Modeling & Computation

Shusen Wang

Learning from Data

Future Data

Model

Predict

“car”

Train

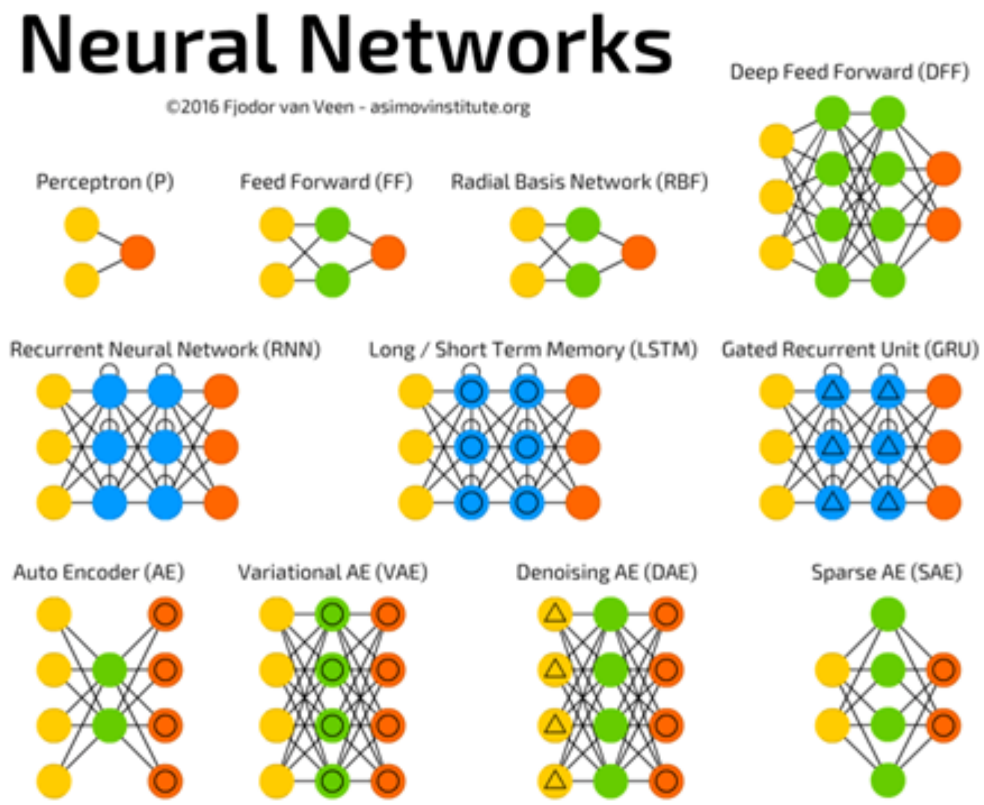
Past Data

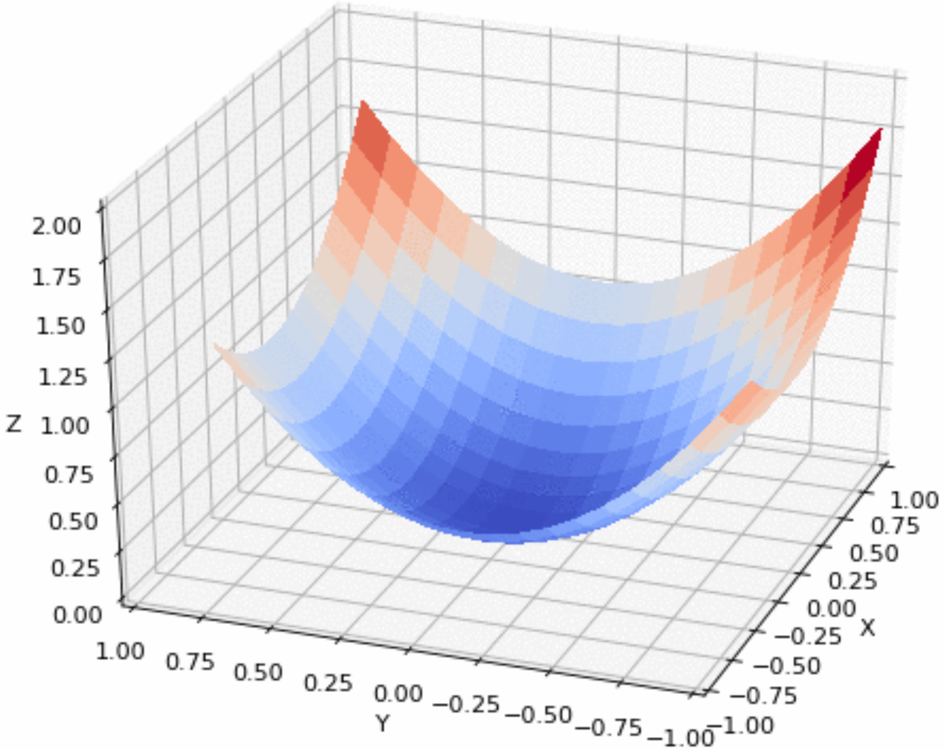
(perhaps with **labels**)

Machine Learning in Practice

Modeling

Computation





Modeling

Linear Models

- Feature vector: $\mathbf{x} \in \mathbb{R}^d$ (e.g., features of a house).
- Prediction: $f(\mathbf{x}) = \mathbf{x}^T \mathbf{w}$ (e.g., housing price).

Linear Models

- Feature vector: $\mathbf{x} \in \mathbb{R}^d$ (e.g., features of a house).
- Prediction: $\boxed{f(\mathbf{x}) = \mathbf{x}^T \mathbf{w}}$ (e.g., housing price).

- $f(\mathbf{x}) = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d$
- $w_1, w_2, \cdots, w_d$: weights
- x_1: # of bedrooms
- x_2: # of bathroom
- x_3: square feet
- …

Linear Models

- Feature vector: $\mathbf{x} \in \mathbb{R}^d$ (e.g., features of a house).
- Prediction: $f(\mathbf{x}) = \mathbf{x}^T \mathbf{w}$ (e.g., housing price).

Question: How to find $\mathbf{w}$?

Price = $0.5M

Features of a House

$\mathbf{x} \in \mathbb{R}^d$

Prediction:

$f(\mathbf{x}) = \mathbf{x}^T \mathbf{w}$

Linear Models

- Feature vector: $\mathbf{x} \in \mathbb{R}^d$ (e.g., features of a house).
- Prediction: $f(\mathbf{x}) = \mathbf{x}^T \mathbf{w}$ (e.g., housing price).

Question: How to find $\mathbf{w}$?

- Training features: $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{R}^d$.
- Training targets: $y_1, \cdots, y_n \in \mathbb{R}$.


$\bullet \bullet \bullet$


totally n houses

Linear Models

- Feature vector: $\mathbf{x} \in \mathbb{R}^d$ (e.g., features of a house).
- Prediction: $f(\mathbf{x}) = \mathbf{x}^T \mathbf{w}$ (e.g., housing price).

Question: How to find $\mathbf{w}$?

- Training features: $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{R}^d$.
- Training targets: $y_1, \cdots, y_n \in \mathbb{R}$.
- Loss function: $L(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{x}_i^T \mathbf{w} - y_i \right)^2$.

Linear Models

- Feature vector: $\mathbf{x} \in \mathbb{R}^d$ (e.g., features of a house).
- Prediction: $f(\mathbf{x}) = \mathbf{x}^T \mathbf{w}$ (e.g., housing price).

Question: How to find $\mathbf{w}$?

- Training features: $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{R}^d$.
- Training targets: $y_1, \cdots, y_n \in \mathbb{R}$.
- Loss function: $L(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{x}_i^T \mathbf{w} - y_i \right)^2$.
- Least squares regression: $\mathbf{w}^\star = \min_{\mathbf{w}} L(\mathbf{w})$.

Linear Models Are Not Expressive

Example: Given a person's photo, predict her/his age.

Photo (features)

Age = 36

Linear Models Are Not Expressive

Example: Given a person's photo, predict her/his age.

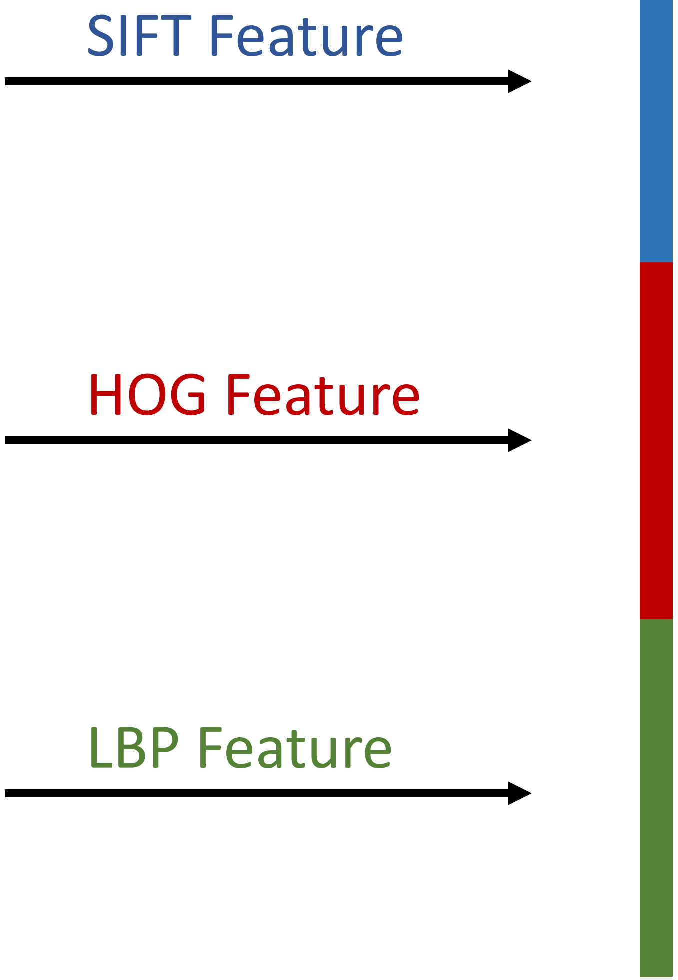
Age = 36

Photo (features)

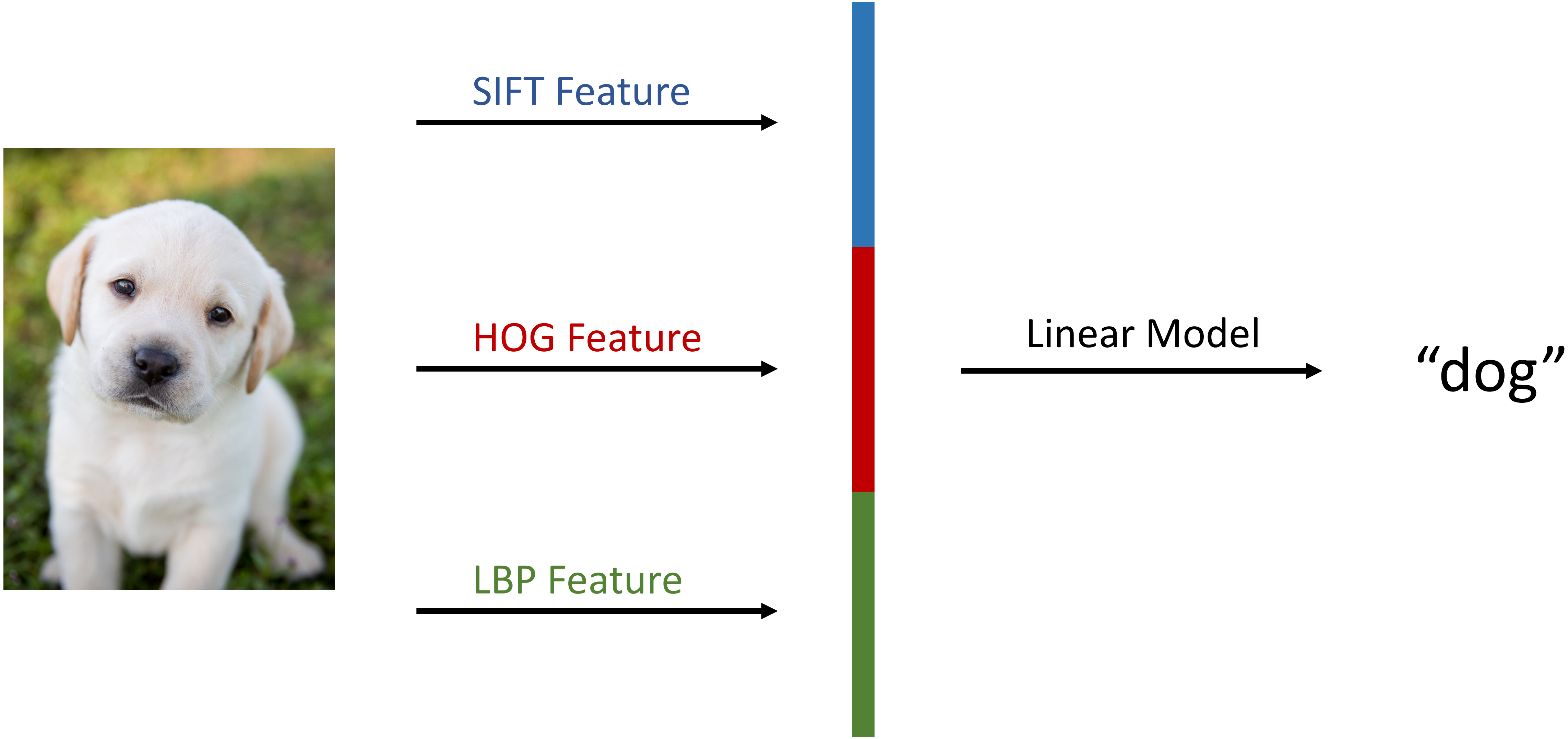
Question: Can we use linear regression?

- Linear regression assumes the target is a weighted average of every pixel in the photo.
- Linear regression works poorly for age prediction.

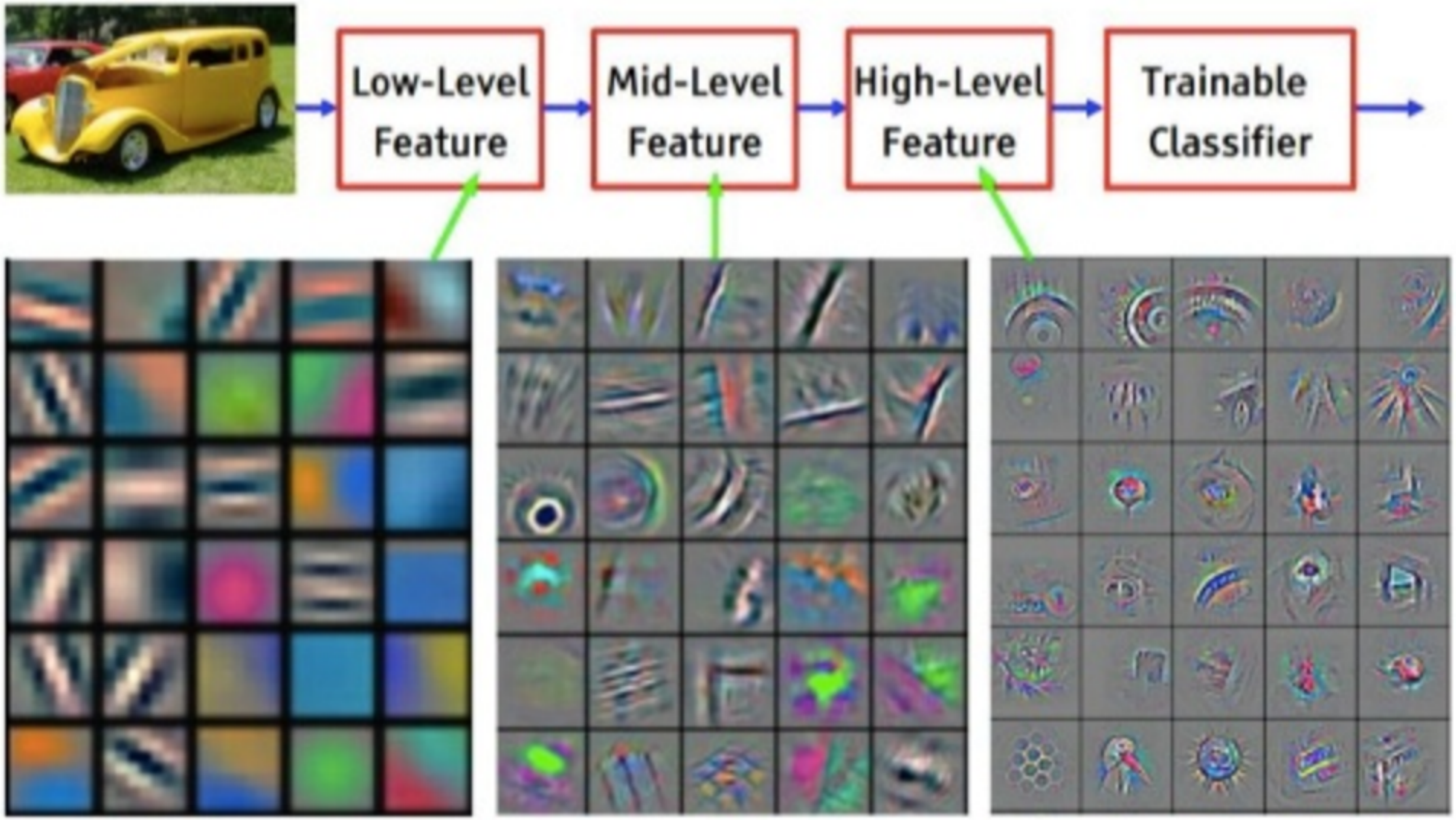
Traditional Approaches



Traditional Approaches



Convolutional Neural Networks (CNNs)

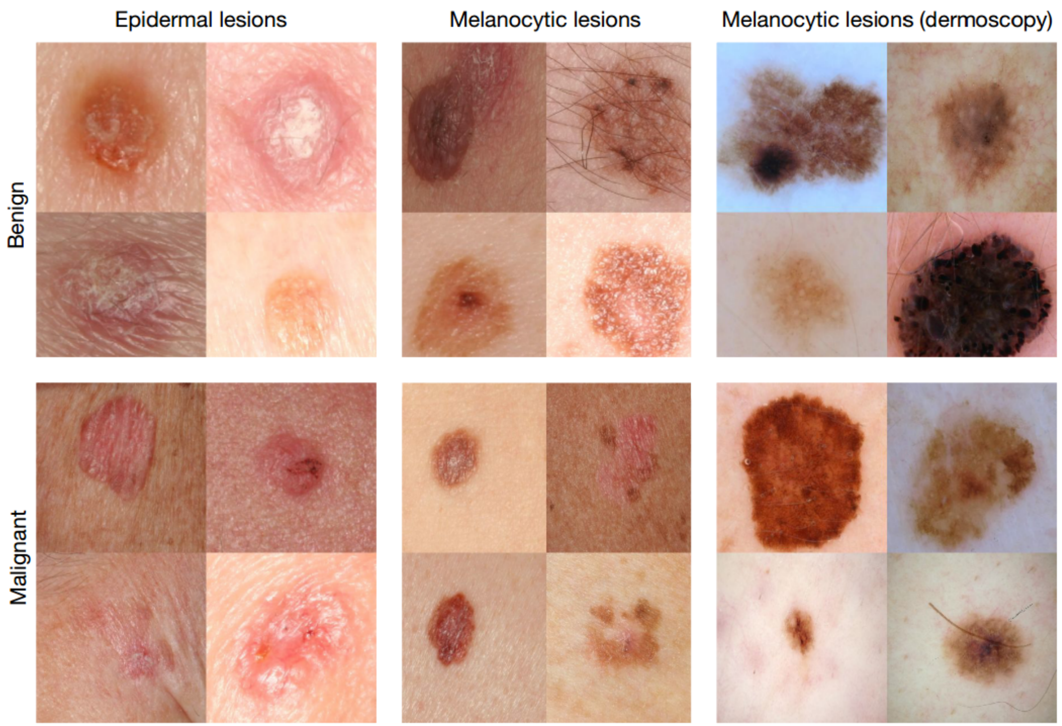


Applications of CNNs

- CNNs are suitable for image data.
- CNNs convert images to effective representations. (Feature extraction.)

- Applications:
  - Image/video recognition.
  - Face recognition.
  - Image generation.
  - …

Applications of CNNs: Medical Diagnosis



Example: Skin cancer diagnosis

- Input: an image.
- Outputs:
    - Is it skin cancer?
    - Benign or malignant?
- The same accuracy as human experts.

Reference

[1] Esteva et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017.

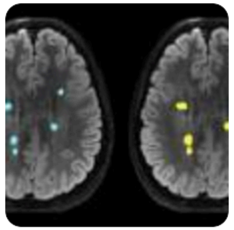
Applications of CNNs: Medical Diagnosis

Healthcare IT News

Australia's SNAC develops AI tools to improve brain scan analysis

... algorithms using the NVIDIA Clara suite of medical imaging tools, as well ... SNAC is building its deep learning algorithms using the PyTorch ...

10 hours ago



Data Center Frontier (blog)

Startups Target AI Opportunities to Disrupt Medical Imaging

AI, or more specifically machine learning and deep learning algorithms, are learning how to analyze the images produced by radiology scans.

2 weeks ago

Medical Xpress

Deep learning model detects diabetic eye diseases accurately

The deep learning model identified referable diabetic retinopathy ... Currently, retinal imaging is the most widely used method for screening and detecting retinopathy, and medical experts evaluate the severity and the

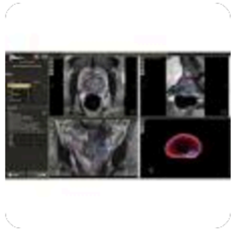
2 days ago

Imaging Technology News (press release) (blog)

Improved Imaging Technique Could Increase Chances of Prostate Cancer Survival

"Imaging has become an essential component of modern medicine," said ... Using a deep learning approach, Yan plans to simplify the process.

20 hours ago



Xtelligent Healthcare Media

Deep Learning Tool Detects Cancer in Radiology Reports

Human reviewers analyzed the imaging text reports and noted whether cancer ... The team then used these reports to train a deep learning algorithm to ... Amazon's Comprehend Medical, an advanced machine

3 weeks ago

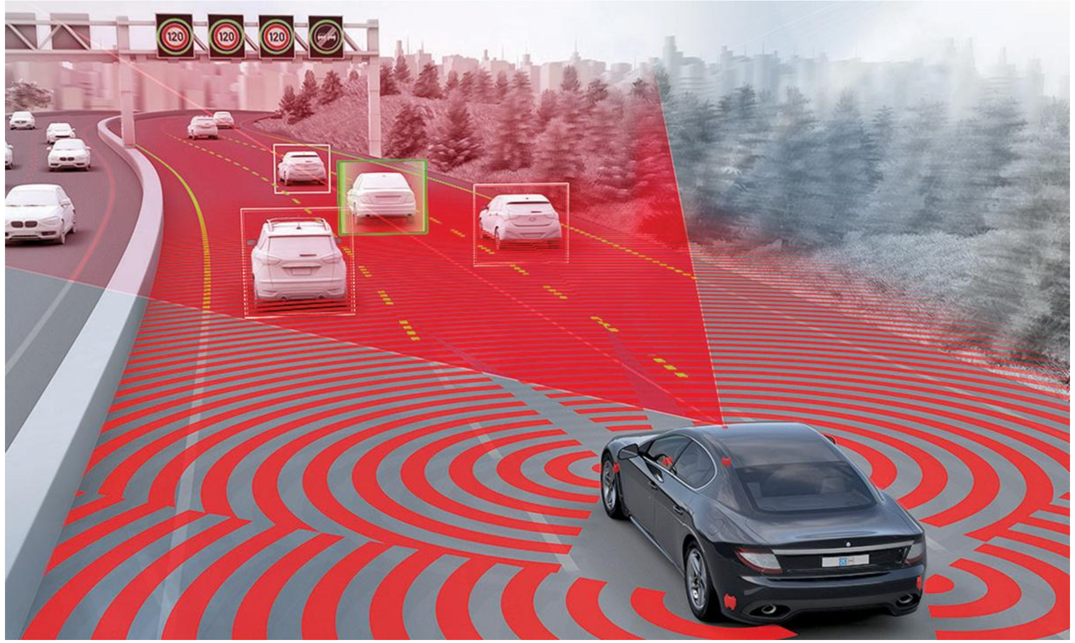
Yahoo Finance

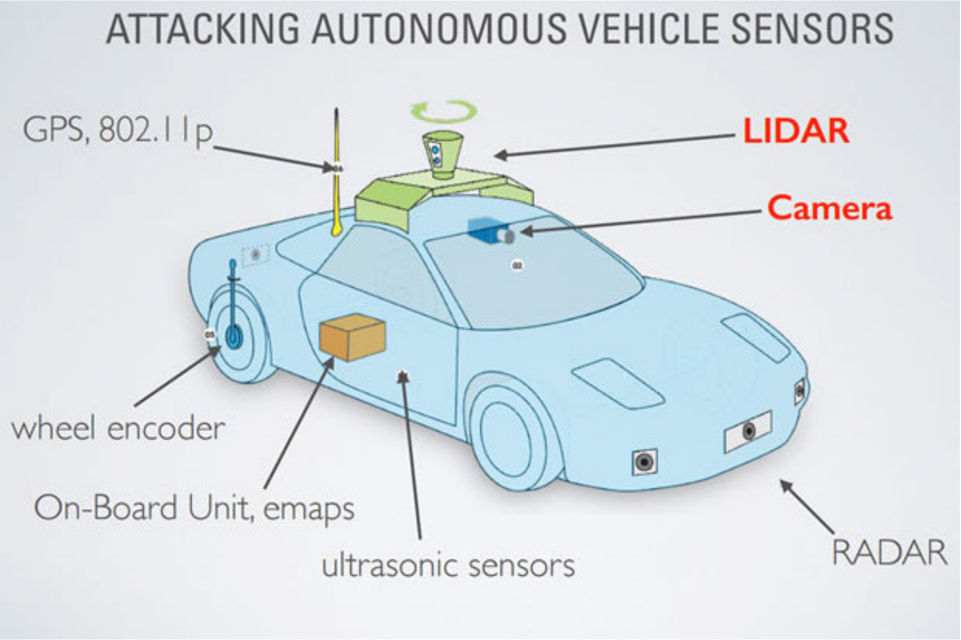
Progenics Pharmaceuticals Announces Collaboration with Veterans Affairs on the AI Research Program for Medical Image Analysis

The collaboration with VA Greater Los Angeles Healthcare System is nation's first to validate deep learning algorithms in medical imaging of ...

4 weeks ago

Applications of CNNs: Self Driving Cars

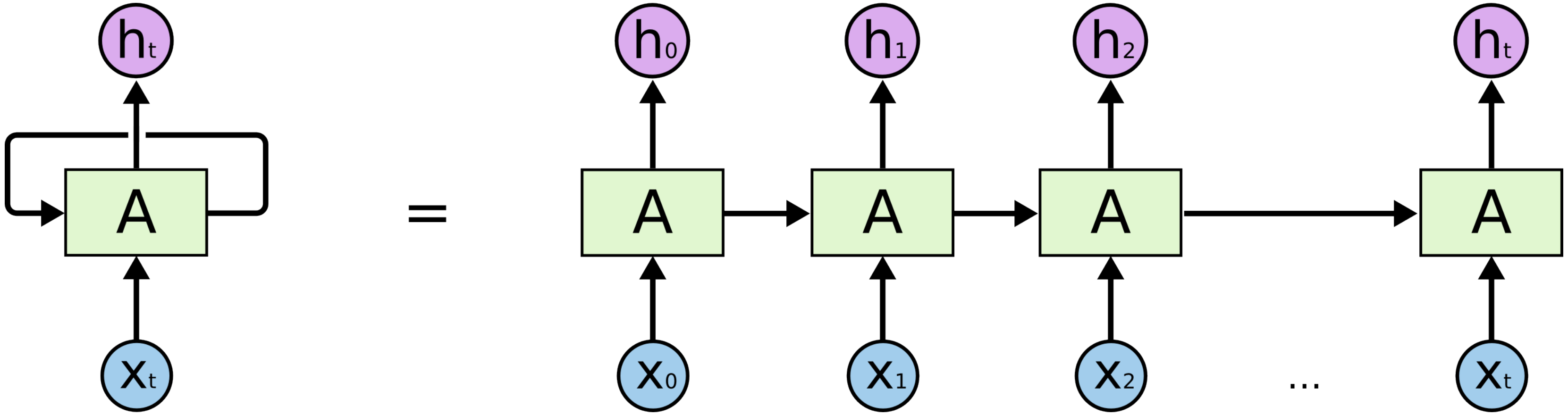




- CNNs play an import role in self driving cars.
  - Understand images taken by the cameras.
  - Recognize signs, cars, pedestrians, and obstacles.
- CNN is not everything; self-driving car is a sophisticated system.

Recurrent Neural Networks (RNNs)

- RNNs naturally fit sequence data, e.g.,
  - time series data,
  - text data,
  - speech data…



Applications of RNNs: Machine Translation

Applications of RNNs: Machine Translation

Applications of RNNs: Speech Recognition

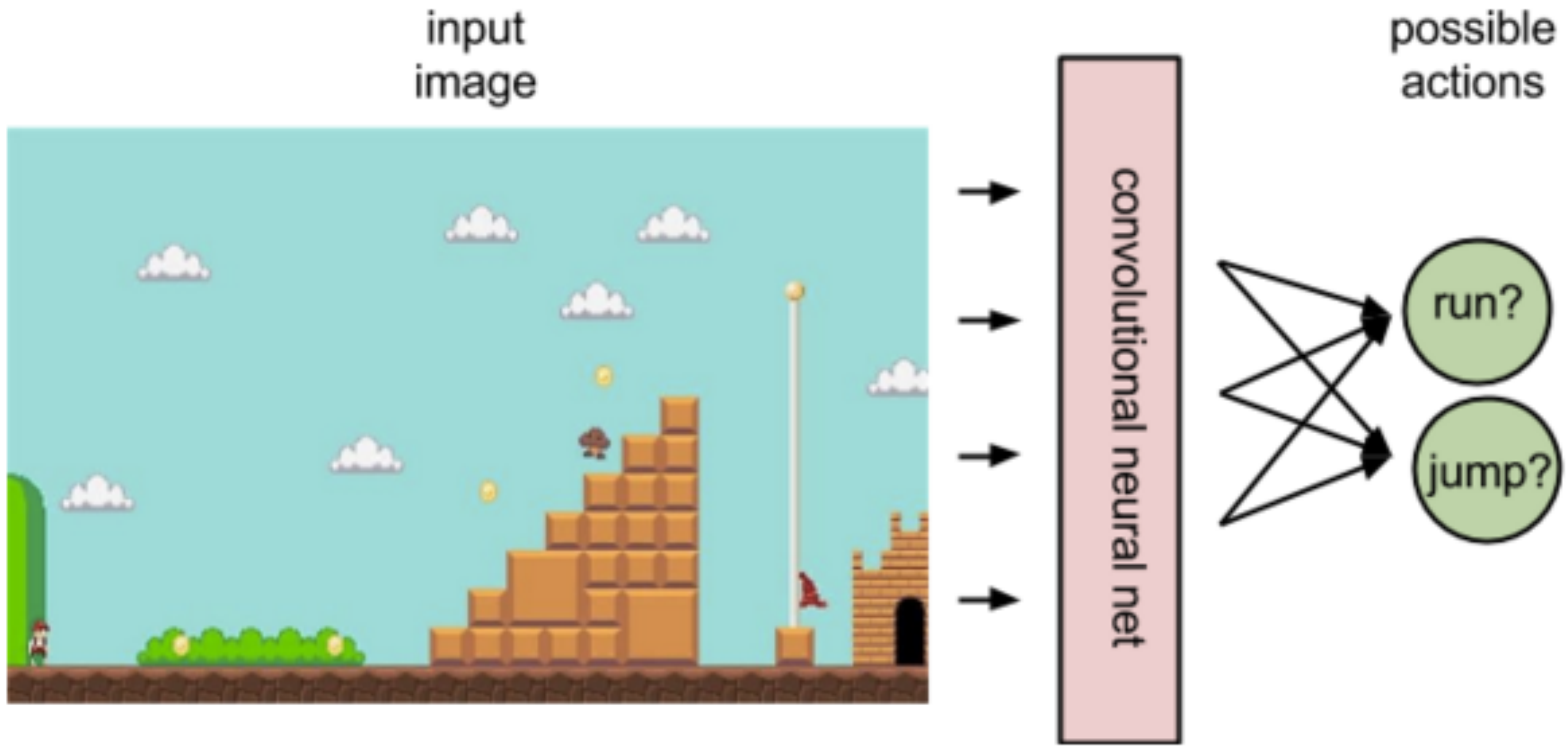
amazon alexa

Hi, I'm Cortana.

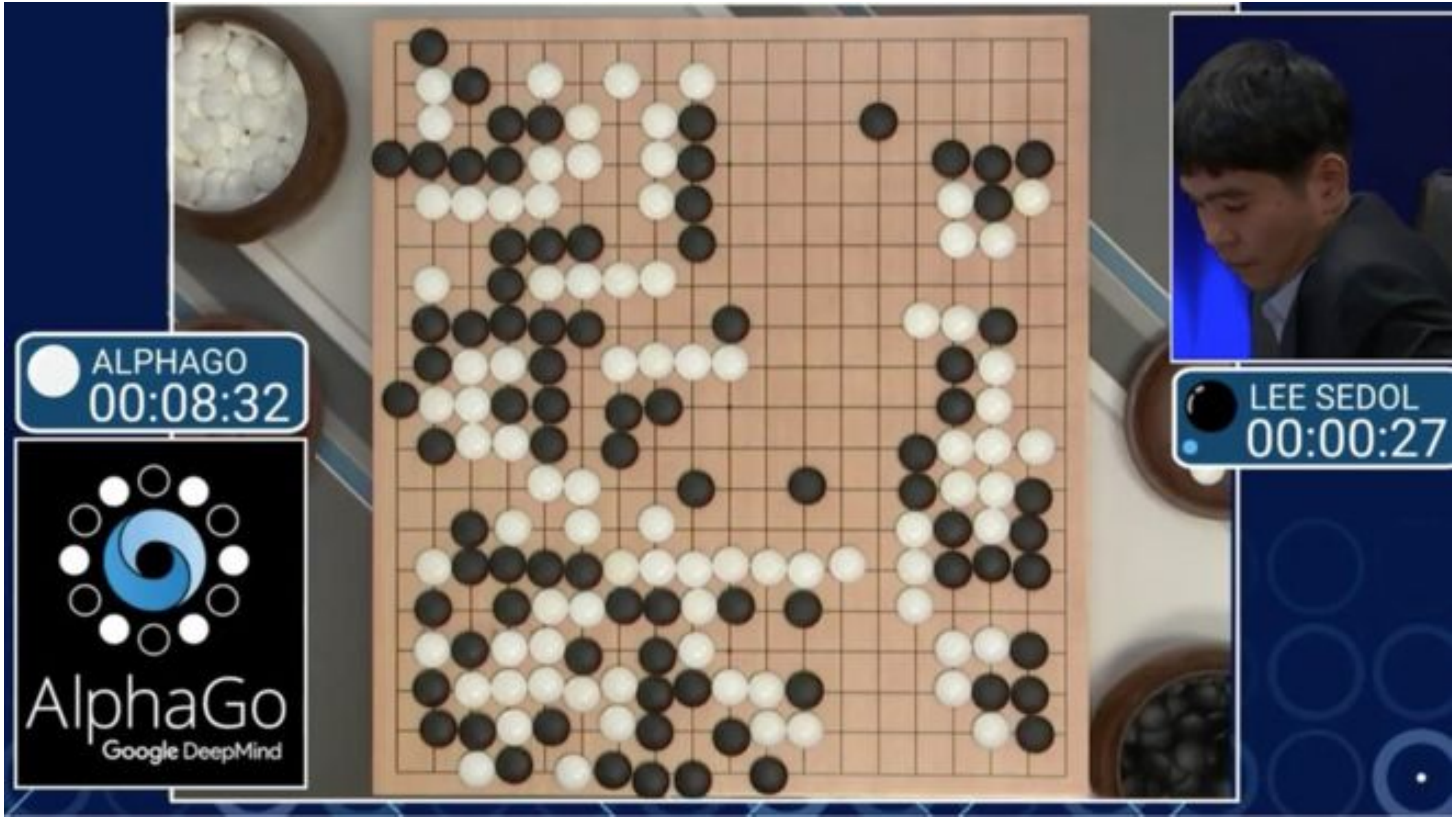
Hey Siri

Deep Reinforcement Learning (DFL)

- DFL has applications in robotics, video game, and finance.



Applications of DRL: Games



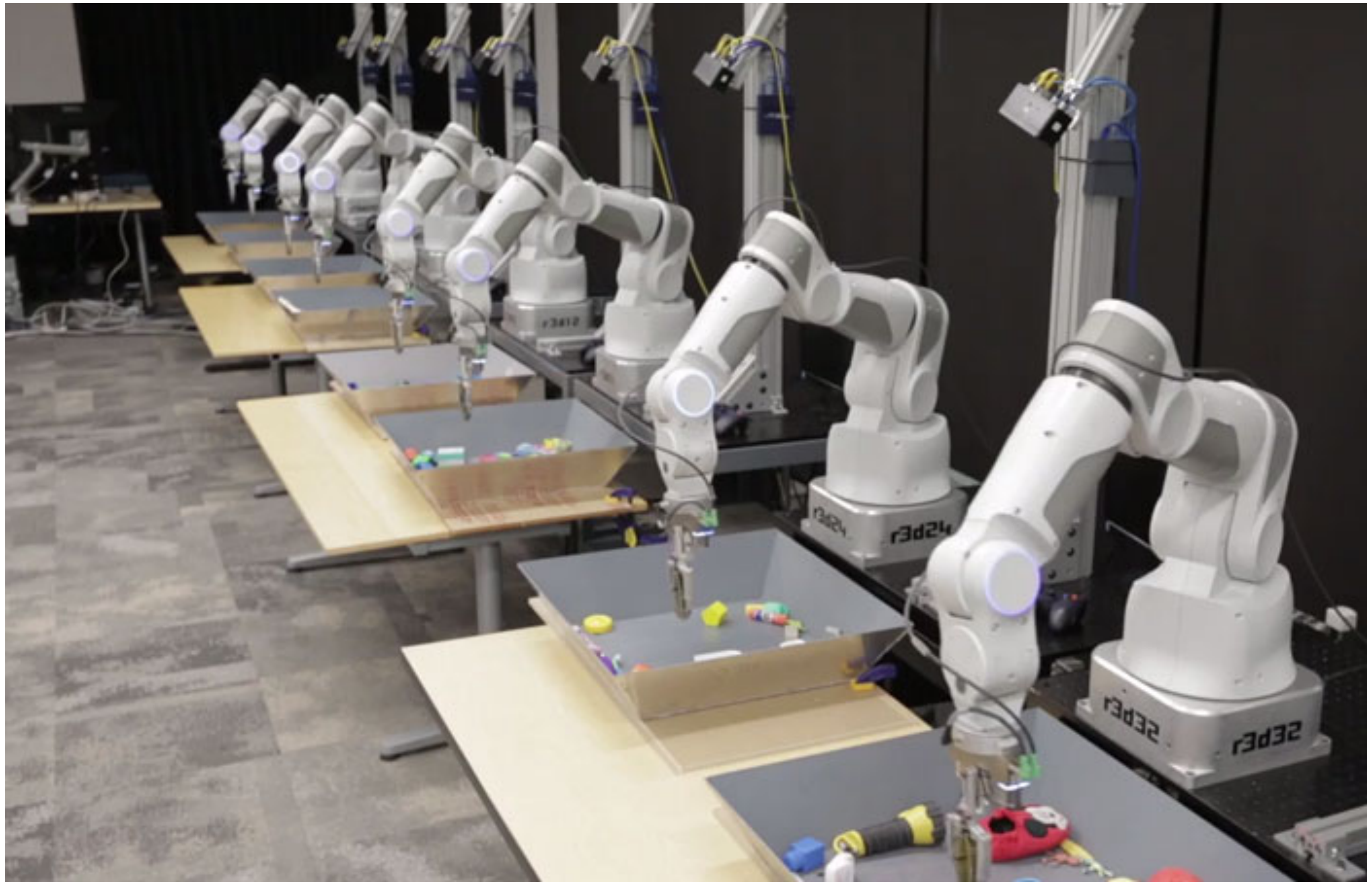
Applications of DRL: Robotics

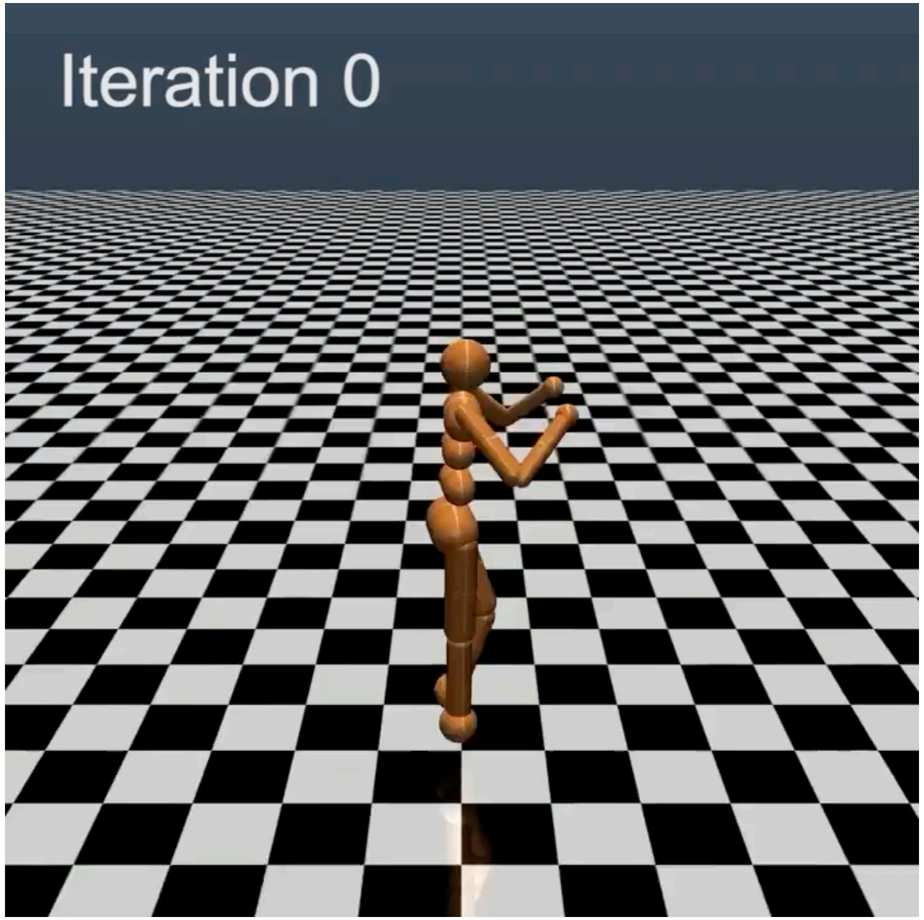
Control Theory v.s. DRL

Boston Dynamics' Atlas

Applications of DRL: Robotics

Control Theory v.s. DRL





What we have learned so far…

- **ML tasks**: regression, classification, …
- **ML models**: linear models, CNNs, RNNs, …

Example: least squares regression model

- Loss function: $L(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^T\mathbf{w} - y_i\right)^2.$
- Optimization: $\mathbf{w}^\star = \min_{\mathbf{w}} L(\mathbf{w}).$

- How to solve the model?

Computations

Computational Methods

- **ML tasks**: regression, classification, …
- **ML models**: linear models, CNNs, RNNs, …

Example: least squares regression model

- Loss function: $L(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^T\mathbf{w} - y_i\right)^2.$
- Optimization: $\mathbf{w}^\star = \min_{\mathbf{w}} L(\mathbf{w}).$

- **Computations**: solve the model using numerical algorithms, e.g., gradient descent (GD) or stochastic descent (SGD).

Gradient Descent

Example: least squares regression model

- Loss function: $L(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^T\mathbf{w} - y_i\right)^2$.
- Optimization: $\mathbf{w}^\star = \min_{\mathbf{w}} L(\mathbf{w})$.

Gradient: $\frac{\partial\, L}{\partial\, \mathbf{w}}$

- $\mathbf{w}$ is a d-dimensional vector.
- $L(\mathbf{w})$ is a scalar.
- Thus $\frac{\partial\, L}{\partial\, \mathbf{w}}$ is a d-dimensional vector.

Gradient Descent

Example: least squares regression model

- Loss function: $L(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^T\mathbf{w} - y_i\right)^2.$
- Optimization: $\mathbf{w}^\star = \min_{\mathbf{w}} L(\mathbf{w}).$

Gradient: $\frac{\partial\, L}{\partial\, \mathbf{w}}$

- $\mathbf{w}$ is a d-dimensional vector.
- $L(\mathbf{w})$ is a scalar.
- Thus $\frac{\partial\, L}{\partial\, \mathbf{w}}$ is a d-dimensional vector.

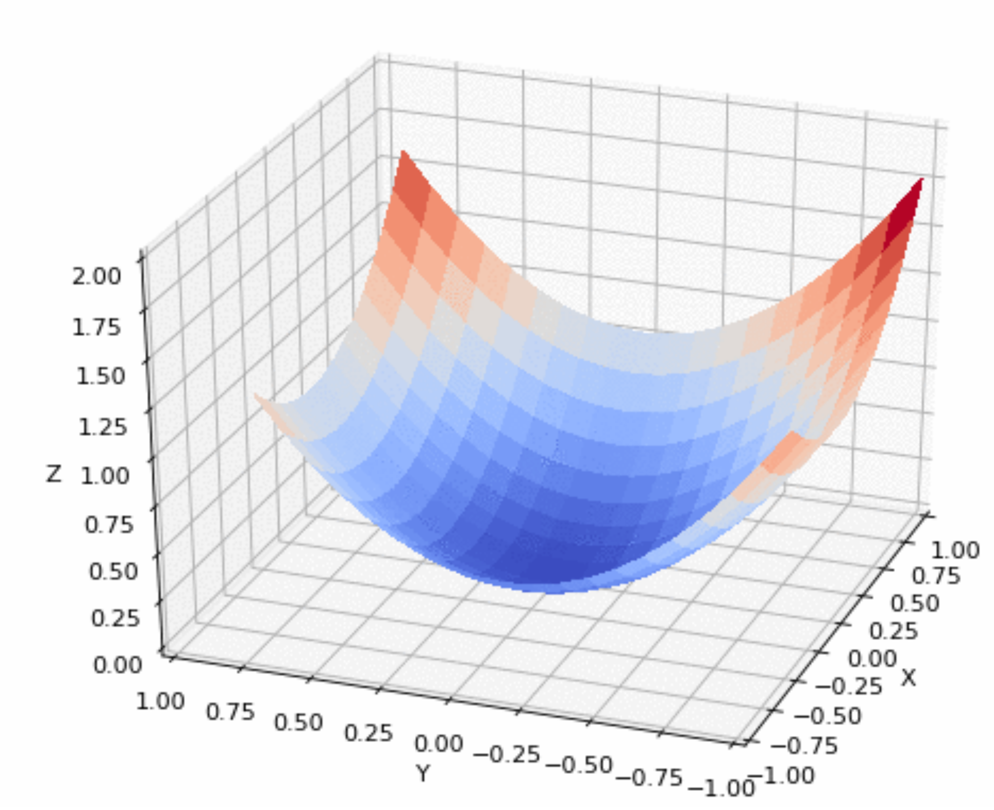
Gradient descent algorithm

- Randomly initialize $\mathbf{w}_0$.
- For $t = 0$ to T:
  - Gradient at $\mathbf{w}_t$: $\mathbf{g}_t = \frac{\partial\, L}{\partial\, \mathbf{w}}\Big|_{\mathbf{w}_t}$;
  - $\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha\, \mathbf{g}_t.$

Gradient Descent

Example: least squares regression model

- Loss function: $L(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^T\mathbf{w} - y_i\right)^2$.
- Optimization: $\mathbf{w}^\star = \min_{\mathbf{w}} L(\mathbf{w})$.



Gradient descent algorithm

- Randomly initialize $\mathbf{w}_0$.
- For $t = 0$ to T:
  - Gradient at $\mathbf{w}_t$: $\mathbf{g}_t = \frac{\partial L}{\partial \mathbf{w}}\Big|_{\mathbf{w}_t}$;
  - $\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha\, \mathbf{g}_t$.

Gradient Descent

Example: least squares regression model

- Loss function: $L(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^T\mathbf{w} - y_i\right)^2.$
- Optimization: $\mathbf{w}^{\star} = \min_{\mathbf{w}} L(\mathbf{w}).$

Variants of Gradient Descent

- Stochastic gradient descent (SGD).
- SGD with momentum.
- RMSProp.
- ADAM…

Computational Challenges

- Big data: too many training samples.
  - ImageNet: 14 million 256×256 images.
- Big model: too many model parameters.
  - ResNet-50 (a very popular CNN architecture) has 25 million parameters.

Computational Challenges

- Big data: too many training samples.
  - ImageNet: 14 million 256×256 images.
- Big model: too many model parameters.
  - ResNet-50 (a very popular CNN architecture) has 25 million parameters.

- Big data + big model bring computational challenges.
- Training ResNet-50 on ImageNet using a single GPU takes around 14 days.

Computational Challenges

- Big data: too many training samples.
    - ImageNet: 14 million 256×256 images.
- Big model: too many model parameters.
    - ResNet-50 (a very popular CNN architecture) has 25 million parameters.

- Big data + big model bring computational challenges.
- Training ResNet-50 on ImageNet using a single GPU takes around 14 days.

→ Efficient algorithms and software systems are necessary.

Machine Learning in Practice

Modeling

Computation

Machine Learning in Practice

Modeling

Computation

- The model that fits the data and problem.
- Decide the network structures, activation functions, loss functions, etc.
- Improve the prediction accuracy.
  - Experience in ML models.
  - Understanding of the problem and data.

Machine Learning in Practice

Modeling

Computation

- Design or apply efficient algorithms.
- Implement the algorithm using systems like TensorFlow, Hadoop, etc.

Machine Learning in Practice

Modeling

Computation

- Design or apply efficient algorithms.
- Implement the algorithm using systems like TensorFlow, Hadoop, etc.
- Optimize your code.
  - Experience in algorithms.
  - Experience in systems.