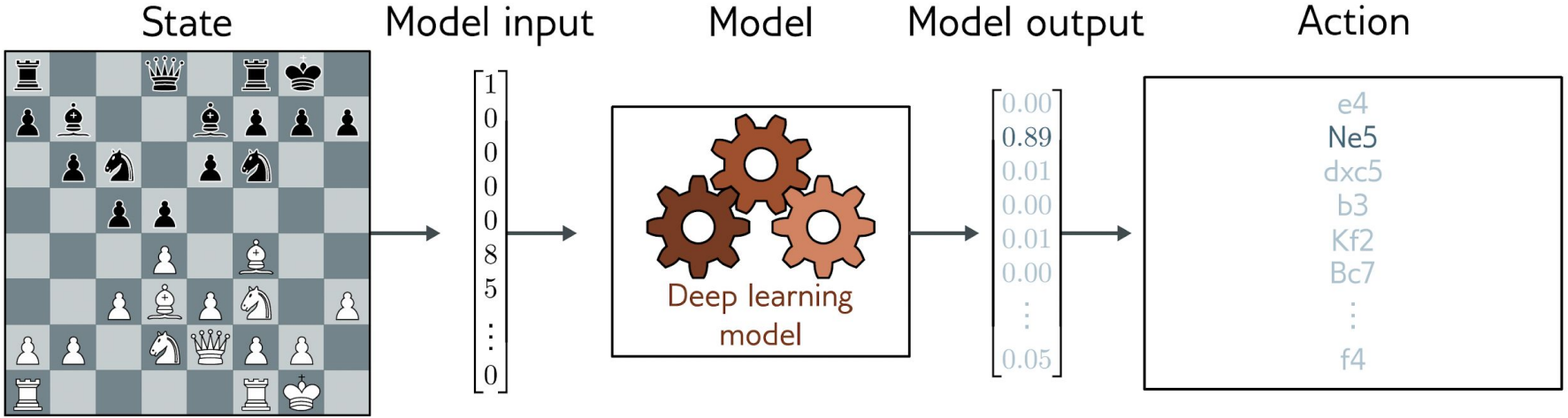
Image source: DALLE 2 (Ramesh et al. 2022)
via Understanding Deep Learning

Reinforcement Learning

Maximize the rewards of an agent over time by picking actions.

- Sometimes, only the final action gets a reward.
- Traditional example was games such as Chess or Go.
- Recently used to
  - Play video games
  - Control robots (mostly lab settings)
  - Improve the behavior of large language models

Reinforcement Learning - Game Playing

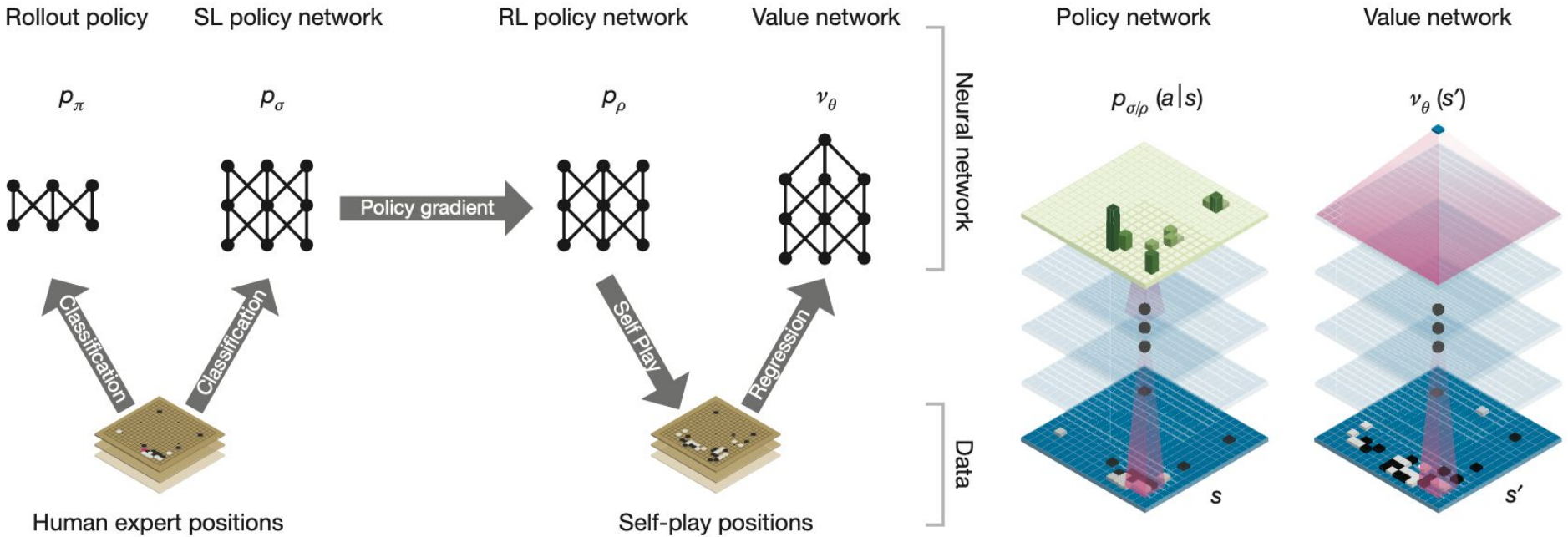


- What are the inputs? Bit vectors of possible piece/position pairs?
- What are the outputs? Action probability here, but could be action/position value too

Image source: Understanding Deep Learning

Reinforcement Learning w/Monte Carlo Tree Search

"Mastering the game of Go with deep neural networks and tree search"
by Silver et al (2016)



Reinforcement Learning - Language Model Tuning

“Training language models to follow instructions with human feedback” by Ouyang et al (2022)

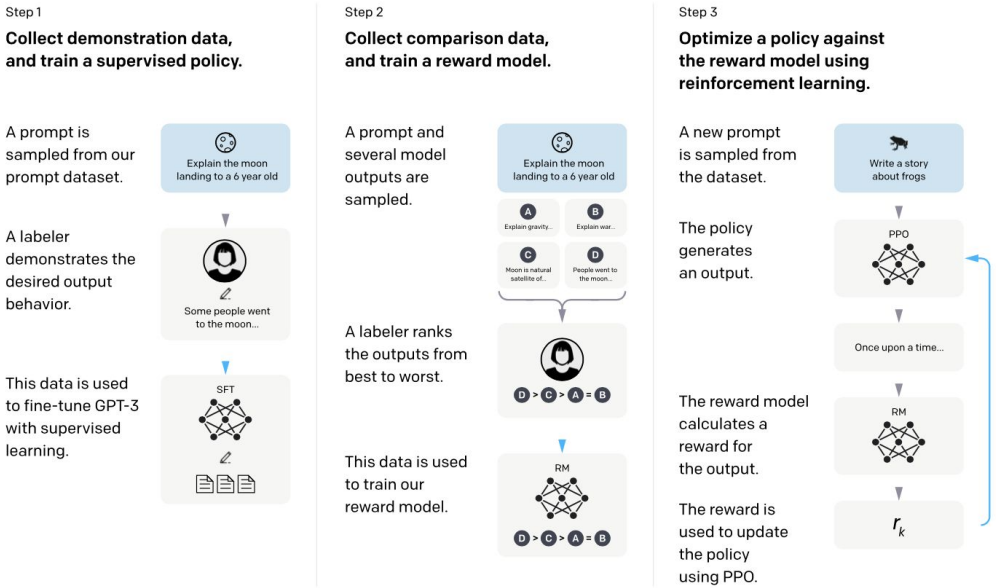


Image source: this paper ^^

Reinforcement Learning - Video Game Playing

Plays and predicts frames… also supports human play.

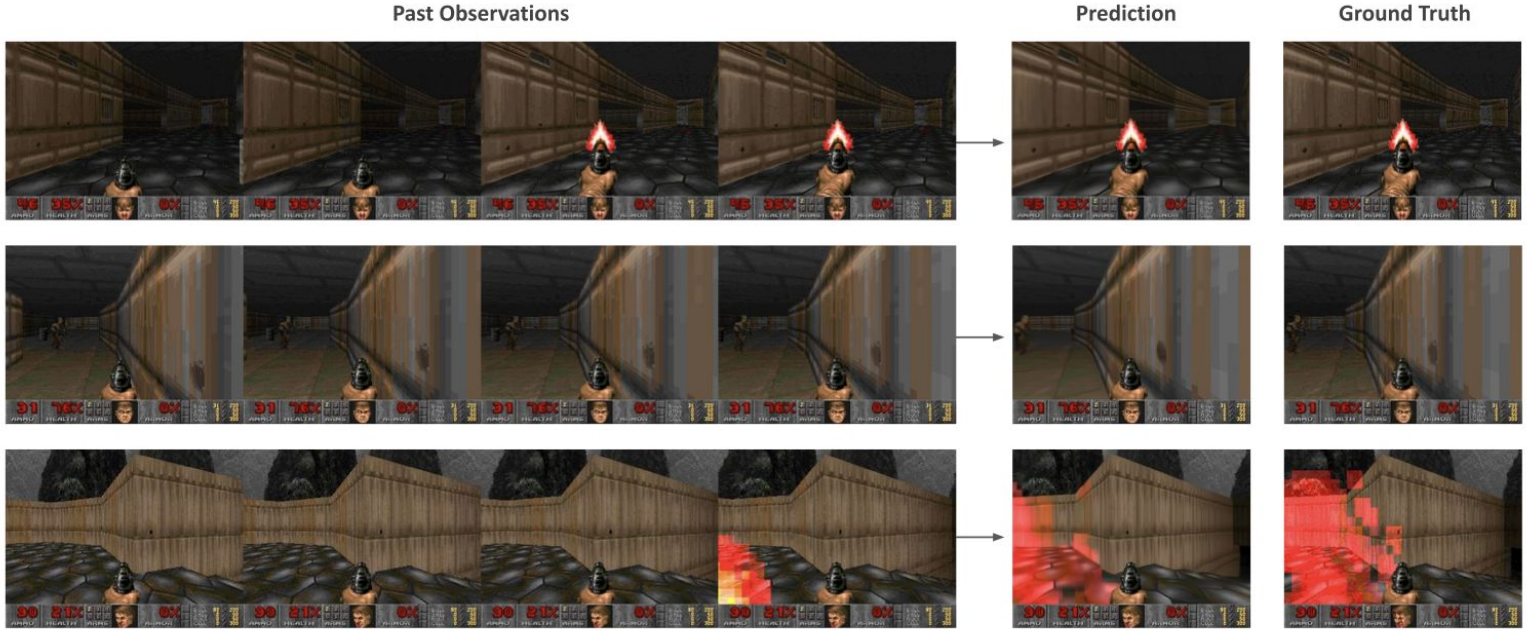


Image source: “DIFFUSION MODELS ARE REAL-TIME GAME ENGINES” by Valevski et al (2024)

Reinforcement Learning - Playing Video Games

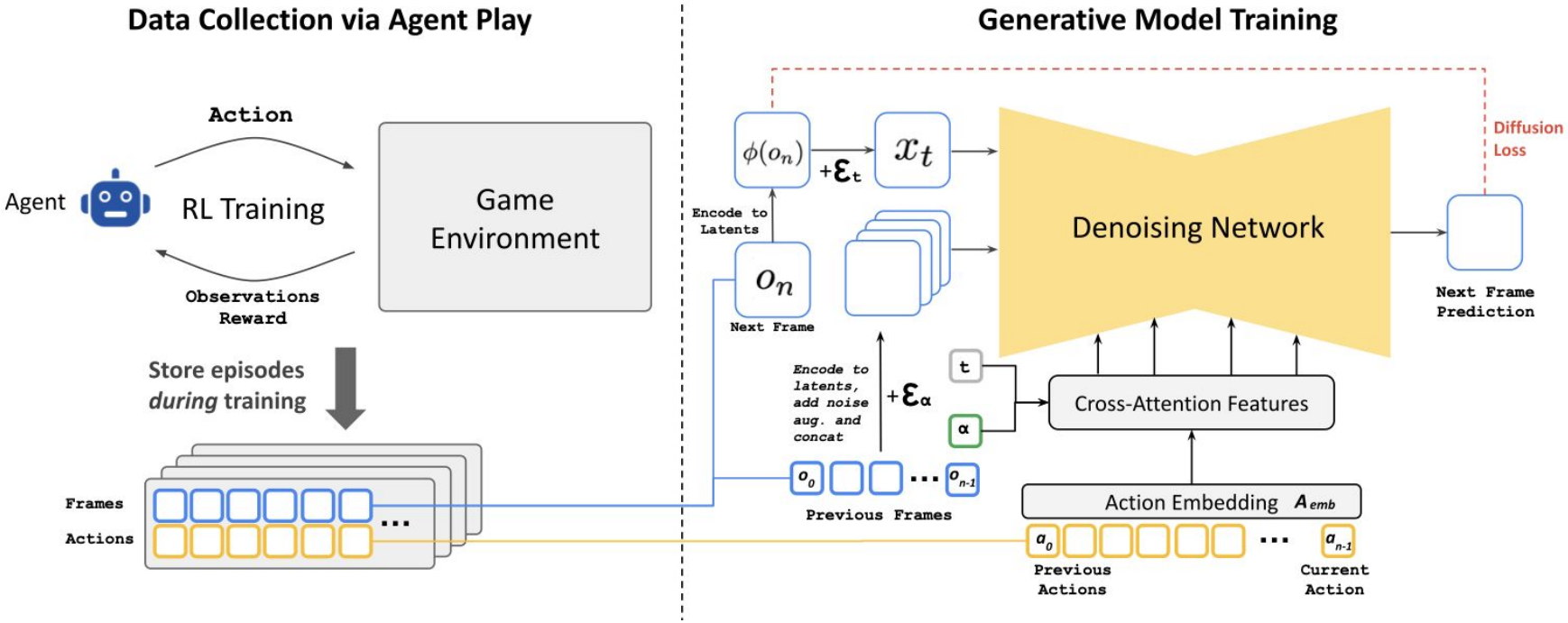


Image source: “DIFFUSION MODELS ARE REAL-TIME GAME ENGINES” by Valevski et al (2024)

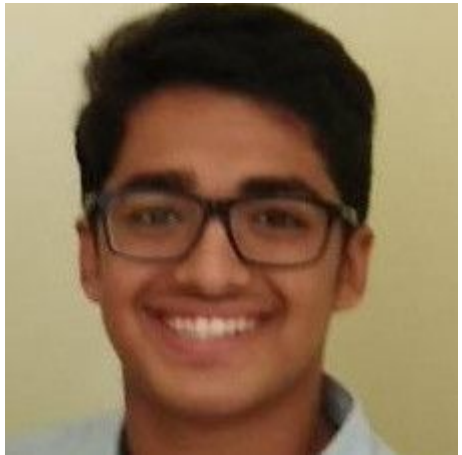
Staff

Instructor Jeffrey Considine

- Email: jconsidi@bu.edu
- Office: CDS 1625
- Office Hours:
  - Tuesdays 11:00am – 12:00pm
  - Wednesday 3:00pm – 4:00pm
  - Thursday 1:00pm – 2:00pm

Teaching Assistant Xavier Thomas

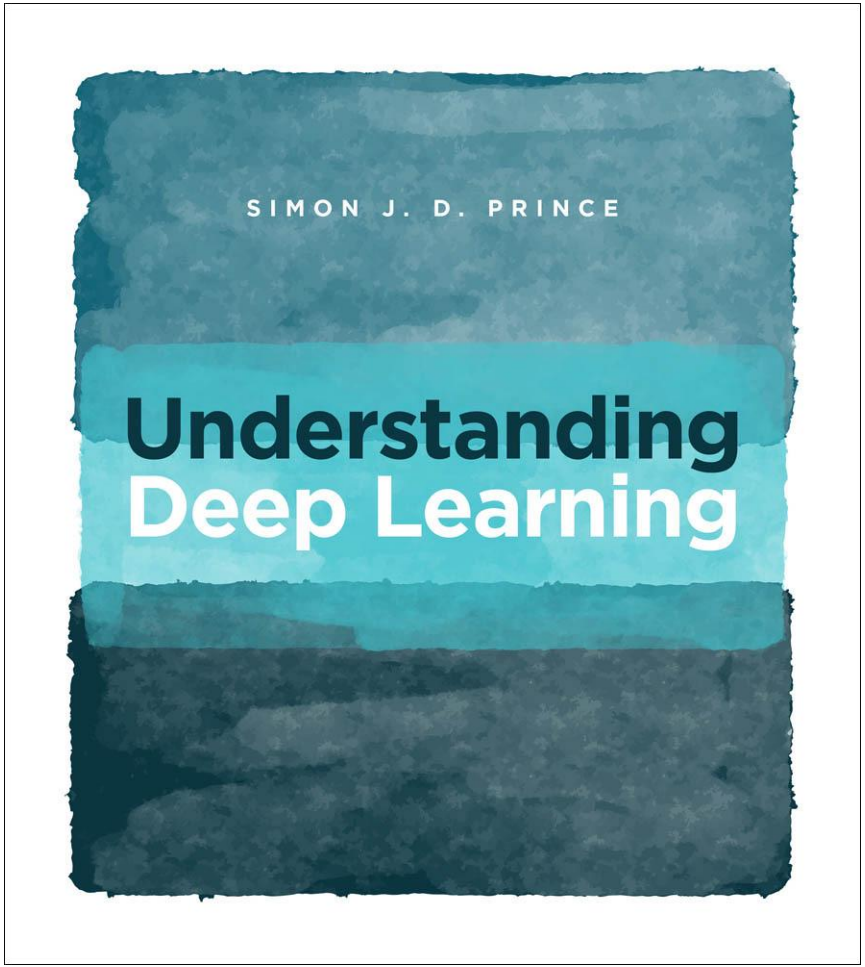
- Email: xthomas@bu.edu
- Office Hours Location: CDS, 16th Floor
- Office Hours:
  - Fridays 3:00pm – 5:00pm



Book

Understanding Deep Learning
by Simon J.D. Prince (2023)

- https://udlbook.com/
- Emphasis on understanding over implementation and history



Course Cadence

First half:

- Roughly covers supervised learning techniques
- One homework per lecture. Mix of Jupyter notebooks and problem sets.
- Midterm will be similar to a Kaggle tournament problem.

Second half:

- “Advanced techniques”. Unsupervised and reinforcement learning.
- Final project spread over ~6 weeks. Many check ins.

Course Logistics

https://dl4ds.github.io/fa2024

Linked there -

- Syllabus
- Schedule w/slides+reading+homework
- Piazza
- Gradescope
- Jupyter notebooks (Google Colab)

First homework is posted and due in a week.