Deep Learning Limitations and New Frontiers

Ava Amini

MIT Introduction to Deep Learning

January 12, 2023

T-shirts! Today!

MIT 6.S191

Program Schedule

Intro to Deep Learning

Lecture 1
Jan. 9, 2023
[Slides] [Video] *coming soon!*

Deep Sequence Modeling

Lecture 2
Jan. 9, 2023
[Slides] [Video] *coming soon!*

Intro to TensorFlow; Music Generation

Software Lab 1
[Code]

Deep Computer Vision

Lecture 3
Jan. 10, 2023
[Slides] [Video] *coming soon!*

Deep Generative Modeling

Lecture 4
Jan. 10, 2023
[Slides] [Video] *coming soon!*

Facial Detection Systems

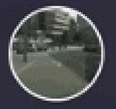
Software Lab 2
[Paper] [Code]

Uncertainty and Bias

Lecture 5
Jan. 11, 2023
[Info] [Slides] [Video] *coming soon!*

Deep Reinforcement Learning

Lecture 6
Jan. 11, 2023
[Slides] [Video] *coming soon!*



Debiasing, Unceratinty, and Robustness

Software Lab 3
[Code]

Limitations and New Frontiers

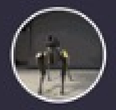
Lecture 7
Jan. 12, 2023
[Slides] [Video] *coming soon!*

The Modern Era of Statistics

Lecture 8
Jan. 12, 2023
[Info][Slides] [Video] *coming soon!*

Final Project

Work on final projects



Robot Learning

Lecture 9
Jan. 13, 2023
[Info] [Slides] [Video] *coming soon!*

Text-to-Image Generation

Lecture 10
Jan. 13, 2023
[Info] [Slides] [Video] *coming soon!*

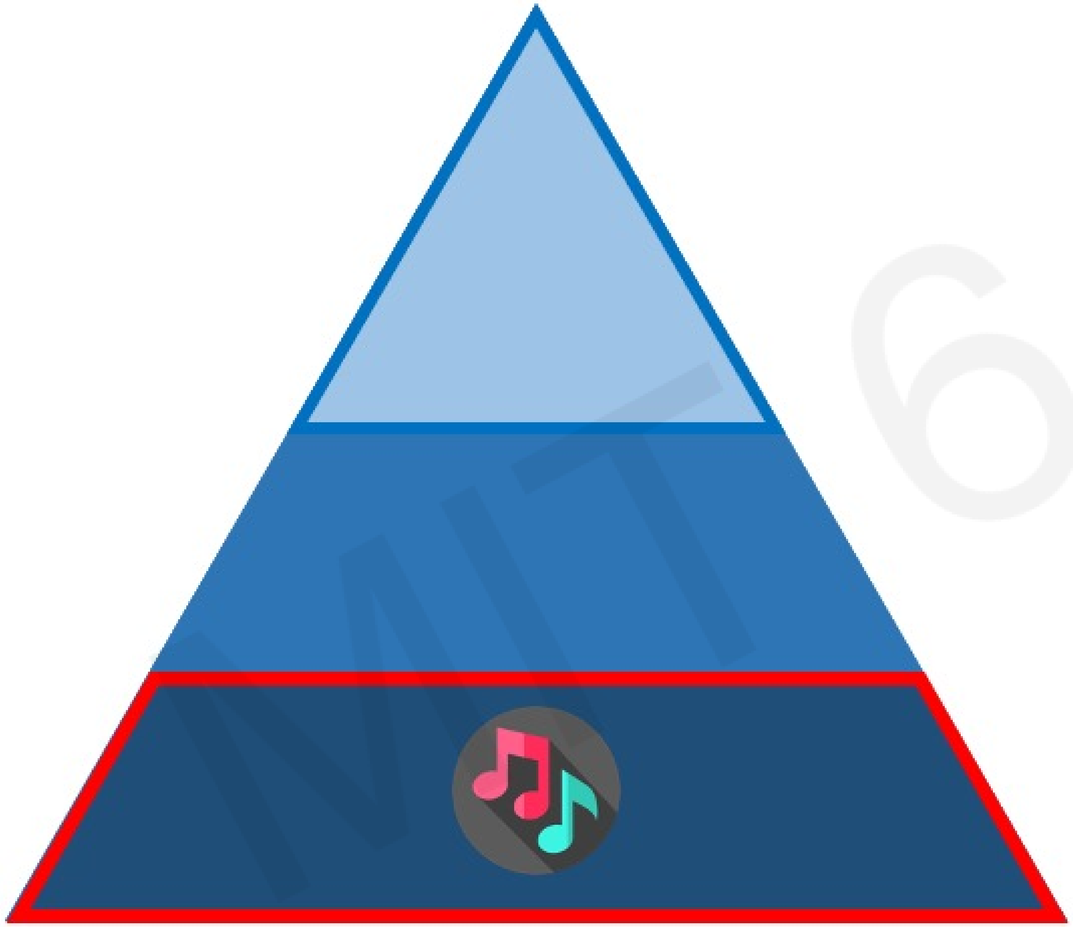
Project Competition

Project pitches and final awards!

- Lab competition: 1/13/23
- Proposal slides: 1/12/23
- Proposal pitch: 1/13/23

Labs and Prizes

Due Friday 1/13 at 12:00pm ET. Instructions: bit.ly/deeplearning-syllabus



Music Generation

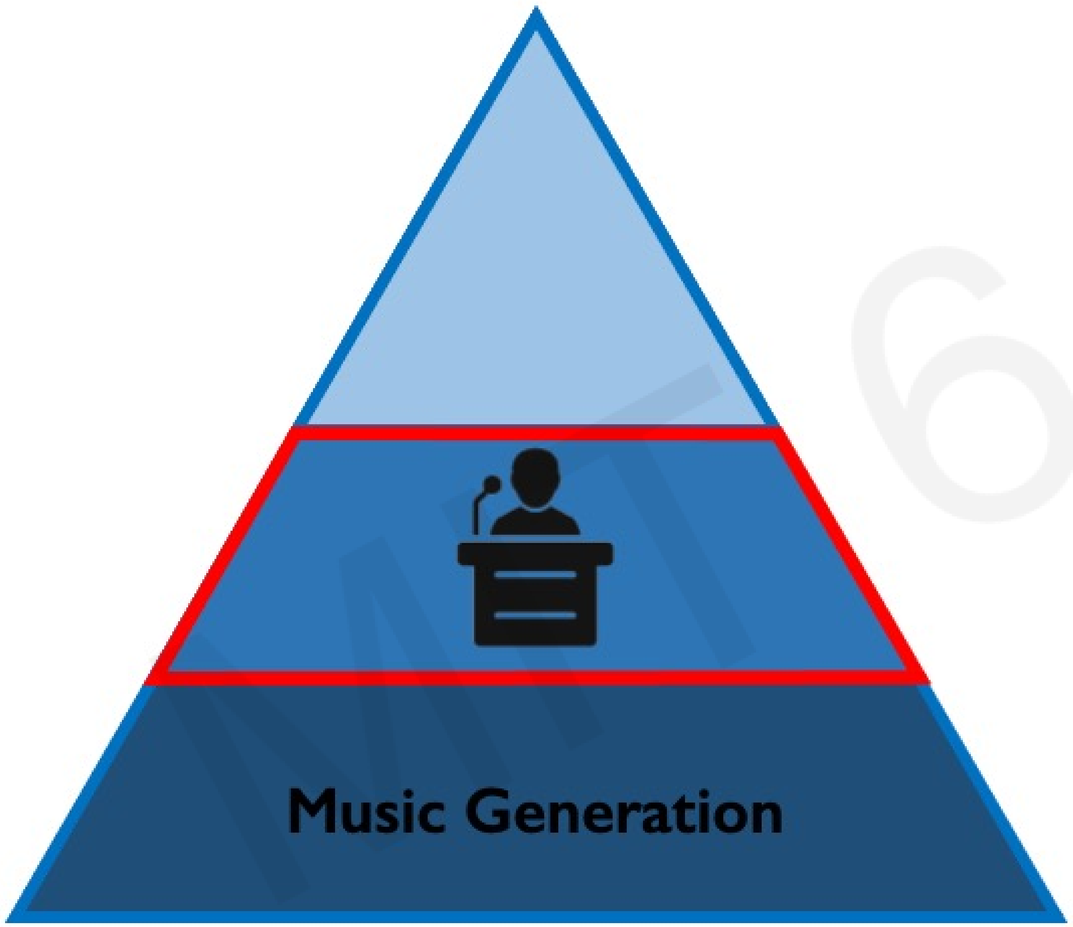
Build a neural network that can learn the genre of Irish folk songs and use it to generate brand new songs!

Prize:

MIT 6.S191

Labs and Prizes

Instructions: bit.ly/deeplearning-syllabus



Project Pitch Competition

Present a novel deep learning research idea or application (3 minutes, strict)

Presentations on **Friday, Jan 13**

Submit groups by **Wed 1/11 11pm ET**

Submit slides by **Thu 1/12 11pm ET**

Instructions: bit.ly/deeplearning-syllabus

Prizes:

Gold:
NVIDIA 3070 GPU

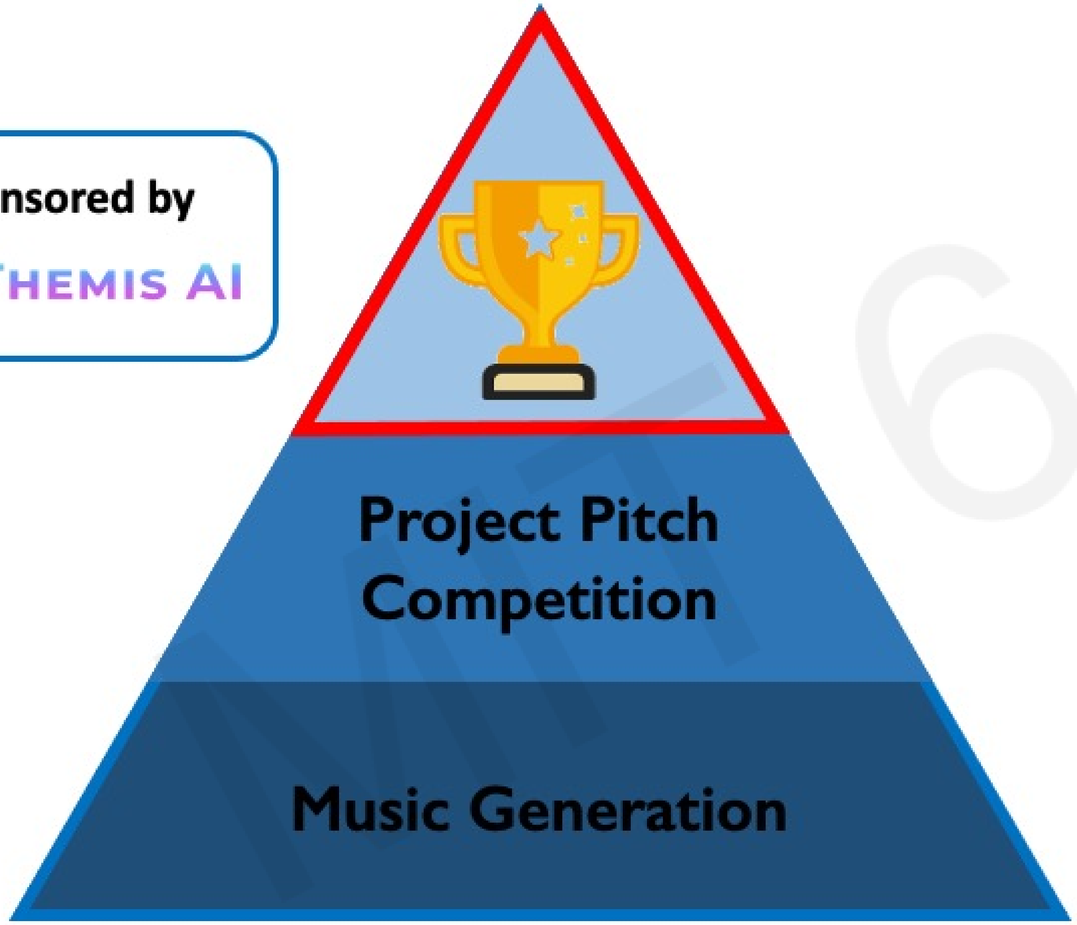
Silver:
Smartwatch

Bronze:
HD Monitor

Labs and Prizes

Due Friday 1/13 at 12:00pm ET. Instructions: bit.ly/deeplearning-syllabus

Sponsored by THEMIS AI



Trustworthy Deep Learning

Build solutions to improve robustness, mitigate bias, and increase accuracy of state-of-the-art vision systems!

(Software labs 2 and 3)

Prizes:

Gold
$1000

Silver
$750

Bronze
$500

Program Guest Lectures

Sadhana Lolla
ThemisAI

Ramin Hasani
Vanguard

Dilip Krishnan
Google

Daniela Rus
MIT CSAIL

So far in Introduction to Deep Learning…

MIT 6.S191

The Rise of Deep Learning

So far in Introduction to Deep Learning…

Data

- Signals
- Images
- Sensors

…

Decision

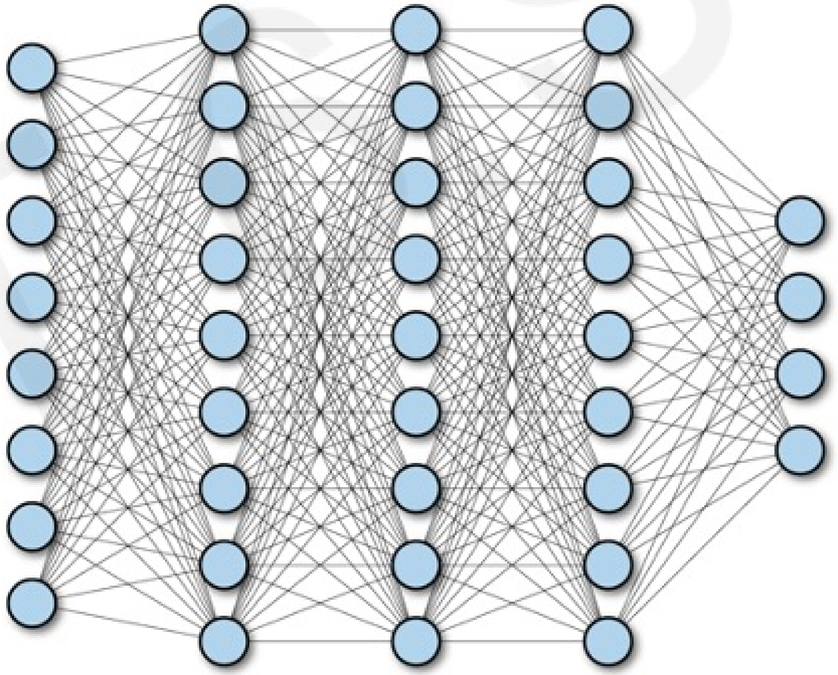
- Prediction
- Detection
- Action

…

Power of Neural Nets

Universal Approximation Theorem

A feedforward network with a single layer is sufficient to approximate, to an arbitrary precision, any continuous function.

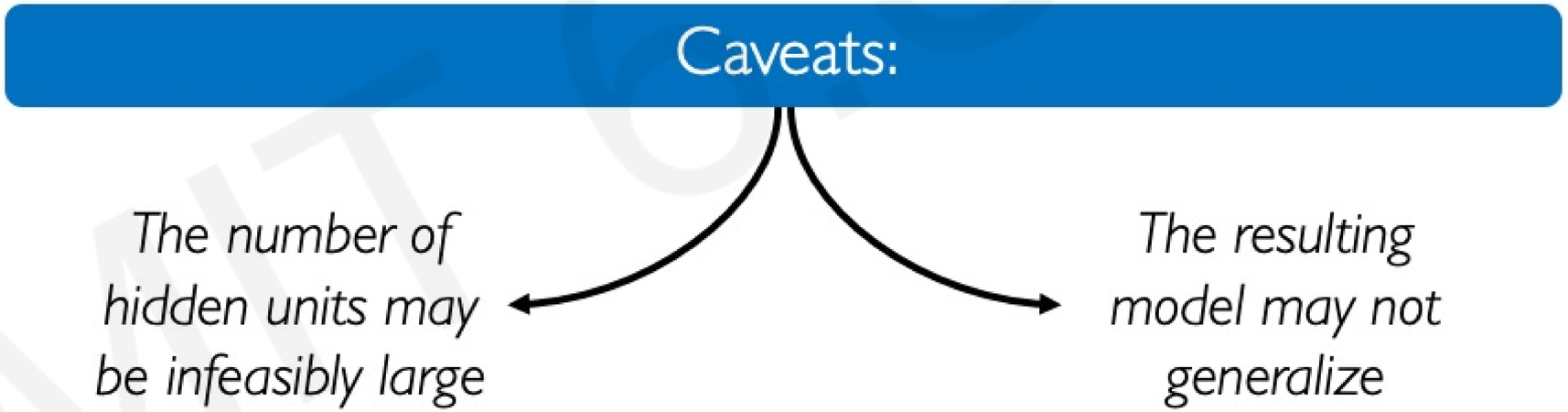


MIT 6S191

Power of Neural Nets

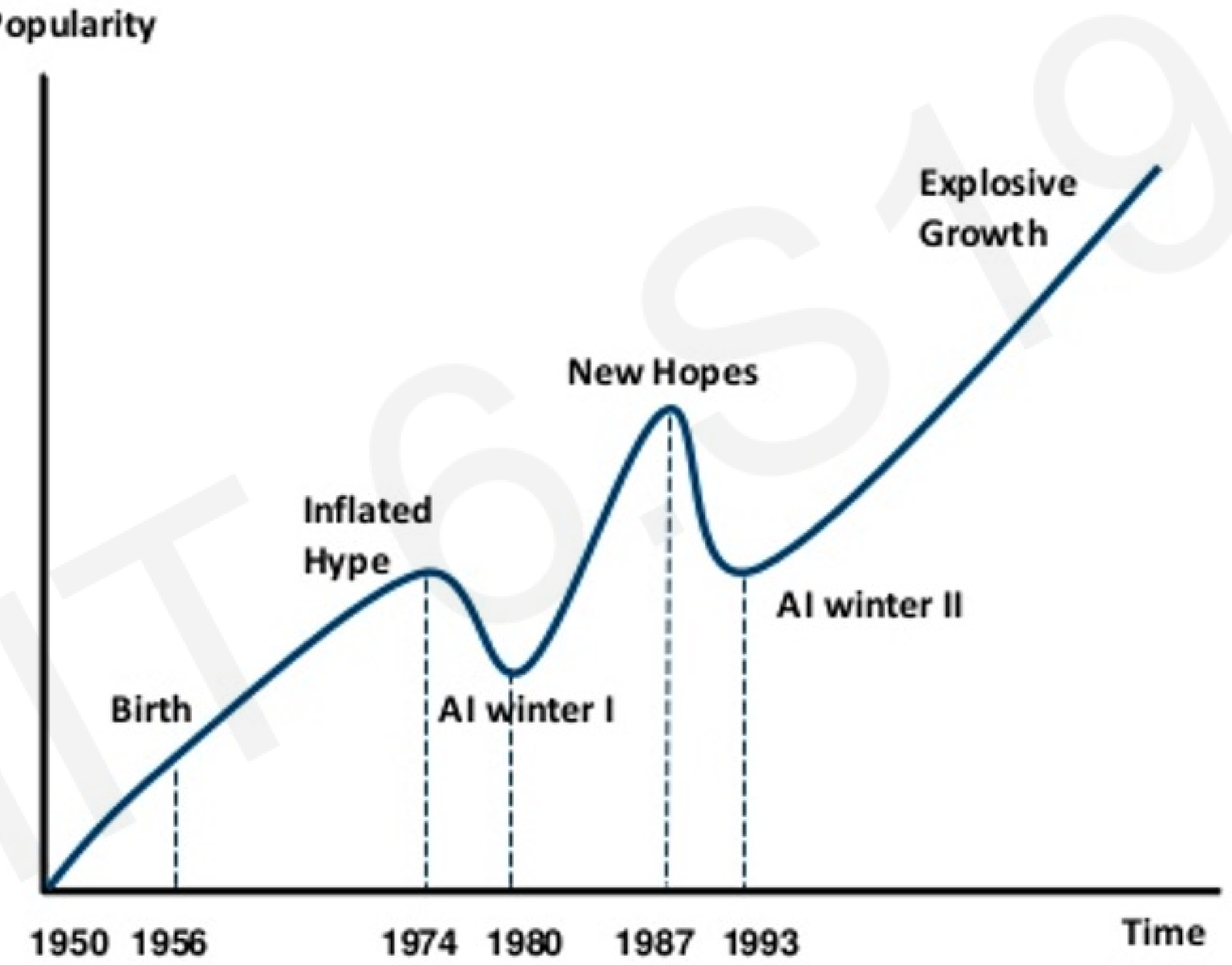
Universal Approximation Theorem

A feedforward network with a single layer is sufficient to approximate, to an arbitrary precision, any continuous function.



MIT 6S191

Artificial Intelligence "Hype": Historical Perspective



MIT 6S191

Limitations

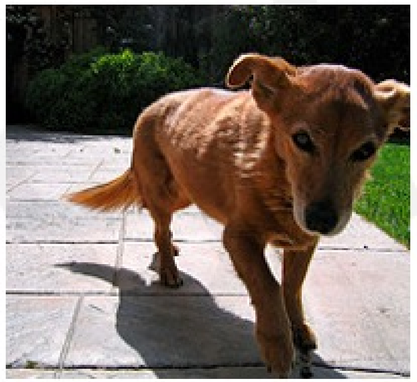
MIT 6.S191

Rethinking Generalization

"Understanding Deep Neural Networks Requires Rethinking Generalization"

dog

banana



dog

tree

MIT 6.S191

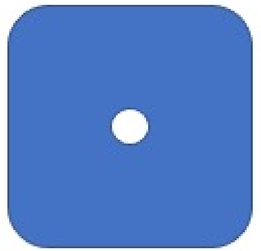
Rethinking Generalization

"Understanding Deep Neural Networks Requires Rethinking Generalization"

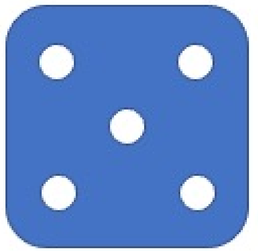
dog

banana

dog



tree



Rethinking Generalization

"Understanding Deep Neural Networks Requires Rethinking Generalization"

dog

banana

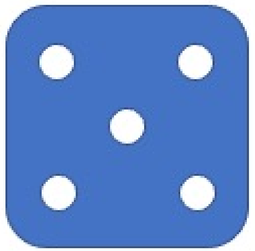
banana

dog

dog

tree

tree



dog

MIT.edu/6.S191

Rethinking Generalization

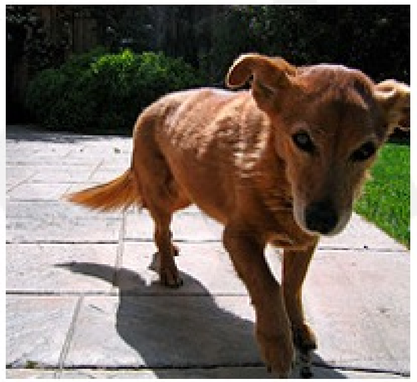
"Understanding Deep Neural Networks Requires Rethinking Generalization"

~~dog~~

banana

~~banana~~

dog



~~dog~~

tree

~~tree~~

dog

MIT.edu

Capacity of Deep Neural Networks

MIT 6.S191

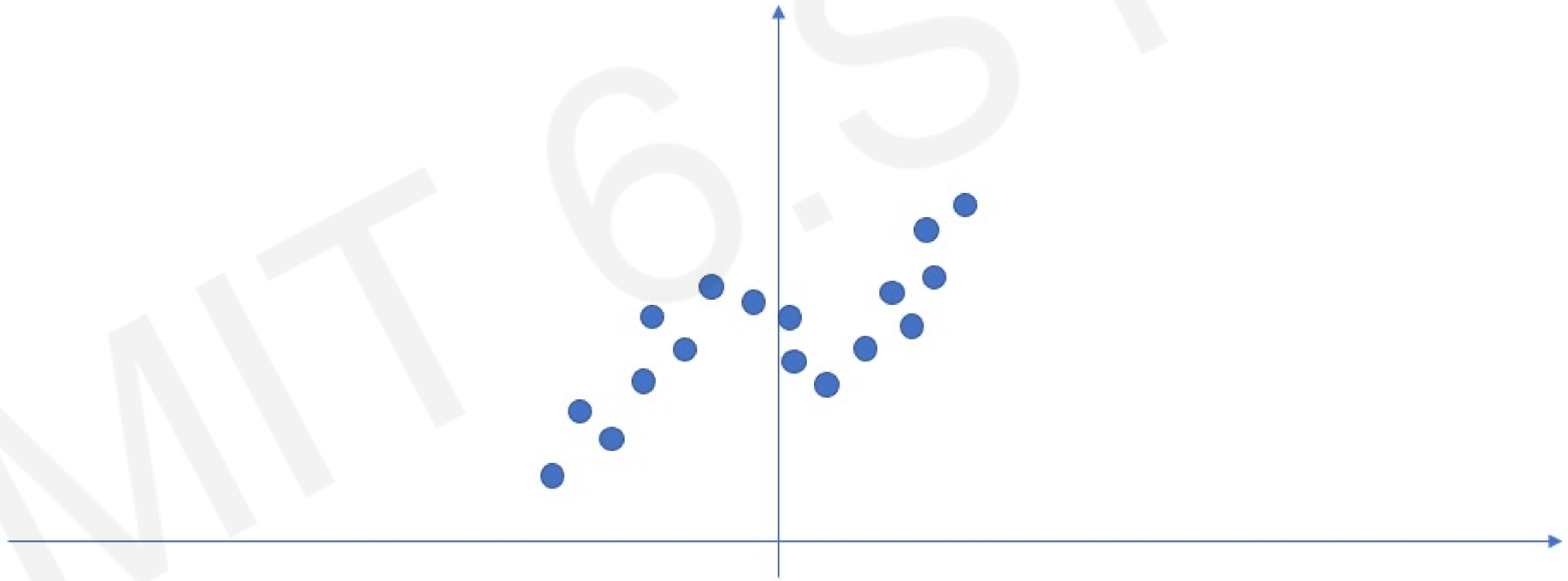
Capacity of Deep Neural Networks

MIT 6.S191

Capacity of Deep Neural Networks

Neural Networks as Function Approximators

Neural networks are **excellent** function approximators



MIT 6.S191

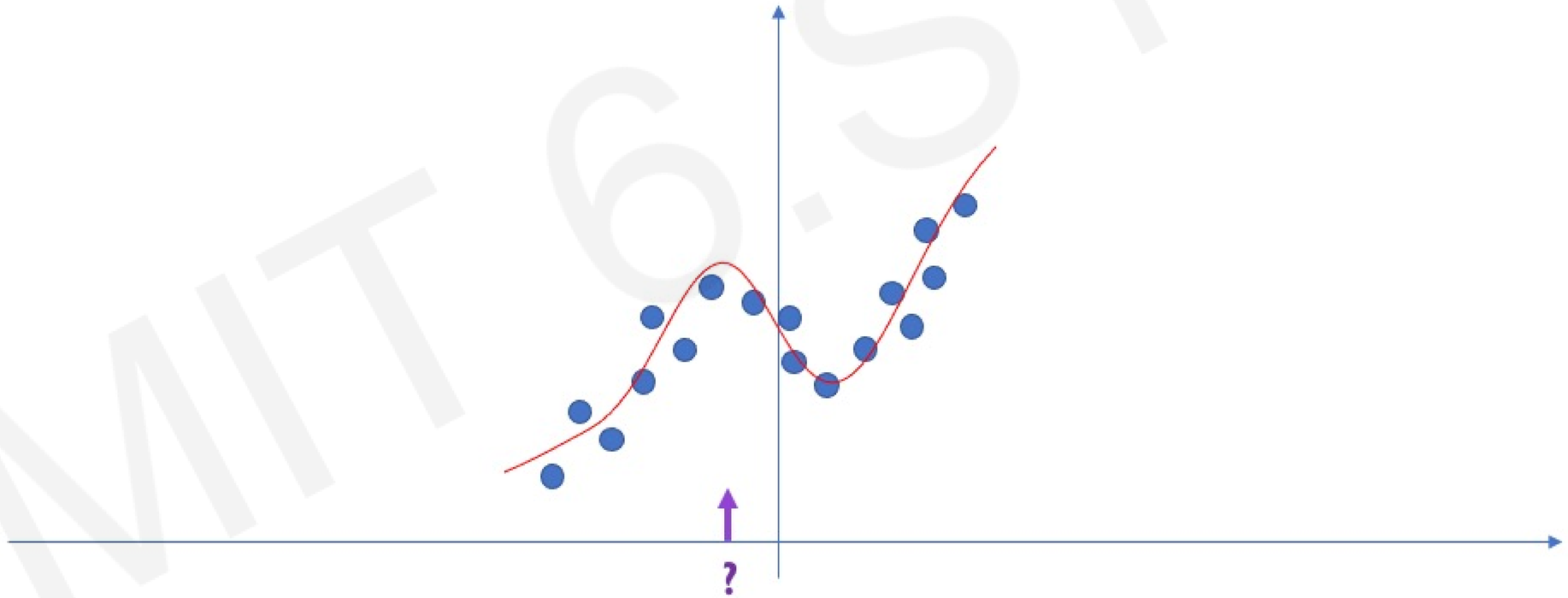
Neural Networks as Function Approximators

Neural networks are **excellent** function approximators

MIT 6.S191

Neural Networks as Function Approximators

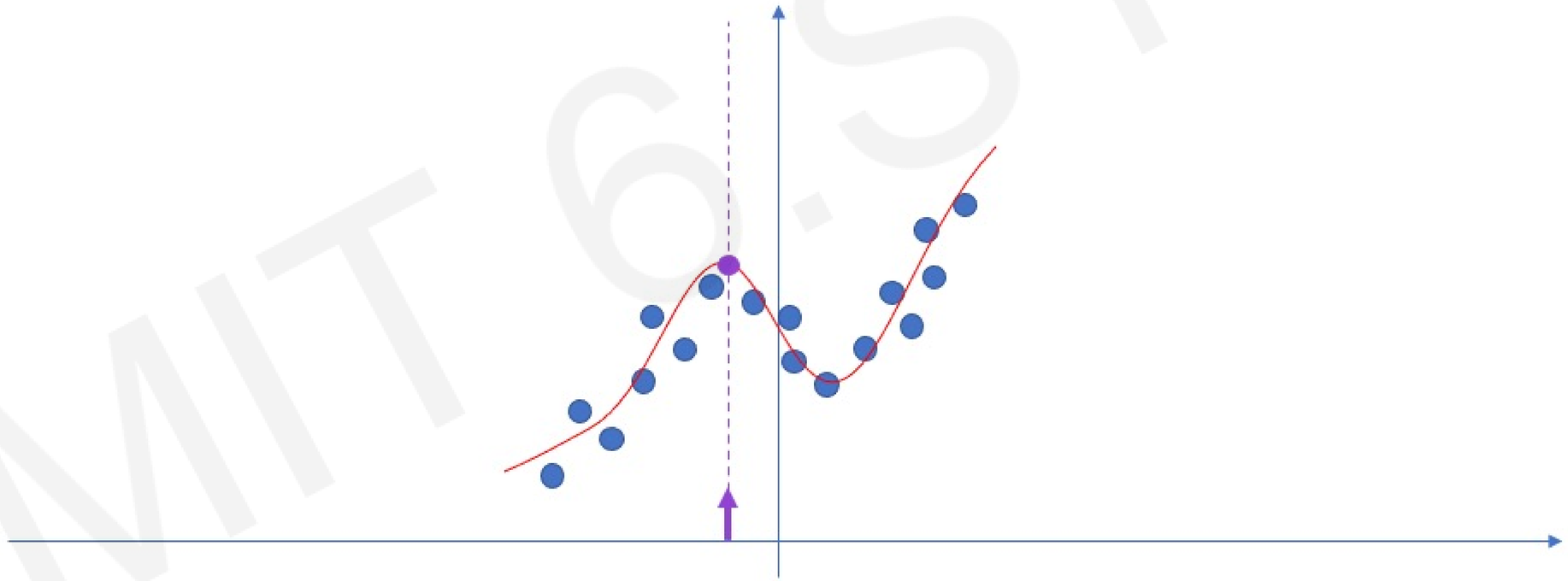
Neural networks are **excellent** function approximators



MIT 6.S191

Neural Networks as Function Approximators

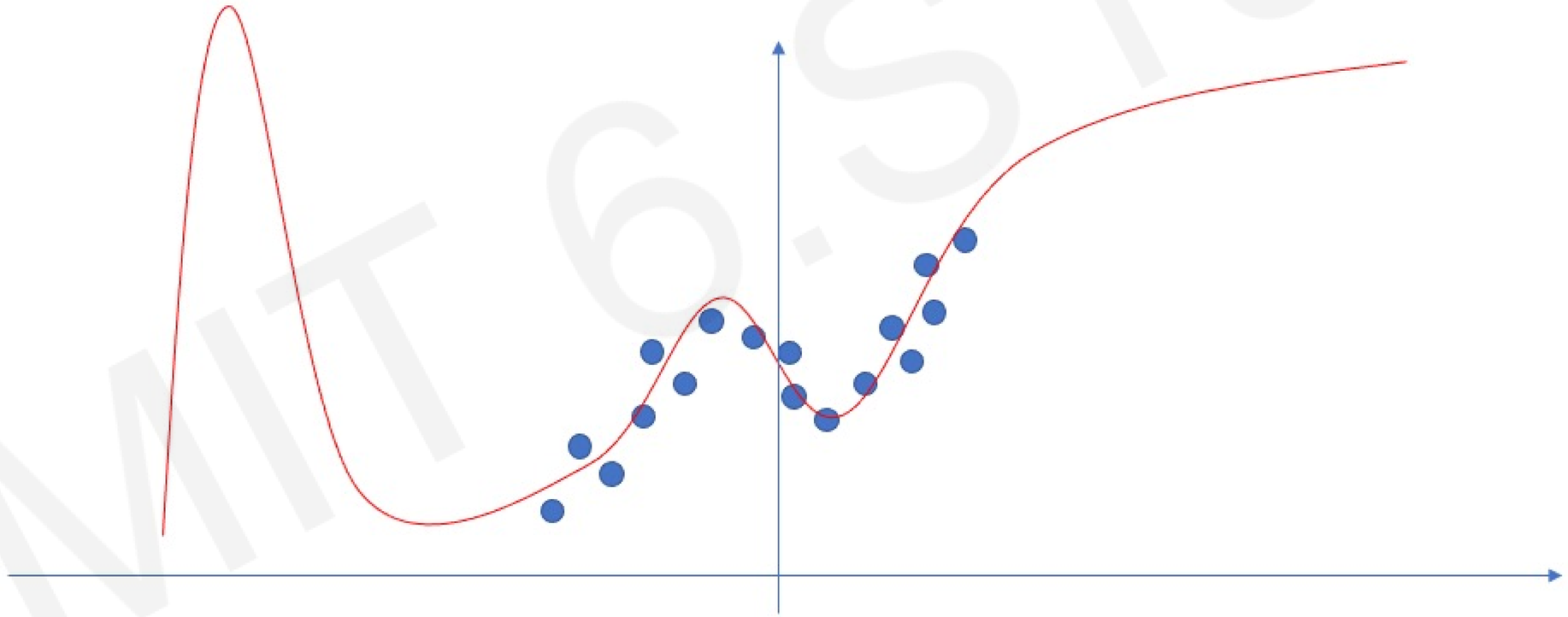
Neural networks are **excellent** function approximators



MIT 6.S191

Neural Networks as Function Approximators

Neural networks are **excellent** function approximators

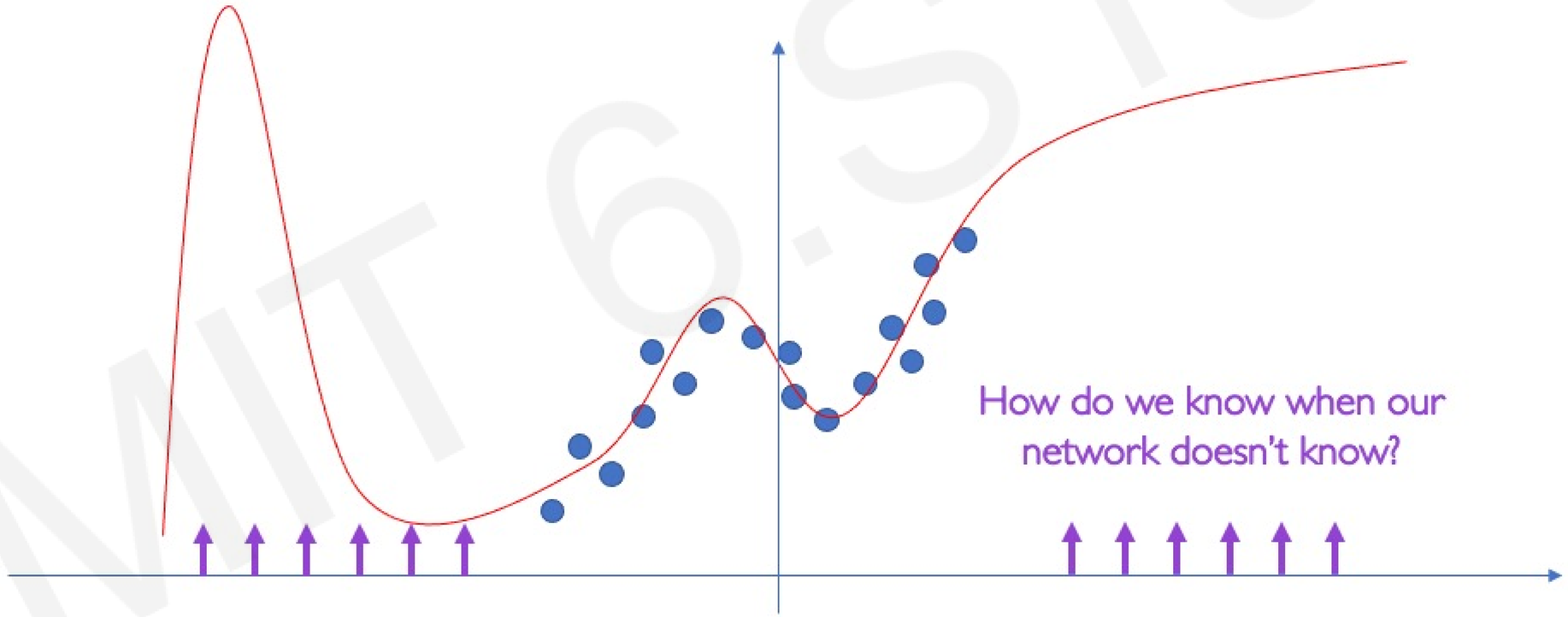


MIT 6.S191

Neural Networks as Function Approximators

Neural networks are **excellent** function approximators

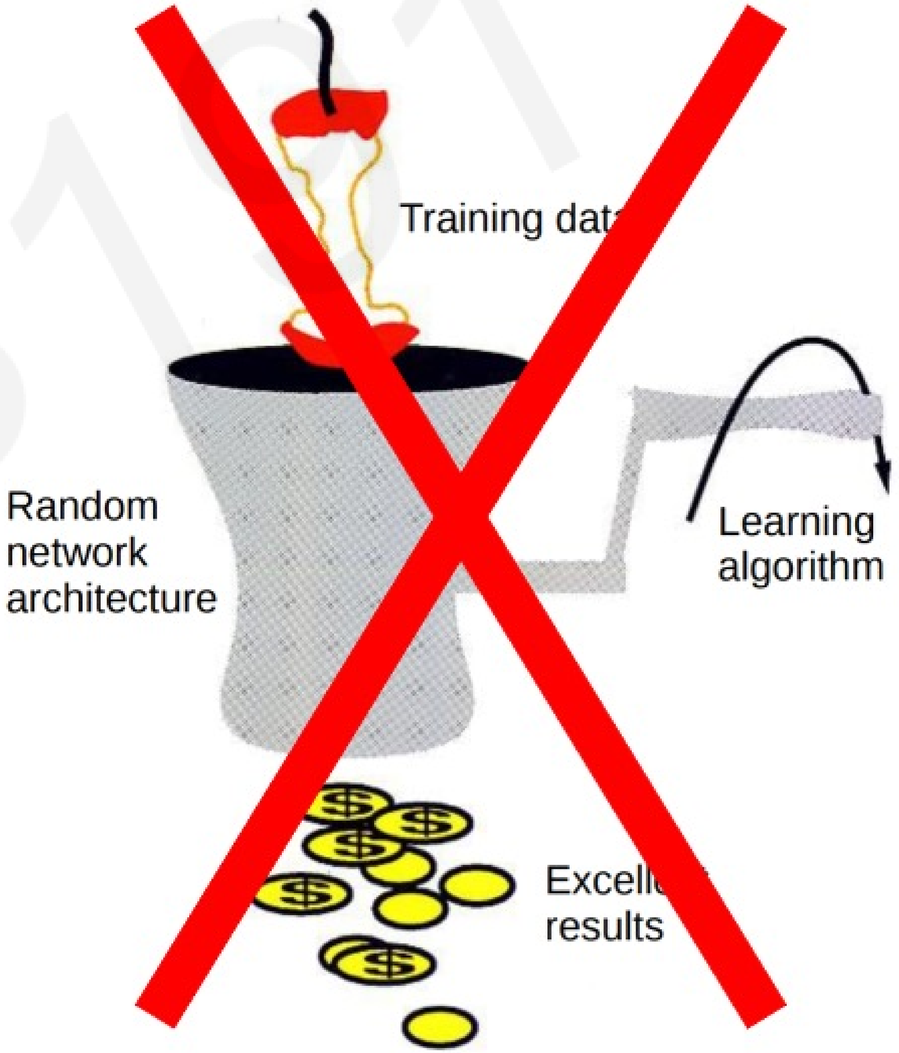
…when they have training data



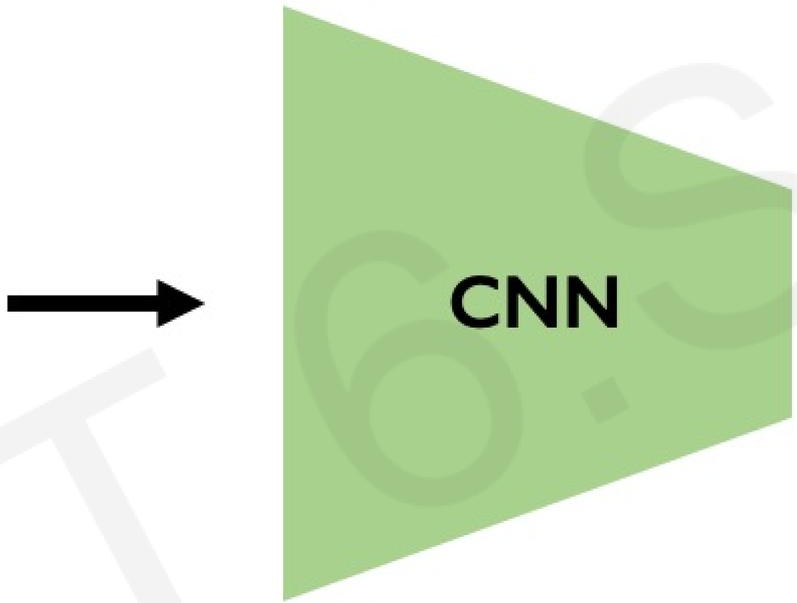
MIT 6.S191

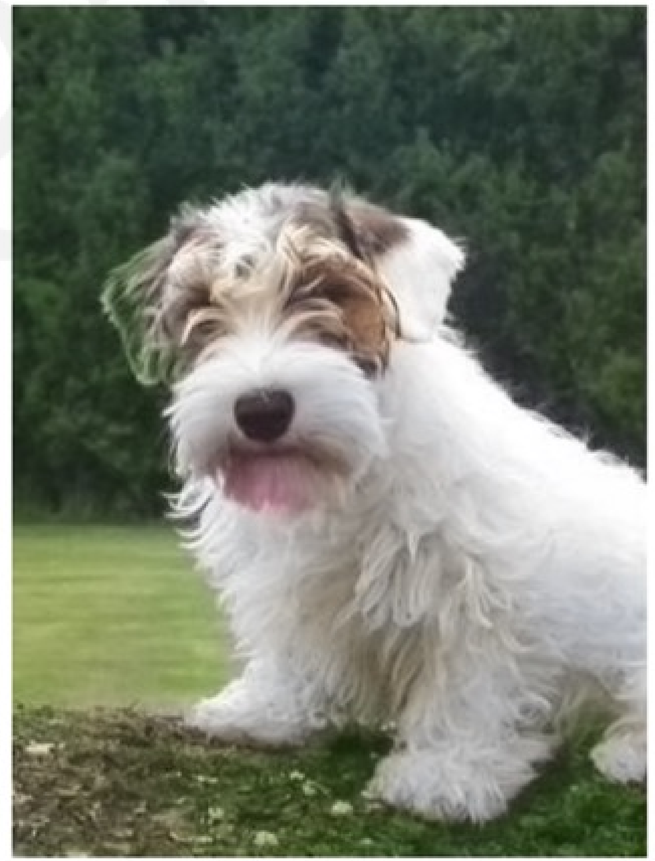
Deep Learning = Alchemy?

NO!



Neural Network Failure Modes, Part 1



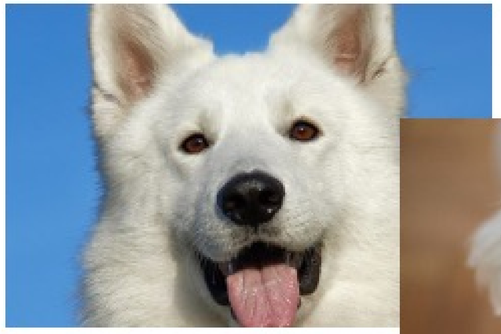
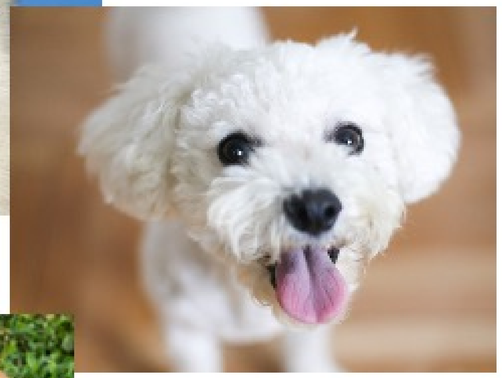


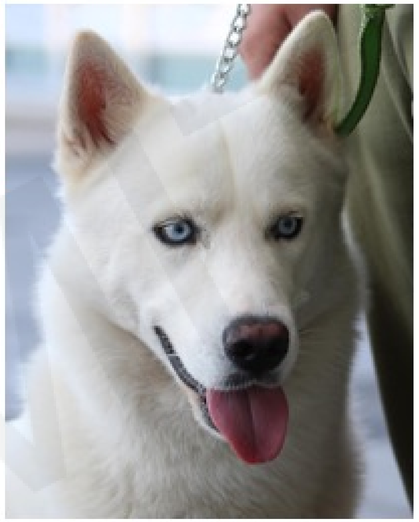
Train network to
colorize BW images.

Why could this be the case?

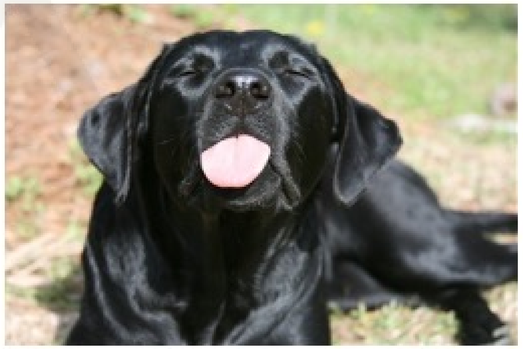
MIT6.S191

What Happens During Training…




CNN


Neural Network Failure Modes, Part II

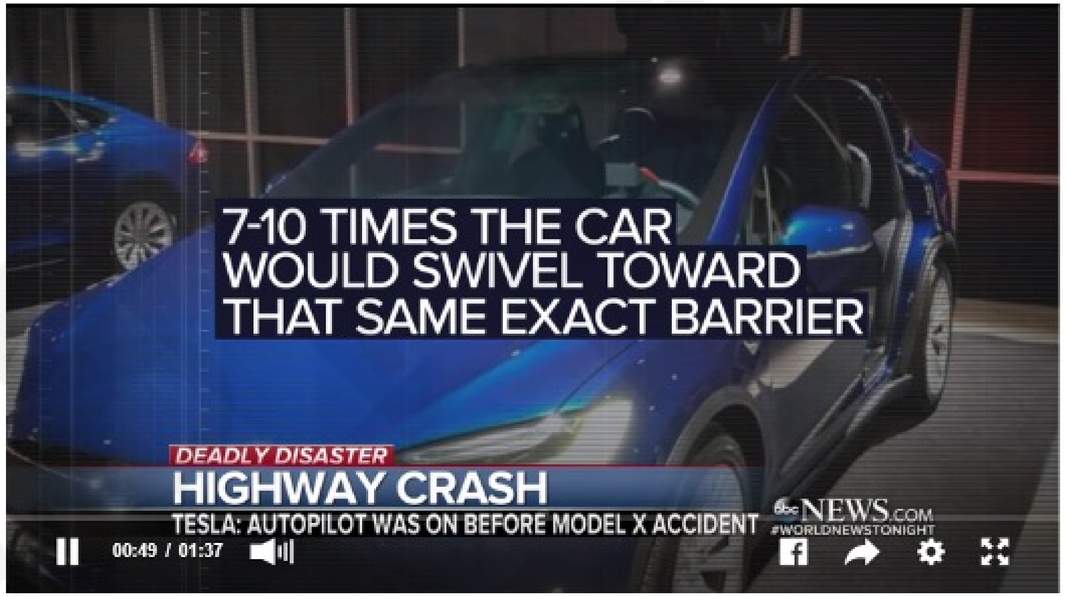
Tesla car was on autopilot prior to fatal crash in California, company says

The crash near Mountain View, California, last week killed the driver.

By **Mark Osborne**

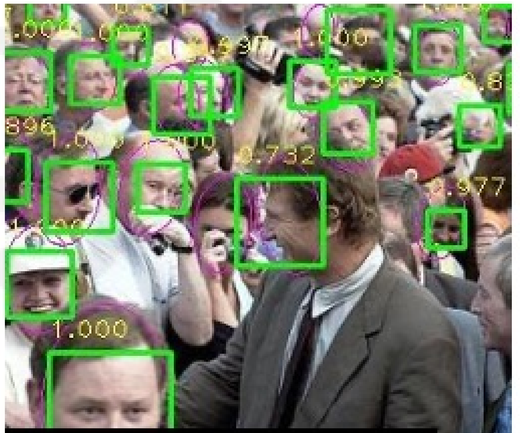
March 31, 2018, 1:57 AM • 5 min read

abc NEWS

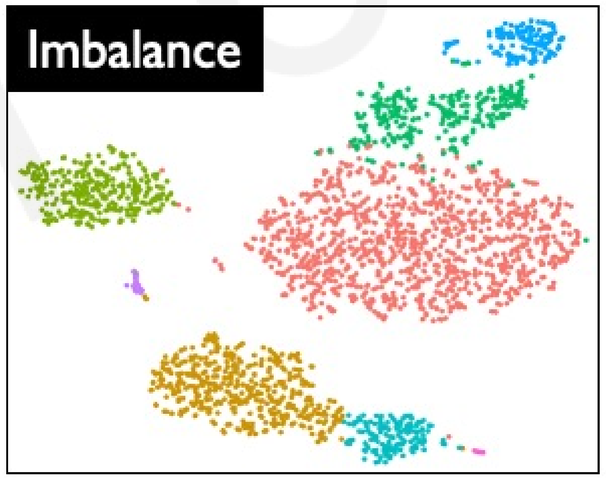


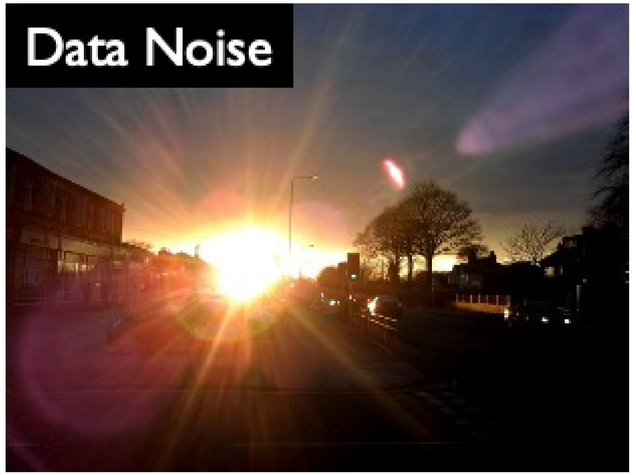
Uncertainty in Deep Learning

Safety-critical applications



Sparse and/or noisy datasets



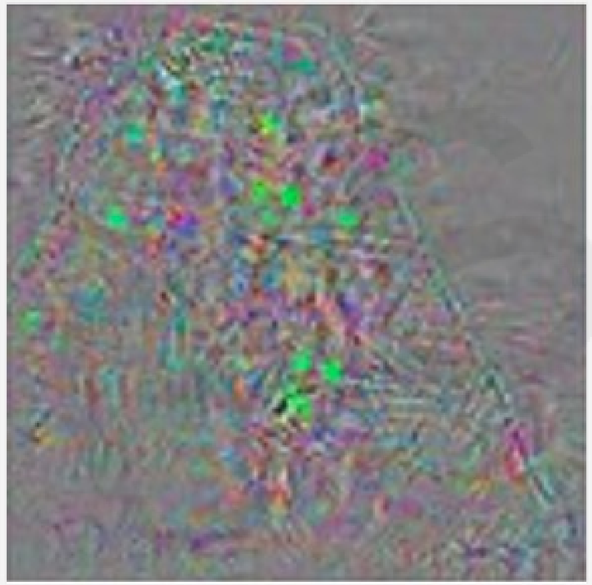


Lab + Lecture

Neural Network Failure Modes, Part III

Original image

Temple (97%)



Perturbations

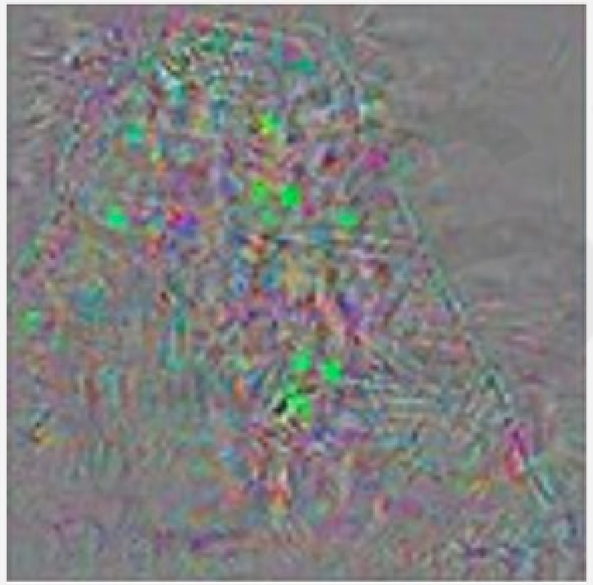
Adversarial example

Ostrich (98%)

Adversarial Attacks on Neural Networks

Original image

Temple (97%)



Perturbations

Adversarial example

Ostrich (98%)

Adversarial Attacks on Neural Networks

Remember:

We train our networks with gradient descent

$$W \leftarrow W - \eta \frac{\partial J(W, x, y)}{\partial W}$$

"How does a small change in weights decrease our loss"

MIT 6S191

Adversarial Attacks on Neural Networks

Remember:

We train our networks with gradient descent

$$W \leftarrow W - \eta \frac{\partial J(W, x, y)}{\partial W}$$

"How does a small change in weights decrease our loss"

MIT 6.S191

Adversarial Attacks on Neural Networks

Remember:

We train our networks with gradient descent

$$W \leftarrow W - \eta \frac{\partial J(W, x, y)}{\partial W}$$

Fix your image x, and true label y

"How does a small change in weights decrease our loss"

MIT 6.S191

Adversarial Attacks on Neural Networks

Adversarial Image:

Modify image to increase error

$$x \leftarrow x + \eta \frac{\partial J(W, x, y)}{\partial x}$$

"How does a small change in the input increase our loss"

MIT 6.S191

Adversarial Attacks on Neural Networks

Adversarial Image:

Modify image to increase error

$$x \leftarrow x + \eta \frac{\partial J(W, x, y)}{\partial x}$$

"How does a small change in the input increase our loss"

MIT 6.S191

Adversarial Attacks on Neural Networks

Adversarial Image:

Modify image to increase error

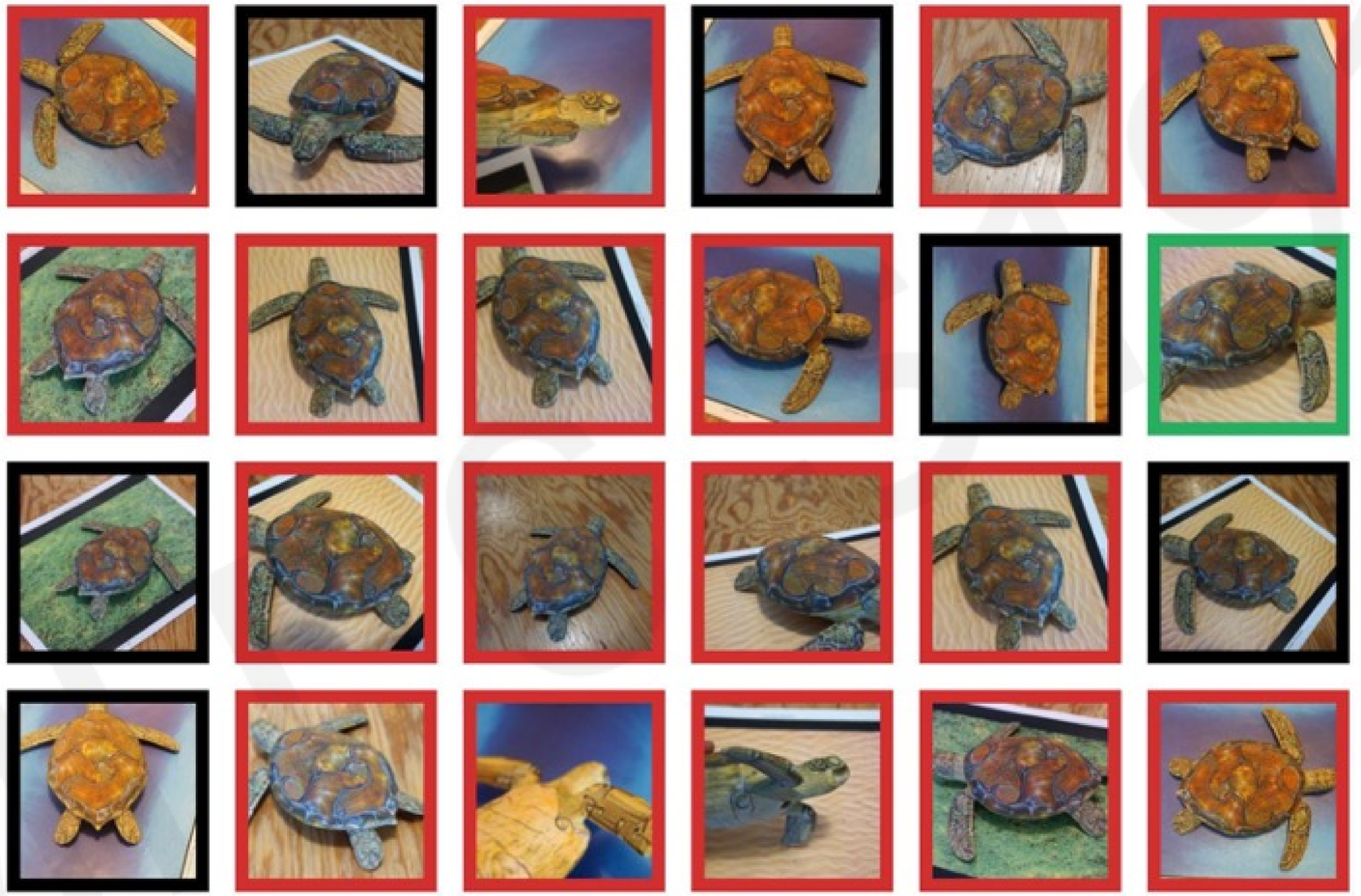
$$x \leftarrow x + \eta \frac{\partial J(W, x, y)}{\partial x}$$

Fix your weights θ, and true label y

"How does a small change in the input increase our loss"

MIT 6.S191

Synthesizing Robust Adversarial Examples



classified as turtle classified as rifle

classified as other

Algorithmic Bias

Overcoming Racial Bias In AI Systems And Startlingly Even In AI Self-Driving Cars

Racial bias in a medical algorithm favors white patients over sicker black patients

AI expert calls for end to UK use of 'racially biased' algorithms

AI Bias Could Put Women's Lives At Risk – A Challenge For Regulators

Gender bias in AI: building fairer algorithms

Bias in AI: A problem recognized but still unresolved

Amazon, Apple, Google, IBM, and Microsoft worse at transcribing black people's voices than white people's with AI voice recognition, study finds

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

When It Comes to Gorillas, Google Photos Remains Blind

Google promised a fix after its photo-categorization software labeled black people as gorillas in 2015. More than two years later, it hasn't found one.

The Week in Tech: Algorithmic Bias Is Bad. Uncovering It Is Good.

Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech

The Best Algorithms Struggle to Recognize Black Faces Equally

US government tests find even top-performing facial recognition systems misidentify blacks at rates five to 10 times higher than they do whites.

Artificial Intelligence has a gender bias problem – just ask Siri

Lab + Lecture

Neural Network Limitations…

- Very **data hungry** (eg. often millions of examples)
- **Computationally intensive** to train and deploy (tractably requires GPUs)
- Easily fooled by **adversarial examples**
- Can be subject to **algorithmic bias**
- Poor at **representing uncertainty** (how do you know what the model knows?)
- Uninterpretable **black boxes**, difficult to trust
- Often require **expert knowledge** to design, fine tune architectures
- Difficult to **encode structure** and prior knowledge during learning
- **Extrapolation**: struggle to go beyond the data

MIT 6.S191

Neural Network Limitations…

- Very **data hungry** (eg. often millions of examples)
- **Computationally intensive** to train and deploy (tractably requires GPUs)
- Easily fooled by **adversarial examples**
- Can be subject to **algorithmic bias**
- Poor at **representing uncertainty** (how do you know what the model knows?)
- Uninterpretable **black boxes**, difficult to trust
- Often require **expert knowledge** to design, fine tune architectures
- Difficult to **encode structure** and prior knowledge during learning
- **Extrapolation**: struggle to go beyond the data

Lab + Lecture

MIT 6.S191

Neural Network Limitations…

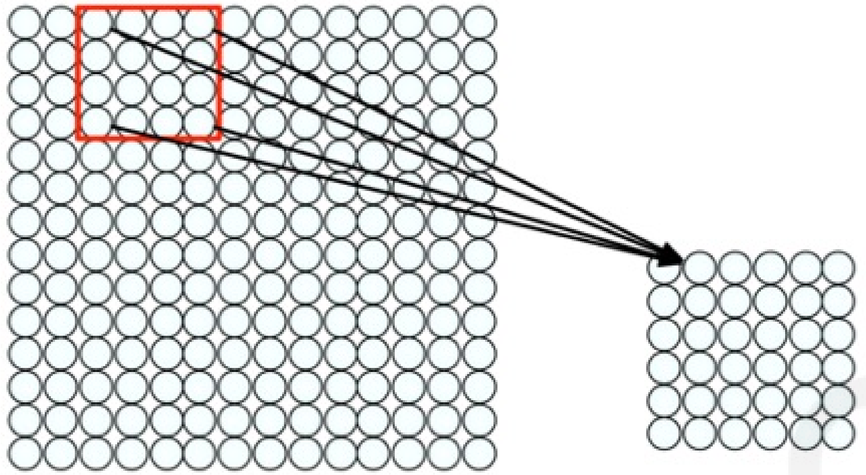
- Very **data hungry** (eg. often millions of examples)
- **Computationally intensive** to train and deploy (tractably requires GPUs)
- Easily fooled by **adversarial examples**
- Can be subject to **algorithmic bias**
- Poor at **representing uncertainty** (how do you know what the model knows?)
- Uninterpretable **black boxes**, difficult to trust
- Often require **expert knowledge** to design, fine tune architectures
- Difficult to **encode structure** and prior knowledge during learning
- **Extrapolation**: struggle to go beyond the data

MIT 6.S191

New Frontiers 1:
Encoding Structure into Deep Learning

MIT 6.S191

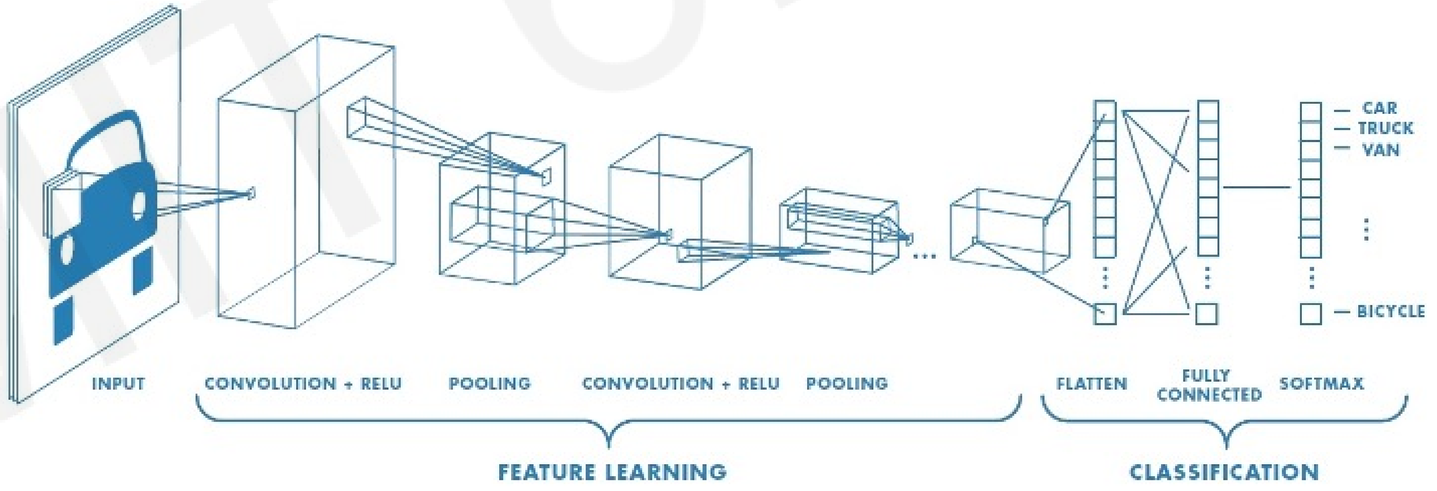
CNNs: Using Spatial Structure



1) Apply a set of weights to extract **local features**

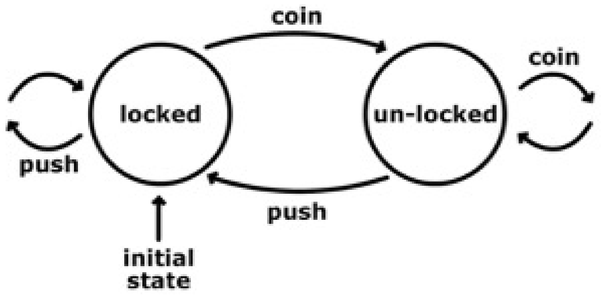
2) Use **multiple filters** to extract different features

3) **Spatially share** parameters of each filter

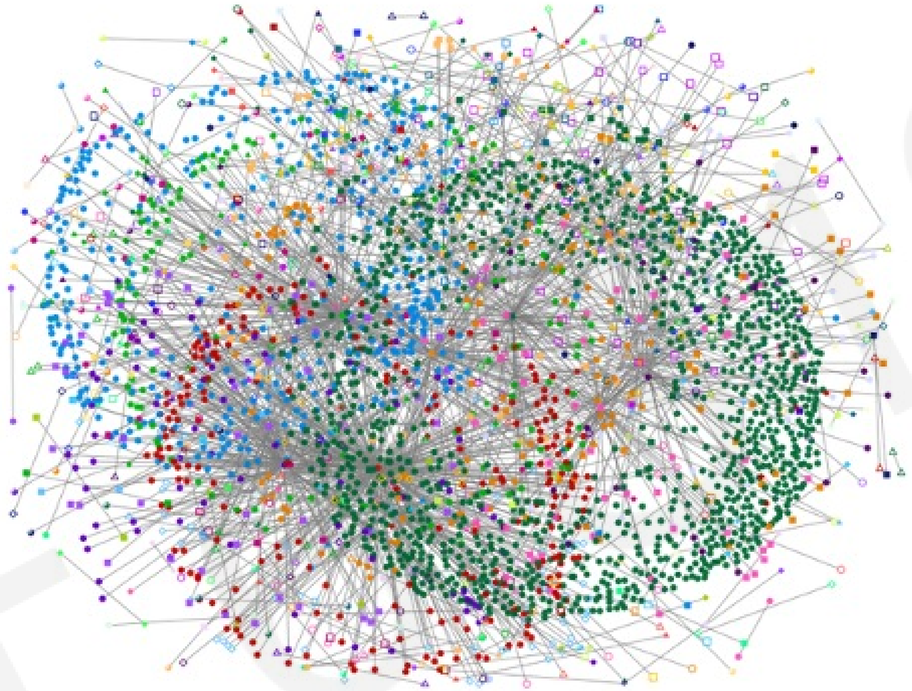


MIT 6.S191

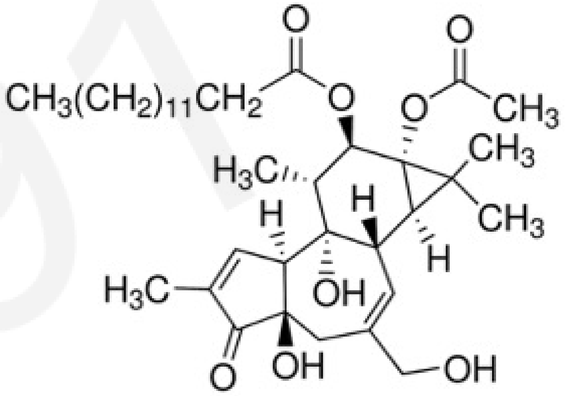
Graphs as a Structure for Representing Data



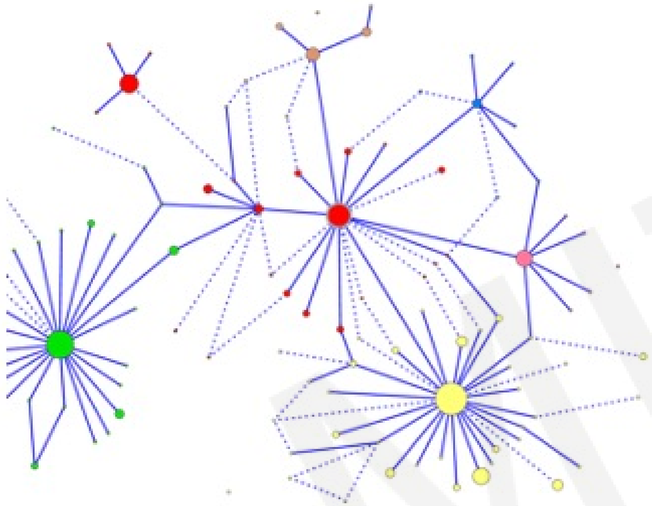
State Machines



Social Networks



Molecules



Biological Networks

Many real-world data – such as networks – cannot be captured by "standard" encodings or Euclidean geometries

Mobility & Transport

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks

Convolutional Networks

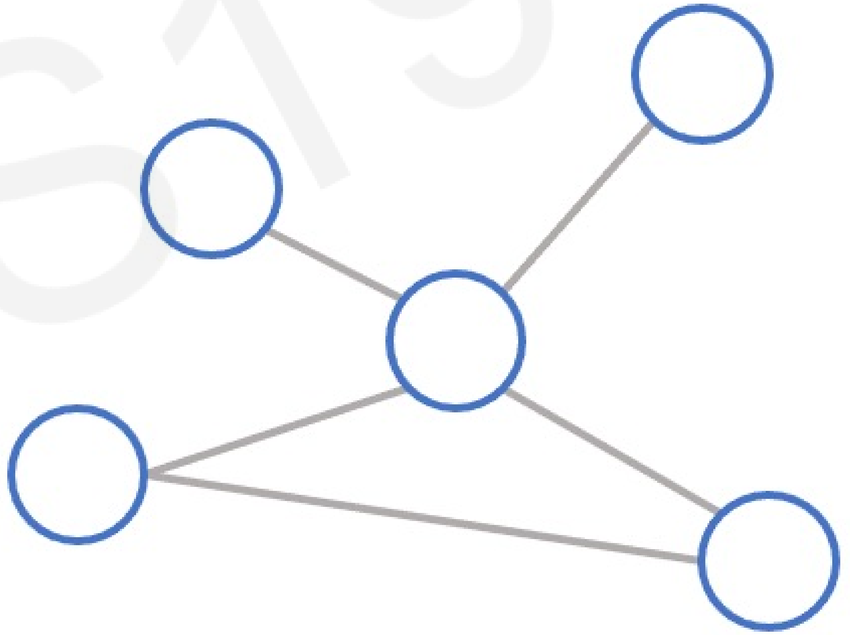
Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks

Convolutional Networks

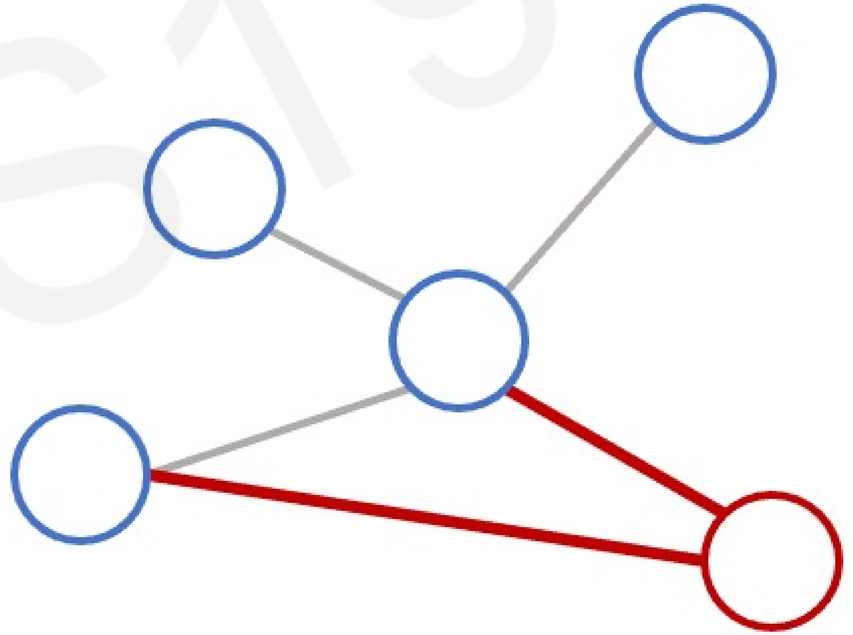
Graph Convolutional Networks (GCNs)



Graph Convolutional Networks

Convolutional Networks

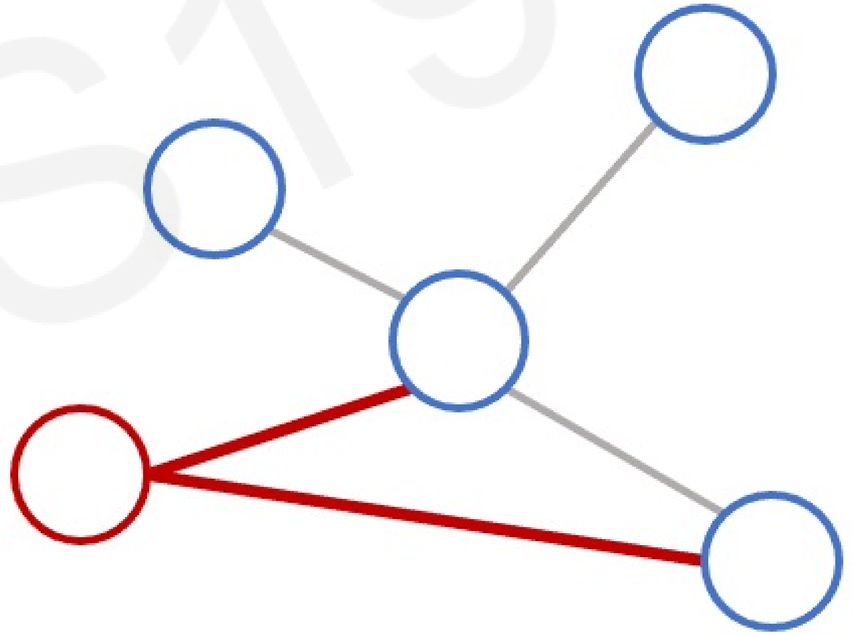
Graph Convolutional Networks (GCNs)



Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)

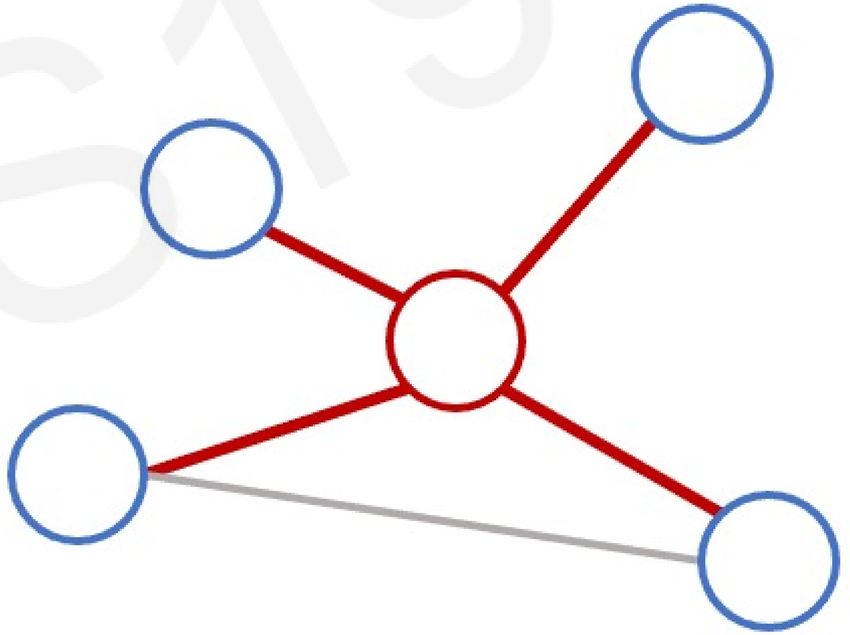


MIT 6.S191

Graph Convolutional Networks

Convolutional Networks

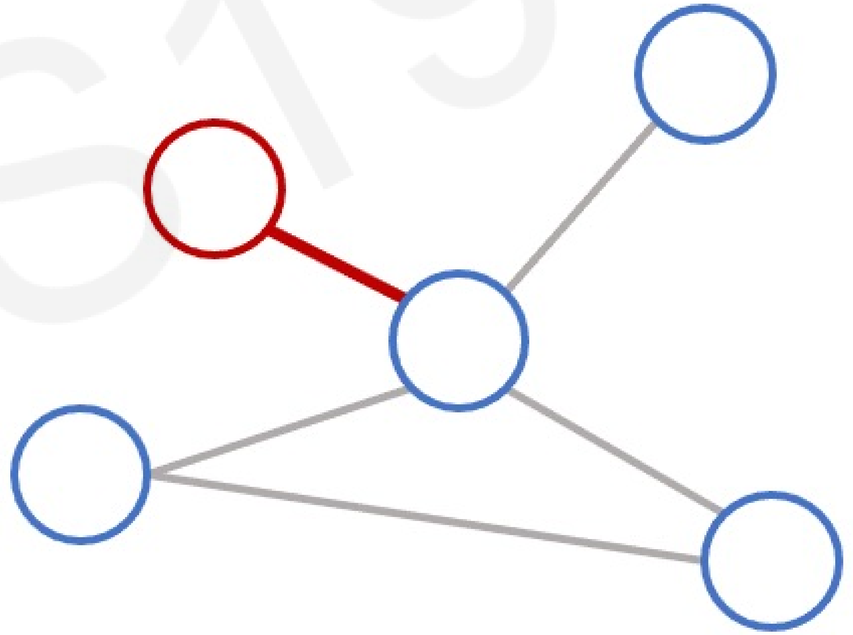
Graph Convolutional Networks (GCNs)



Graph Convolutional Networks

Convolutional Networks

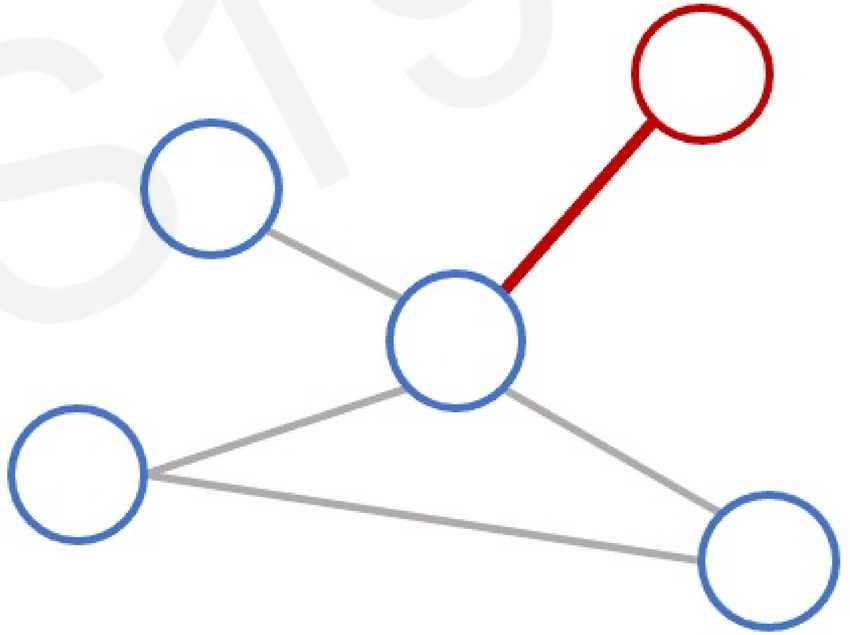
Graph Convolutional Networks (GCNs)



Graph Convolutional Networks

Convolutional Networks

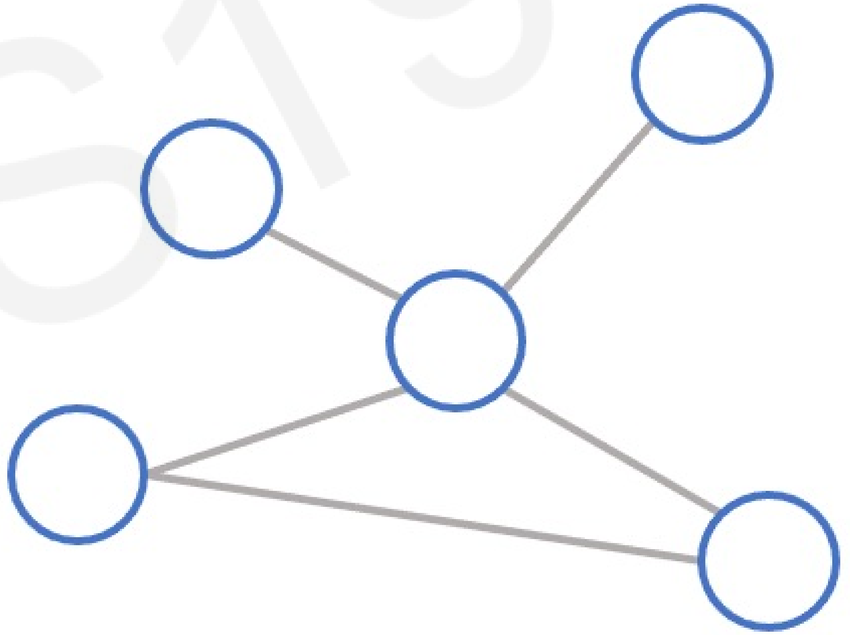
Graph Convolutional Networks (GCNs)



Graph Convolutional Networks

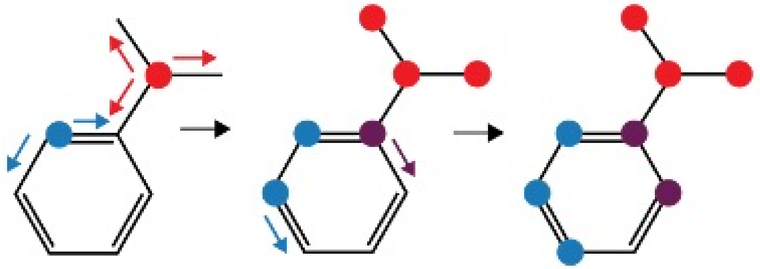
Convolutional Networks

Graph Convolutional Networks (GCNs)



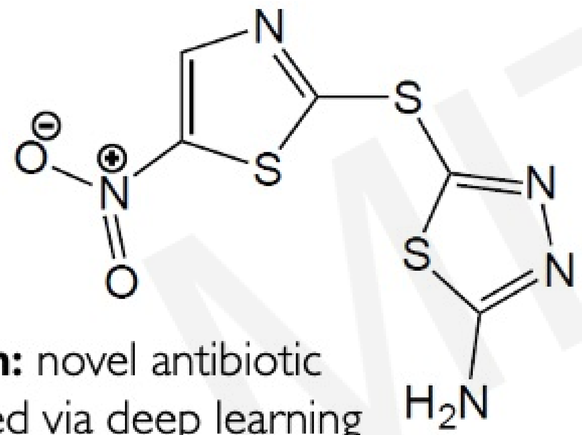
Applications of Graph Neural Networks

Molecular Discovery



Message-passing neural network

Jin+ *JCIM* 2019; Soleimany+ *ACS Cent. Sci.* 2021



Halicin: novel antibiotic discovered via deep learning

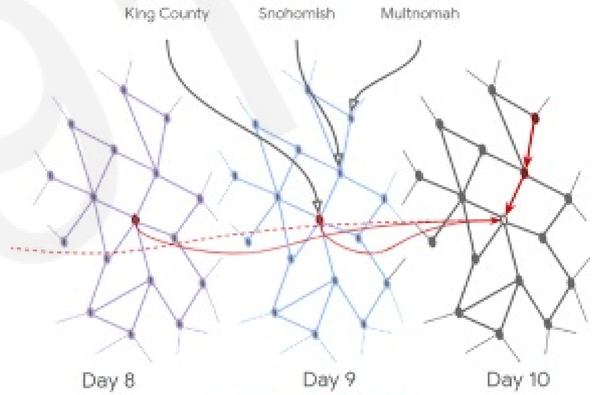
Stokes+ *Cell* 2020

Traffic Prediction

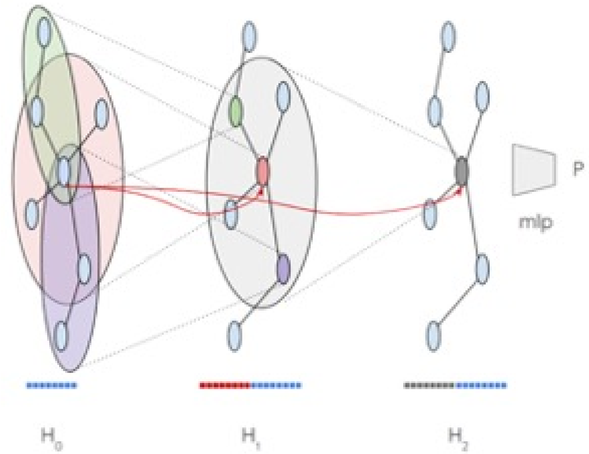
ETA Improvements with GoogleMaps

DeepMind + GoogleMaps

COVID-19 Forecasting



Spatio-temporal data



Graph network + temporal embedding

Kapoor+ *KDD* 2020

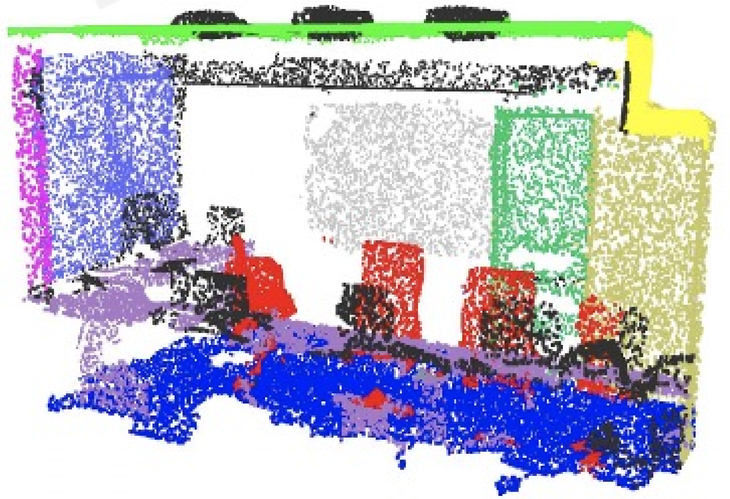
MIT 6.S191

Learning From 3D Data

Point clouds are **unordered sets** with **spatial dependence** between points

Classification

Part Segmentation

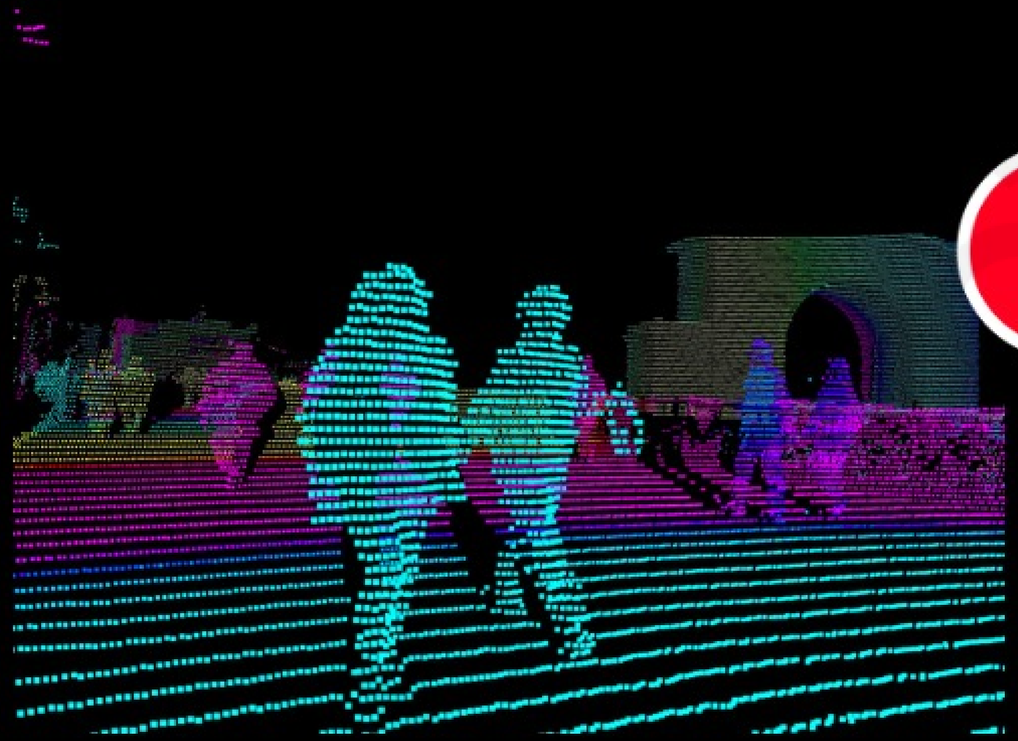


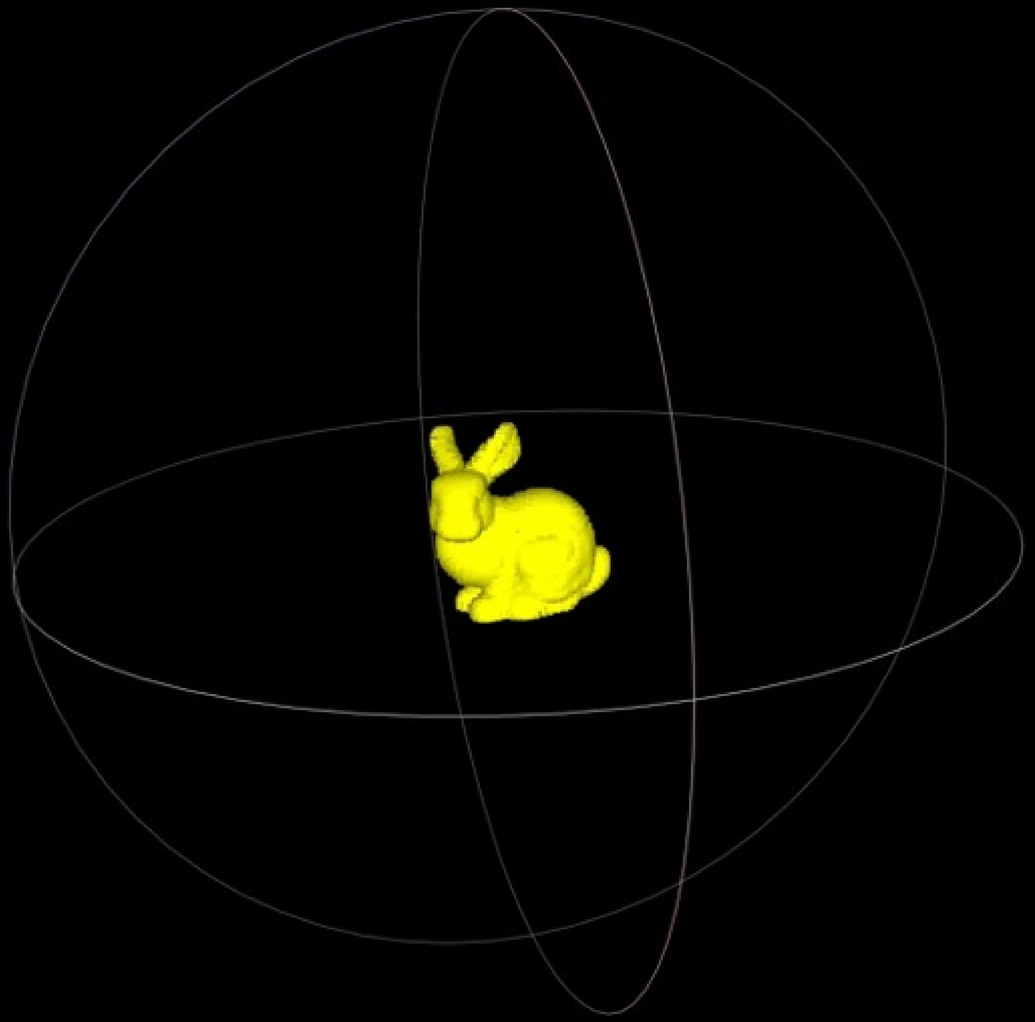
Semantic Segmentation

MIT6S191

Extending Graph CNNs to Pointclouds

Capture local geometric features of point clouds while maintaining order invariance



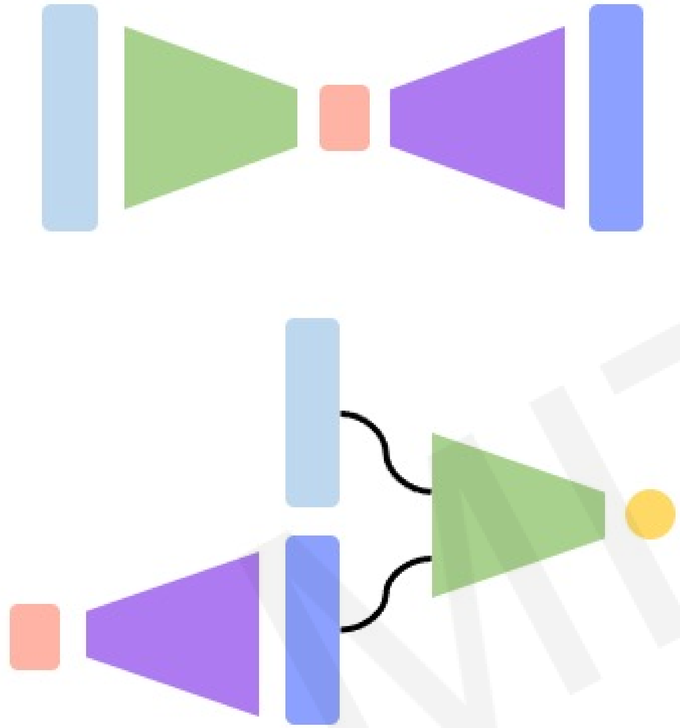


New Frontiers II:
Diffusion Models & Generative AI

MIT 6.S191

The Landscape of Generative Modeling

Lecture 4:
VAEs and GANs



Limitations

- Mode collapse
- Generating OOD
- Hard to train

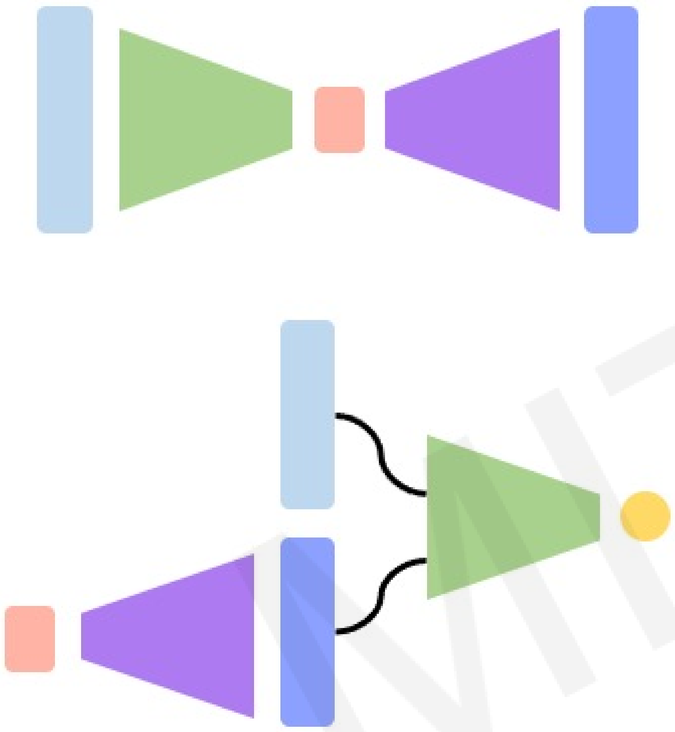
Challenges

- Stability
- Efficiency
- Quality
- Novelty

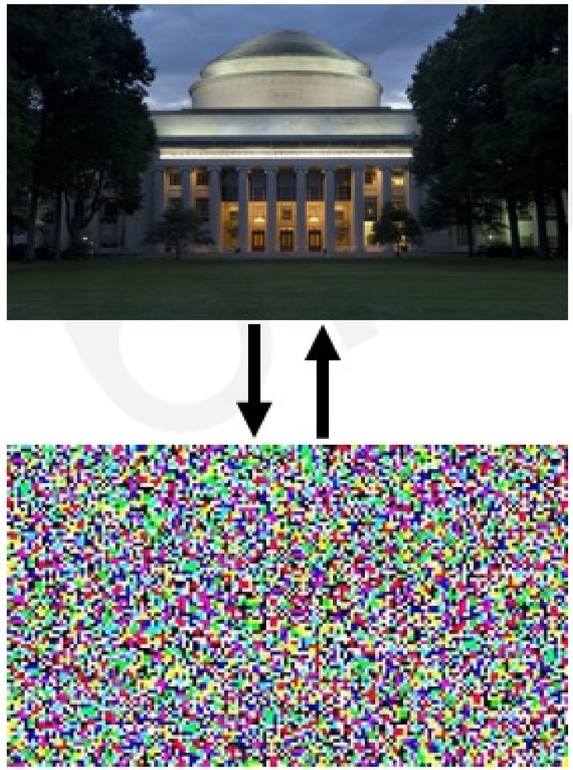
MIT 6.S191

The Landscape of Generative Modeling

Lecture 4:
VAEs and GANs



Diffusion Models

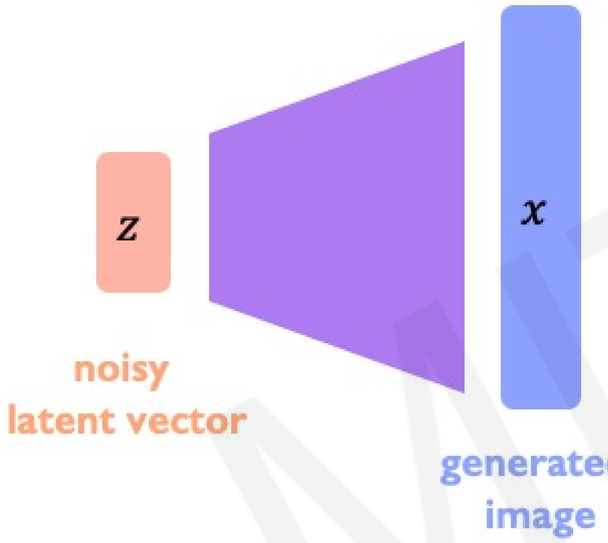


Text-to-Image

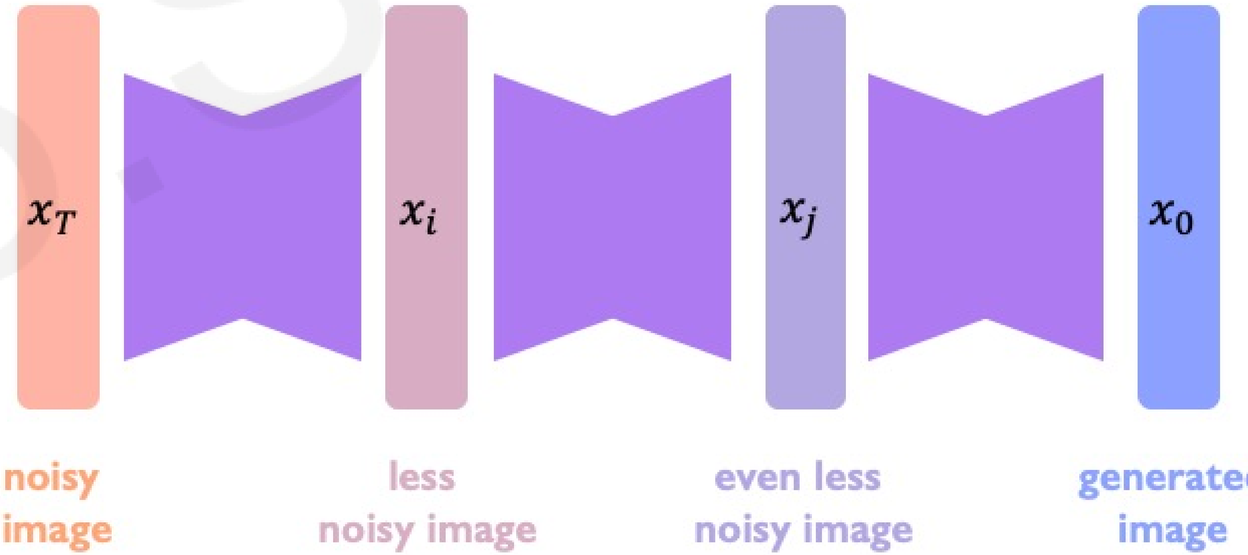
"Two cats doing research"

Diffusion Models

VAEs/GANs: Generating images <u>in one-shot</u> directly from low-dimensional latent variables



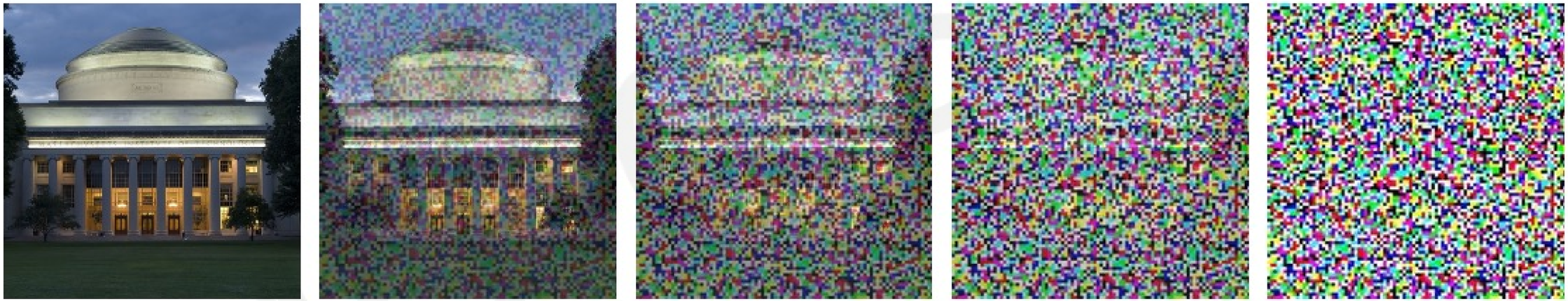
Diffusion: Generating images <u>iteratively</u> by repeatedly refining and removing noise



The Diffusion Process

Forward noising

(data-to-noise)

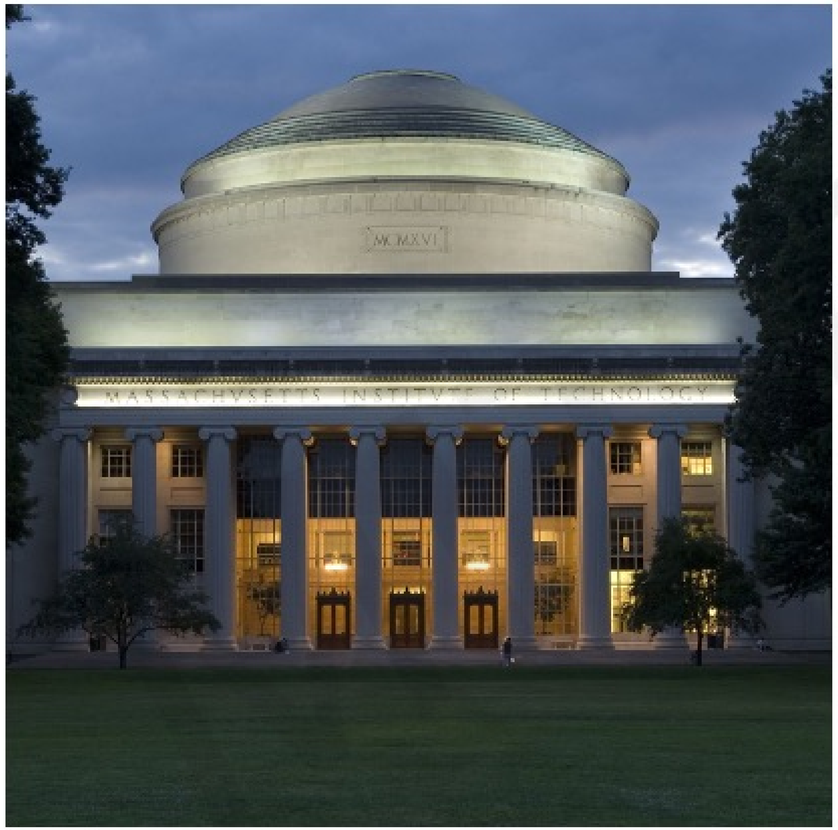


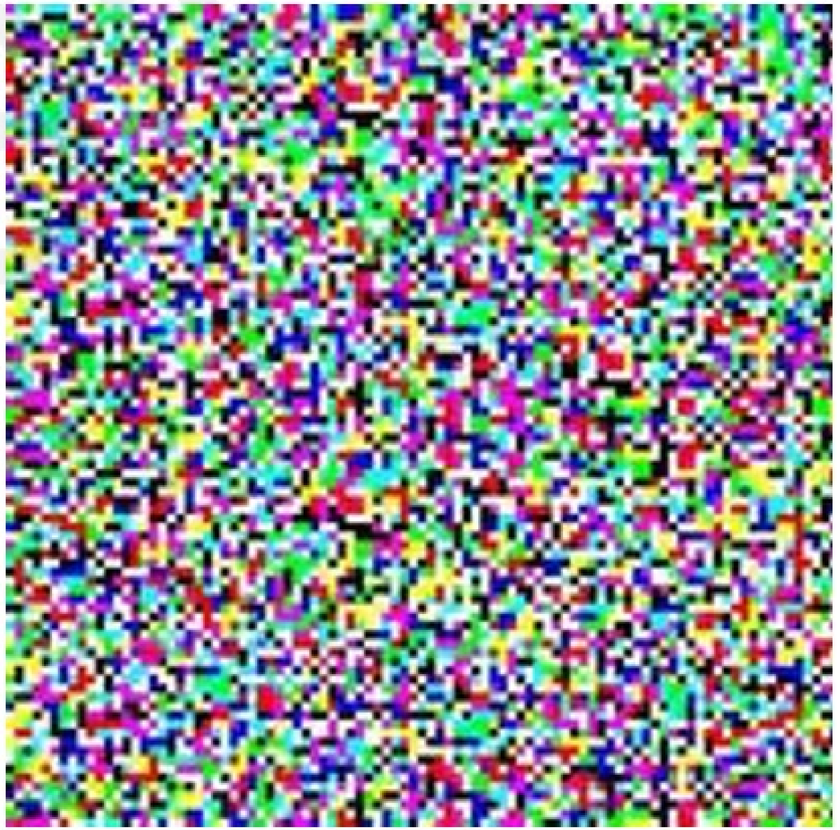
Reverse denoising

(noise-to-data)

Forward Noising

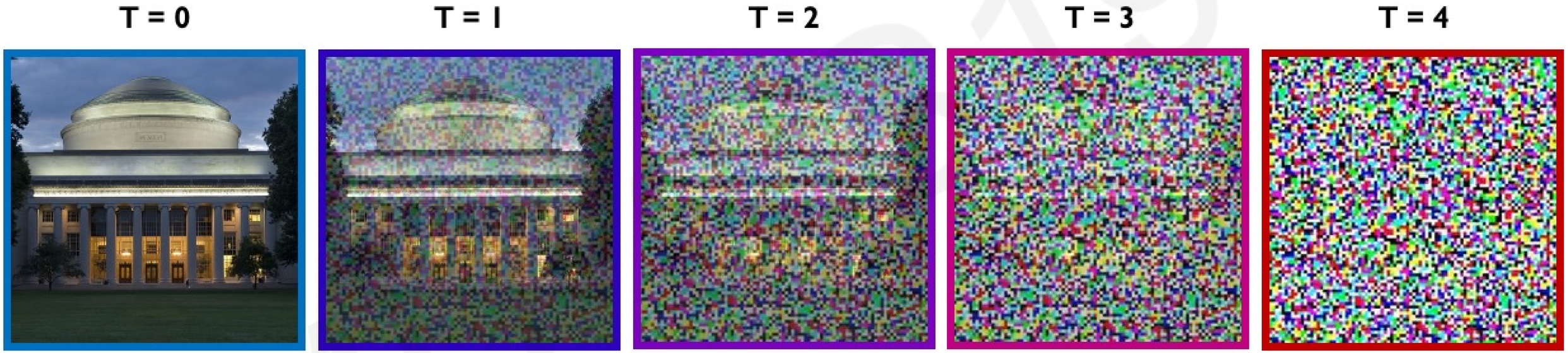
Step 1: Given an image (left), randomly sample a random noise pattern (right)



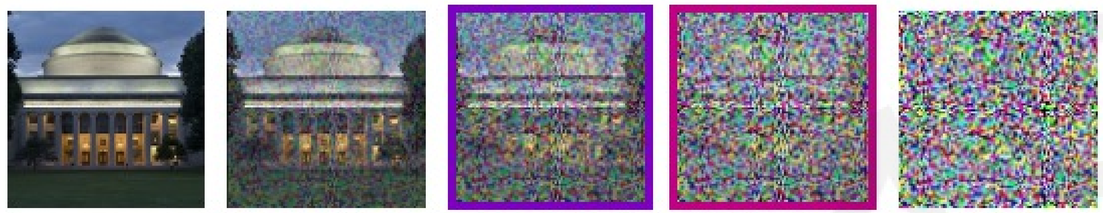


Forward Noising

Step 2: Progressively add more and more of the noise to your image



Reverse Denoising

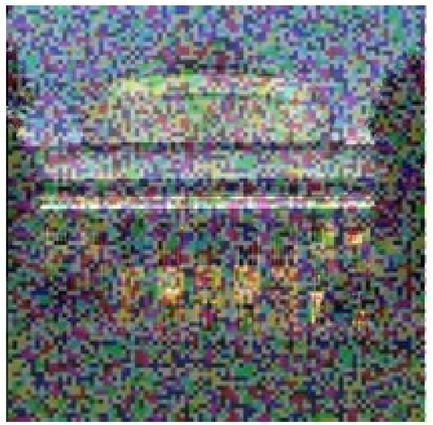


T = 0 ... T = 4

Goal: Given image at **T**, can we learn to estimate image at **T-1**?

T = 3

T = 2

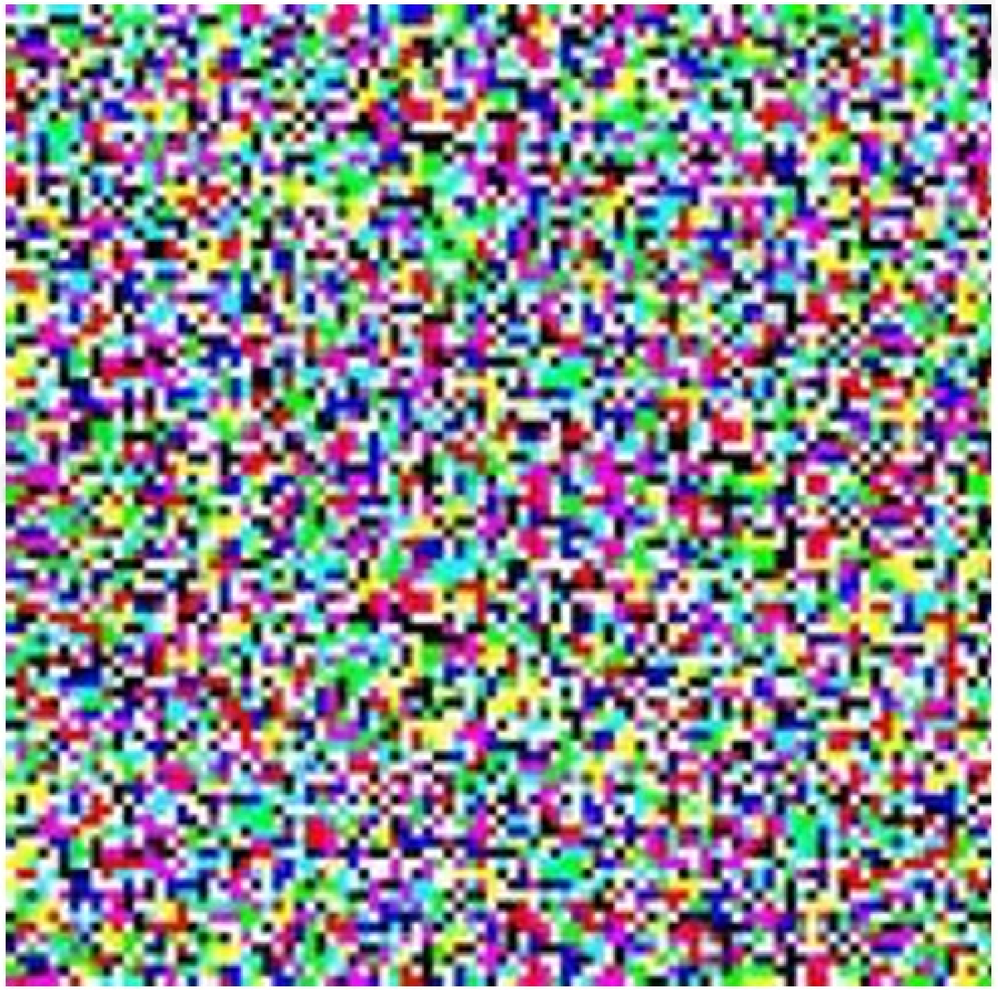


How can we train this network?

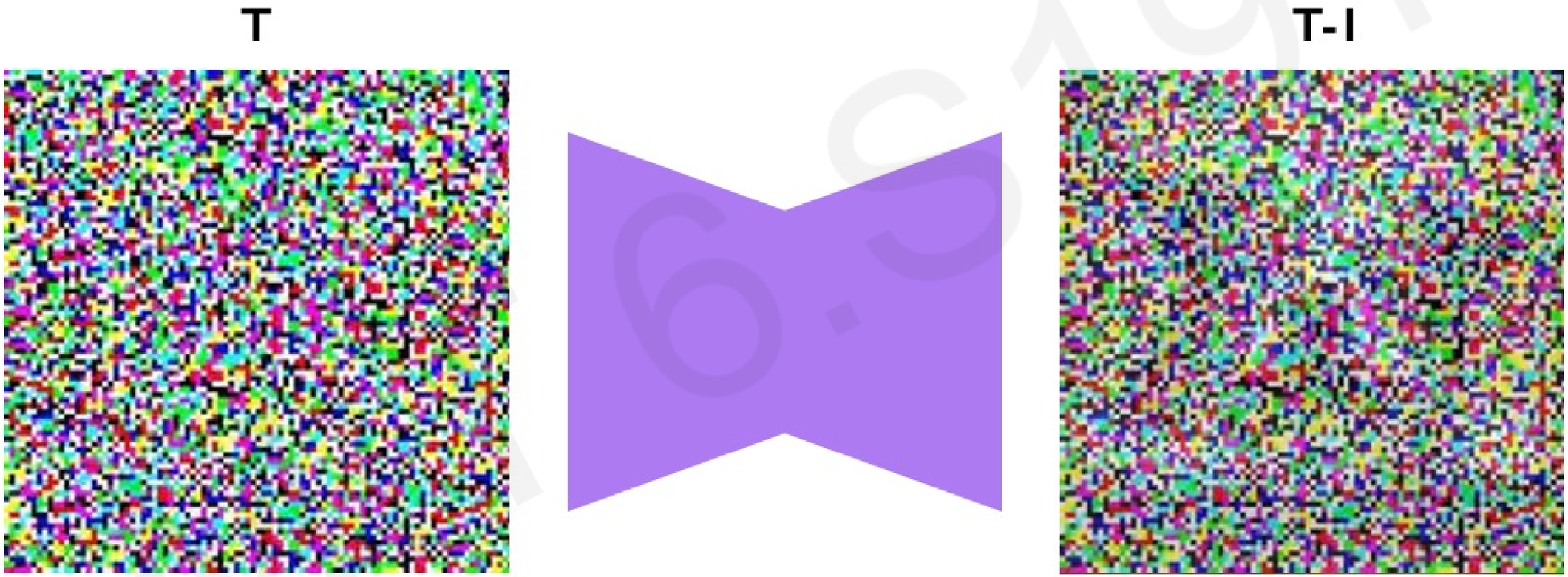
MIT 6.S191

Sampling Brand New Generations

T



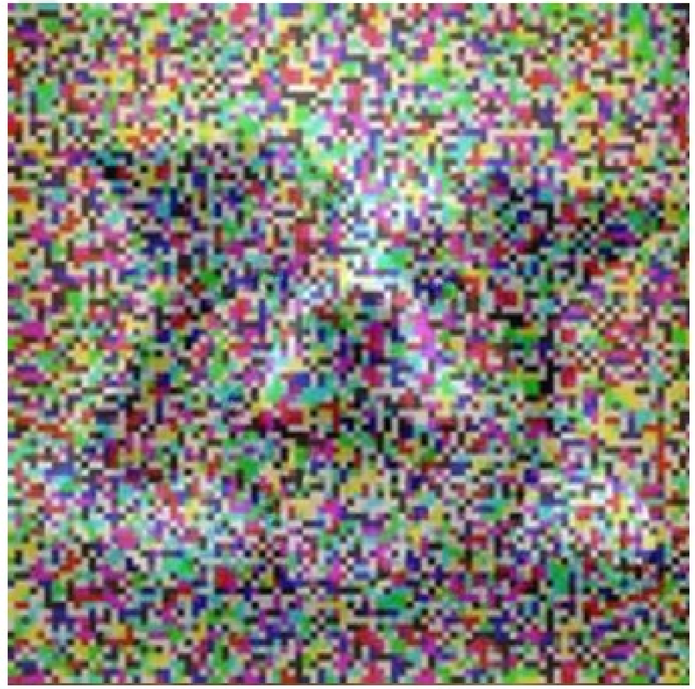
Sampling Brand New Generations



Sampling Brand New Generations

T-1

T-2



Sampling Brand New Generations

T-2

T-3

Sampling Brand New Generations

T-3

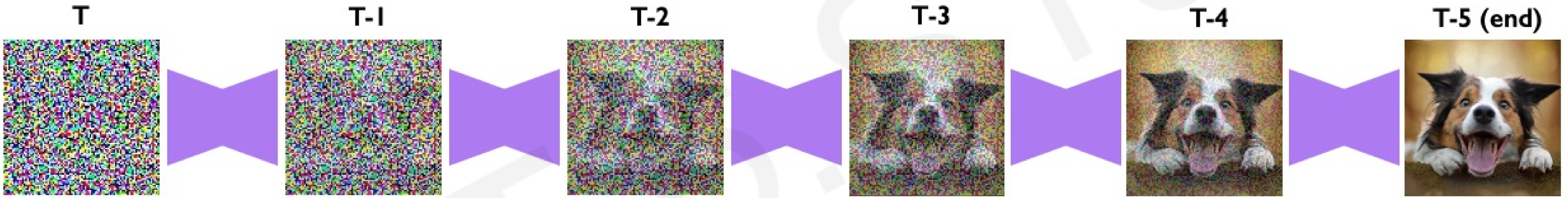
T-4

Sampling Brand New Generations

T-4

T-5 (end)

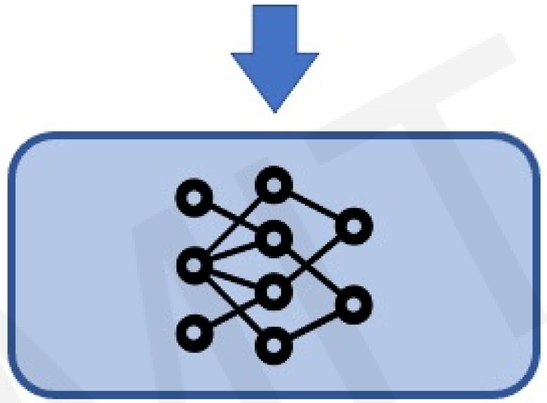
Sampling Brand New Generations



MIT 6.S191

Generating Images from Natural Language

"A photo of an astronaut riding a horse."



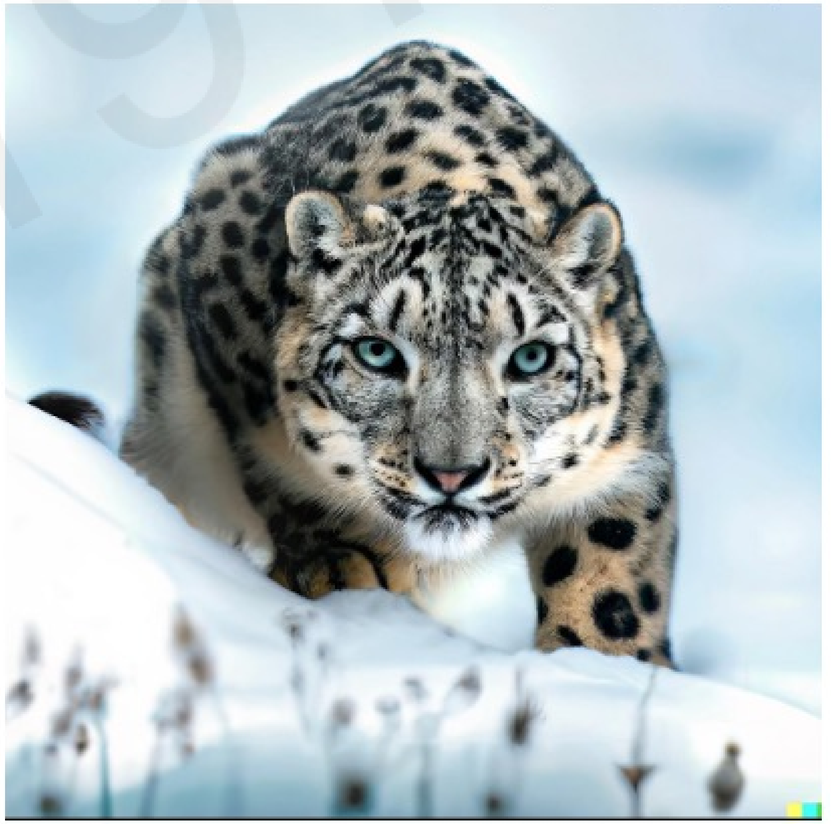
Ramesh+ *arXiv* 2022

Text-to-Image Generation

"a painting of a fox sitting in a field at sunrise in the style of Claude Monet"

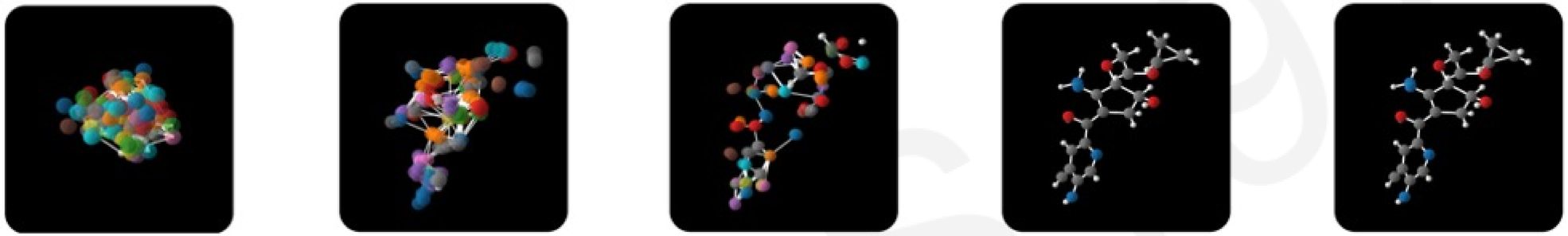
"an ibis in the wild, painted in the style of John Audubon"

"close-up of a snow leopard in the snow hunting, rack focus, nature photography"



Beyond Images: Molecular Design

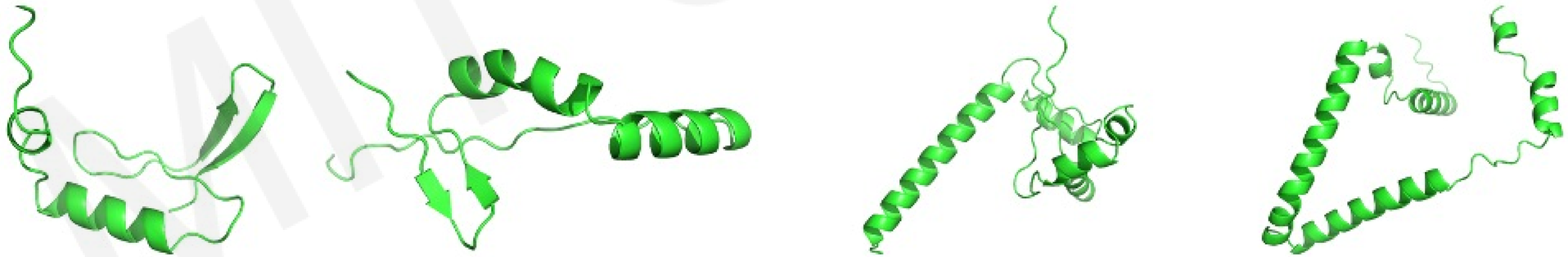
Chemistry: Generating Molecules in 3D



Noise ⟶ Molecule

Hoogeboom+ *ICML* 2022

Biology: Generating Novel Proteins



Wu+ *arXiv* 2022, Anand+ *arXiv* 2022, Trippe+ *arXiv* 2022

Generative Models for Protein Design

Can we design **new proteins** with new biological or therapeutic functions?

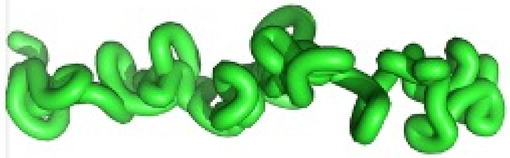
Protein function is defined by structure

Protein folding leads to structure

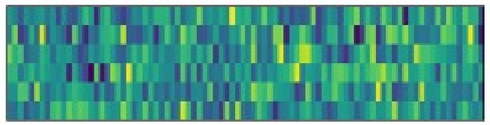
MIT 6.S191

Protein Structure Generation via Folding Diffusion

Structure, unfolded

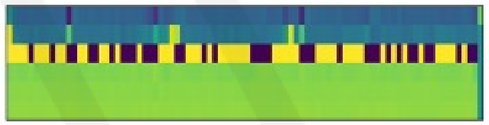


Angles, unfolded

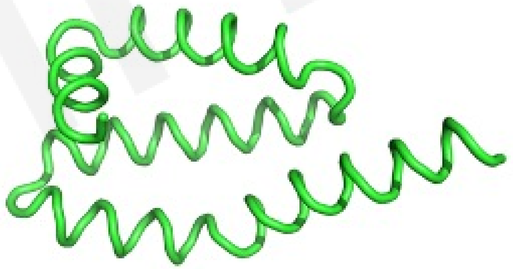


Diffusion Model

Angles, generated



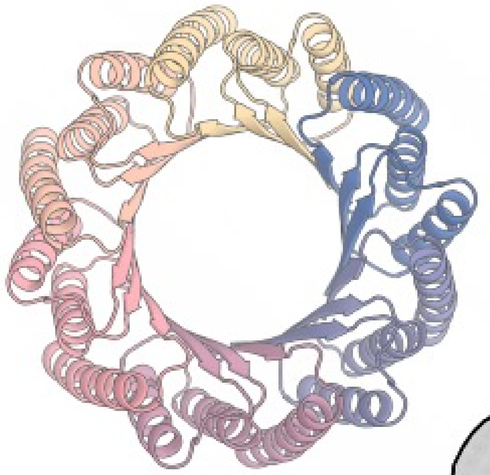
Structure, generated



Denoising to generate a structure!

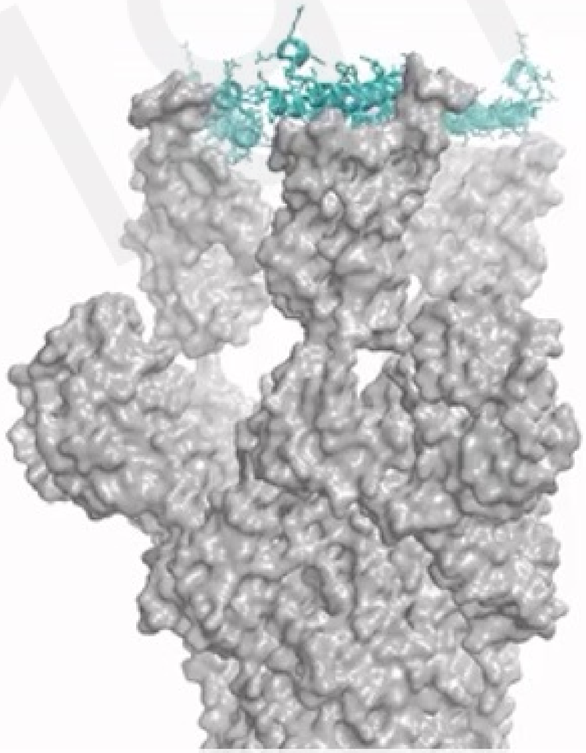
KE Wu, KK Yang, R van den Berg, J Zou, AX Lu, AP Amini *arXiv* 2022.

Diffusion Models: Foundation for Programmable Protein Design



AI-generated protein design

Real-world structure



Generating a novel binder to COVID spike

MIT 6.S191

Generative AI Spawns a Powerful Idea

“What I cannot create, I cannot understand."
Richard Feynman

- Images, language, biology, and more
- Design AI to improve and evolve AI itself
- Power and Caution of Generative AI – and AI at large

Connections and distinctions between artificial and human intelligence