

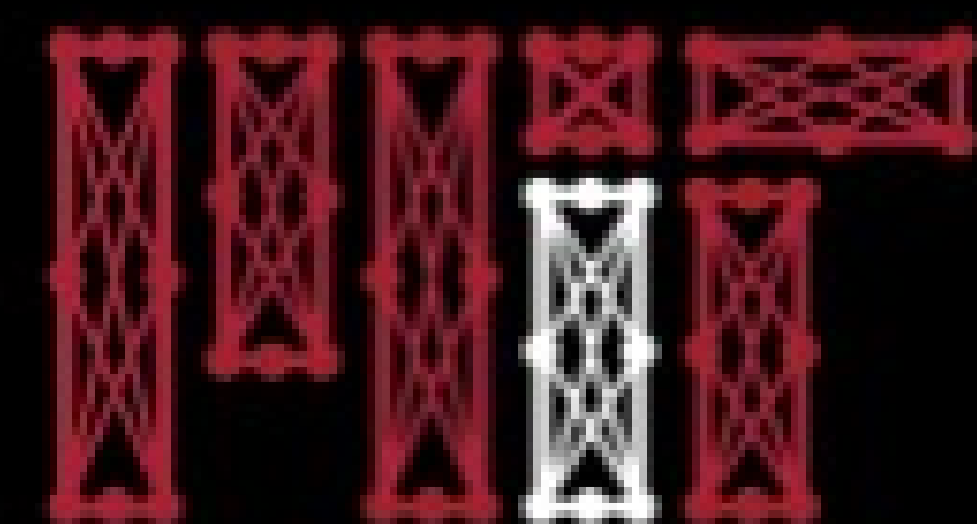


Deep Learning Limitations and New Frontiers

Ava Amini

MIT Introduction to Deep Learning

January 12, 2023



MIT Introduction to Deep Learning

introtodeeplearning.com [@MITDeepLearning](https://twitter.com/MITDeepLearning)



T-shirts! Today!



Program Schedule



Intro to Deep Learning

Lecture 1

Jan. 9, 2023

[\[Slides\]](#) [\[Video\]](#) coming soon!



Deep Computer Vision

Lecture 3

Jan. 10, 2023

[\[Slides\]](#) [\[Video\]](#) coming soon!



Uncertainty and Bias

Lecture 5

Jan. 11, 2023

[\[Info\]](#) [\[Slides\]](#) [\[Video\]](#) coming soon!



Limitations and New Frontiers

Lecture 7

Jan. 12, 2023

[\[Slides\]](#) [\[Video\]](#) coming soon!



Robot Learning

Lecture 9

Jan. 13, 2023

[\[Info\]](#) [\[Slides\]](#) [\[Video\]](#) coming soon!



Deep Sequence Modeling

Lecture 2

Jan. 9, 2023

[\[Slides\]](#) [\[Video\]](#) coming soon!



Deep Generative Modeling

Lecture 4

Jan. 10, 2023

[\[Slides\]](#) [\[Video\]](#) coming soon!



Deep Reinforcement Learning

Lecture 6

Jan. 11, 2023

[\[Slides\]](#) [\[Video\]](#) coming soon!



The Modern Era of Statistics

Lecture 8

Jan. 12, 2023

[\[Info\]](#) [\[Slides\]](#) [\[Video\]](#) coming soon!



Text-to-Image Generation

Lecture 10

Jan. 13, 2023

[\[Info\]](#) [\[Slides\]](#) [\[Video\]](#) coming soon!



Intro to TensorFlow; Music Generation

Software Lab 1

[\[Code\]](#)



Facial Detection Systems

Software Lab 2

[\[Paper\]](#) [\[Code\]](#)



Debiasing, Uncertainty, and Robustness

Software Lab 3

[\[Code\]](#)



Final Project

Work on final projects



Project Competition

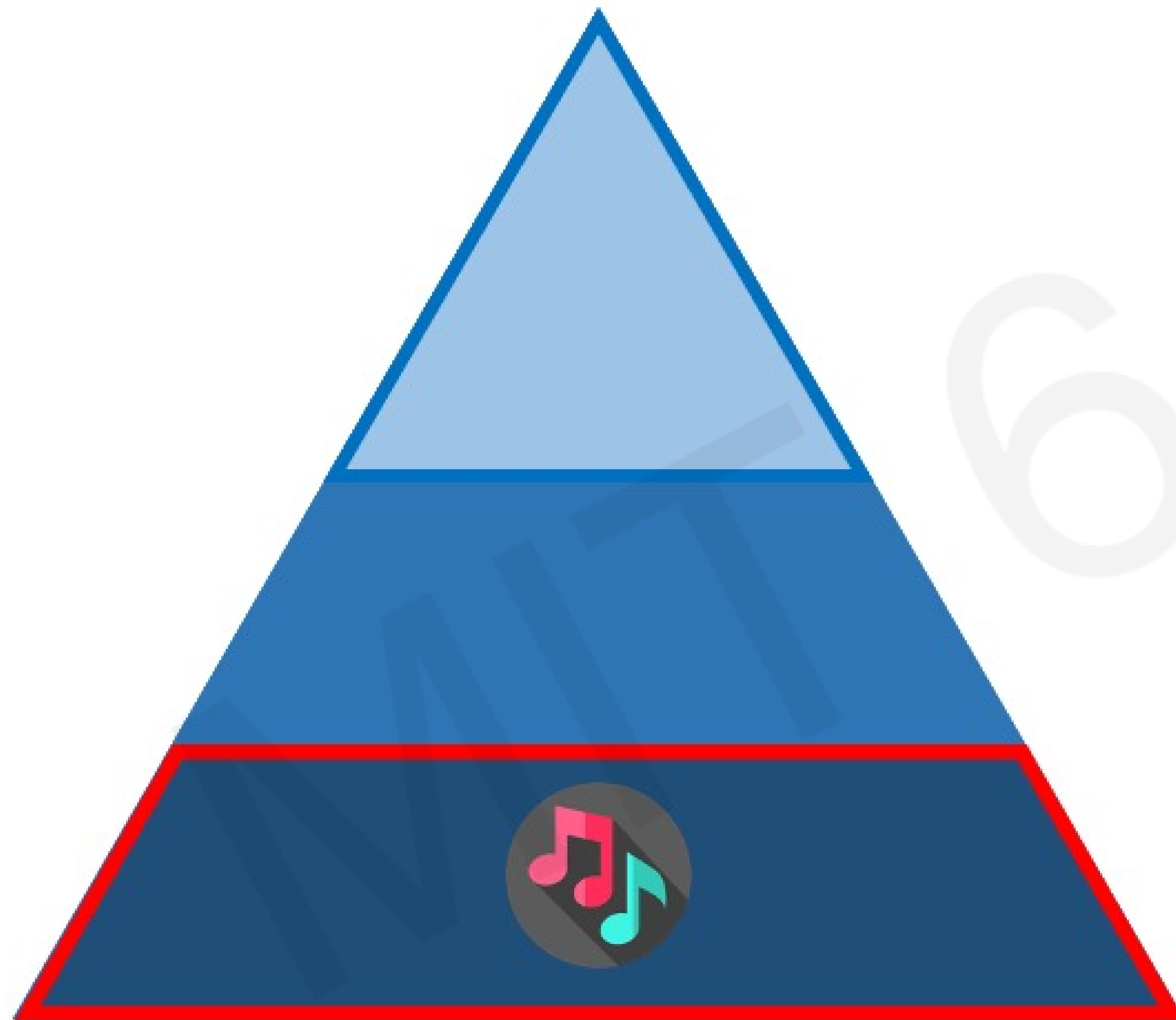
Project pitches and final awards!



- 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

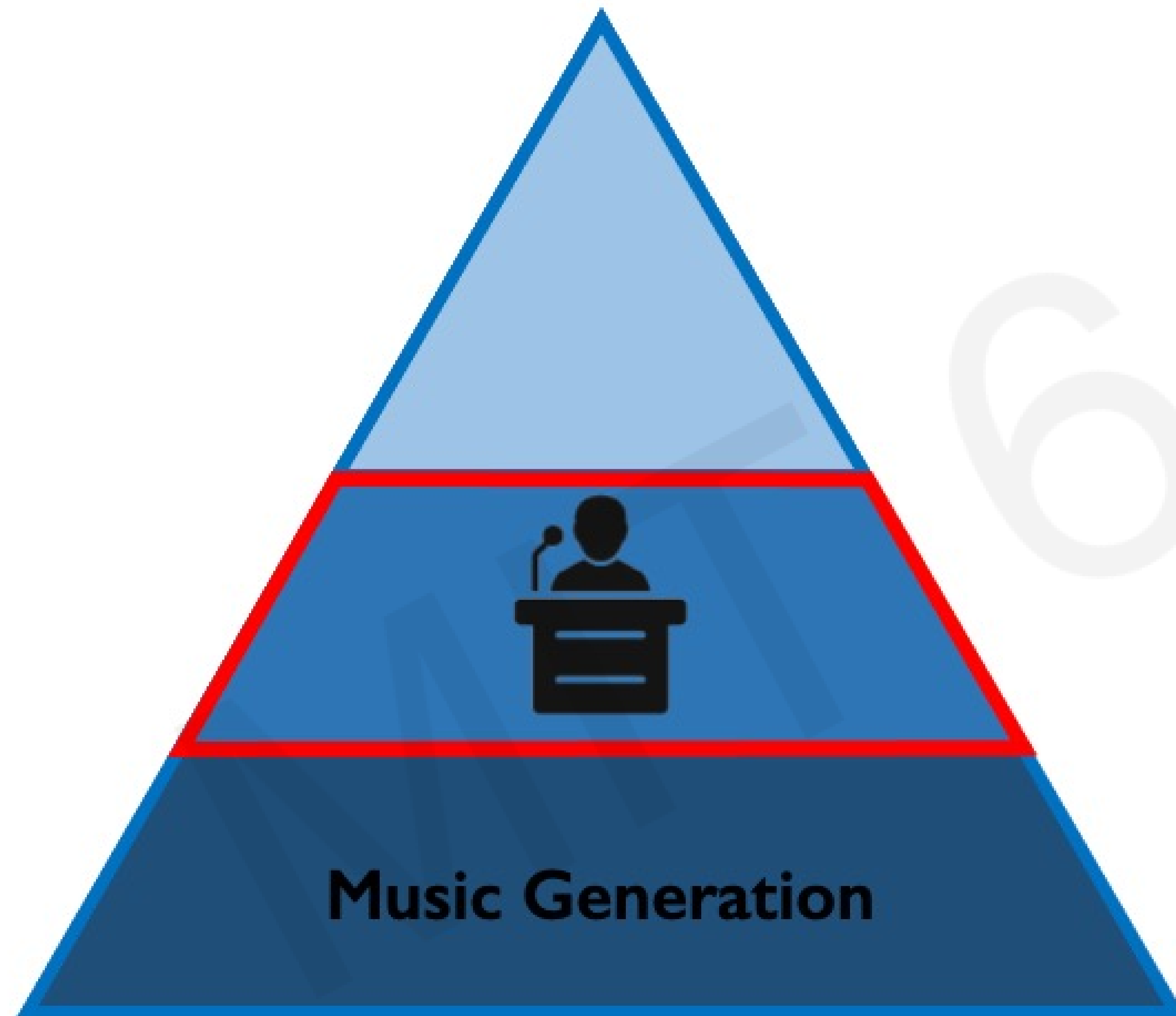
Build a neural network that can learn the genre of Irish folk songs and use it to generate brand new songs!

Prize:



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



Project Pitch
Competition

Music Generation

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...

The Rise of Deep Learning

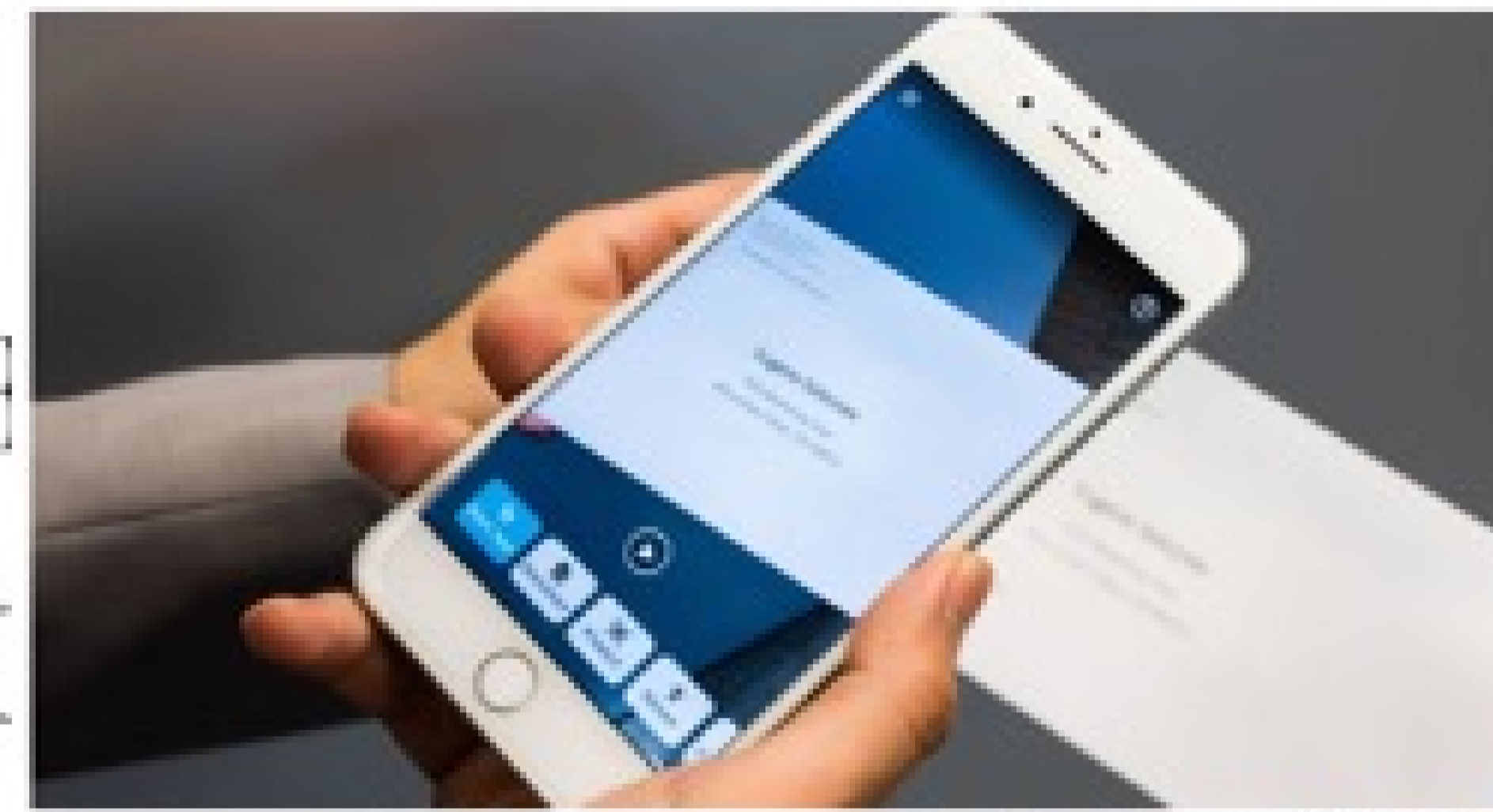
'Deep Voice' Software Can Clone Anyone's Voice With Just 3.7 Seconds of Audio

Using snippets of voices, Baidu's 'Deep Voice' can generate new speech, accents, and tones.



with DEEPMIND'S STARCRAFT TRIUMPH

Let There Be Sight: How Deep Learning Is Helping the Blind 'See'



Technology outpacing security measures

| Facial Recognition | Features and Interviews

AI beats docs in cancer spotting

A new study provides a fresh example of machine learning as an important diagnostic tool. Paul Biegler reports.

AI Can Help In Predicting Cryptocurrency Value



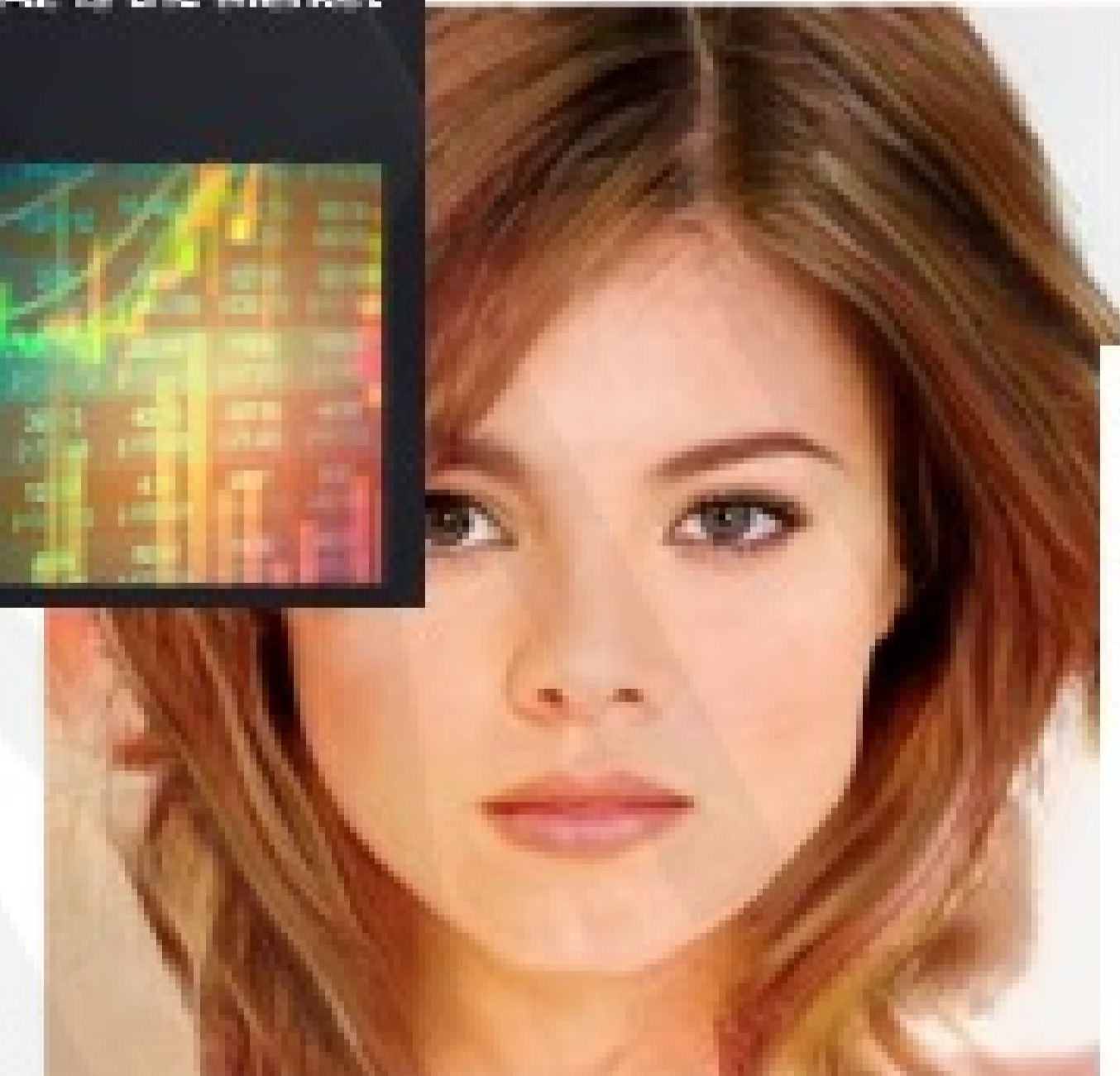
'Creative' AlphaZero leads way for chess computers and, maybe, science

Former chess world champion Garry Kasparov likes what he sees of computer that could be used to find cures for diseases



How an A.I. 'Cat-and-Mouse Game' Generates Believable Fake Photos

By CHRIS MITCHELL and KIRSTY COLLINS | JAN. 4, 2018



Stock Predictions Based On AI: Is the Market Truly Predictable?



Complex of bacteria-infecting viral proteins modeled in CASP 13. The complex contains that were modeled individually. [View on CASP 13](#)

Google's DeepMind aces protein folding

By Robert F. Service | Dec. 6, 2018, 12:05 PM



AI, Faked Data

MIT researchers have developed a special-purpose chip that increases the speed of neural network computations by three to seven times over its predecessors, while reducing power consumption 43 to 56 percent. This could make it practical to run neural networks locally on smartphones or even to embed them in household appliances.



Neural networks everywhere

New chip reduces neural networks' power consumption by up to 95 percent, making them practical for battery-powered devices.

Wed, 06/16/2016 - 8:00am | Comment | by Kerry Walter - Digital Reporter | [@FastCompany](#)

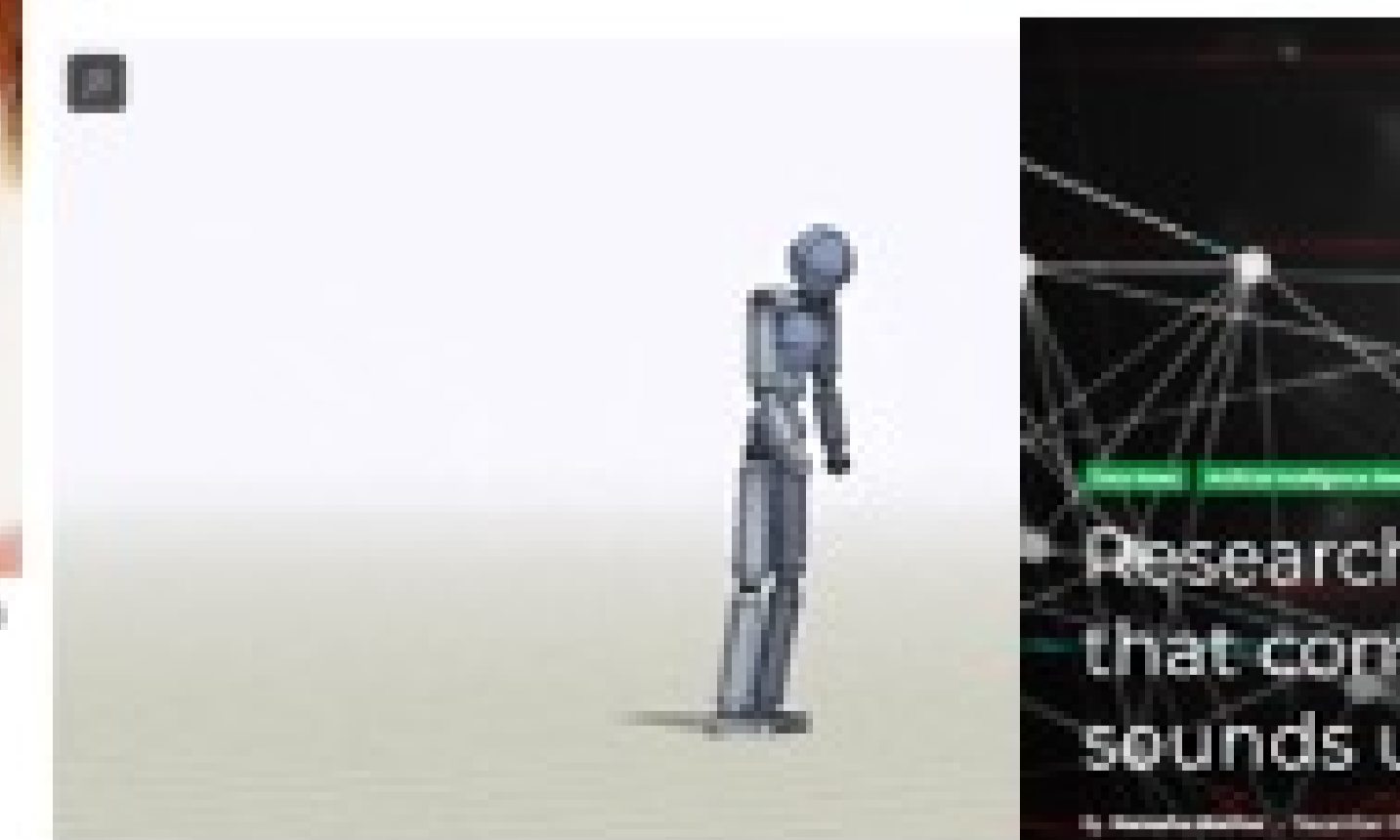
AI faces show how far AI image generation has needed in just four years

People on the right aren't real; they're the product of machine learning



After Millions of Trials, These Simulated Humans Learned to Do Perfect Backflips and Cartwheels

George Siu | 2/16/18 11:56am | Photo: MIT



Researchers introduce a deep learning method that converts mono audio recordings into 3D sounds using video scenes

Automation And Algorithms: De-Risking Manufacturing With Artificial Intelligence

Sarah Goehrie | Contributor | [@SarahGoehrie](#)
Focus on the industrialization of additive manufacturing.

TWEET THIS
The two key applications of AI in manufacturing are pricing and manufacturability feedback

So far in Introduction to Deep Learning...



Data

- Signals
- Images
- Sensors

...



Decision

- Prediction
- Detection
- Action

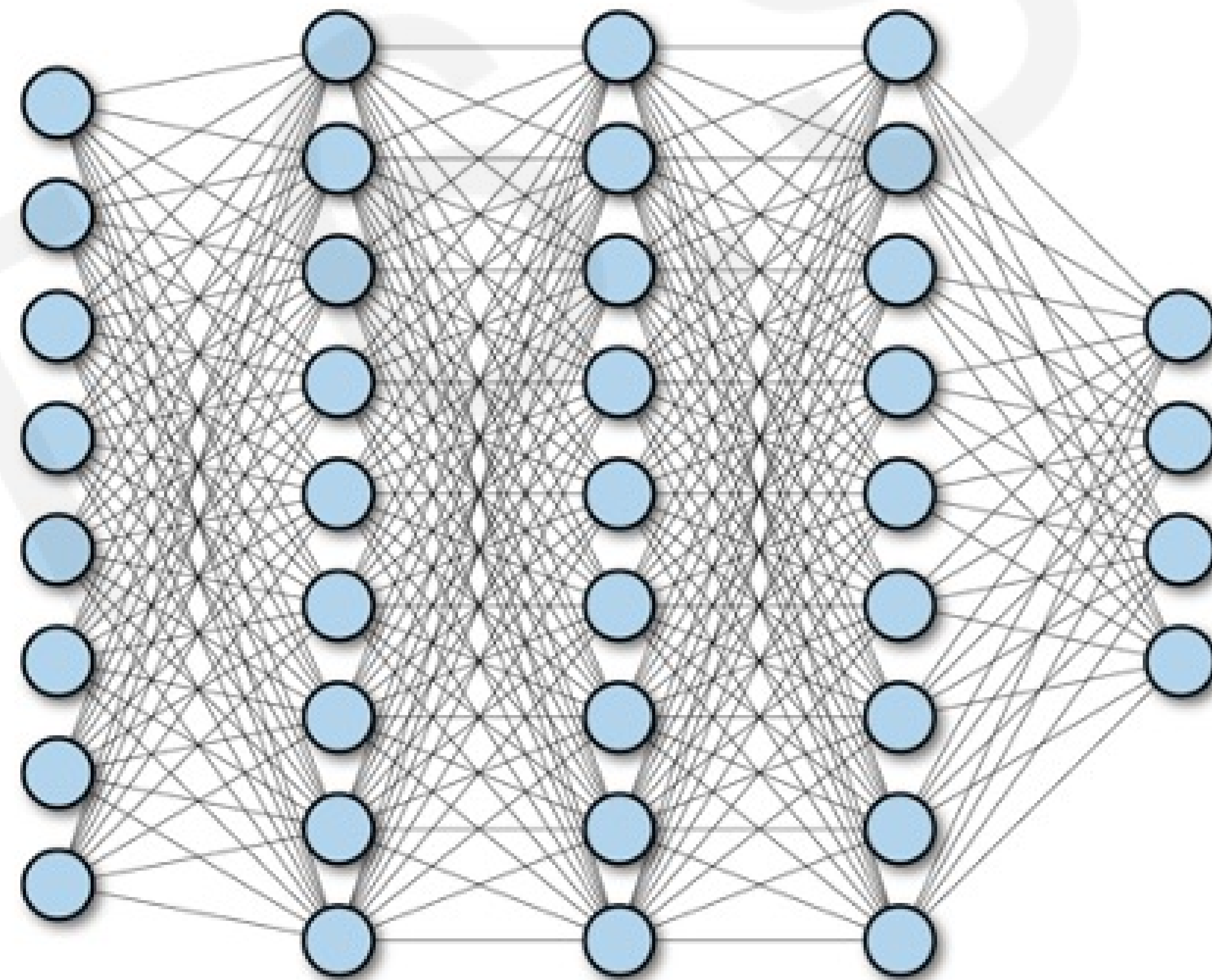
...



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.



Power of Neural Nets

Universal Approximation Theorem

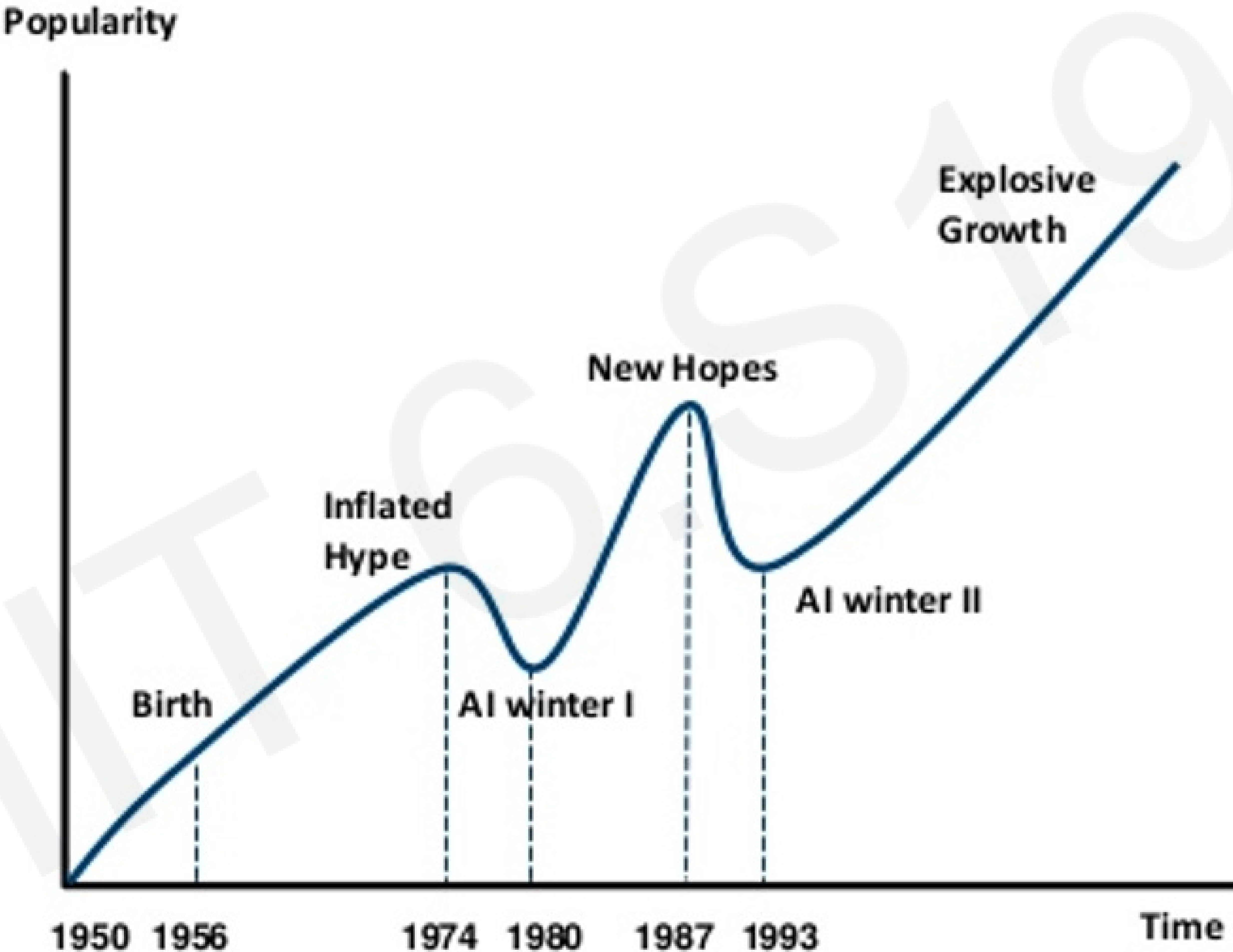
A feedforward network with a single layer is sufficient to approximate, to an arbitrary precision, any continuous function.

Caveats:

The number of hidden units may be infeasibly large

The resulting model may not generalize

Artificial Intelligence “Hype”: Historical Perspective



Limitations

Rethinking Generalization

“Understanding Deep Neural Networks Requires Rethinking Generalization”



dog



banana



dog



tree

Rethinking Generalization

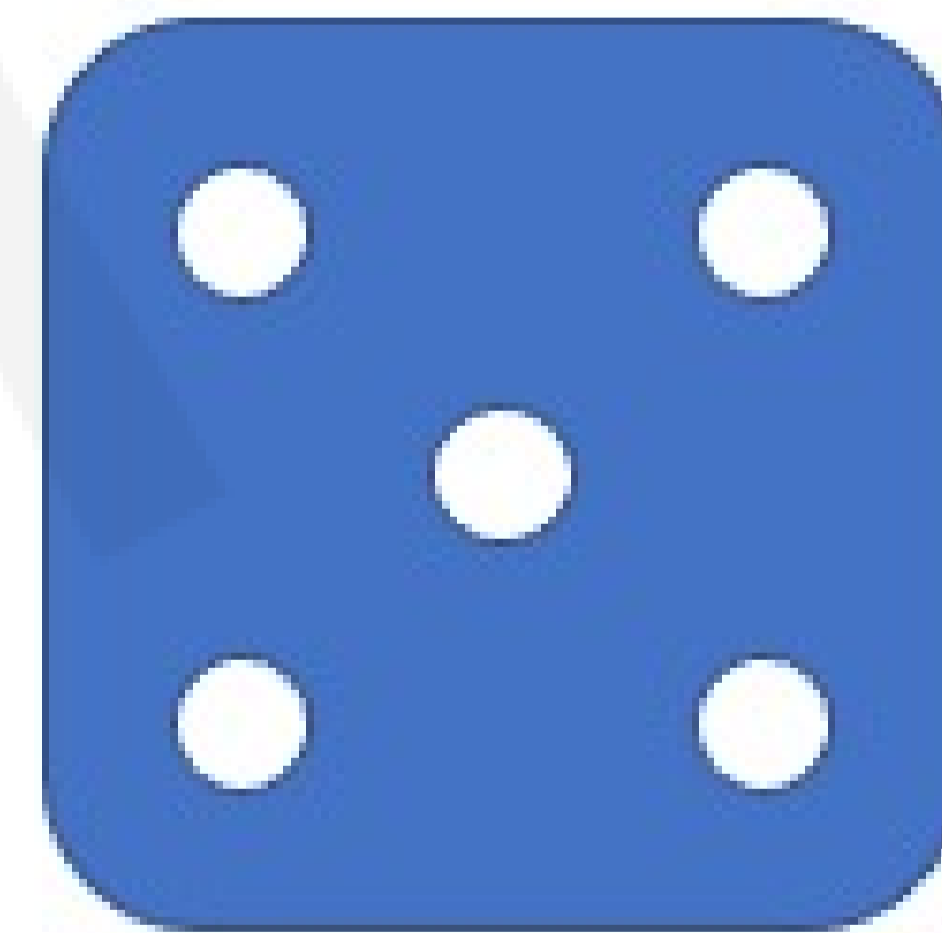
“Understanding Deep Neural Networks Requires Rethinking Generalization”



dog



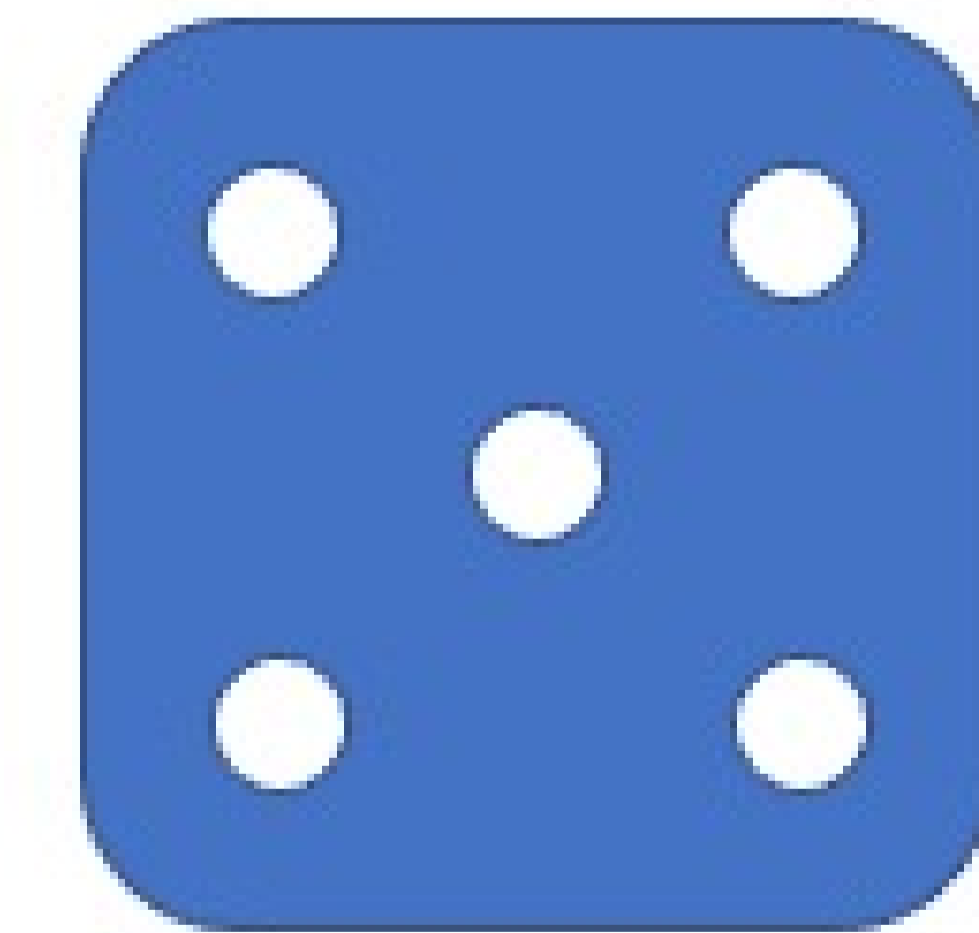
banana



dog



tree



Rethinking Generalization

“Understanding Deep Neural Networks Requires Rethinking Generalization”



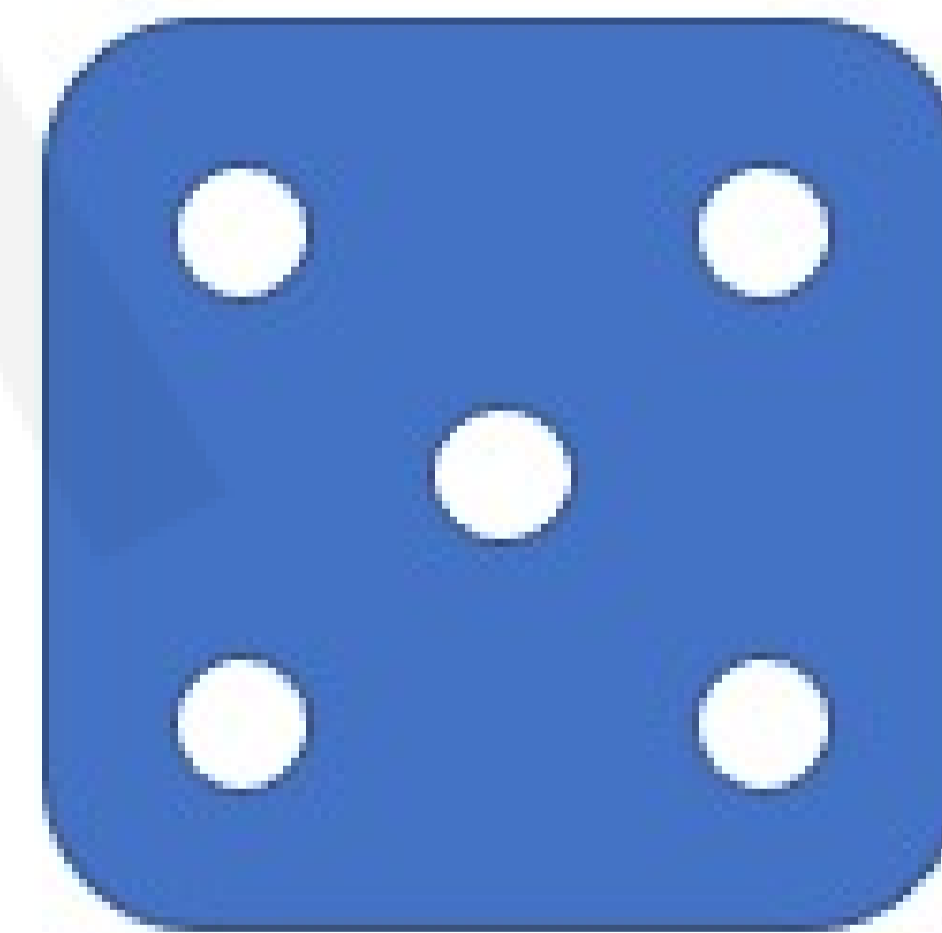
dog



banana



banana



dog



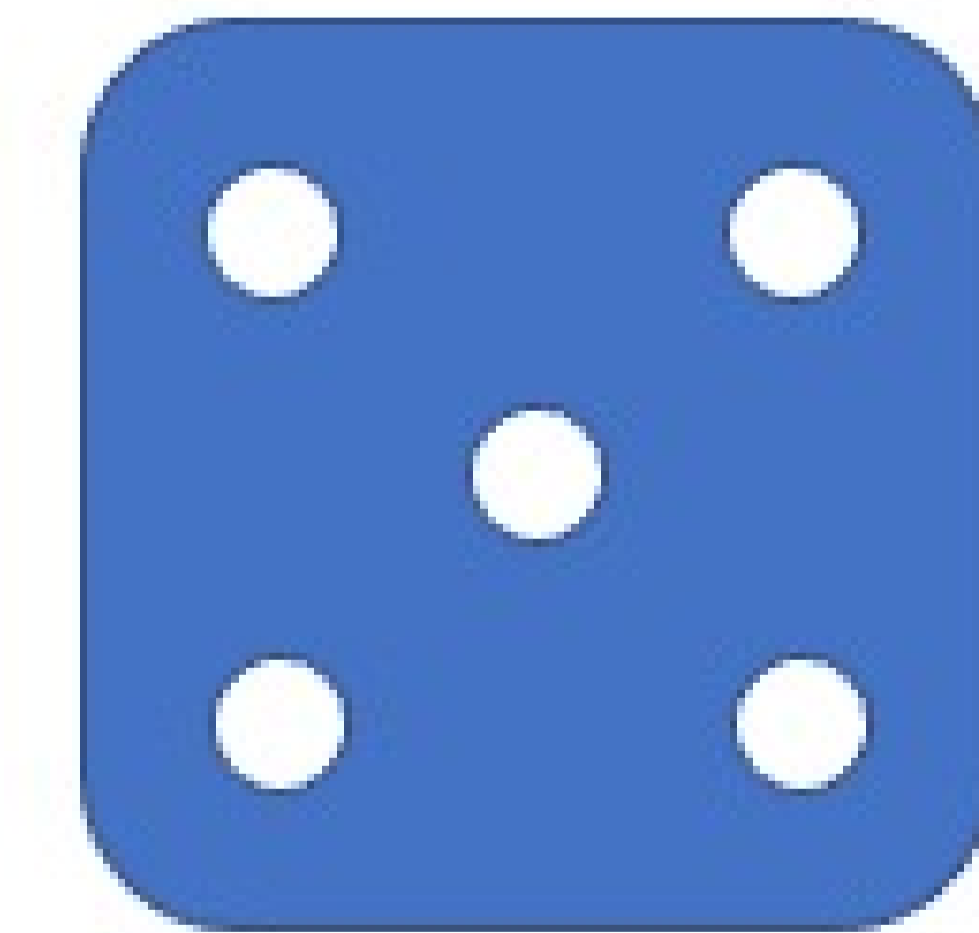
dog



tree



tree



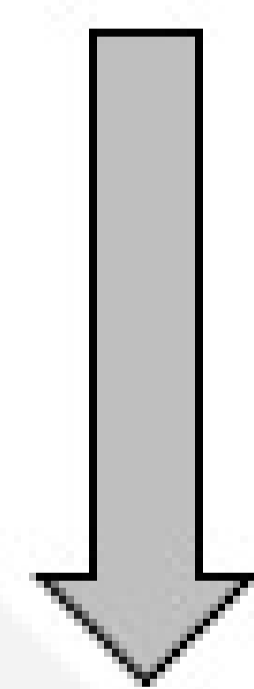
dog

Rethinking Generalization

“Understanding Deep Neural Networks Requires Rethinking Generalization”



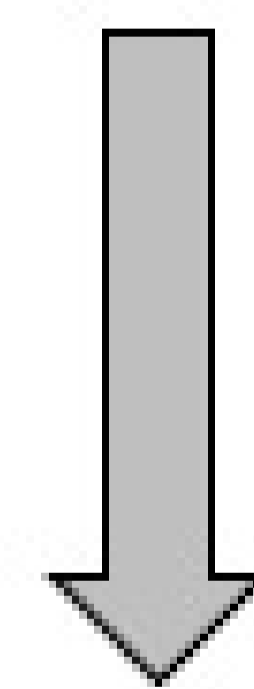
~~dog~~



banana



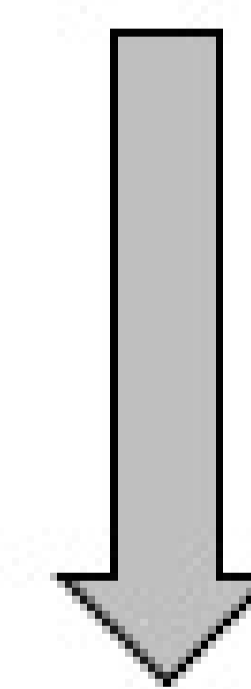
~~banana~~



dog



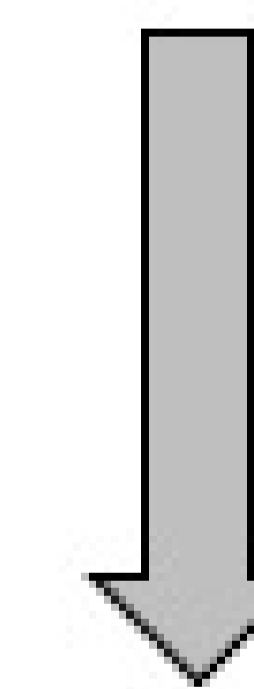
~~dog~~



tree



~~tree~~



dog

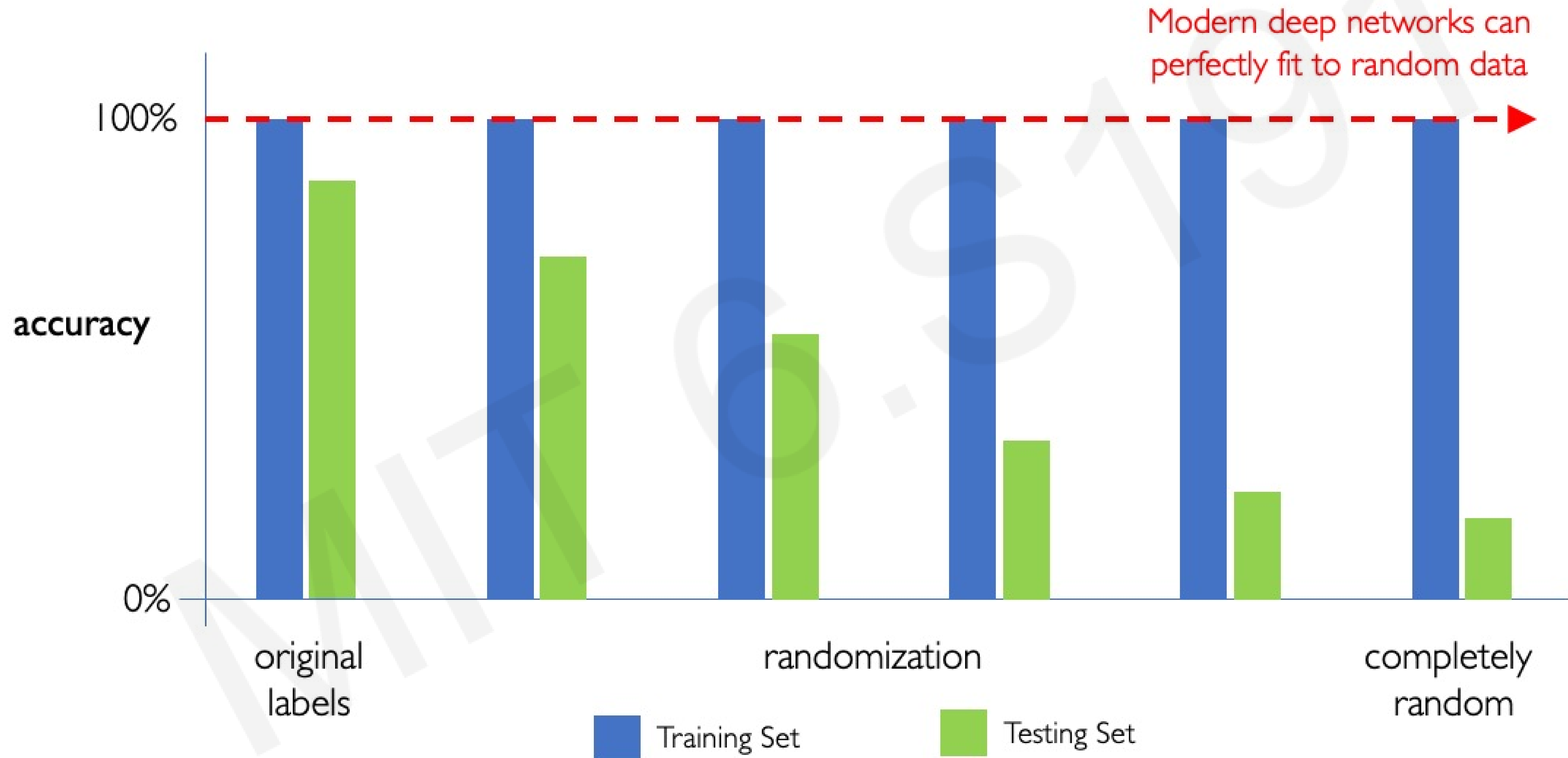
Capacity of Deep Neural Networks



Capacity of Deep Neural Networks

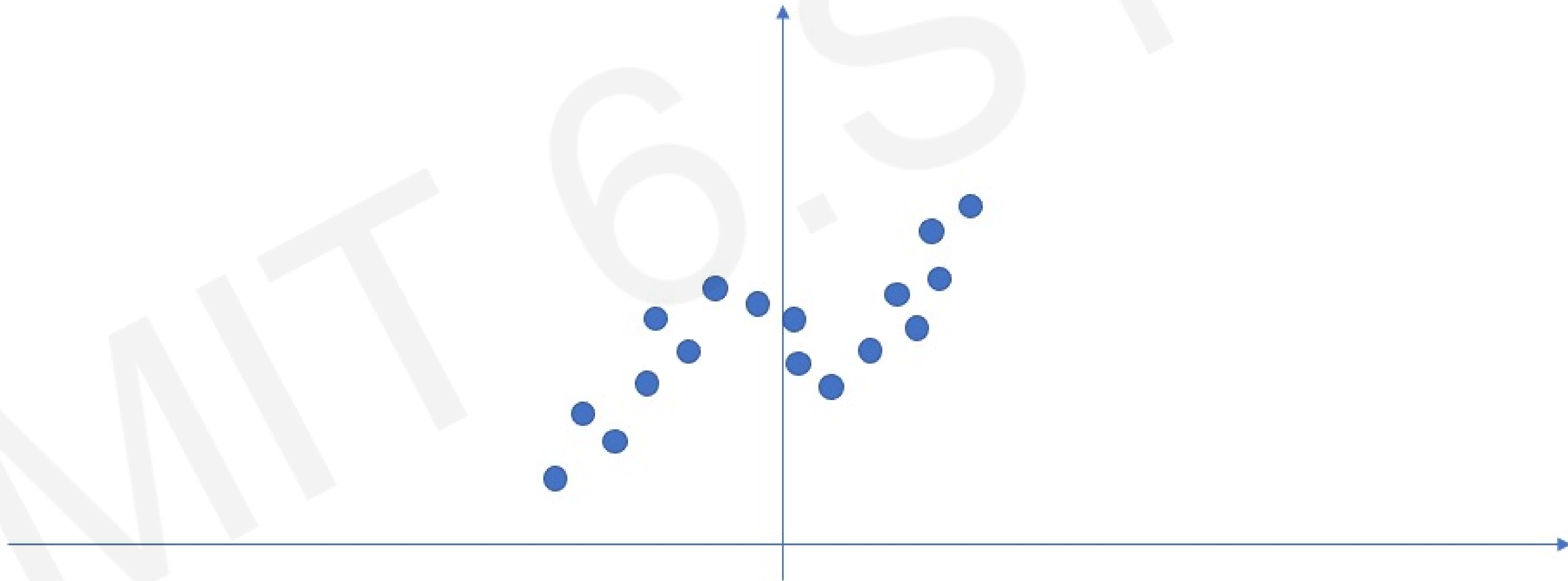


Capacity of Deep Neural Networks



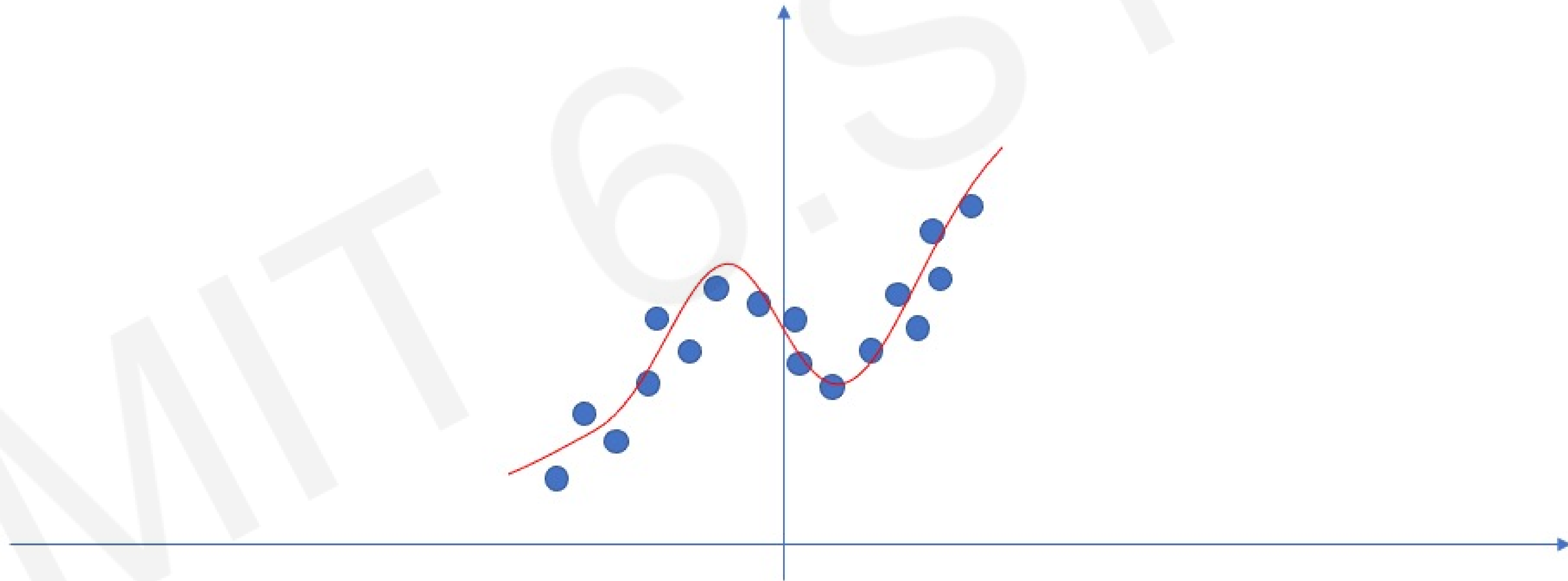
Neural Networks as Function Approximators

Neural networks are excellent function approximators



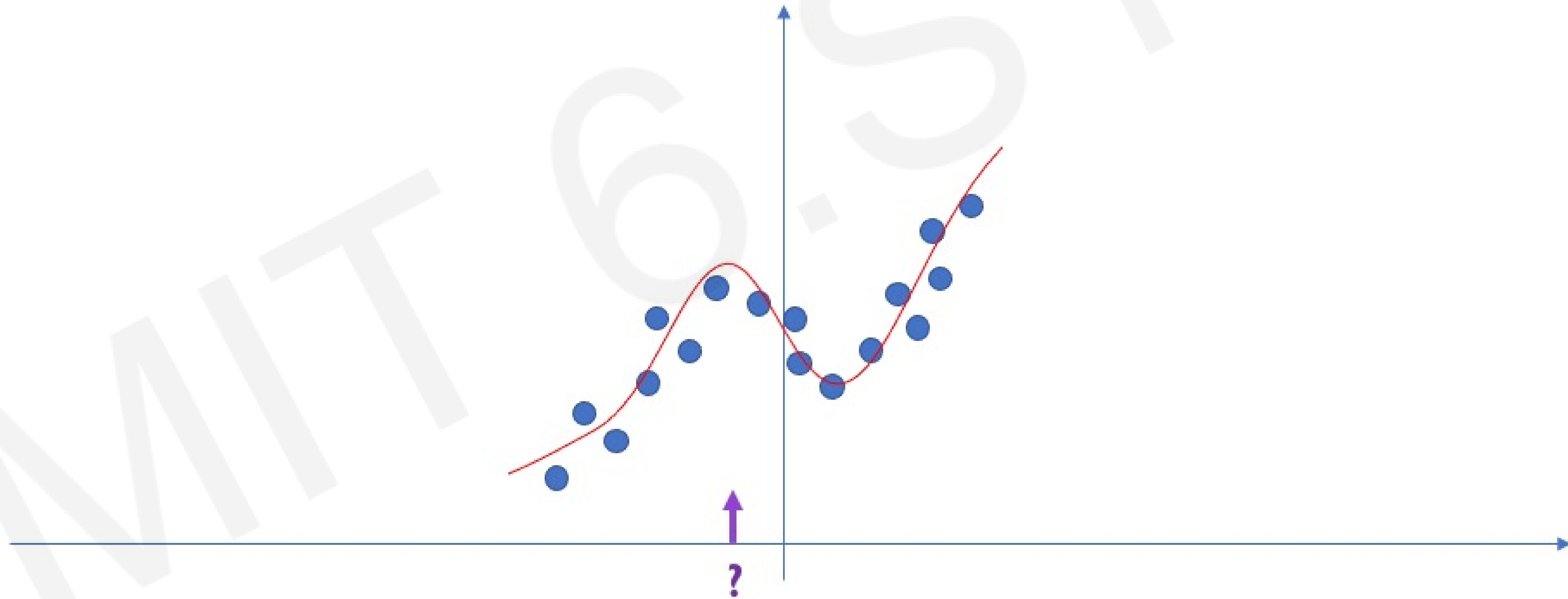
Neural Networks as Function Approximators

Neural networks are excellent function approximators



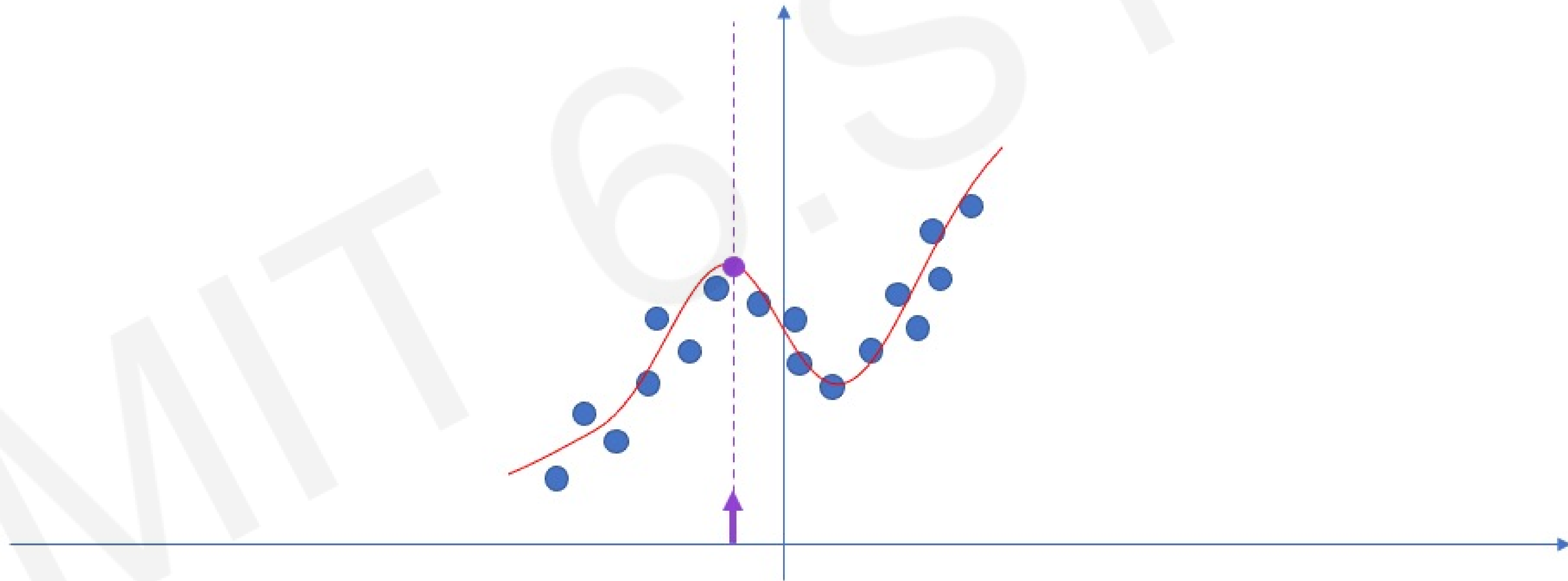
Neural Networks as Function Approximators

Neural networks are excellent function approximators



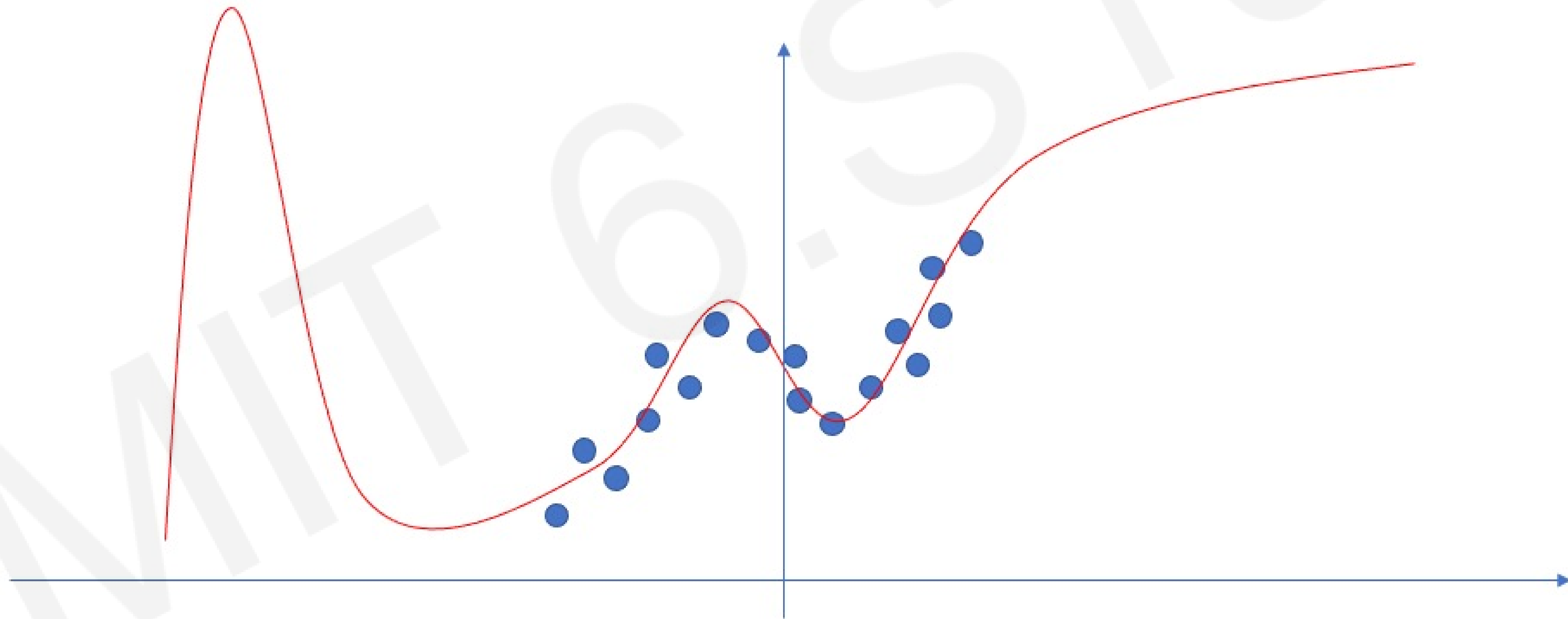
Neural Networks as Function Approximators

Neural networks are excellent function approximators



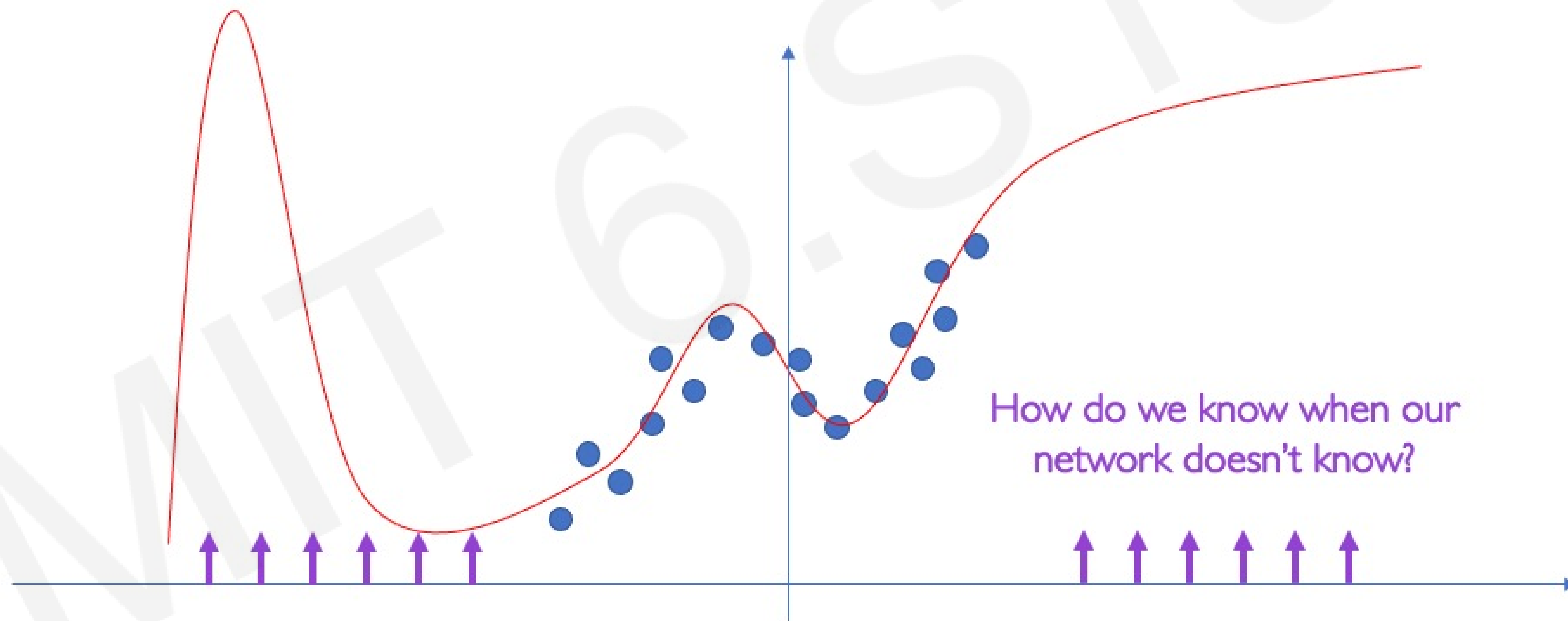
Neural Networks as Function Approximators

Neural networks are excellent function approximators

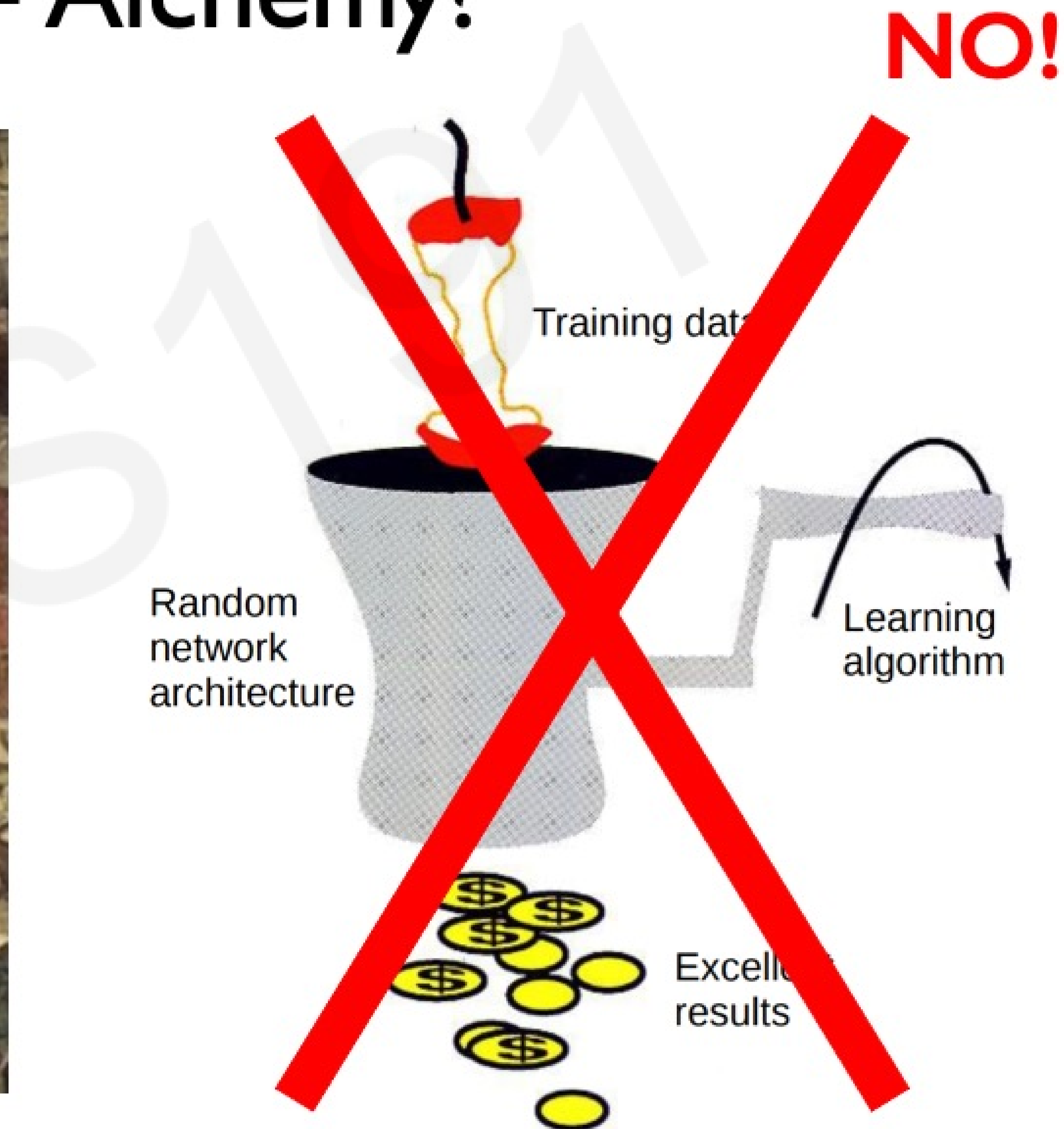
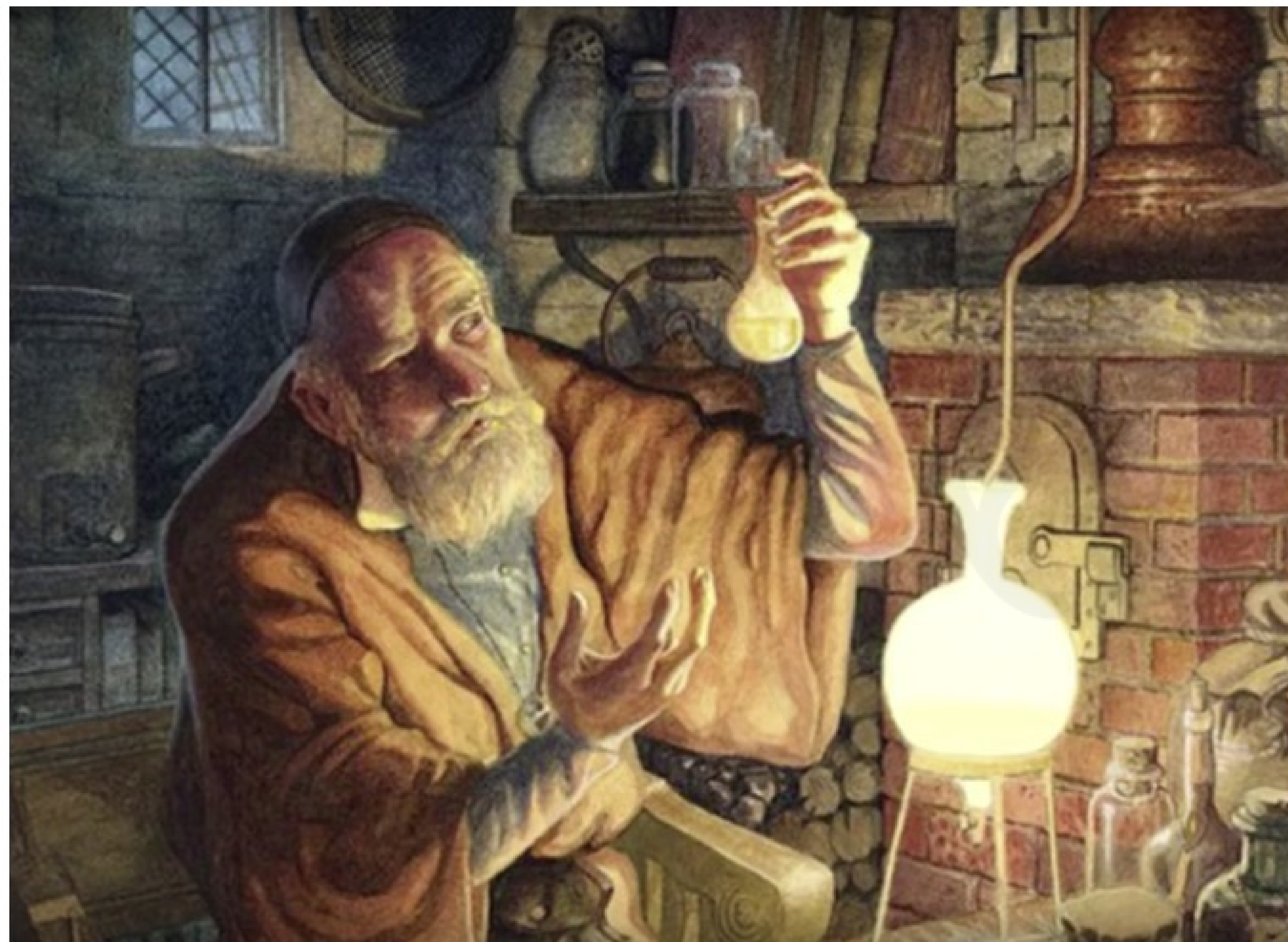


Neural Networks as Function Approximators

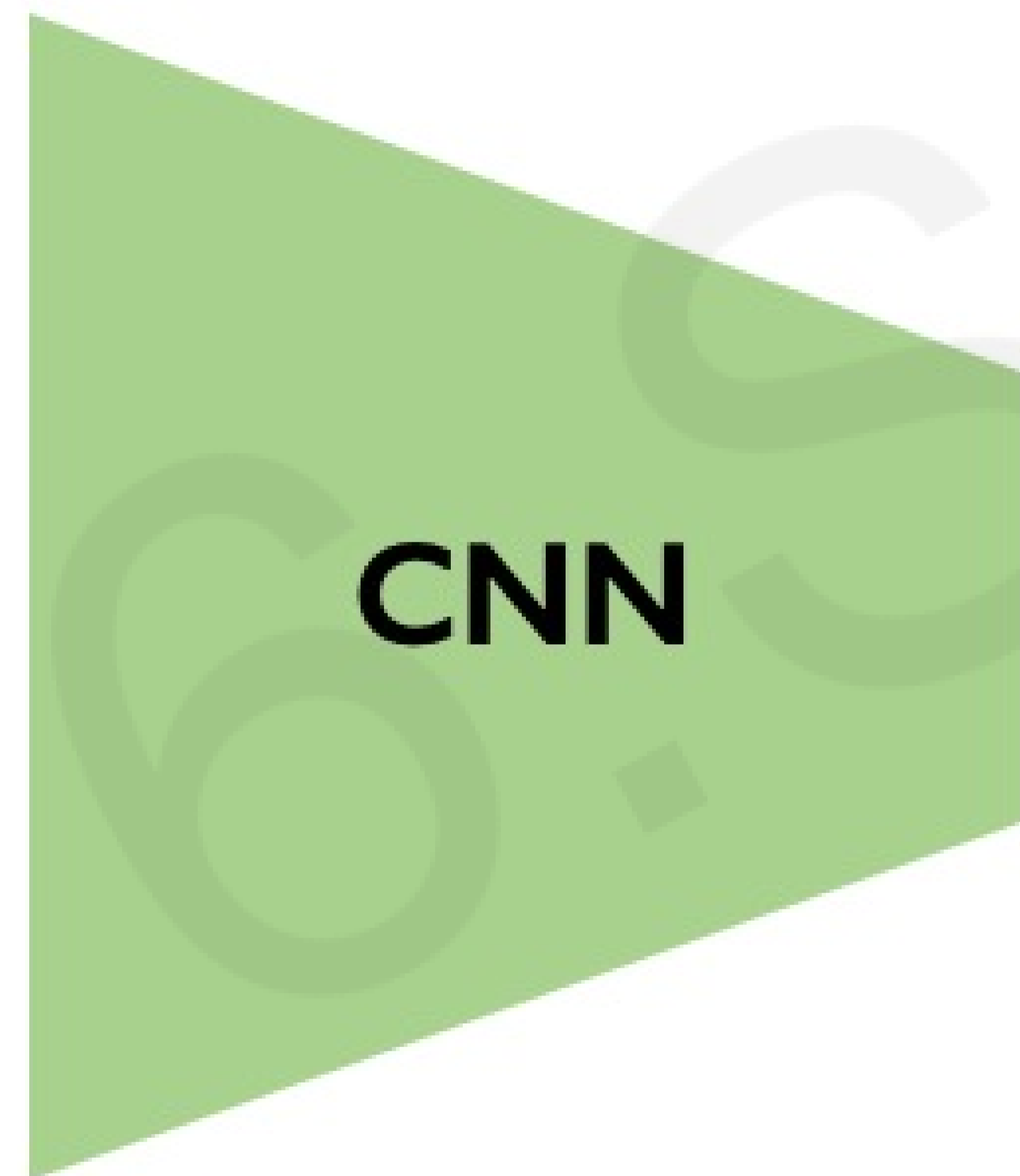
Neural networks are excellent function approximators
...when they have training data



Deep Learning = Alchemy?



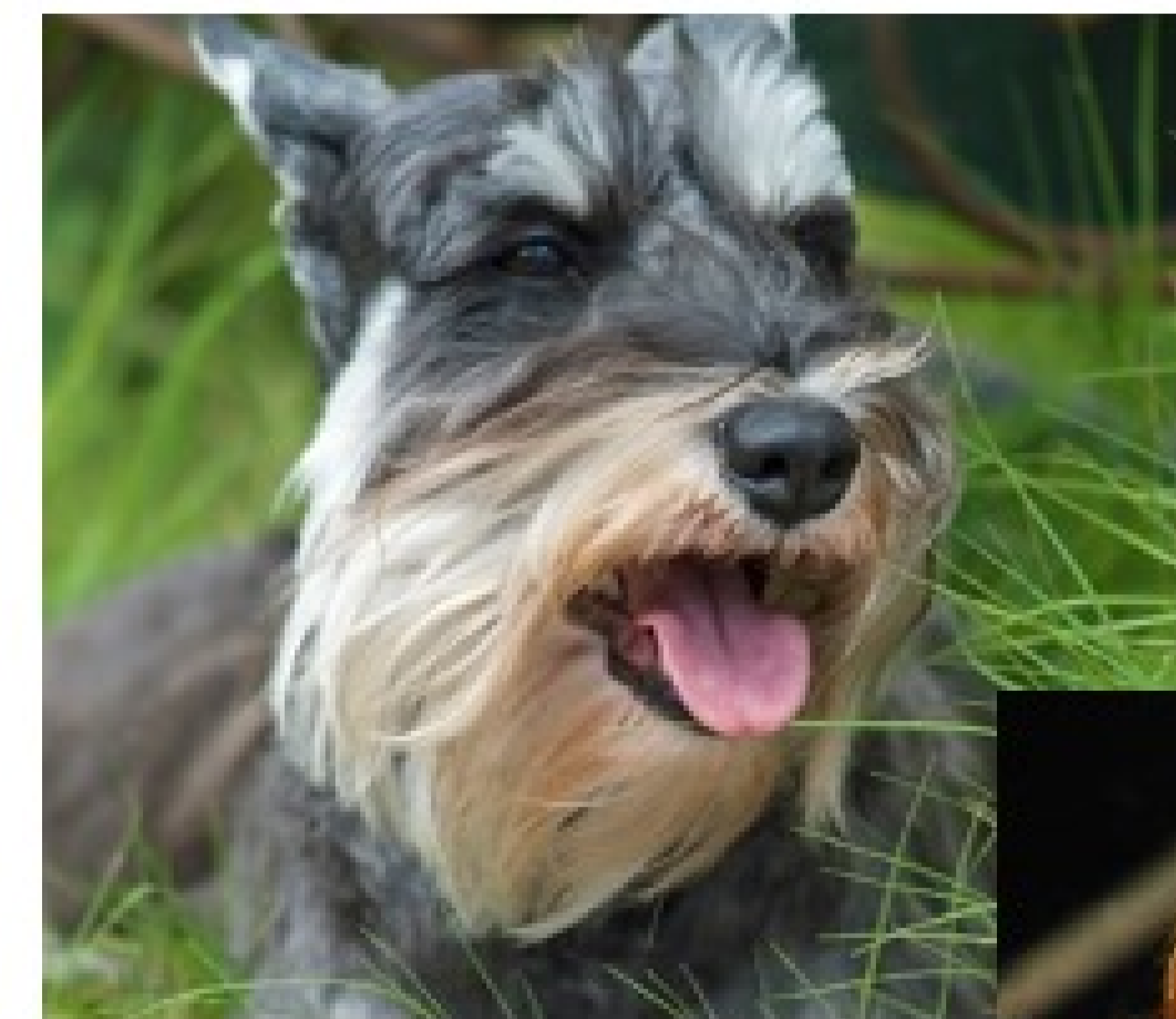
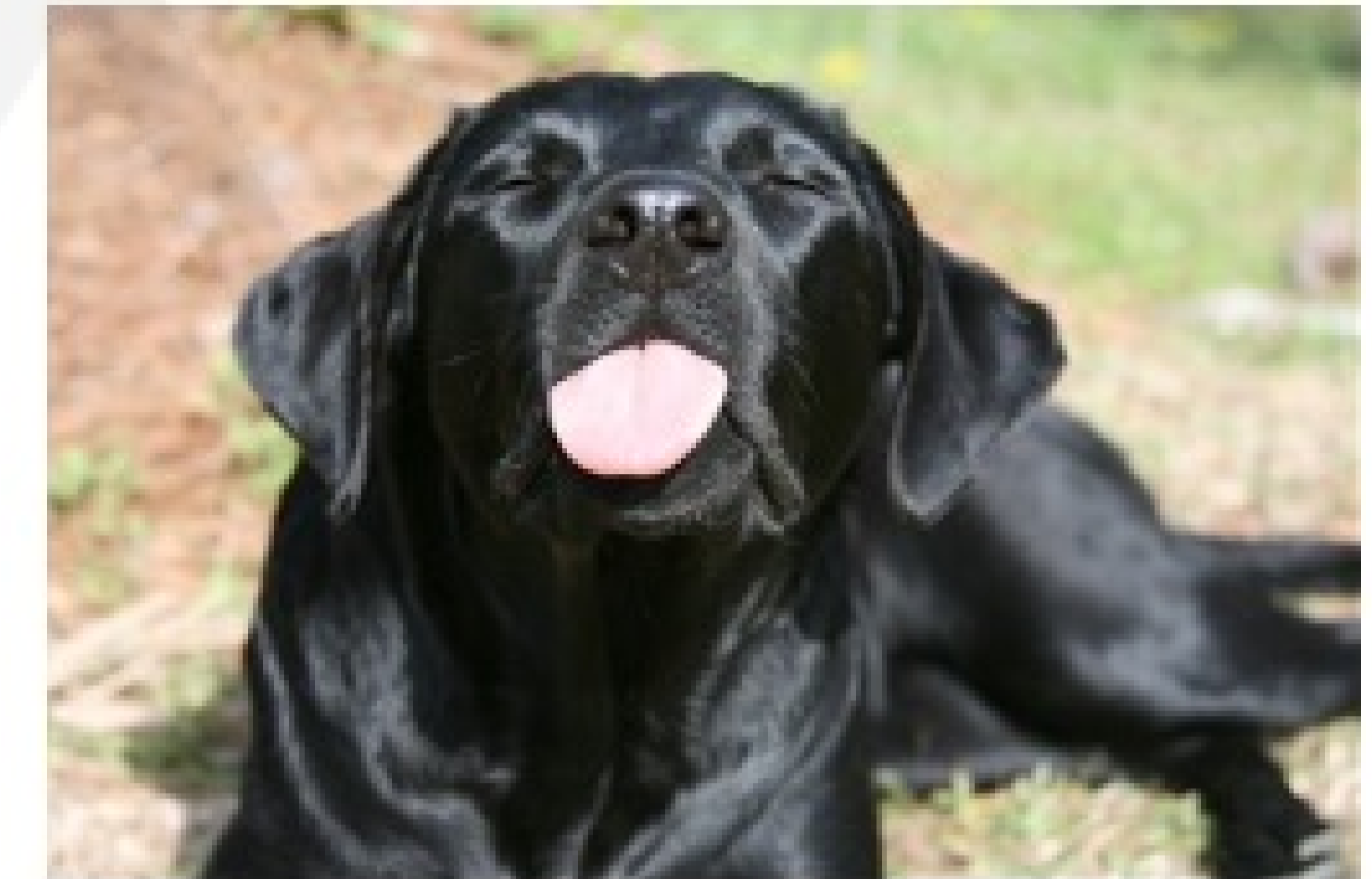
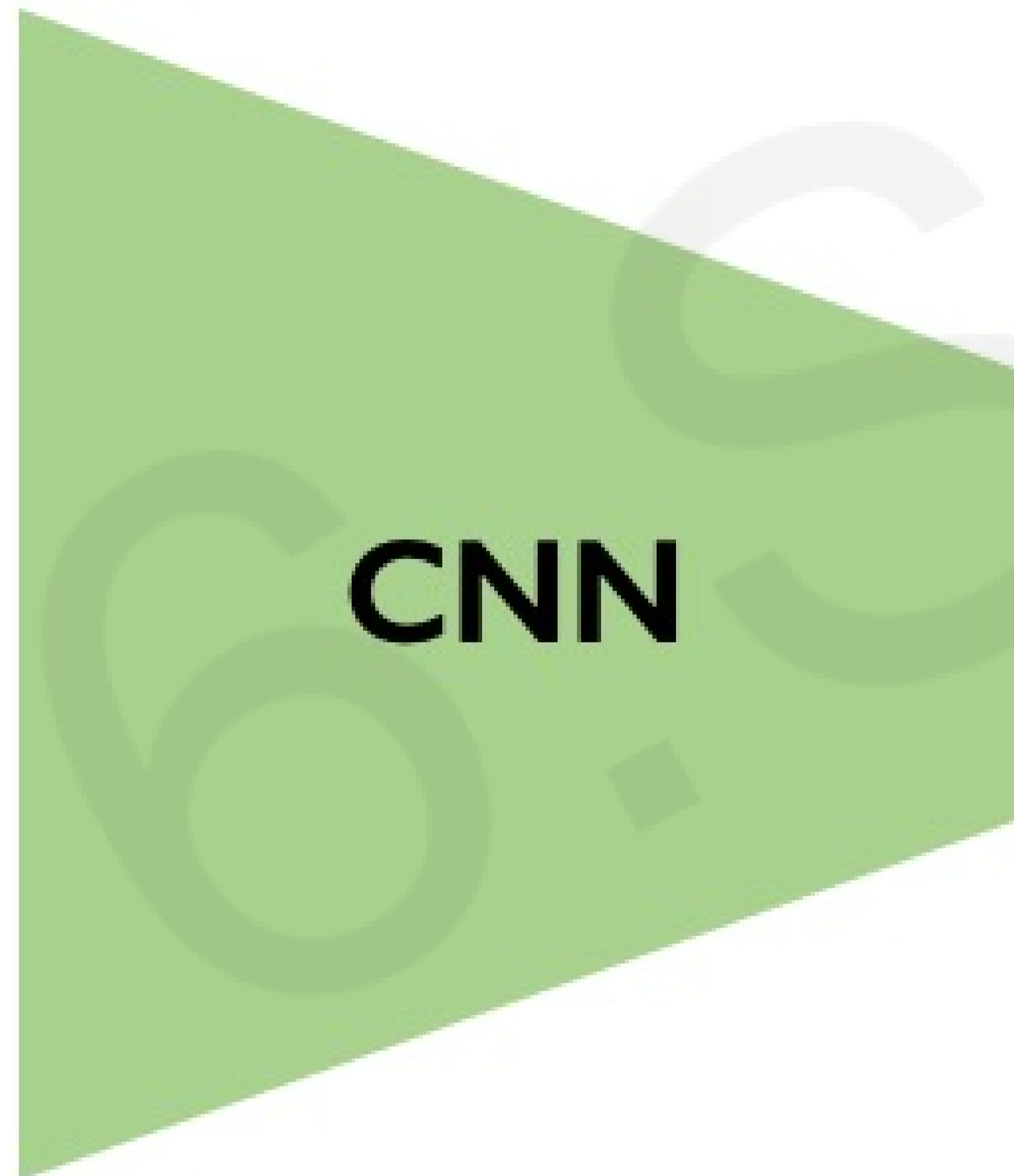
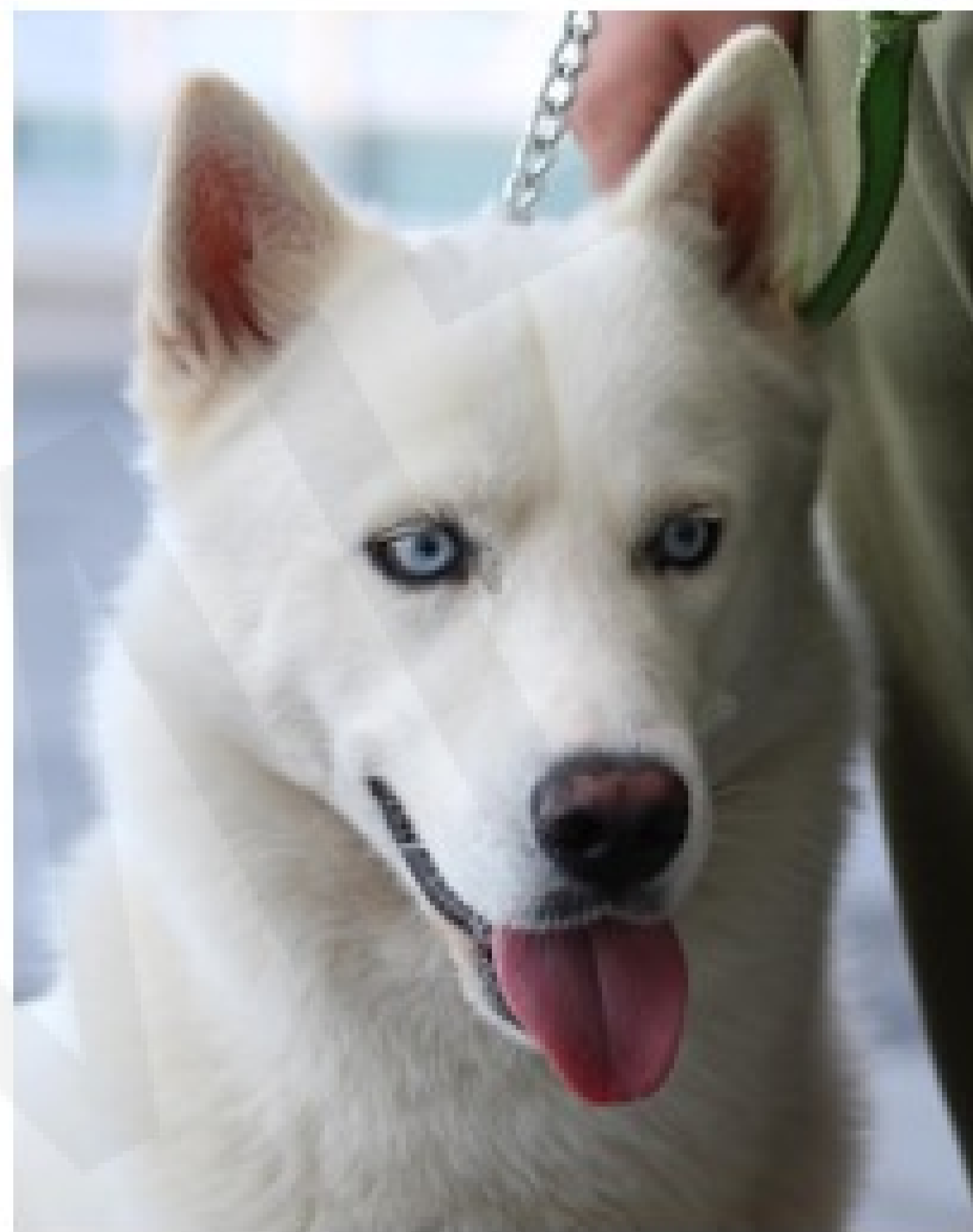
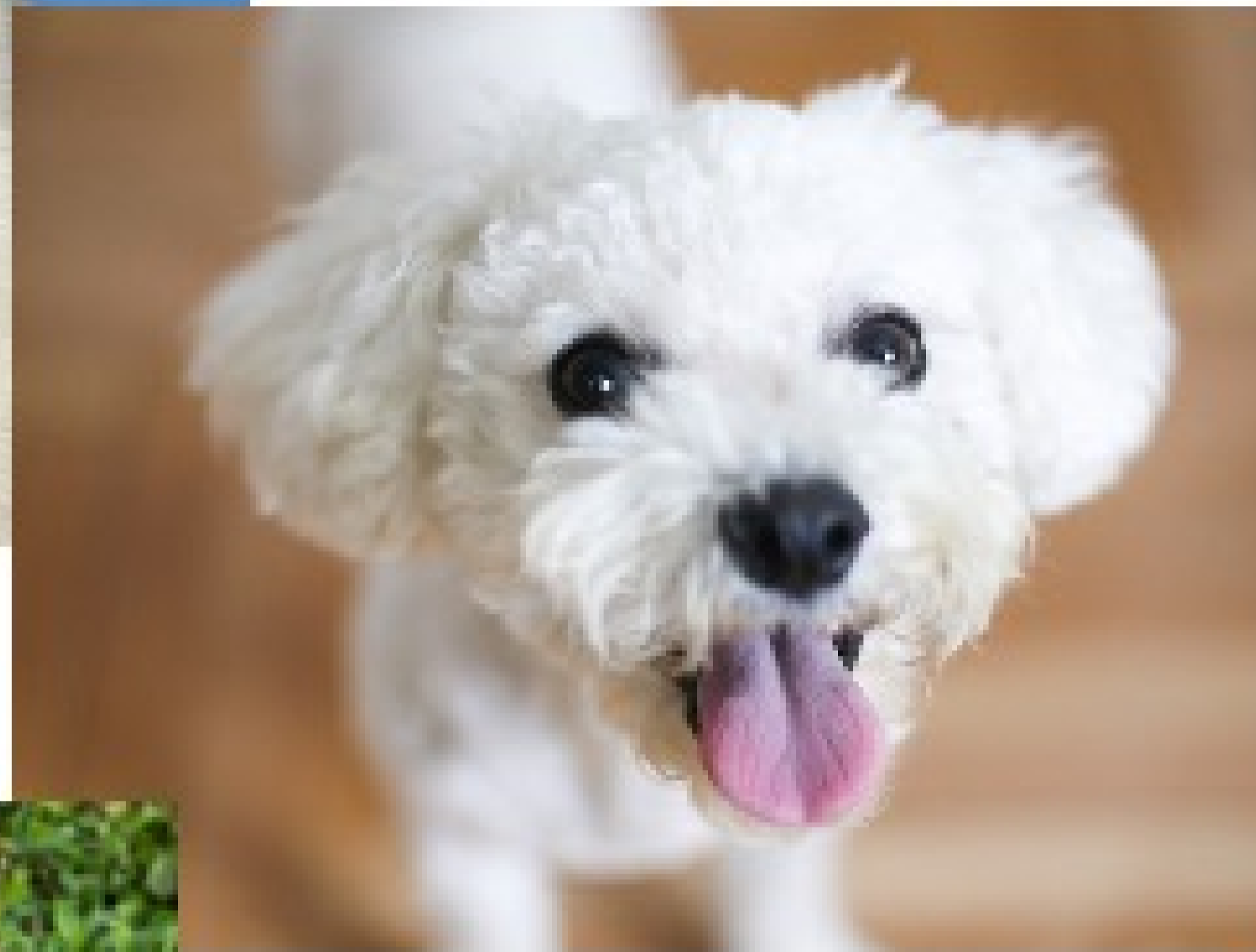
Neural Network Failure Modes, Part I



Train network to
colorize BW images.

Why could this be the case?

What Happens During Training...



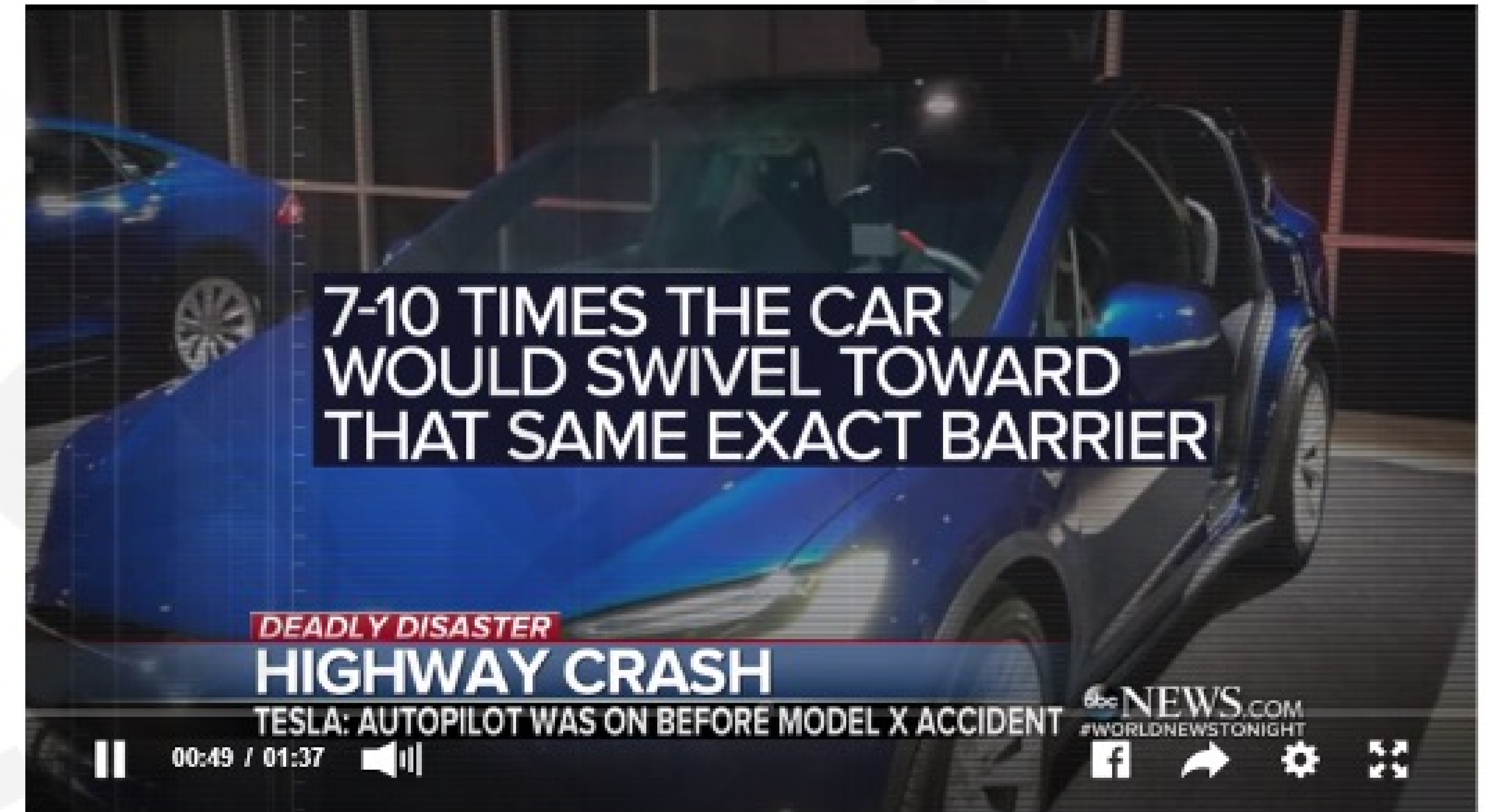
Neural Network Failure Modes, Part II

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



Uncertainty in Deep Learning

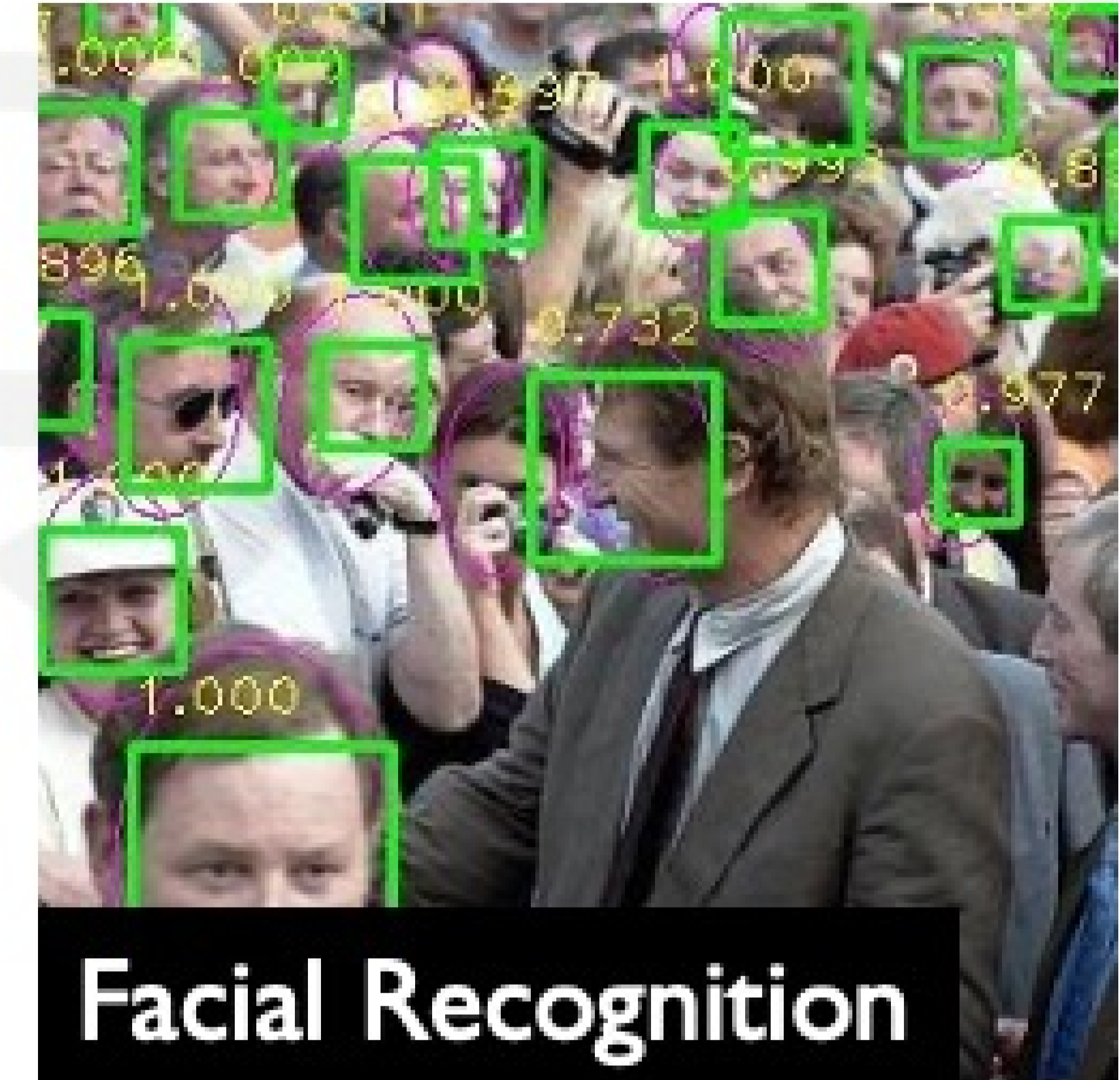
Safety-critical applications



Autonomous Vehicles

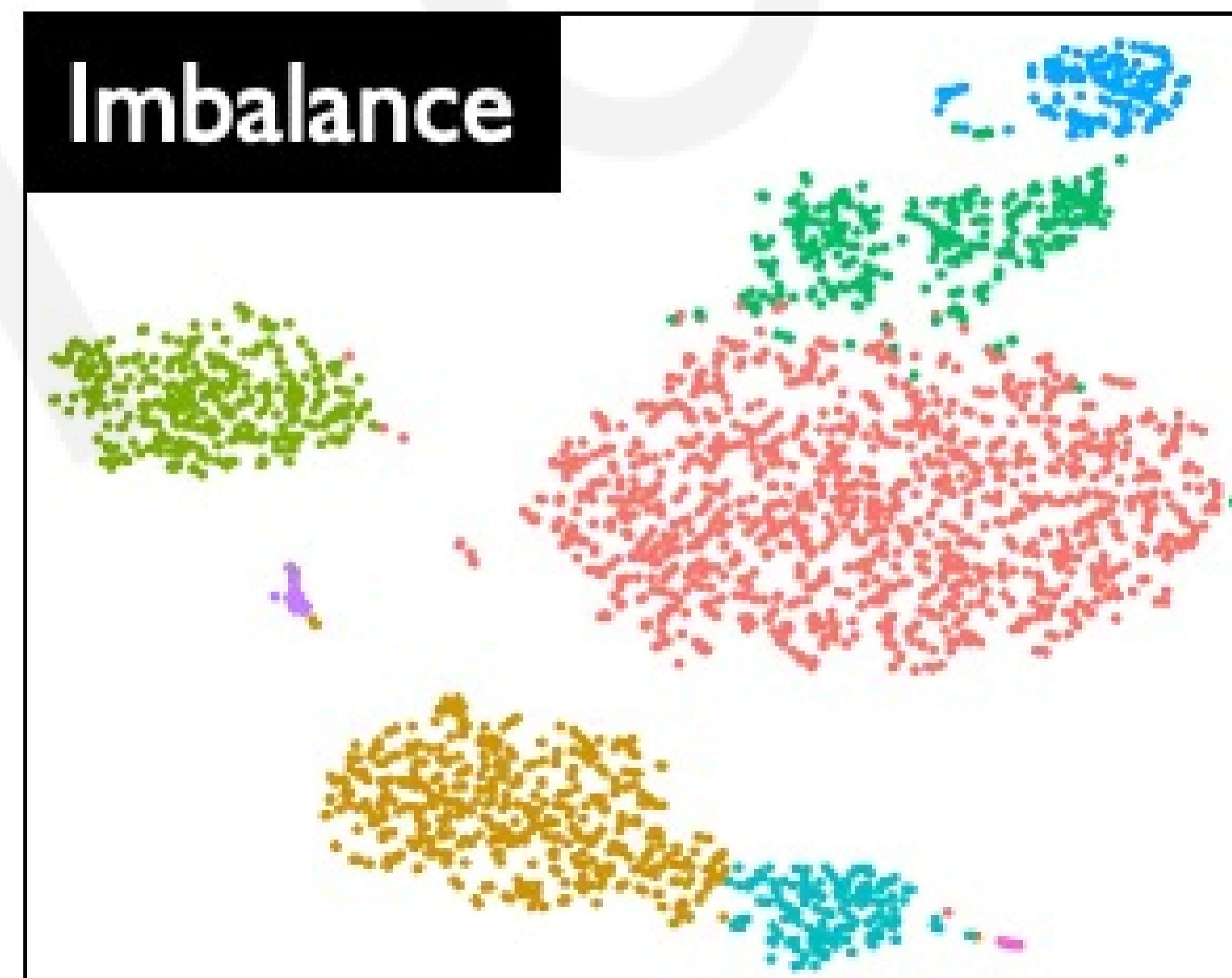


Medicine



Facial Recognition

Sparse and/or noisy datasets



Imbalance

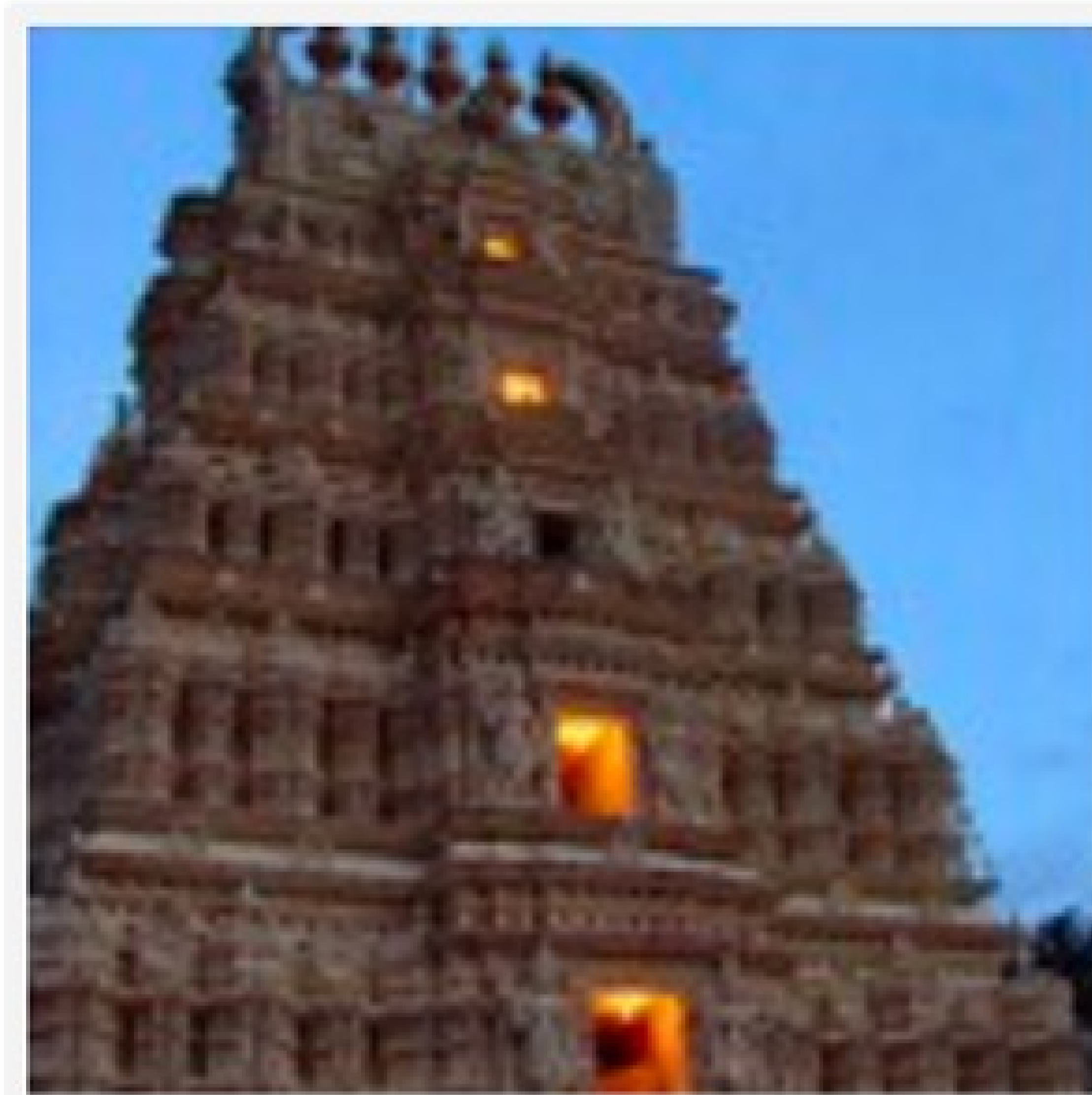


Data Noise



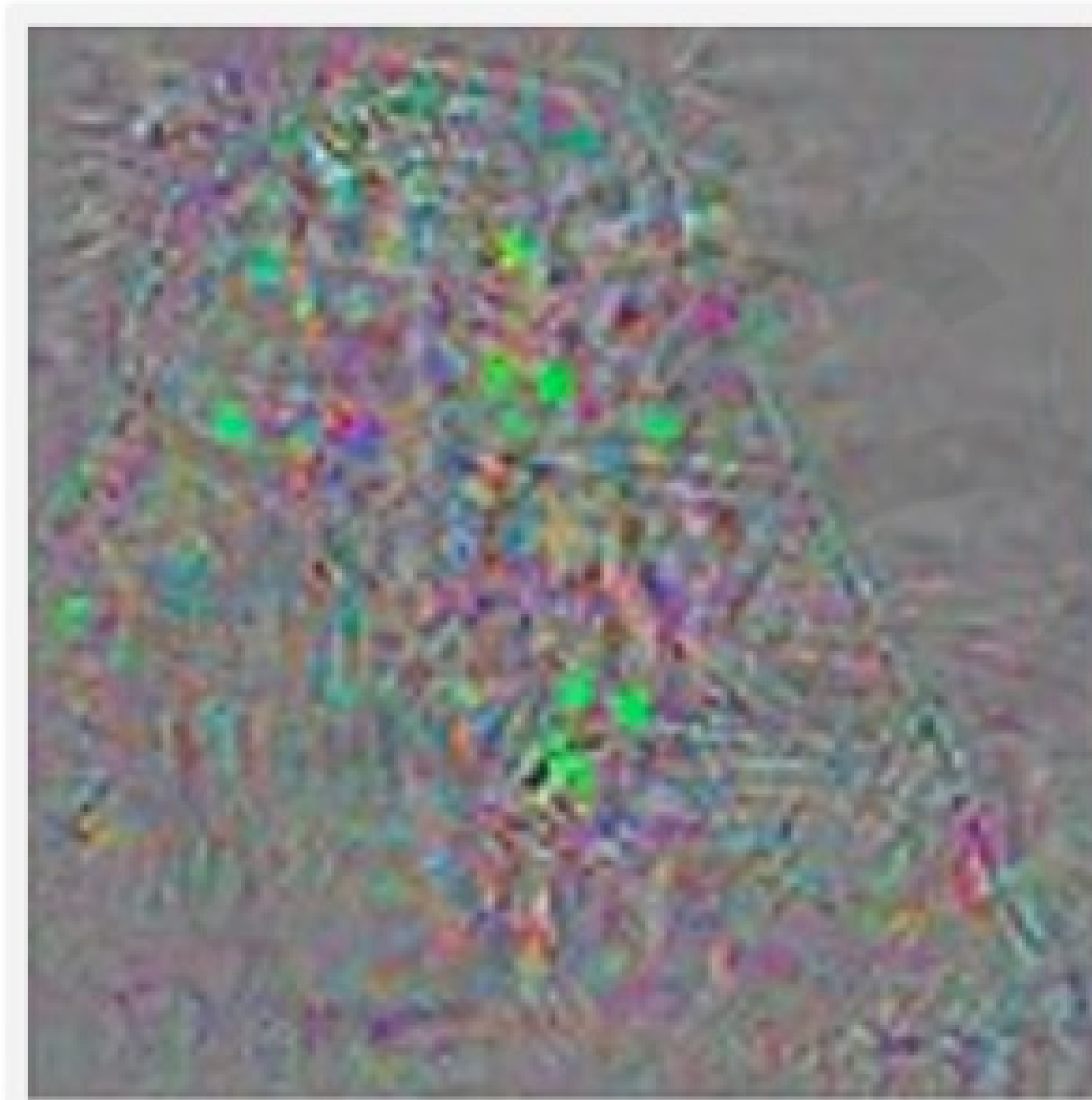
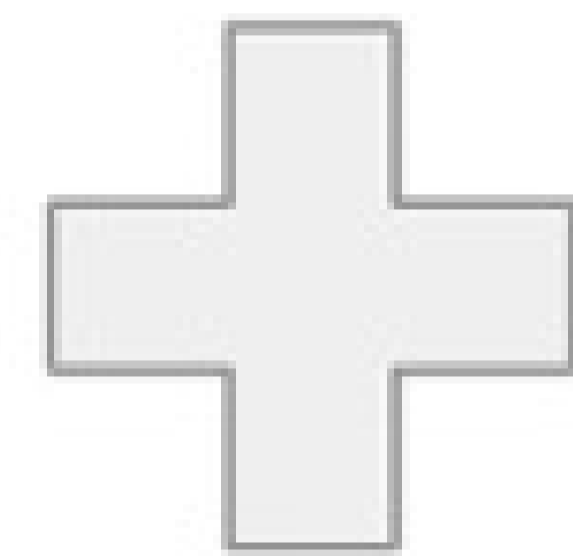
**Lab +
Lecture**

Neural Network Failure Modes, Part III

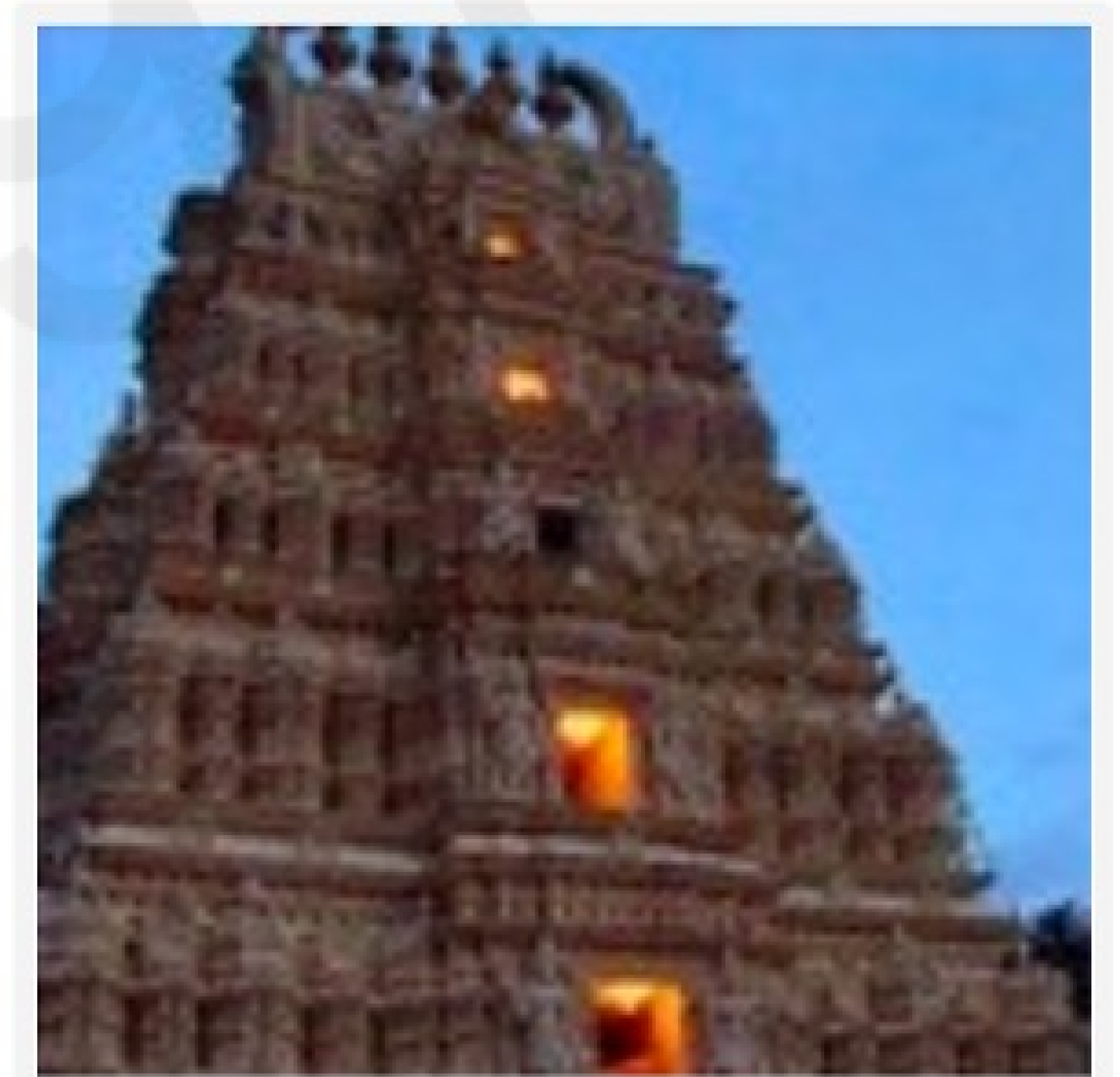


Original image

Temple (97%)



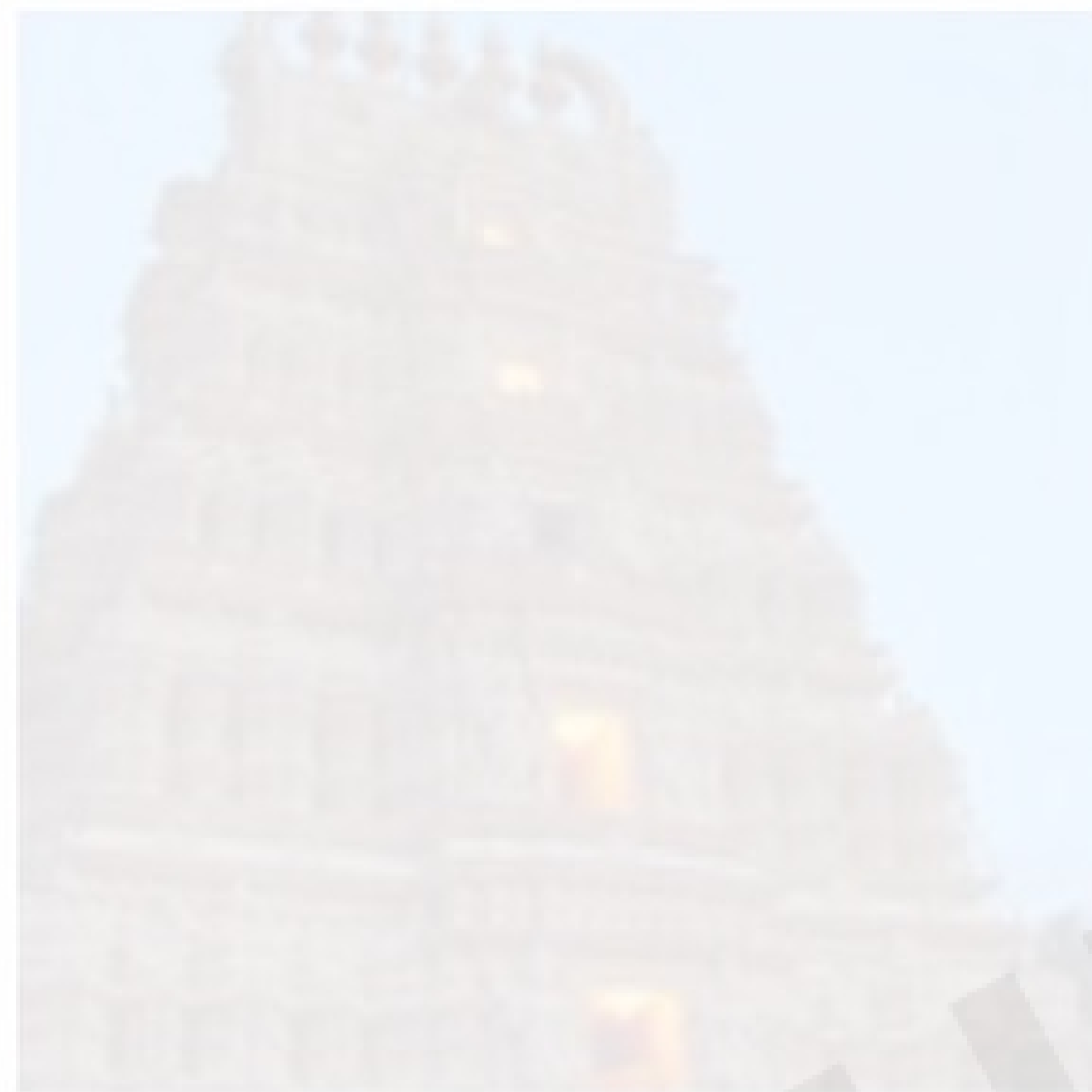
Perturbations



Adversarial example

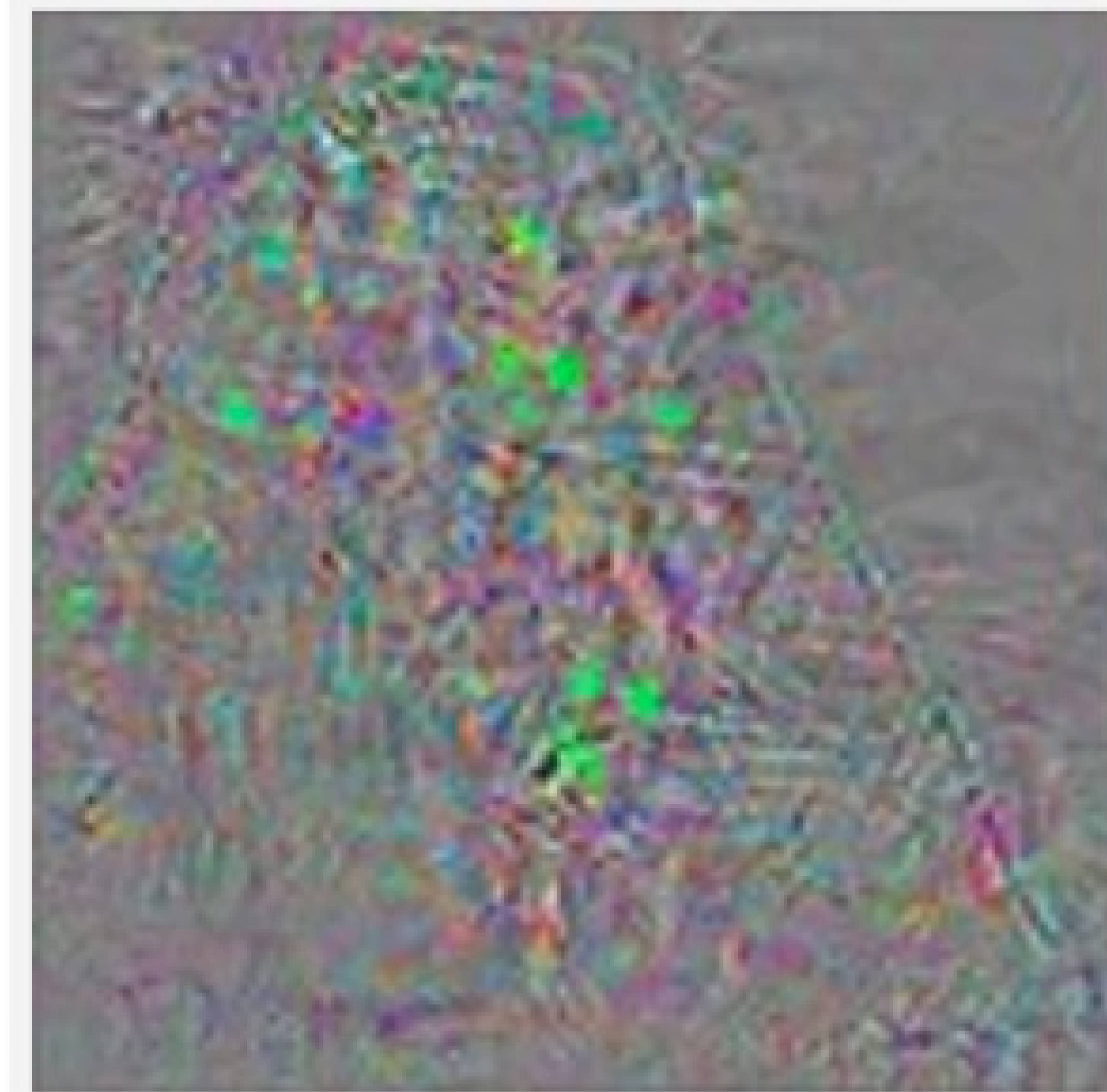
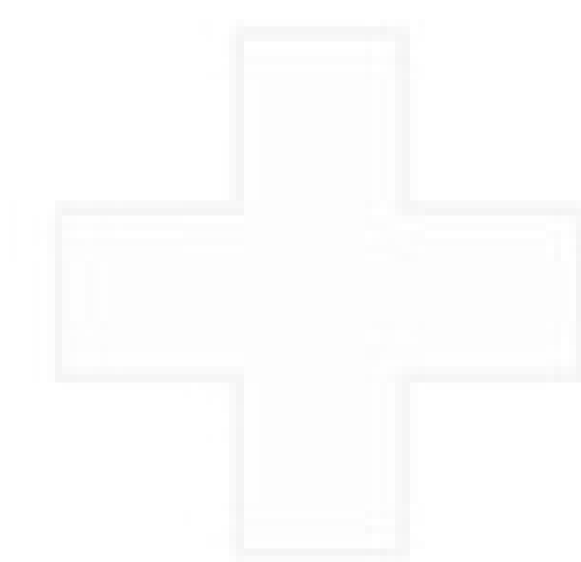
Ostrich (98%)

Adversarial Attacks on Neural Networks

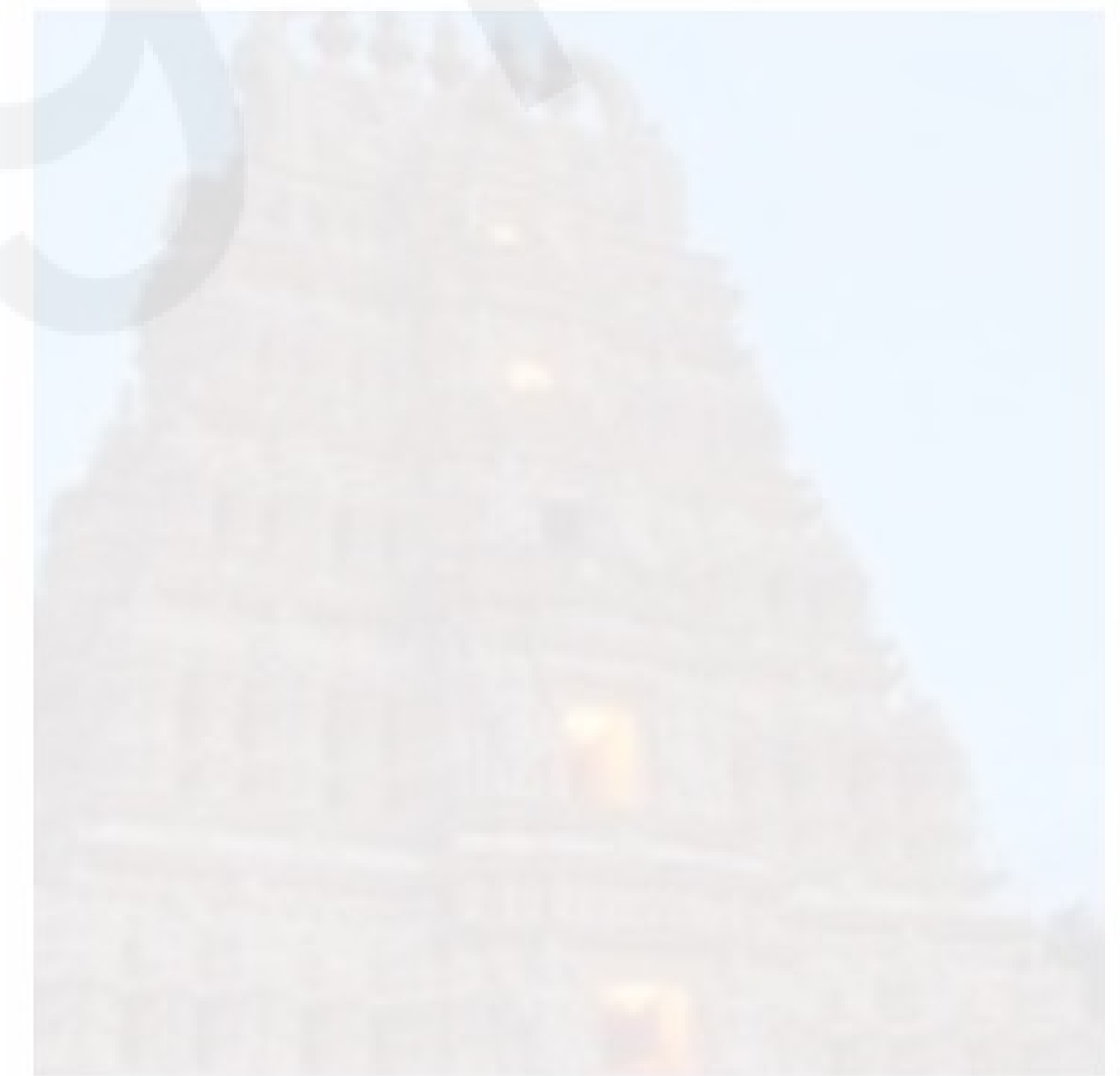
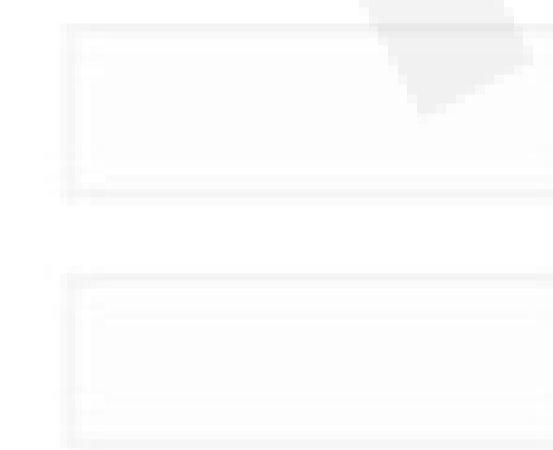


Original image

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”

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”

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”

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”

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”

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”

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

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

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

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

Artificial Intelligence has a gender bias problem – just ask Siri

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.



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

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



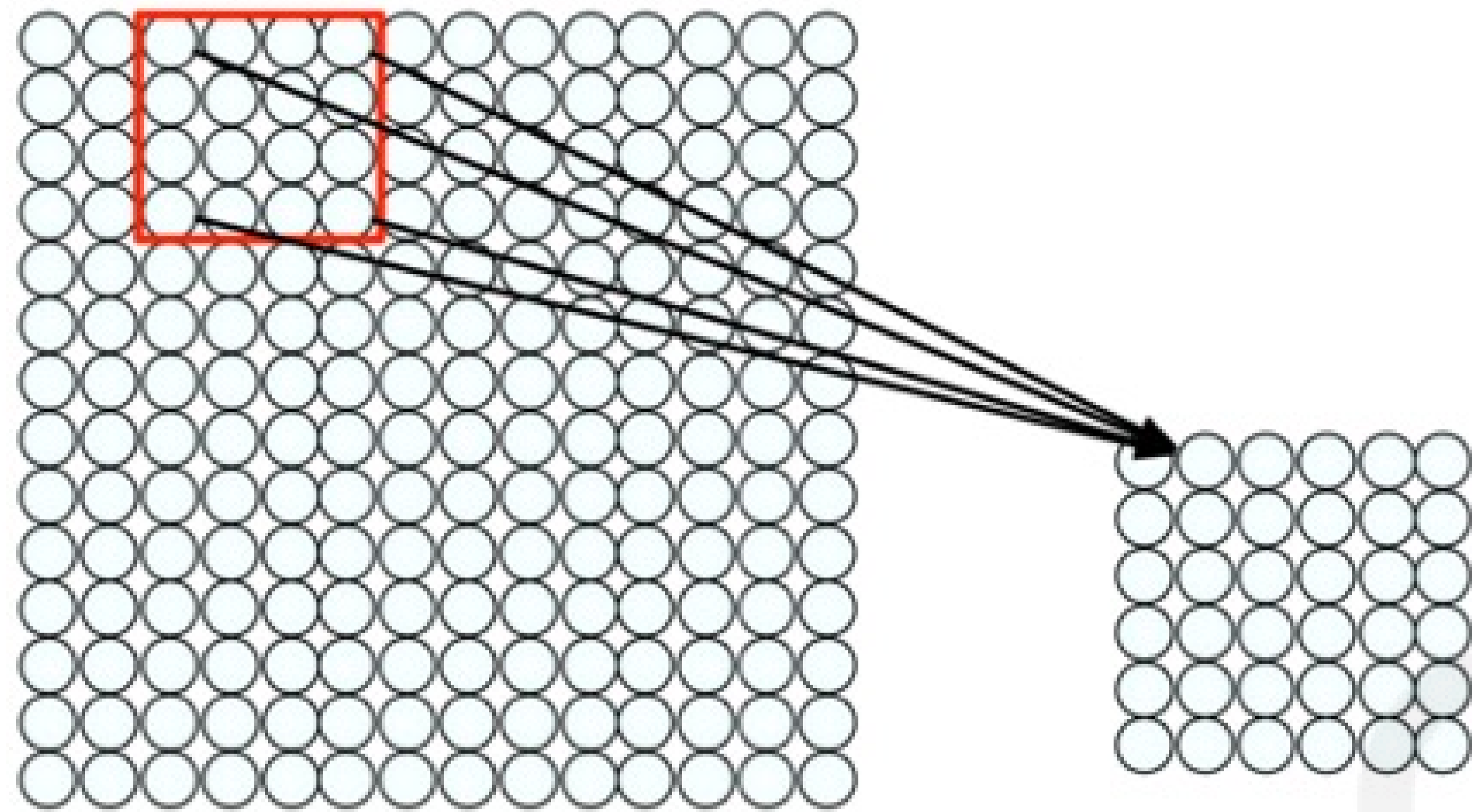
**Lab +
Lecture**

Neural Network Limitations...

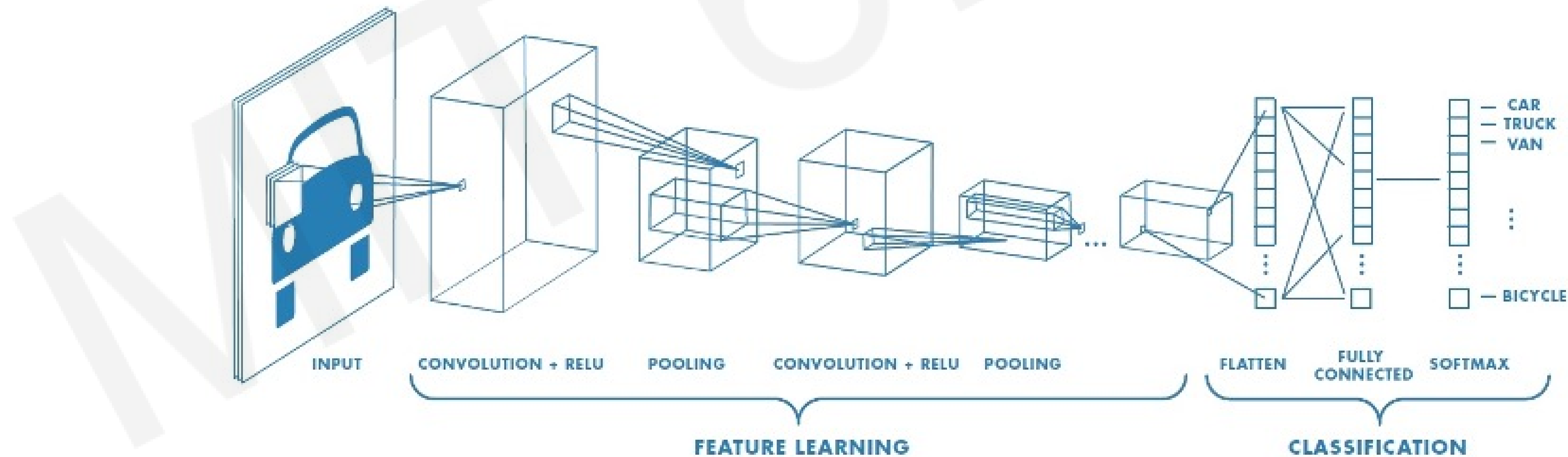
- Very **data hungry** (eg. often millions of examples)
- **Computationally intensive** to train and deploy (tractably requires GPUs)
- Easily fooled by **adversarial examples**
- Can be subject to **algorithmic bias**
- Poor at **representing uncertainty** (how do you know what the model knows?)
- Uninterpretable **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

New Frontiers I: Encoding Structure into Deep Learning

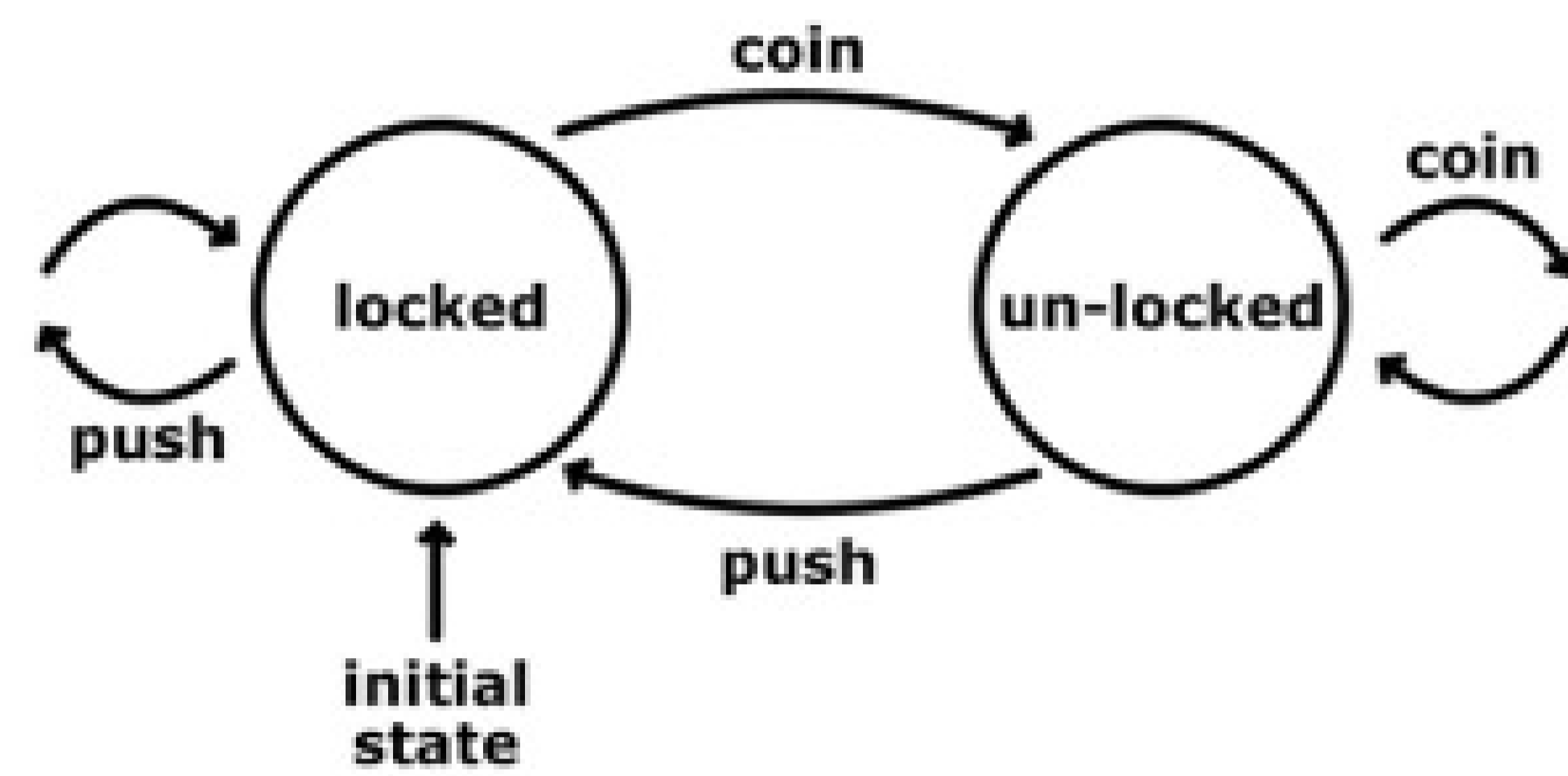
CNNs: Using Spatial Structure



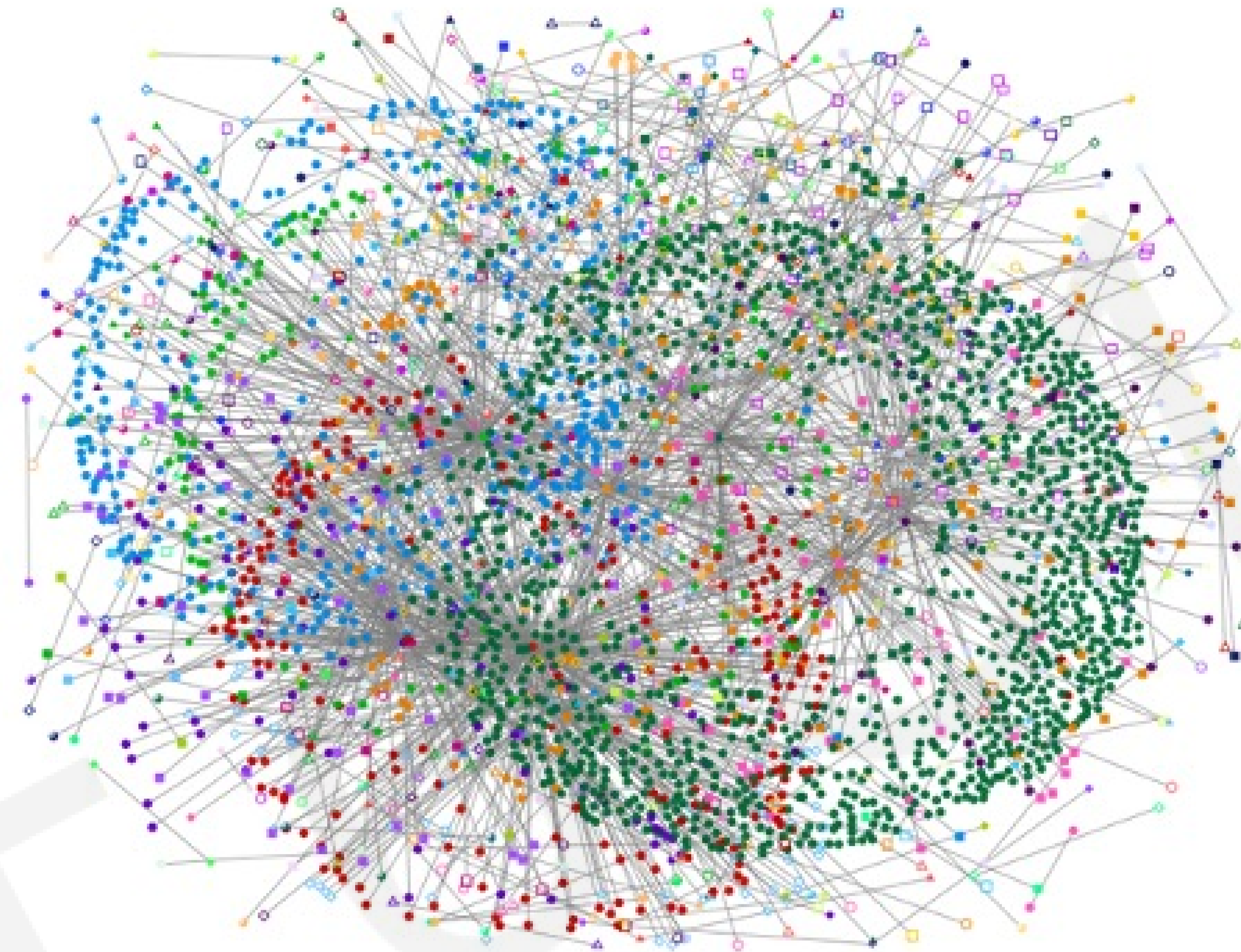
- 1) Apply a set of weights to extract **local features**
- 2) Use **multiple filters** to extract different features
- 3) **Spatially share** parameters of each filter



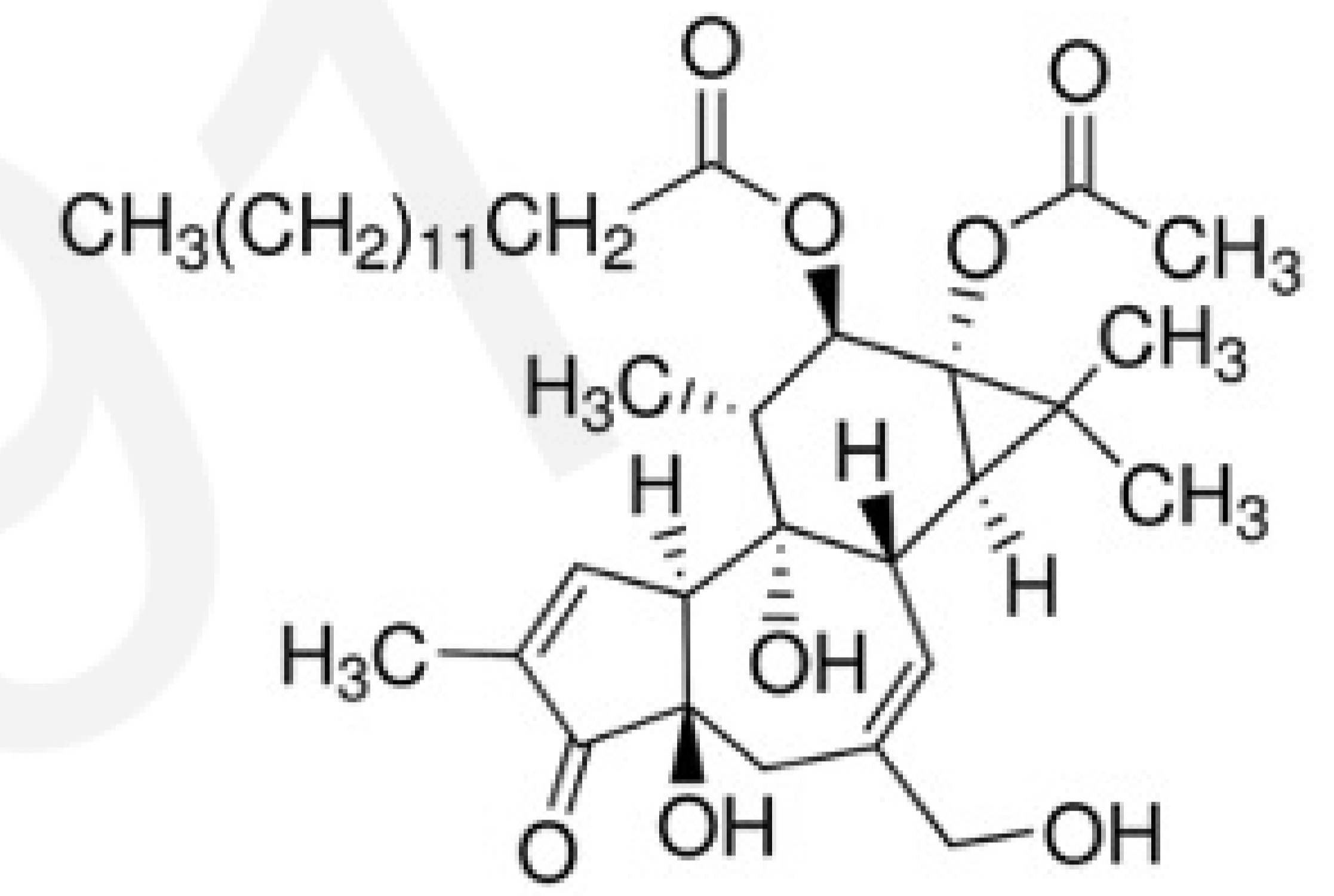
Graphs as a Structure for Representing Data



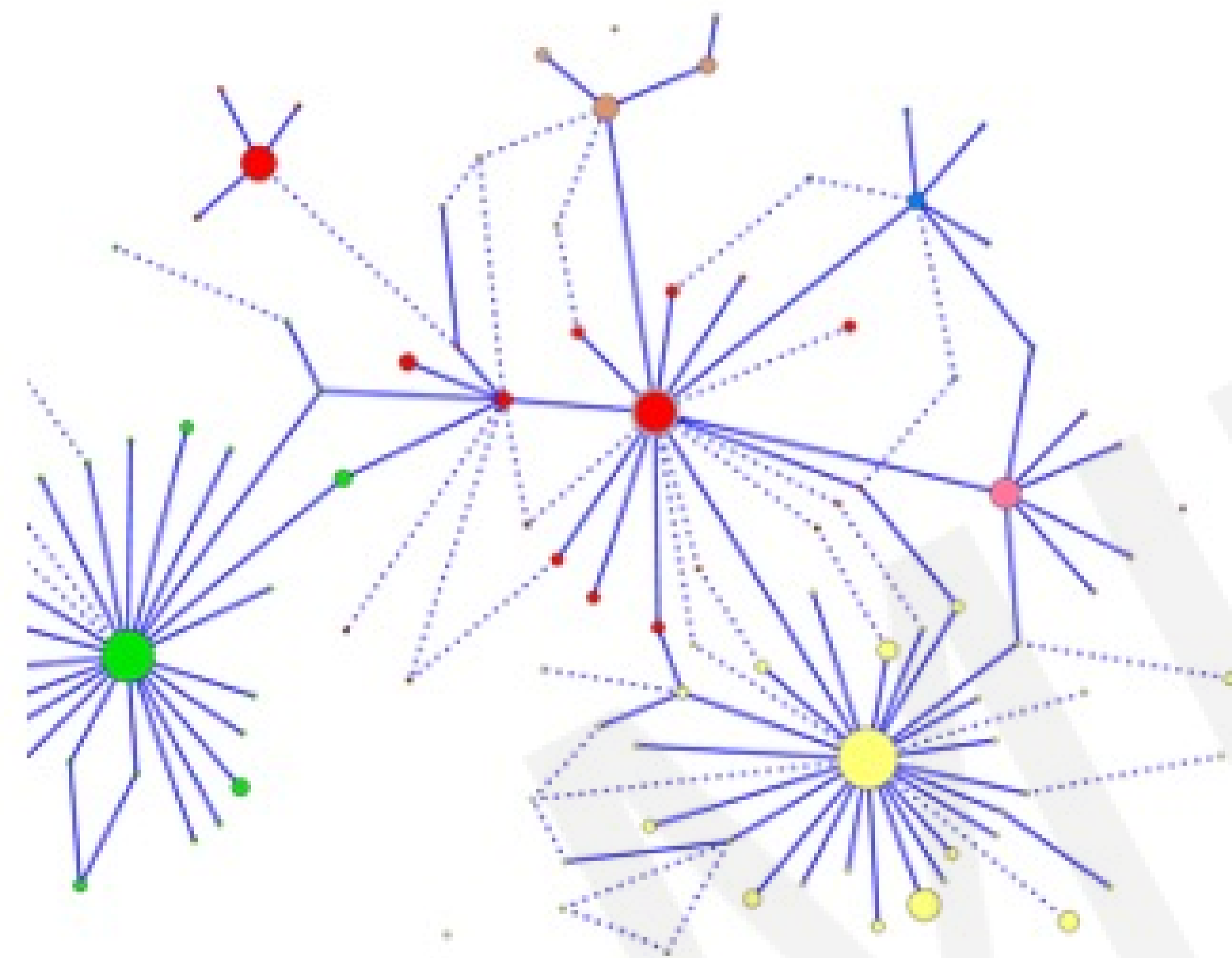
State Machines



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 (GCNs)



Graph Convolutional Networks

Convolutional Networks



Graph Convolutional Networks (GCNs)



Graph Convolutional Networks

Convolutional Networks



Graph Convolutional Networks (GCNs)



Graph Convolutional Networks

Convolutional Networks

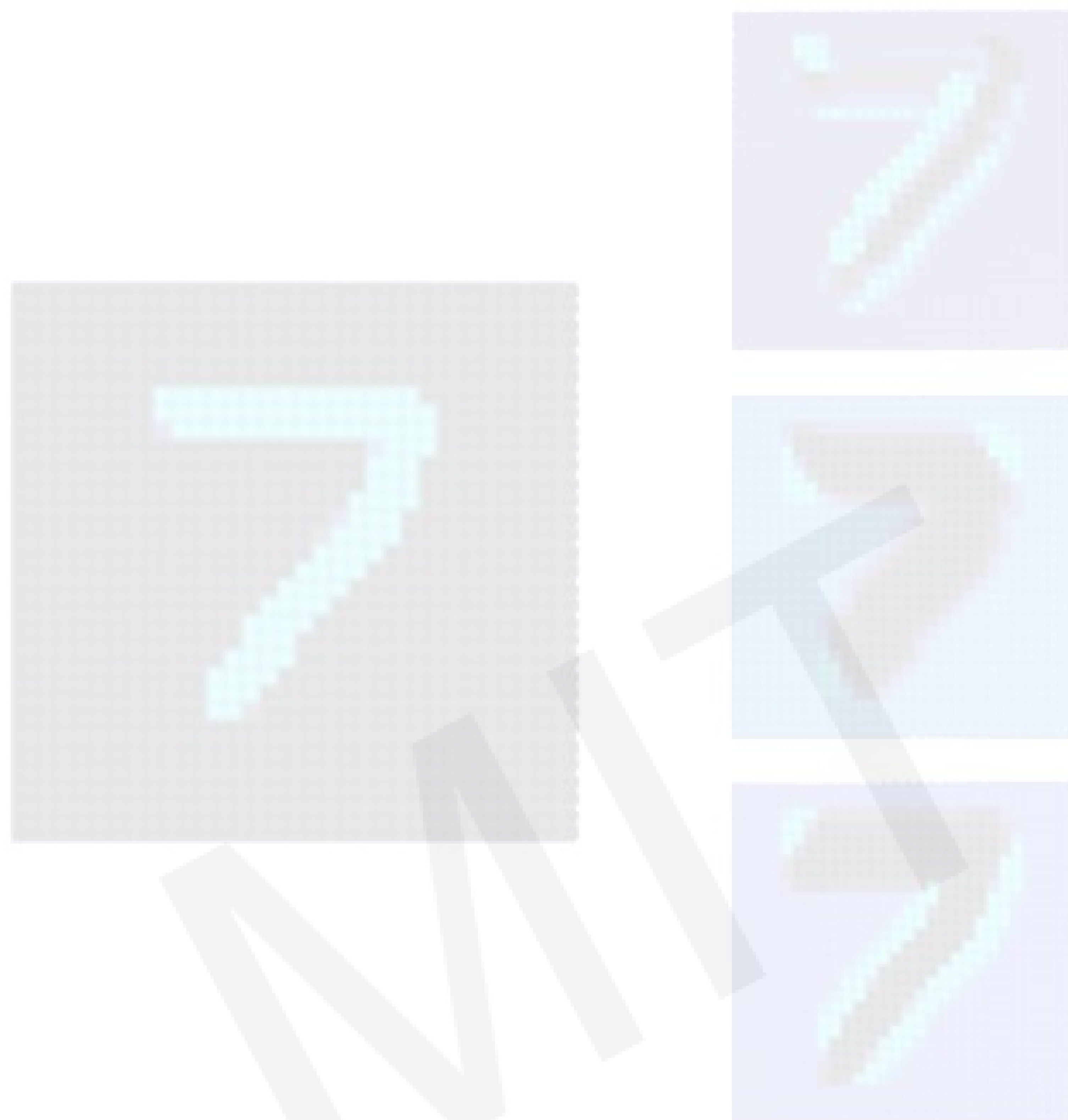


Graph Convolutional Networks (GCNs)

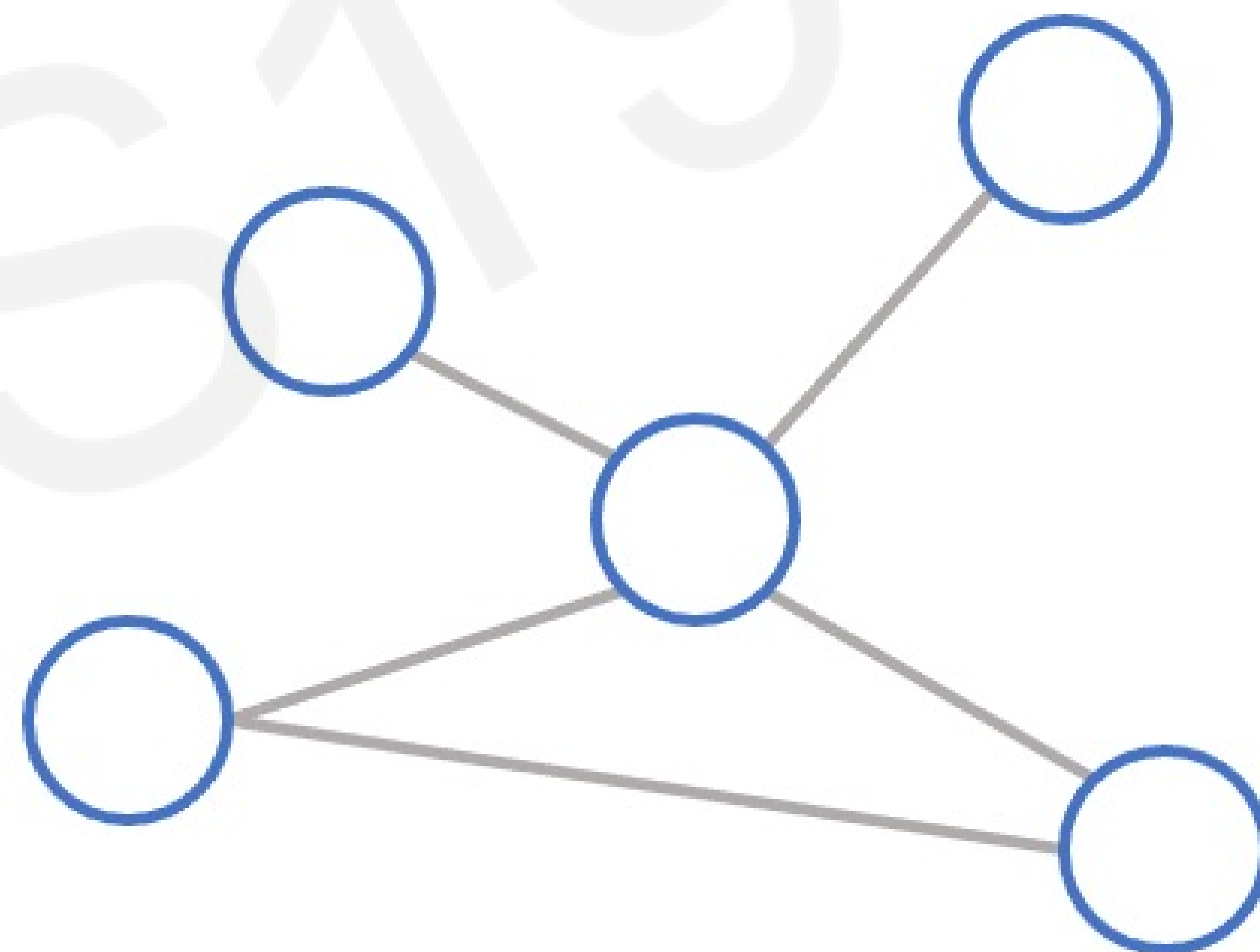


Graph Convolutional Networks

Convolutional Networks

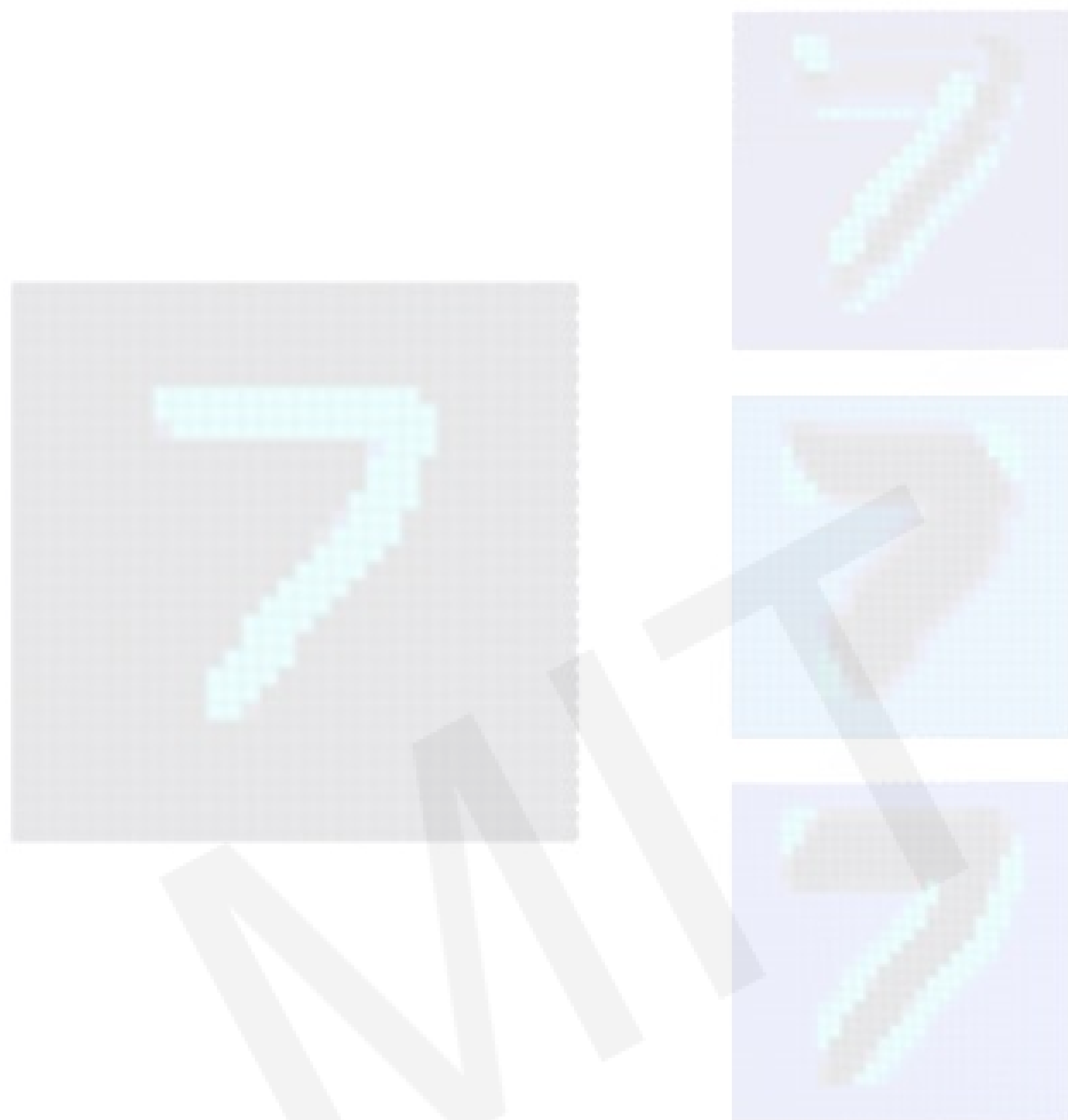


Graph Convolutional Networks (GCNs)

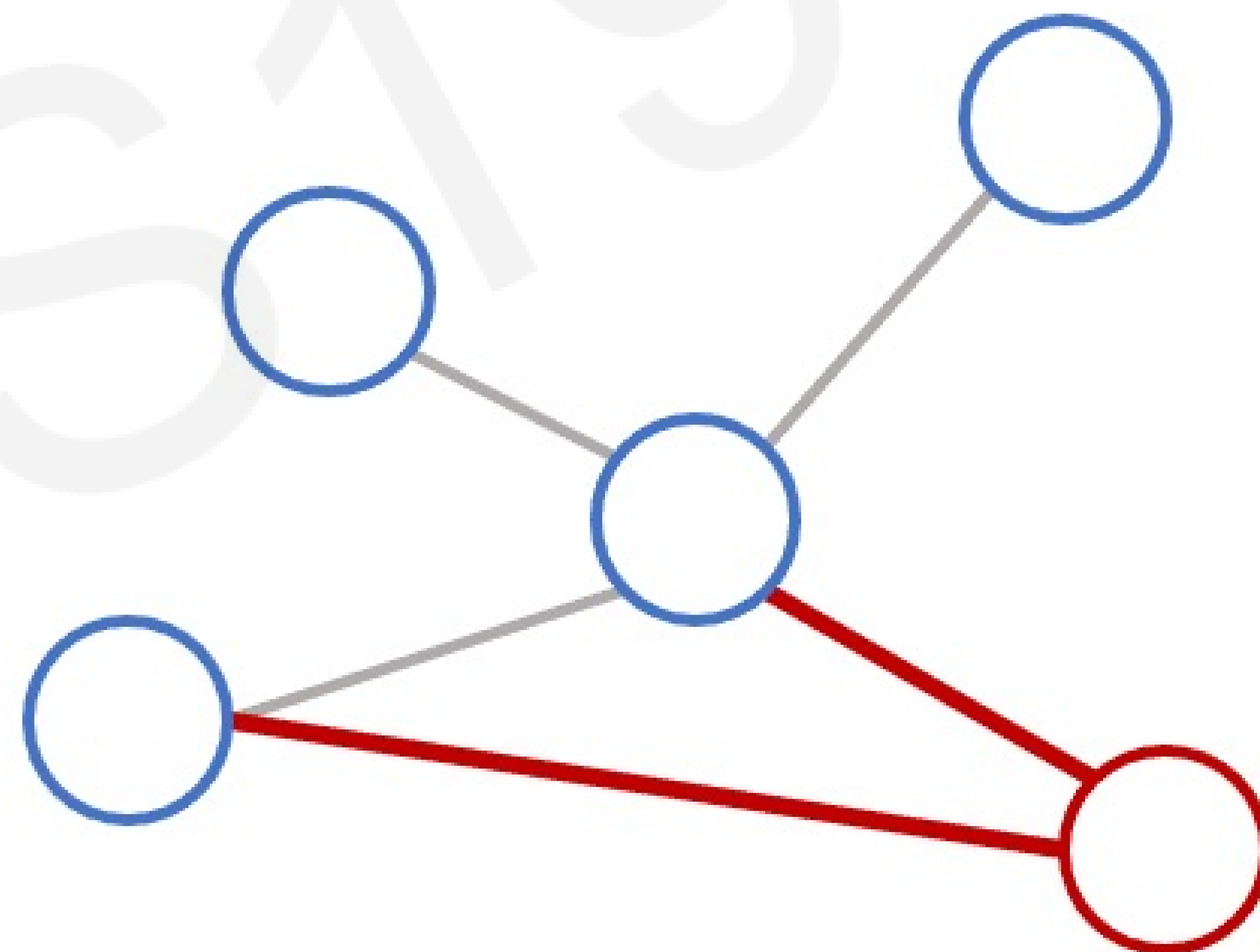


Graph Convolutional Networks

Convolutional Networks

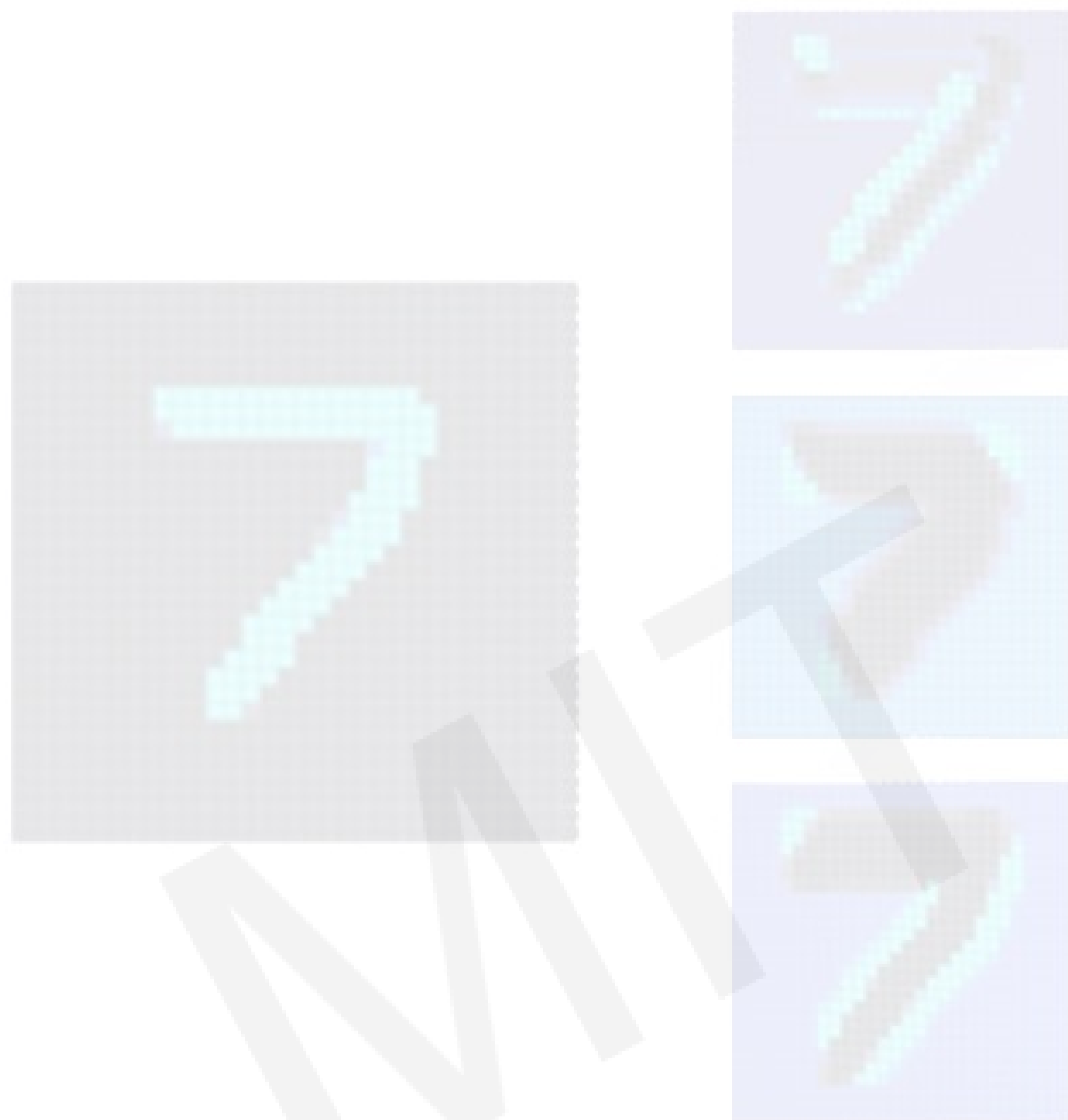


Graph Convolutional Networks (GCNs)

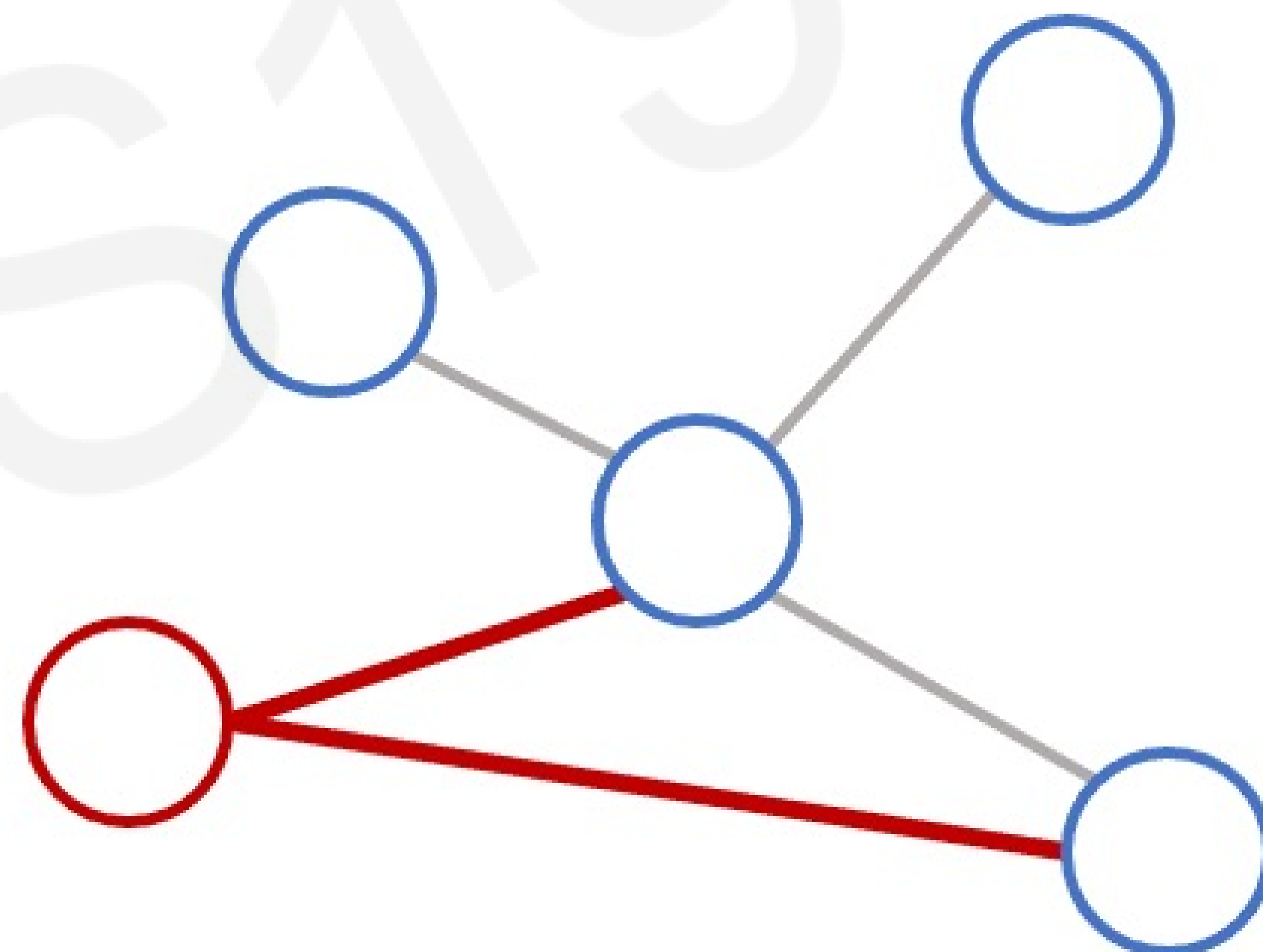


Graph Convolutional Networks

Convolutional Networks

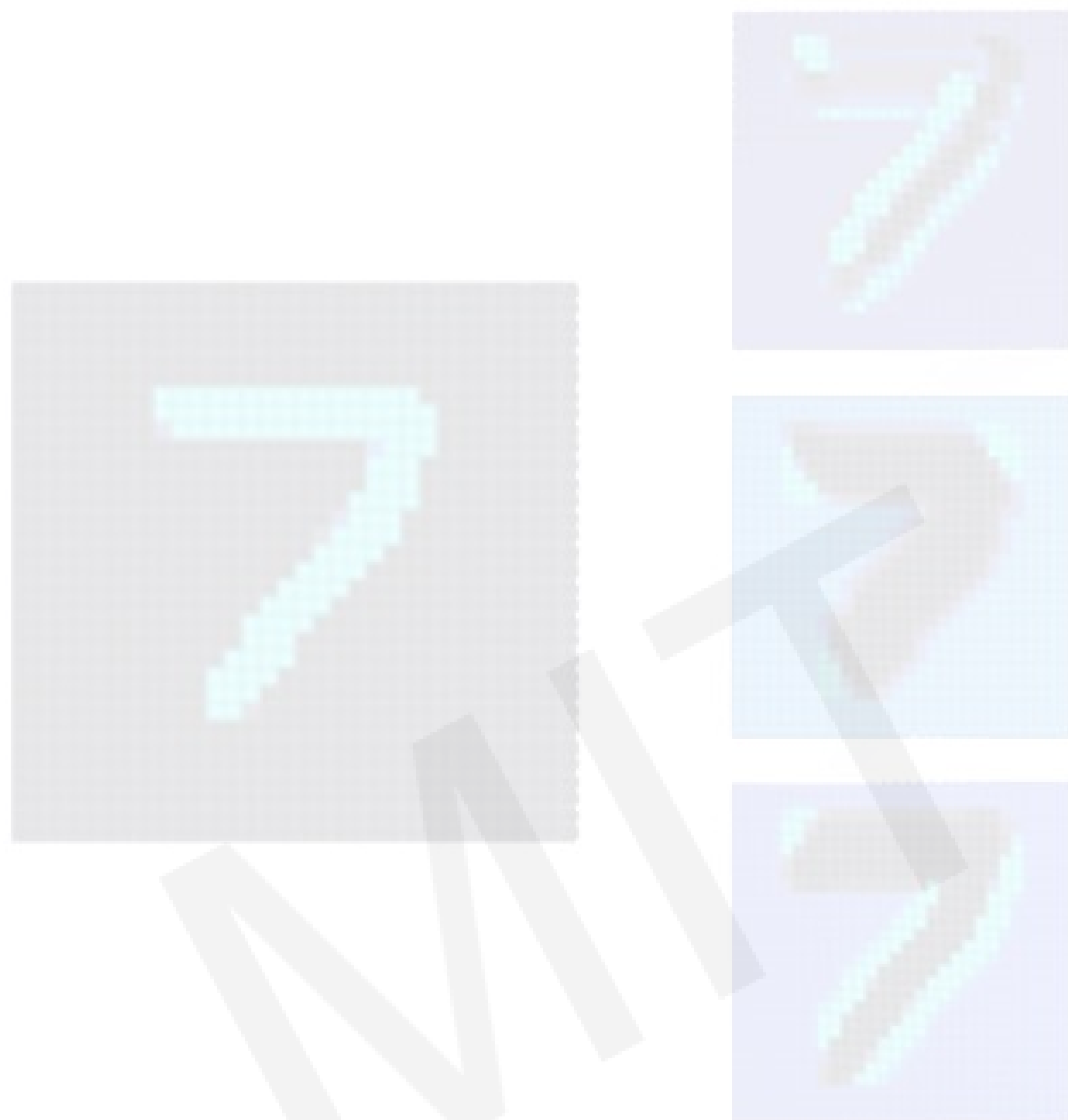


Graph Convolutional Networks (GCNs)

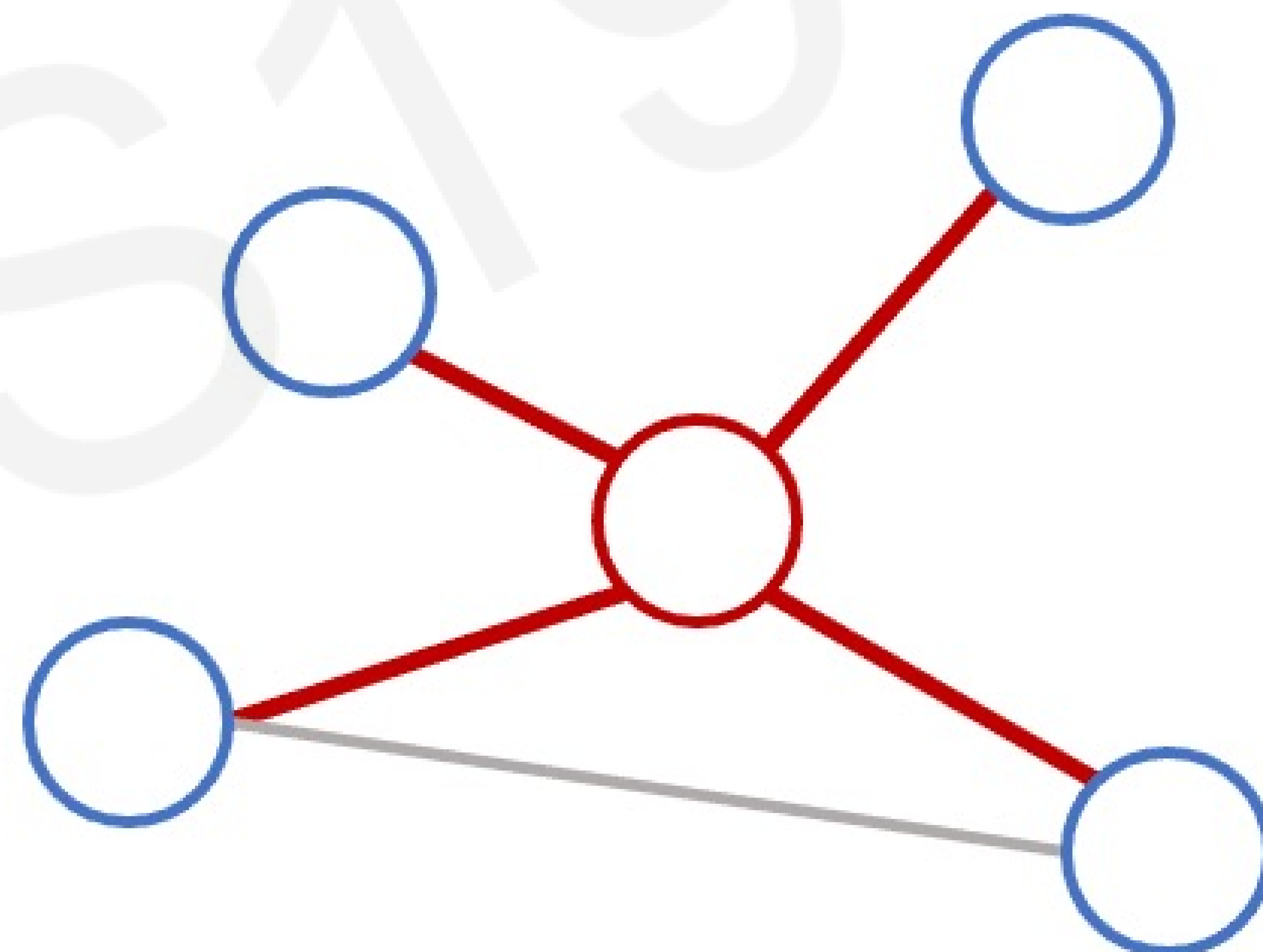


Graph Convolutional Networks

Convolutional Networks

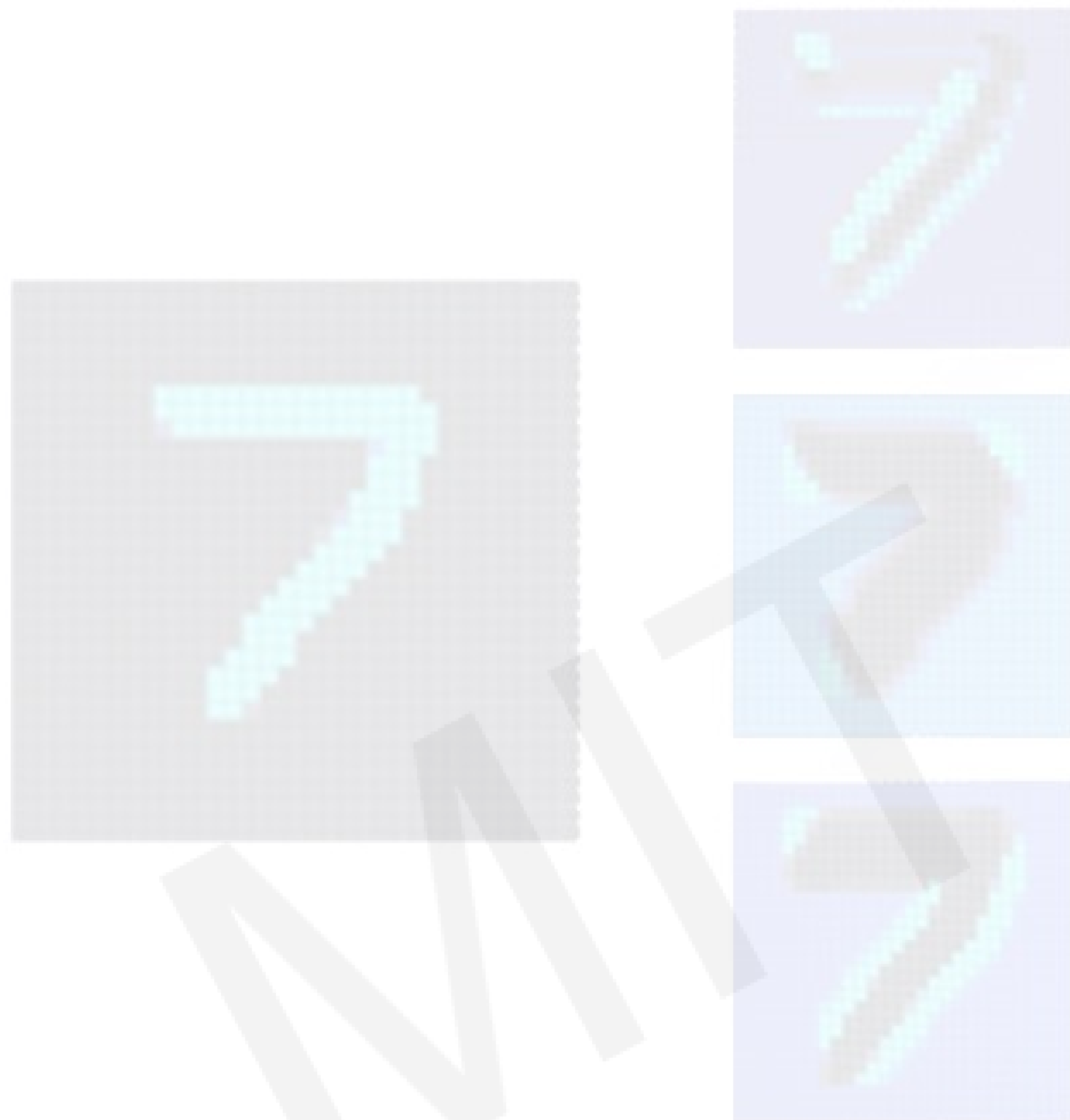


Graph Convolutional Networks (GCNs)

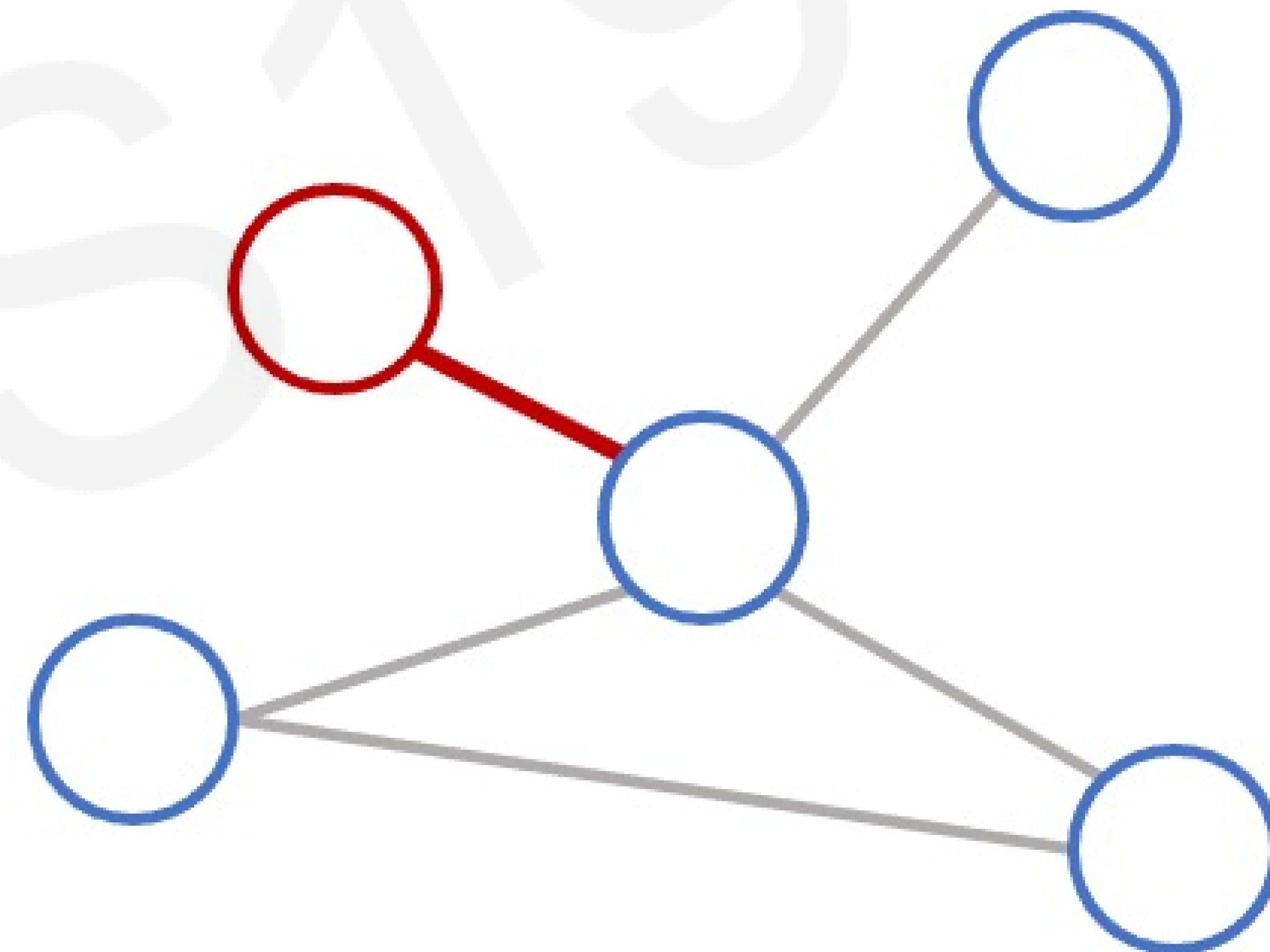


Graph Convolutional Networks

Convolutional Networks

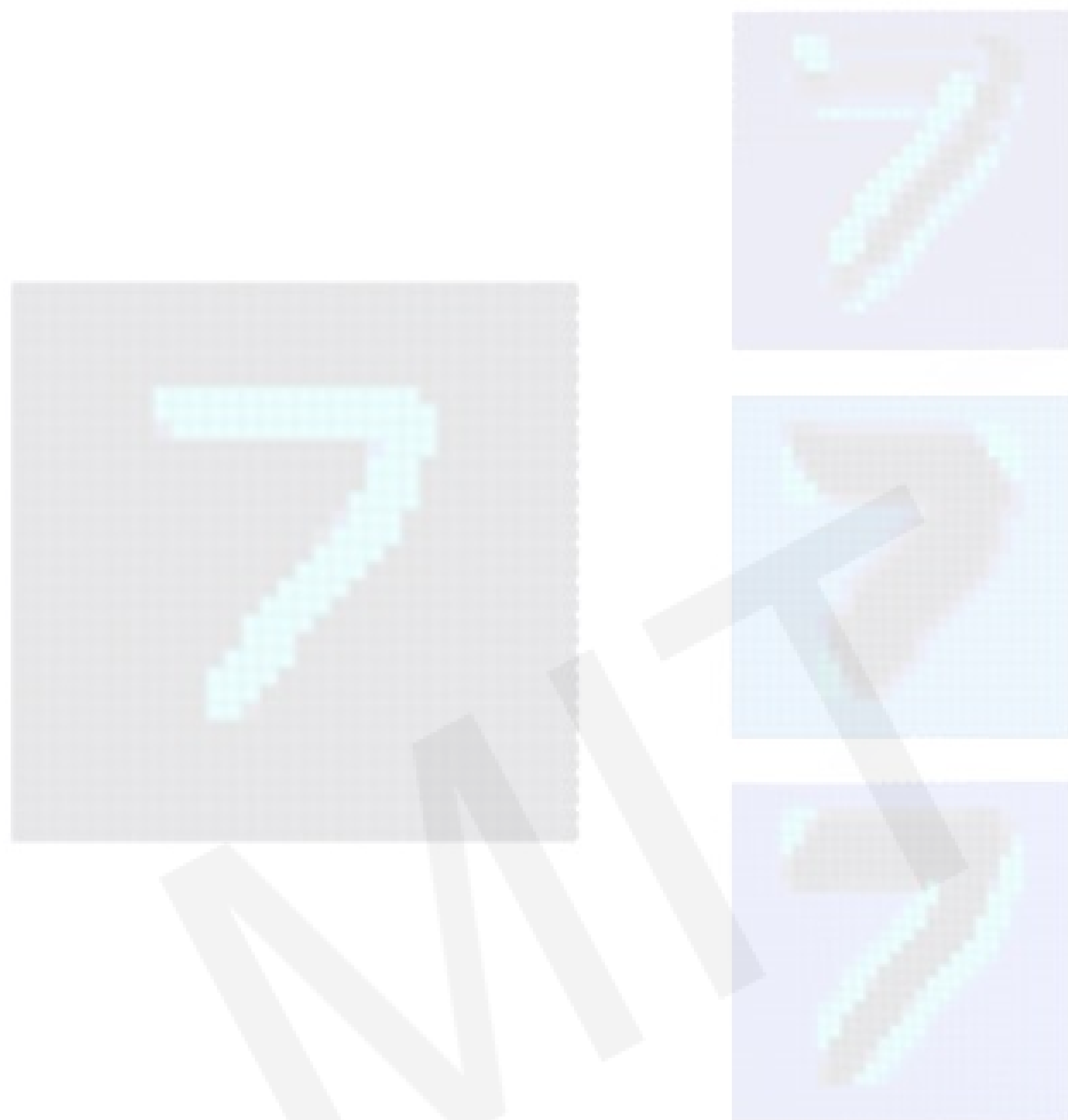


Graph Convolutional Networks (GCNs)

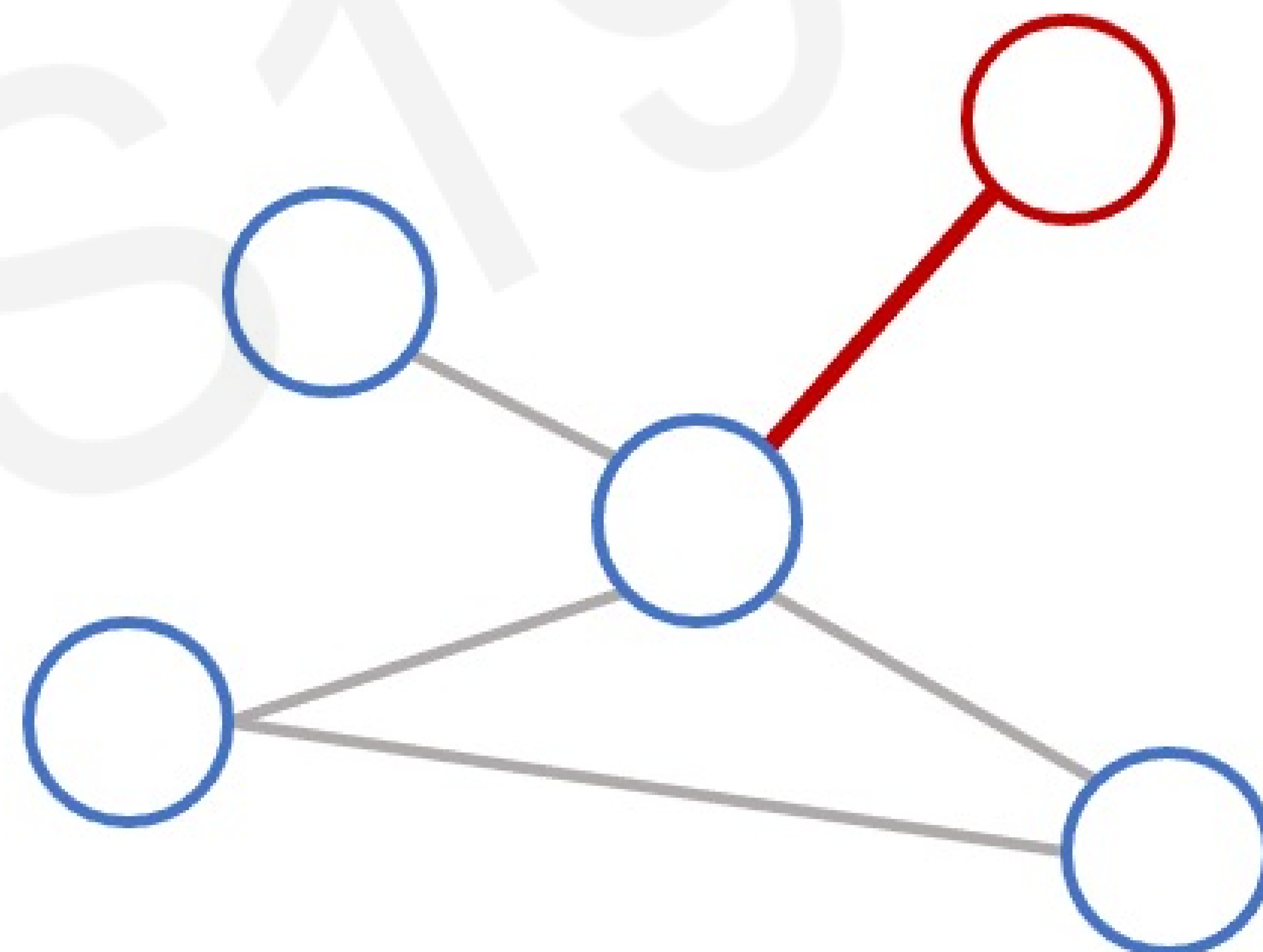


Graph Convolutional Networks

Convolutional Networks

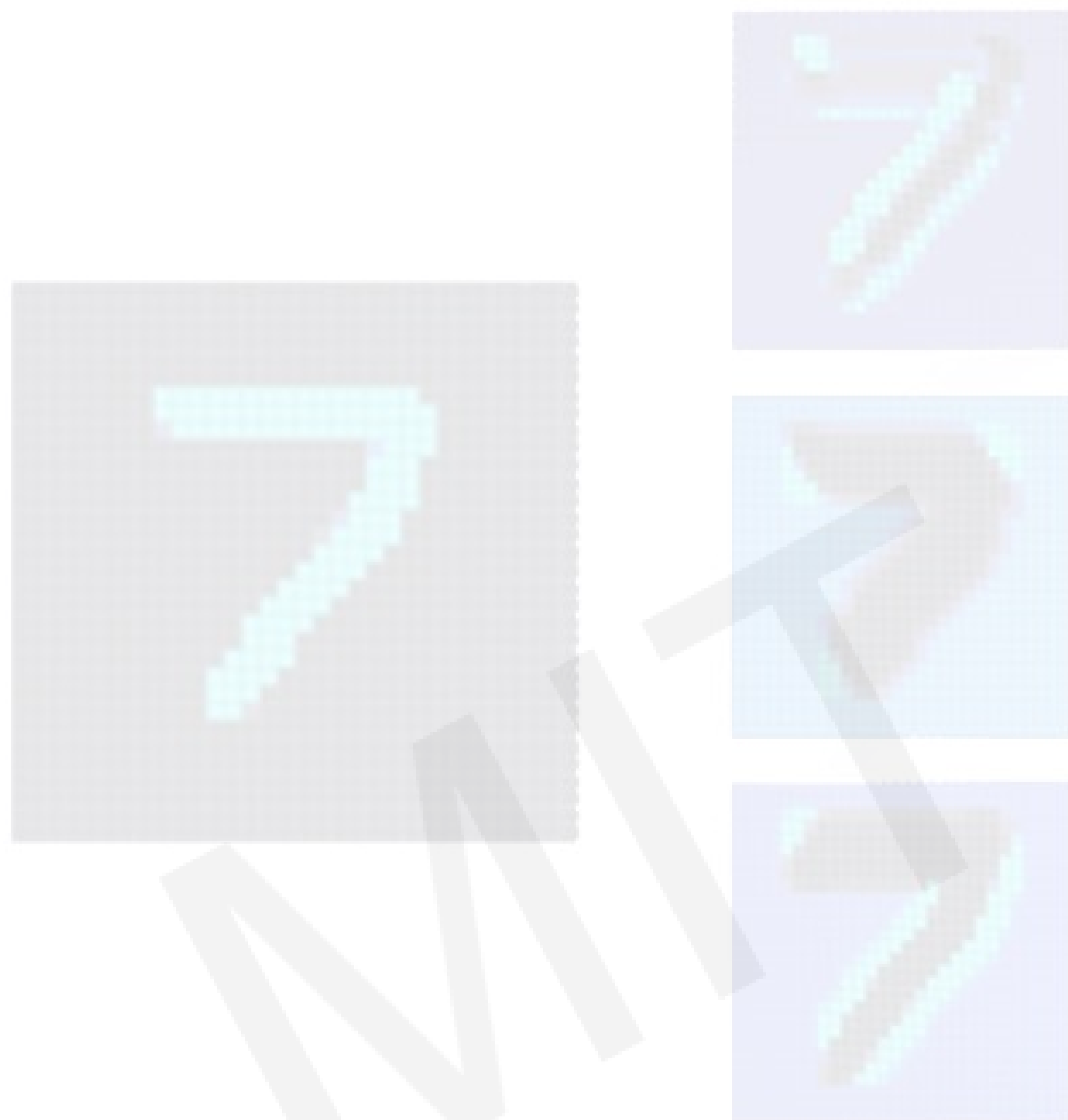


Graph Convolutional Networks (GCNs)

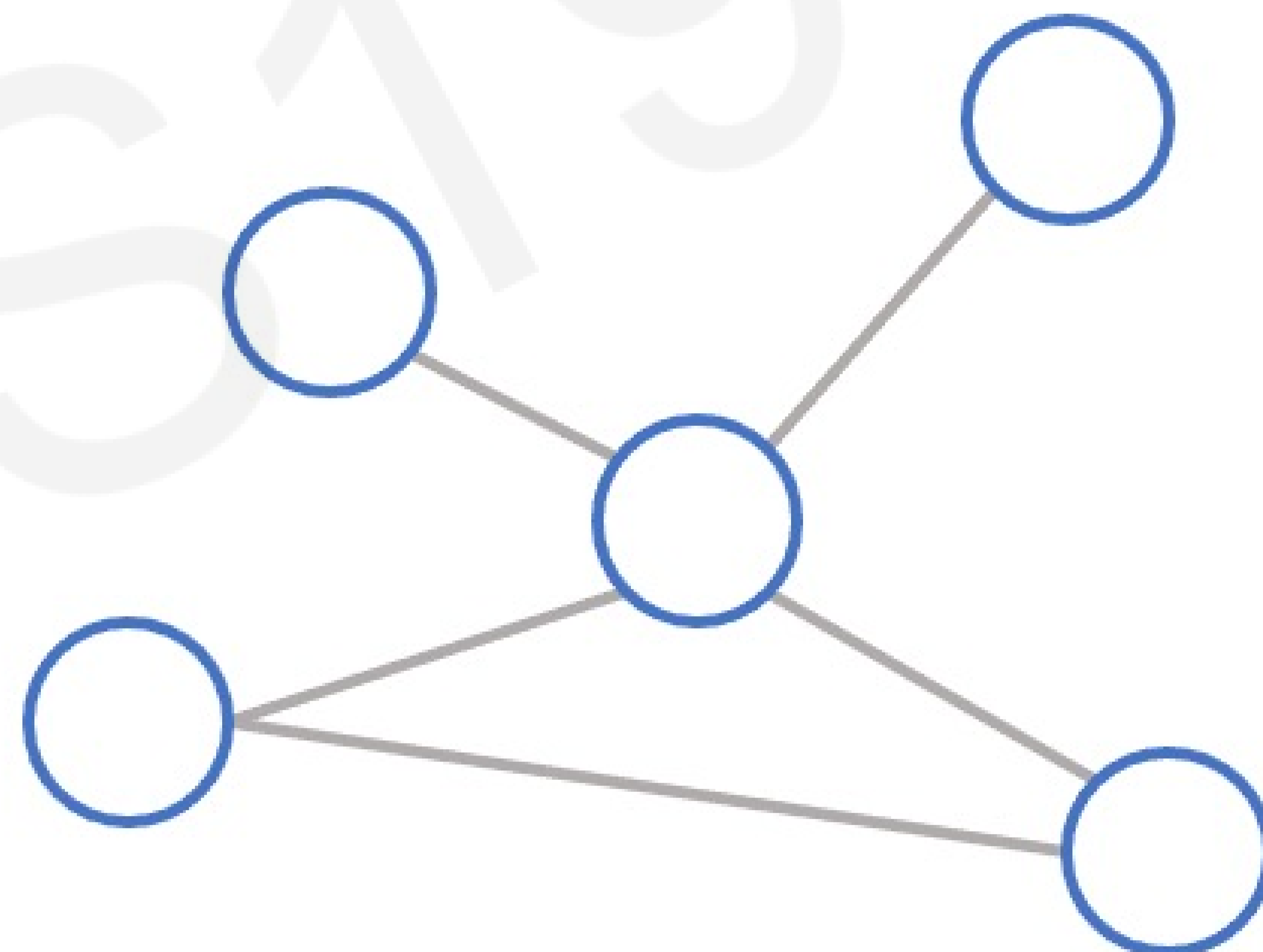


Graph Convolutional Networks

Convolutional Networks

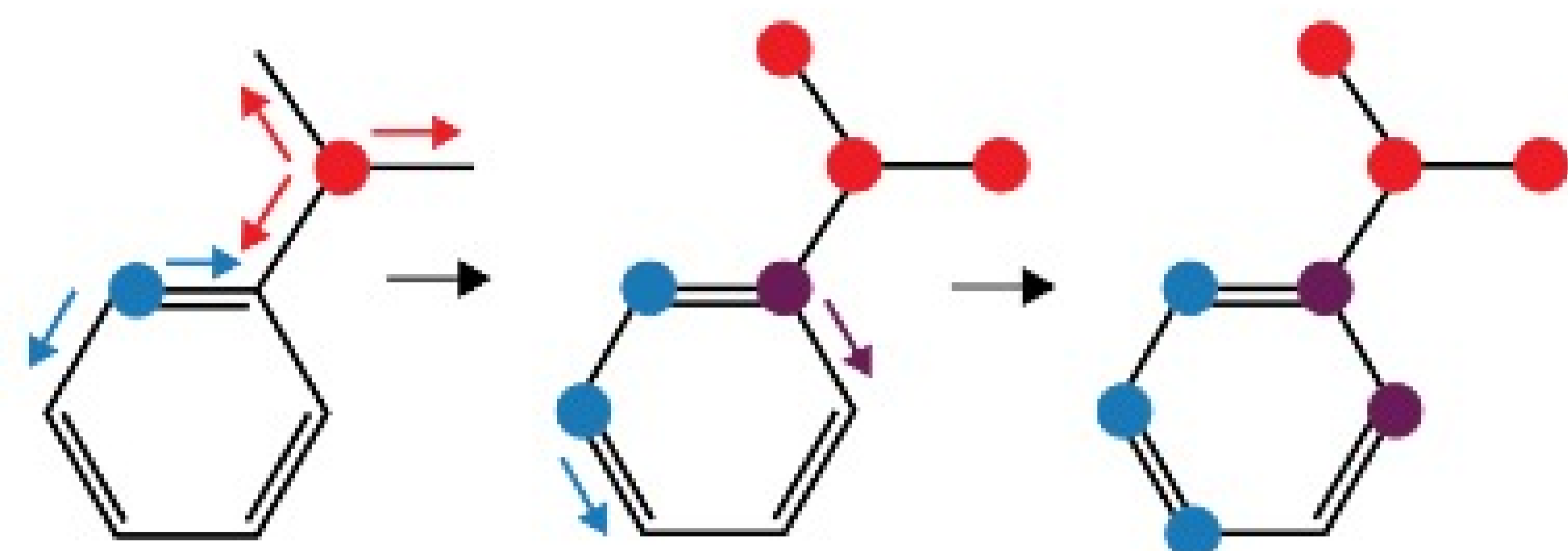


Graph Convolutional Networks (GCNs)



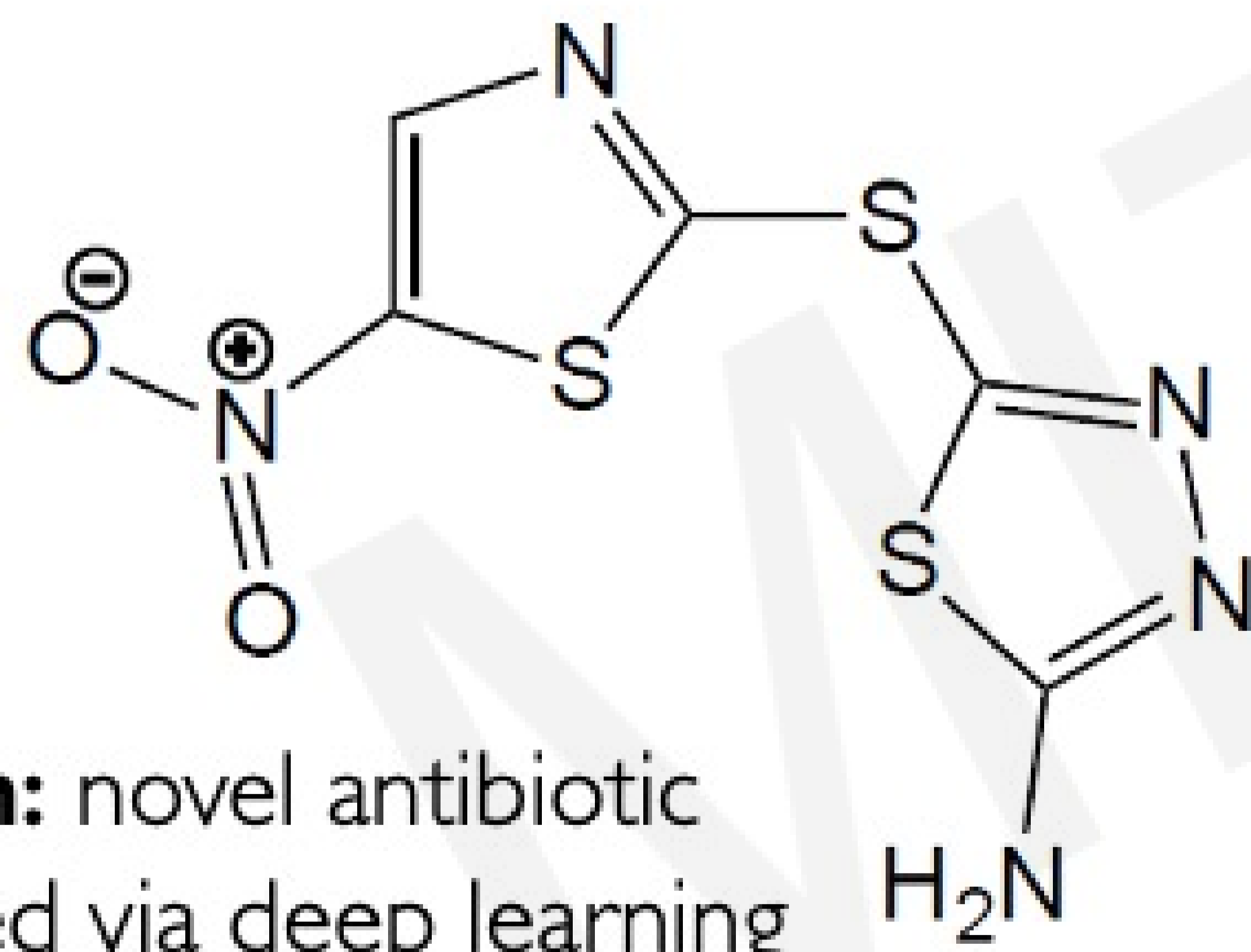
Applications of Graph Neural Networks

Molecular Discovery



Message-passing neural network

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

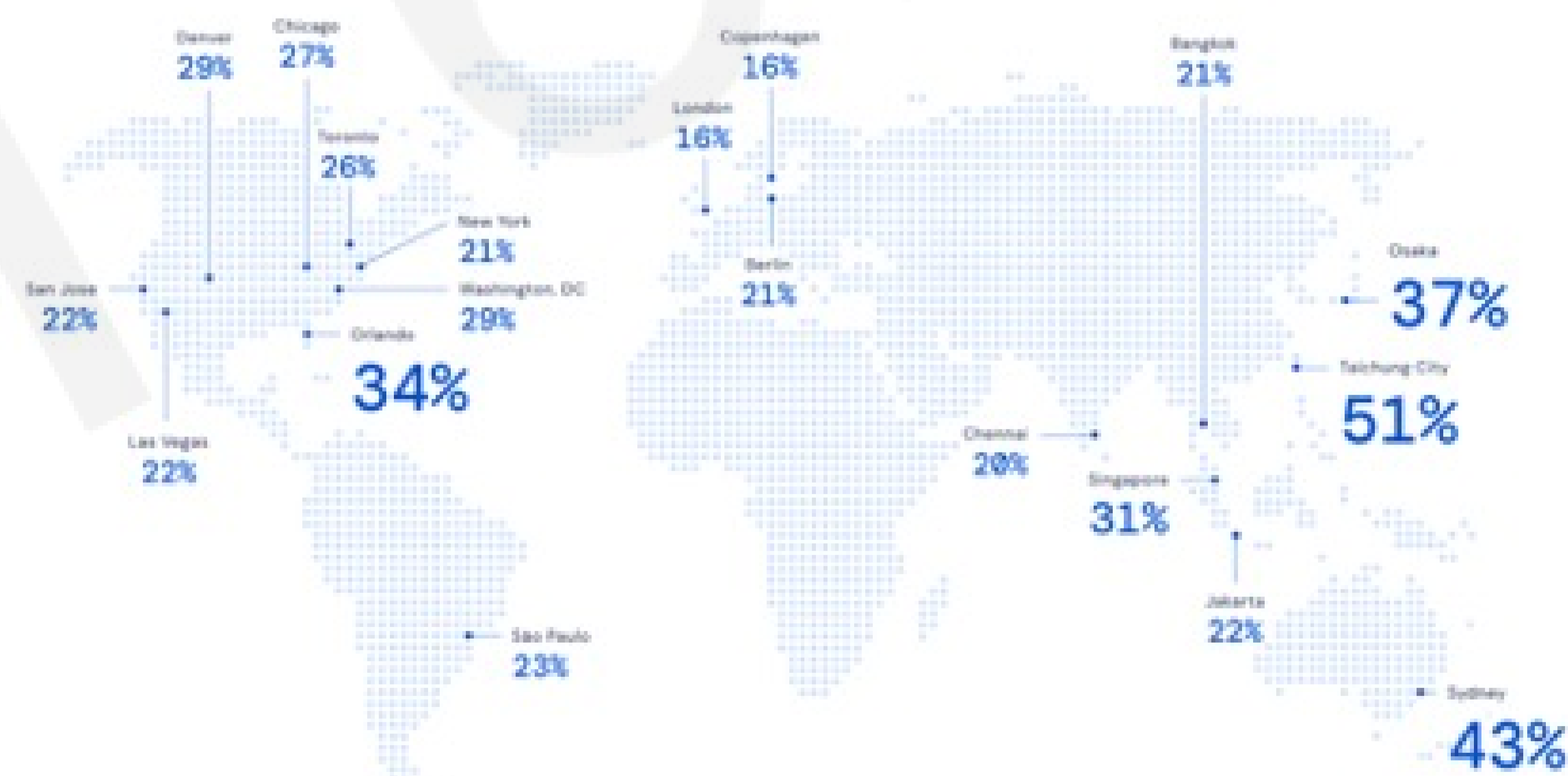


Halicin: novel antibiotic discovered via deep learning

Stokes+ *Cell* 2020

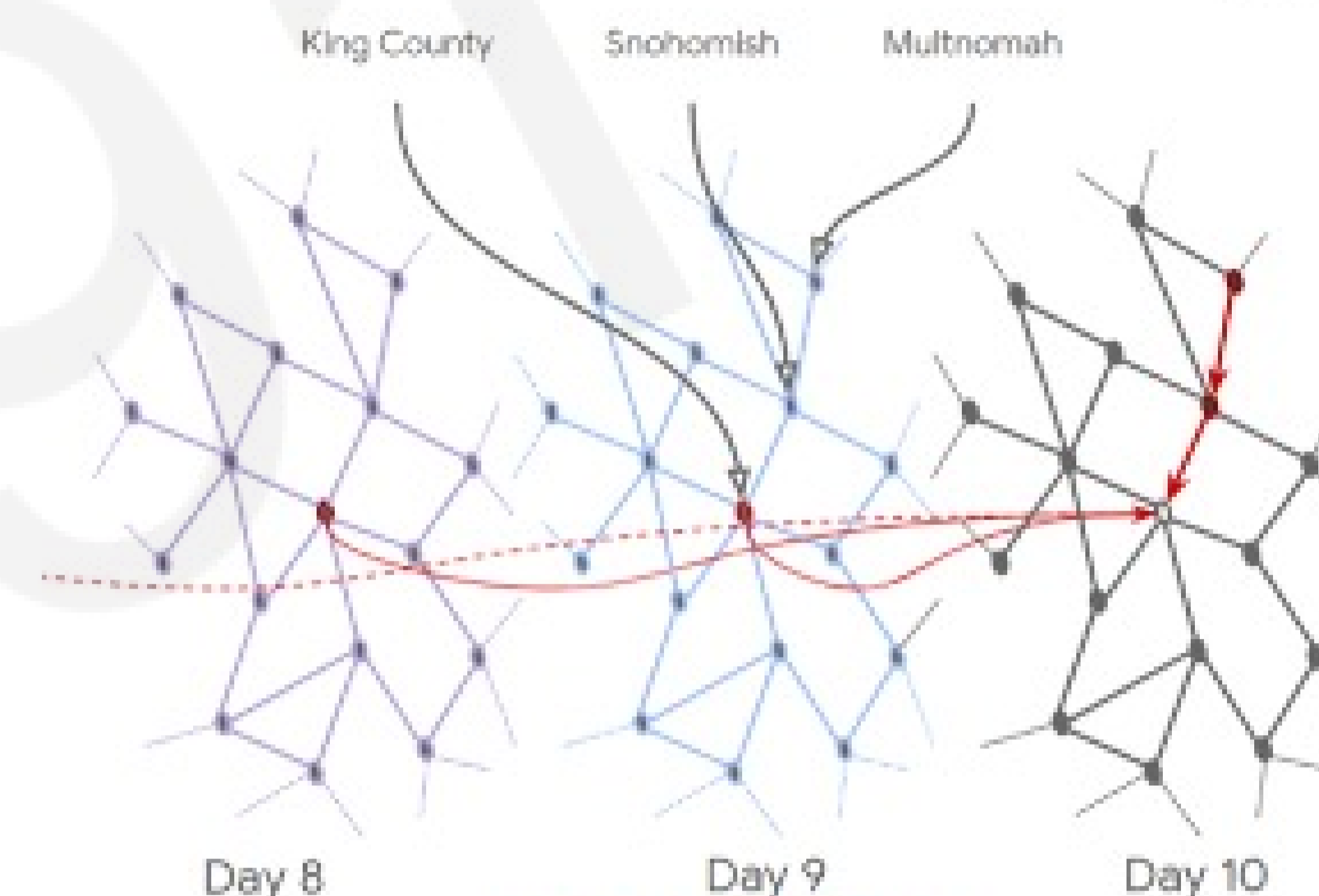
Traffic Prediction

ETA Improvements with GoogleMaps

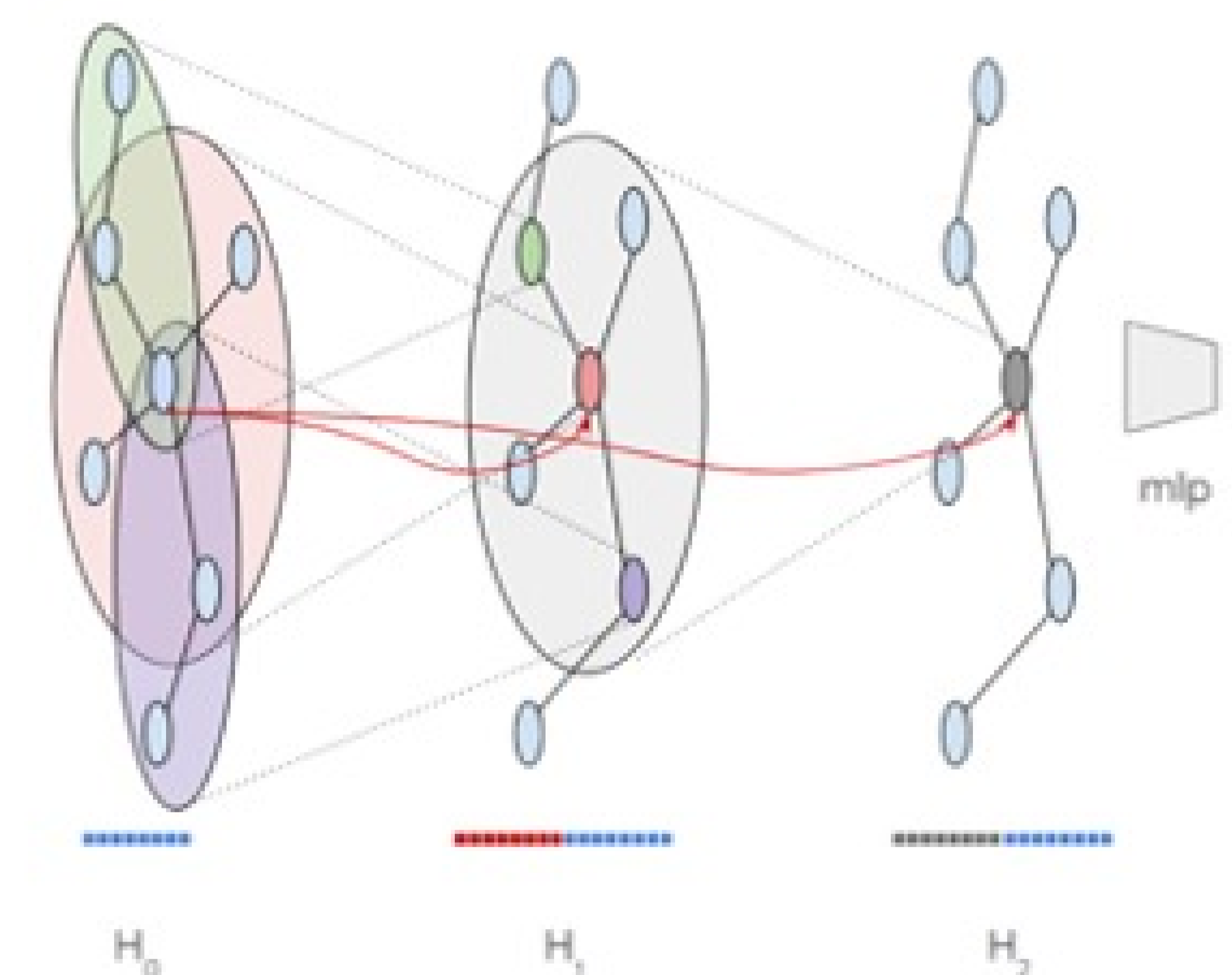


DeepMind + GoogleMaps

COVID-19 Forecasting



Spatio-temporal data

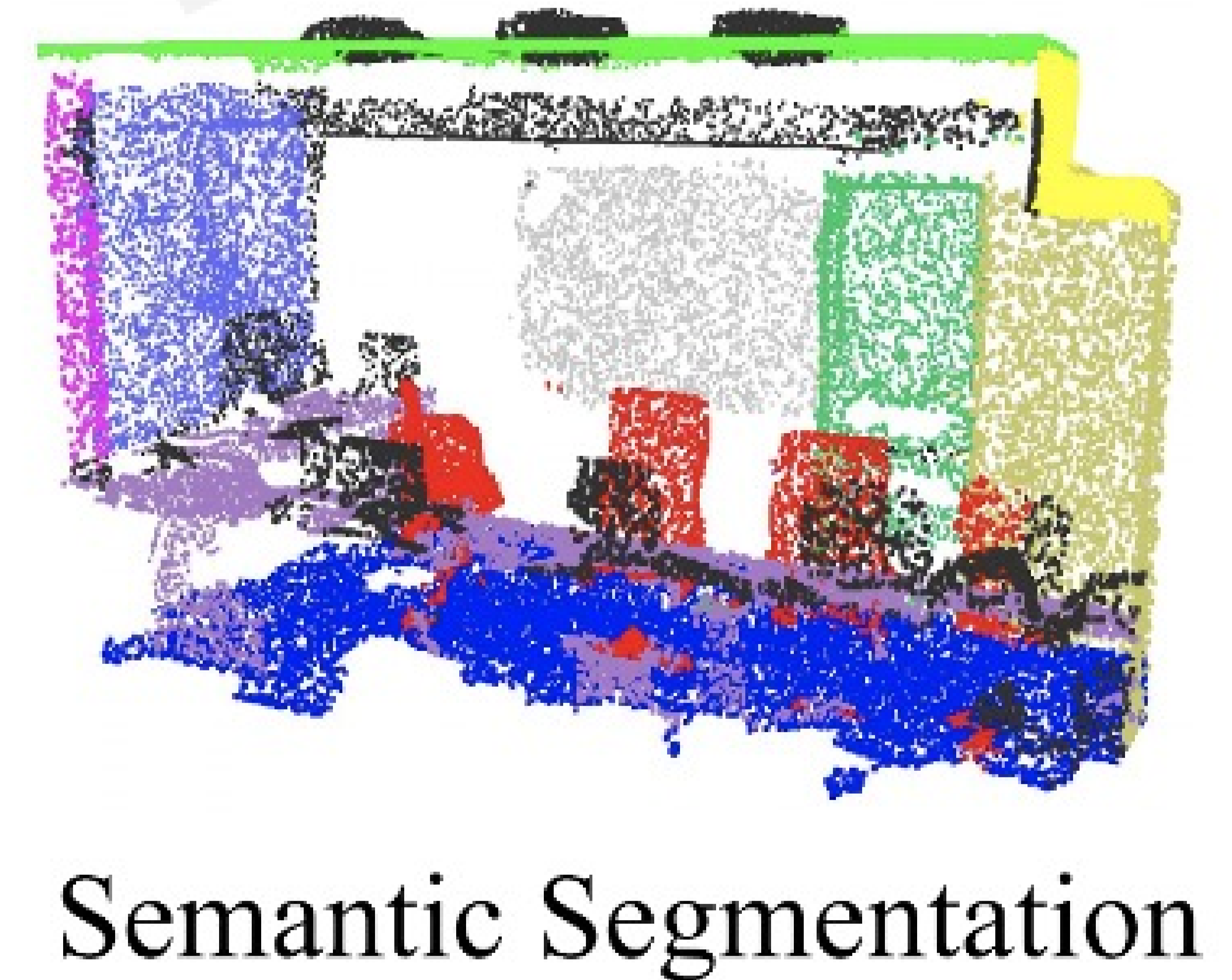
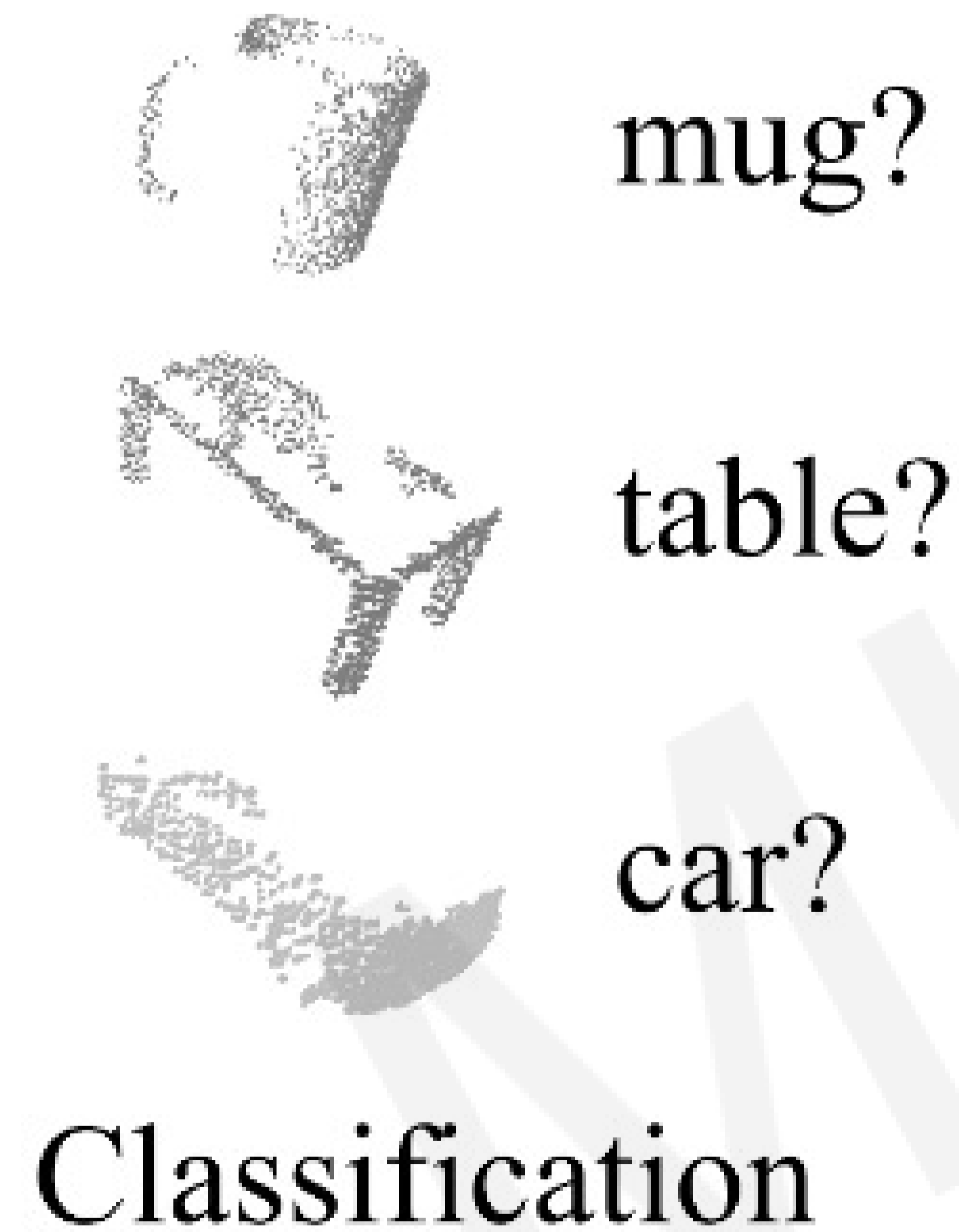


Graph network + temporal embedding

Kapoor+ *KDD* 2020

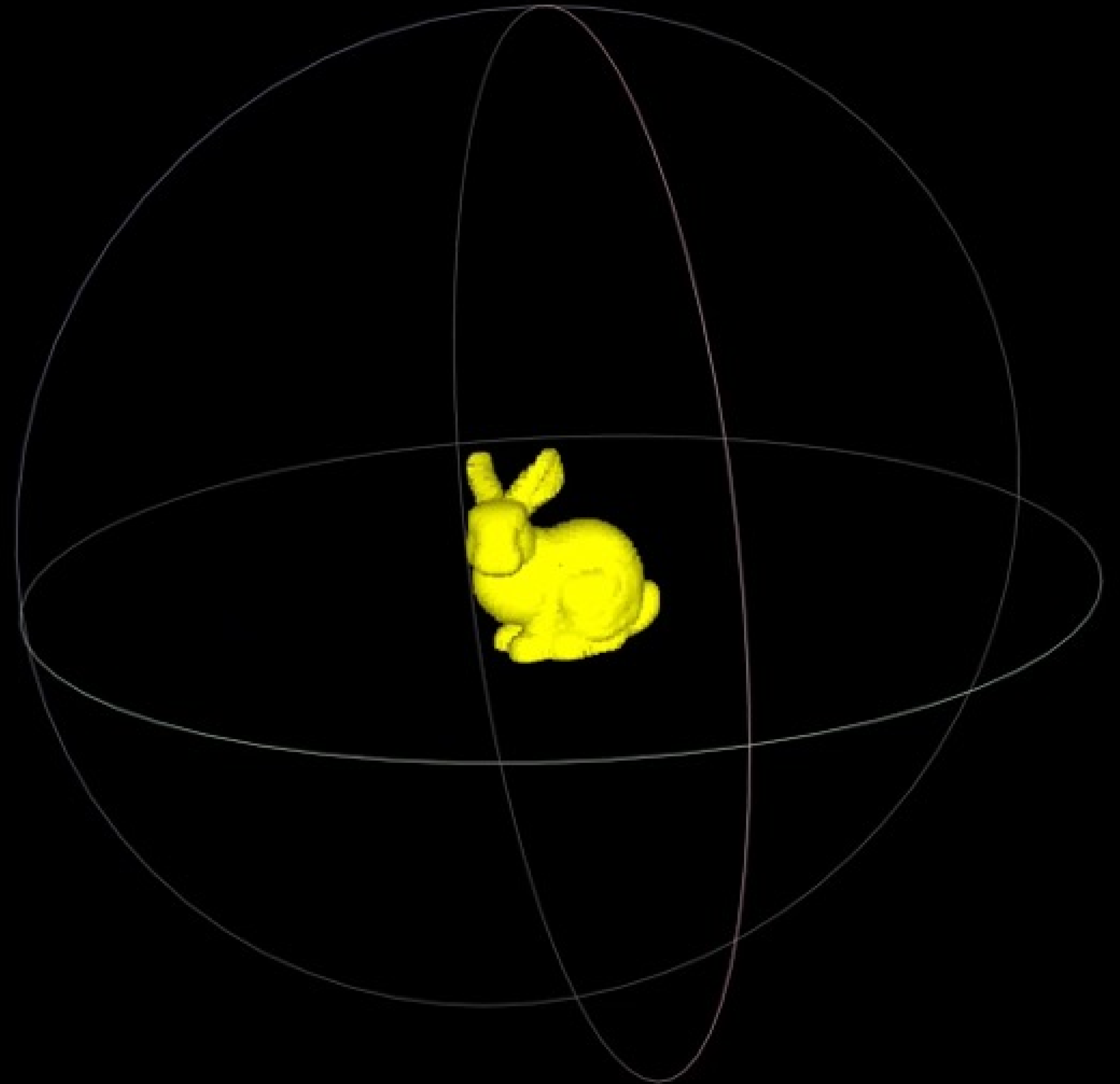
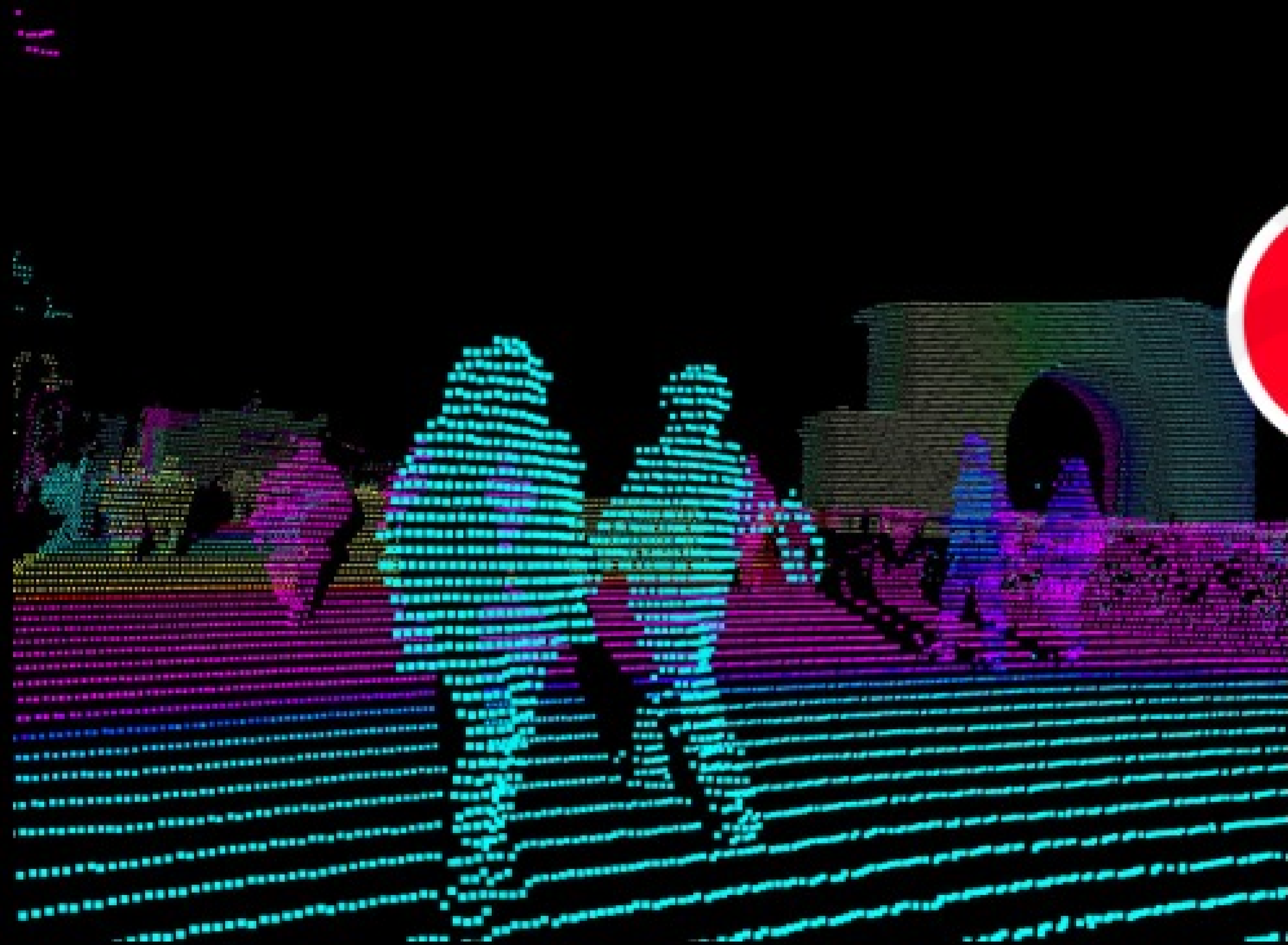
Learning From 3D Data

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



Extending Graph CNNs to Pointclouds

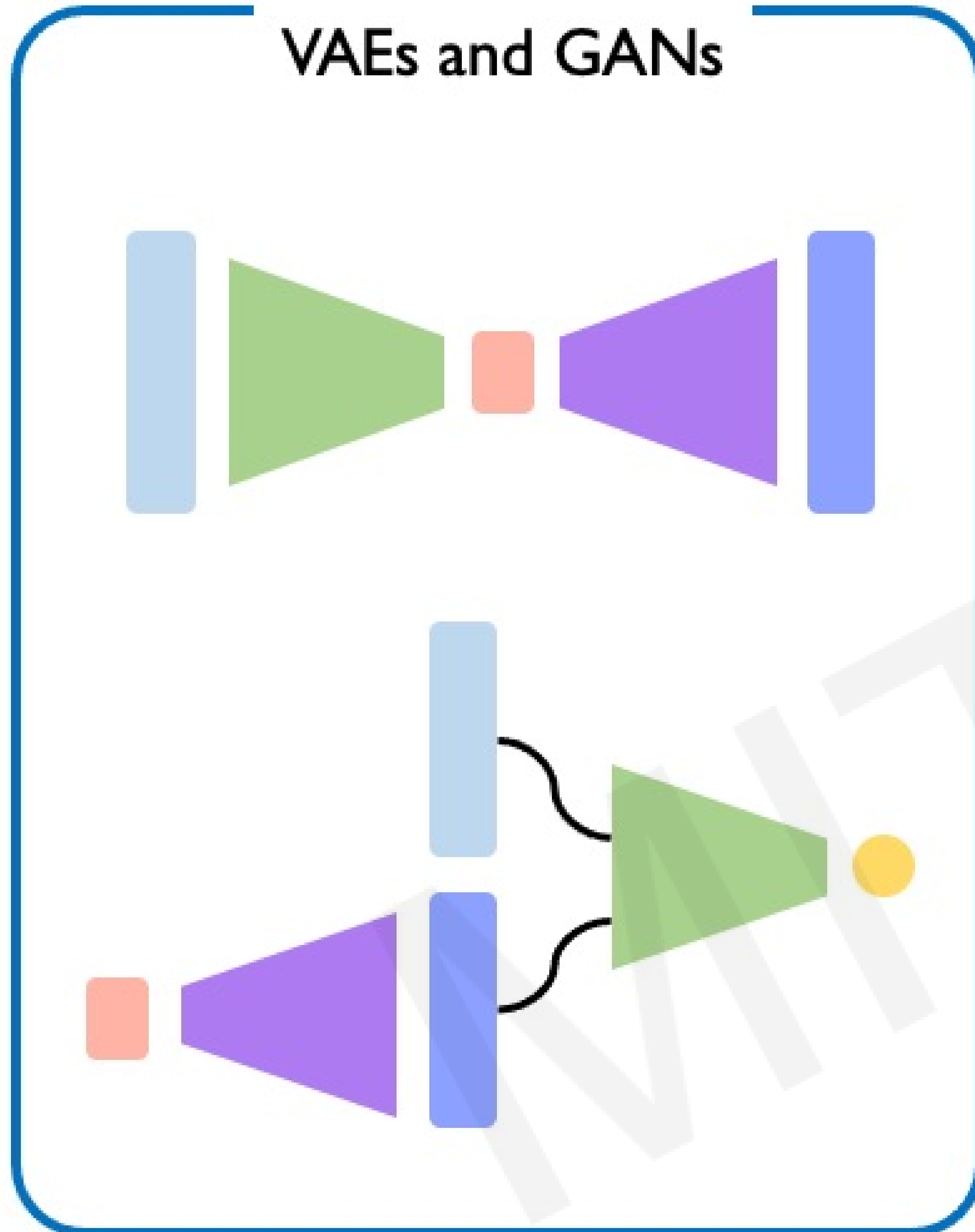
Capture local geometric features of point clouds while maintaining order invariance






New Frontiers II: Diffusion Models & Generative AI

The Landscape of Generative Modeling

Lecture 4: VAEs and GANs



Limitations

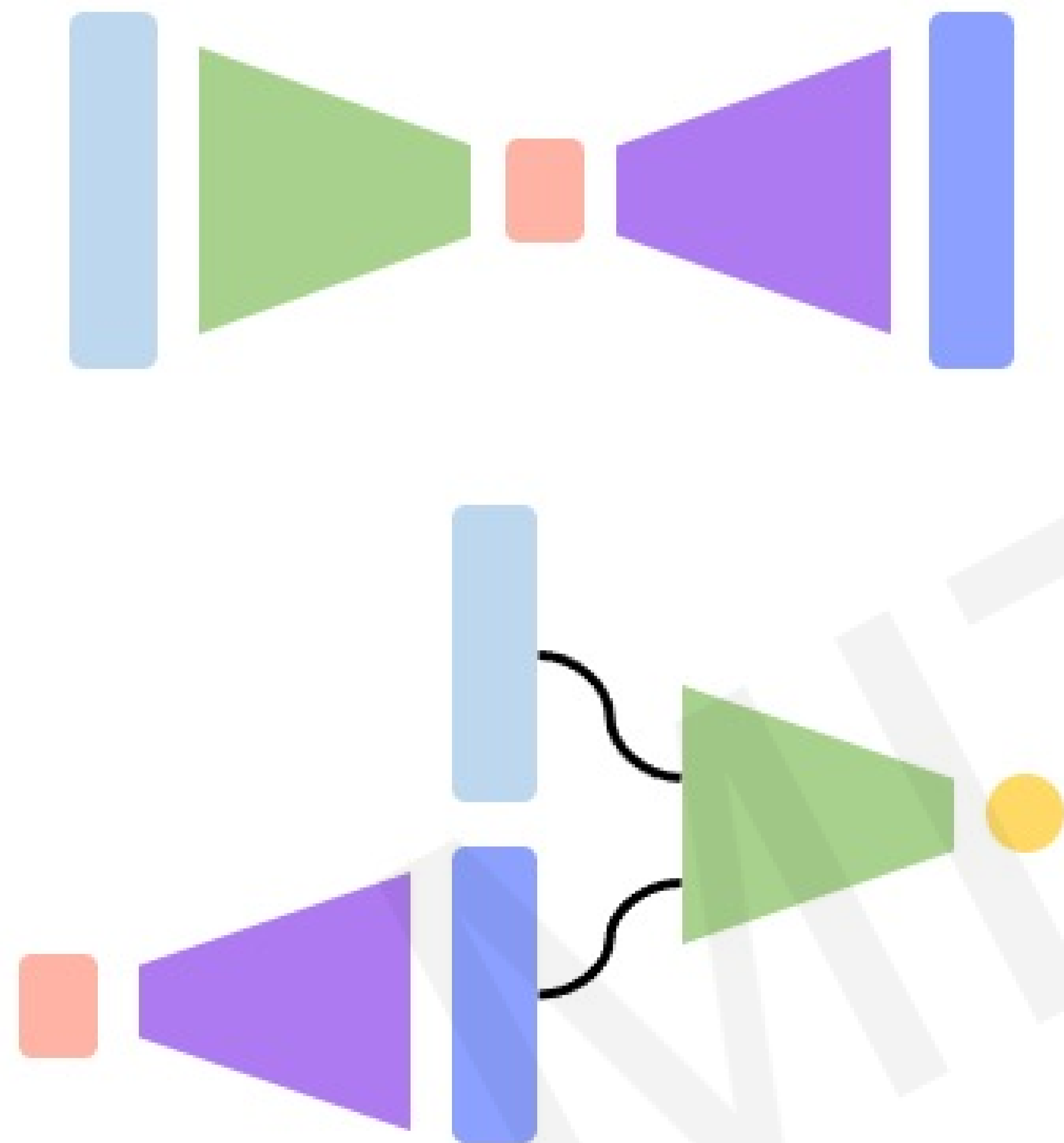
-  Mode collapse
-  Generating OOD
-  Hard to train

Challenges

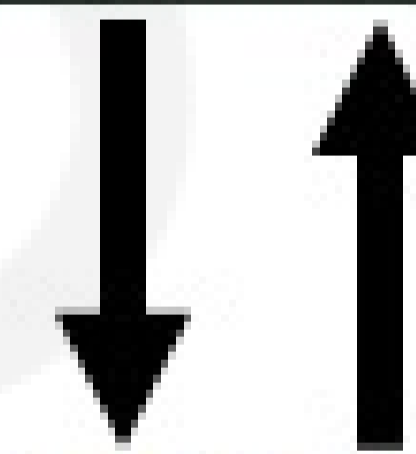
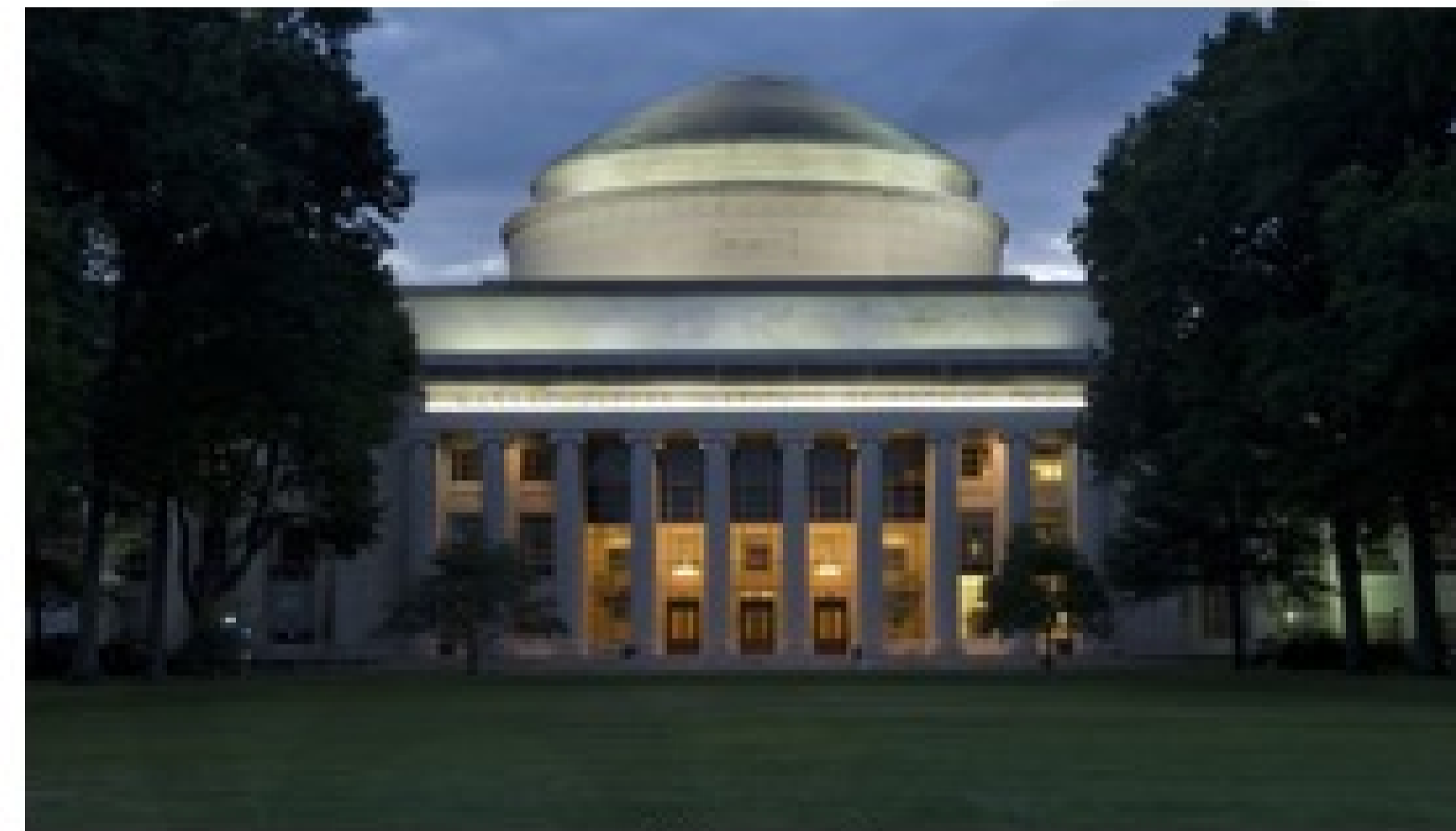
-  Stability
-  Efficiency
-  Quality
-  Novelty

The Landscape of Generative Modeling

Lecture 4: VAEs and GANs



Diffusion Models



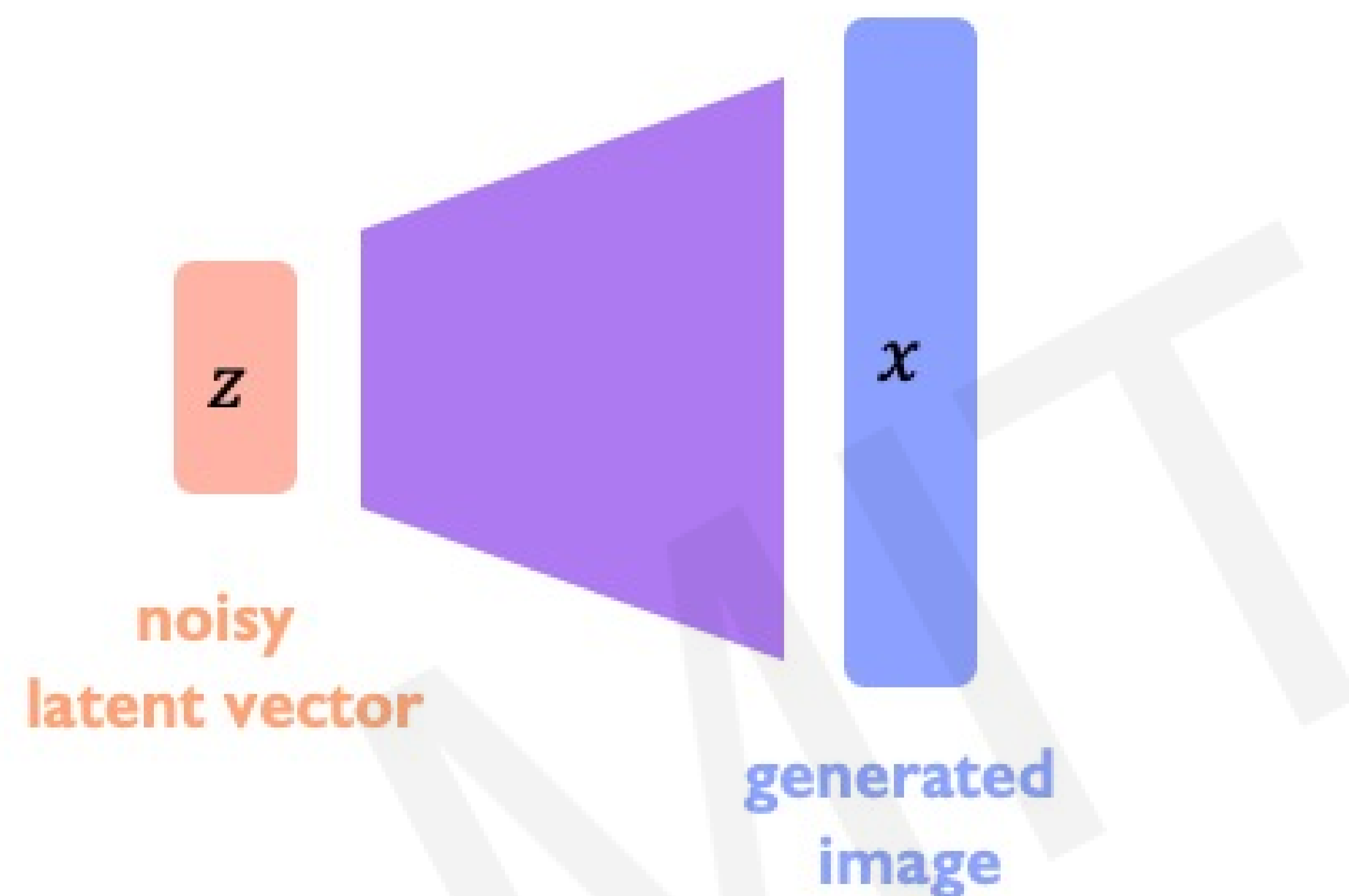
Text-to-Image



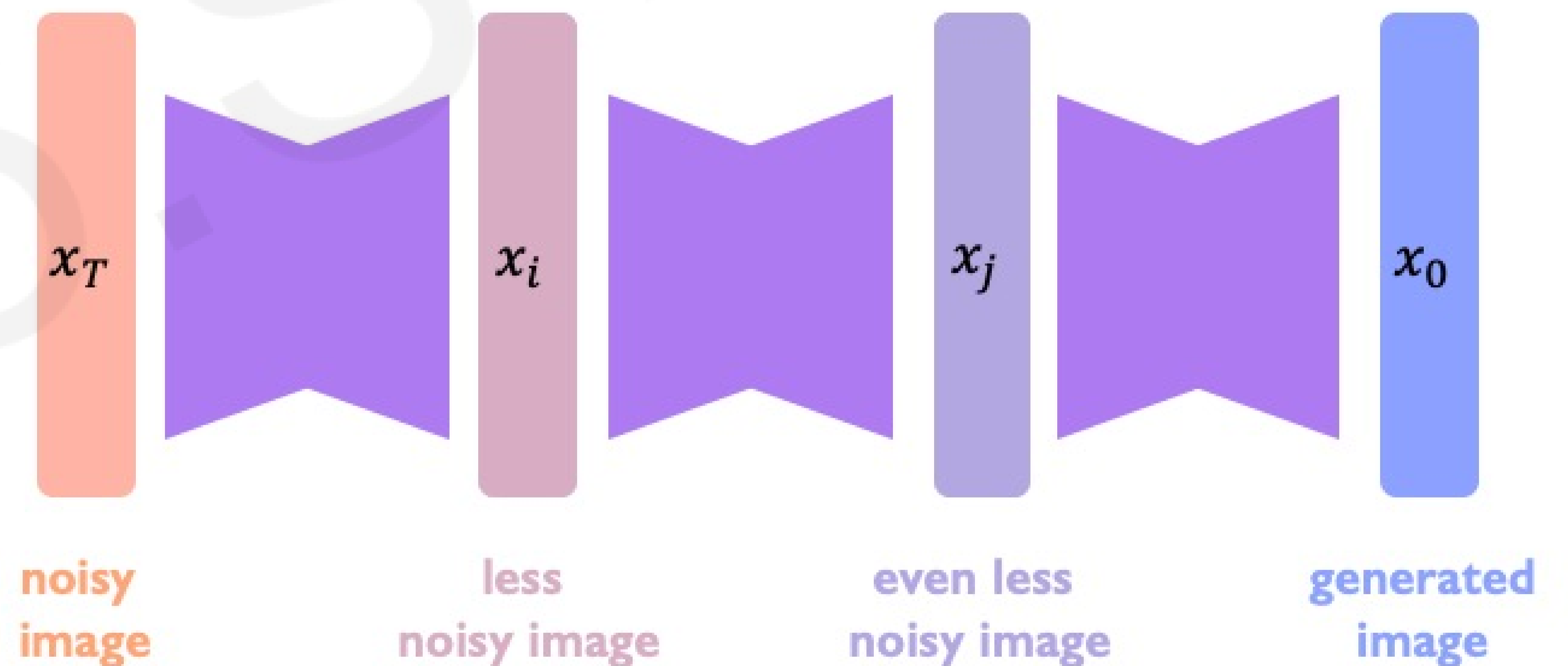
“Two cats doing research”

Diffusion Models

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

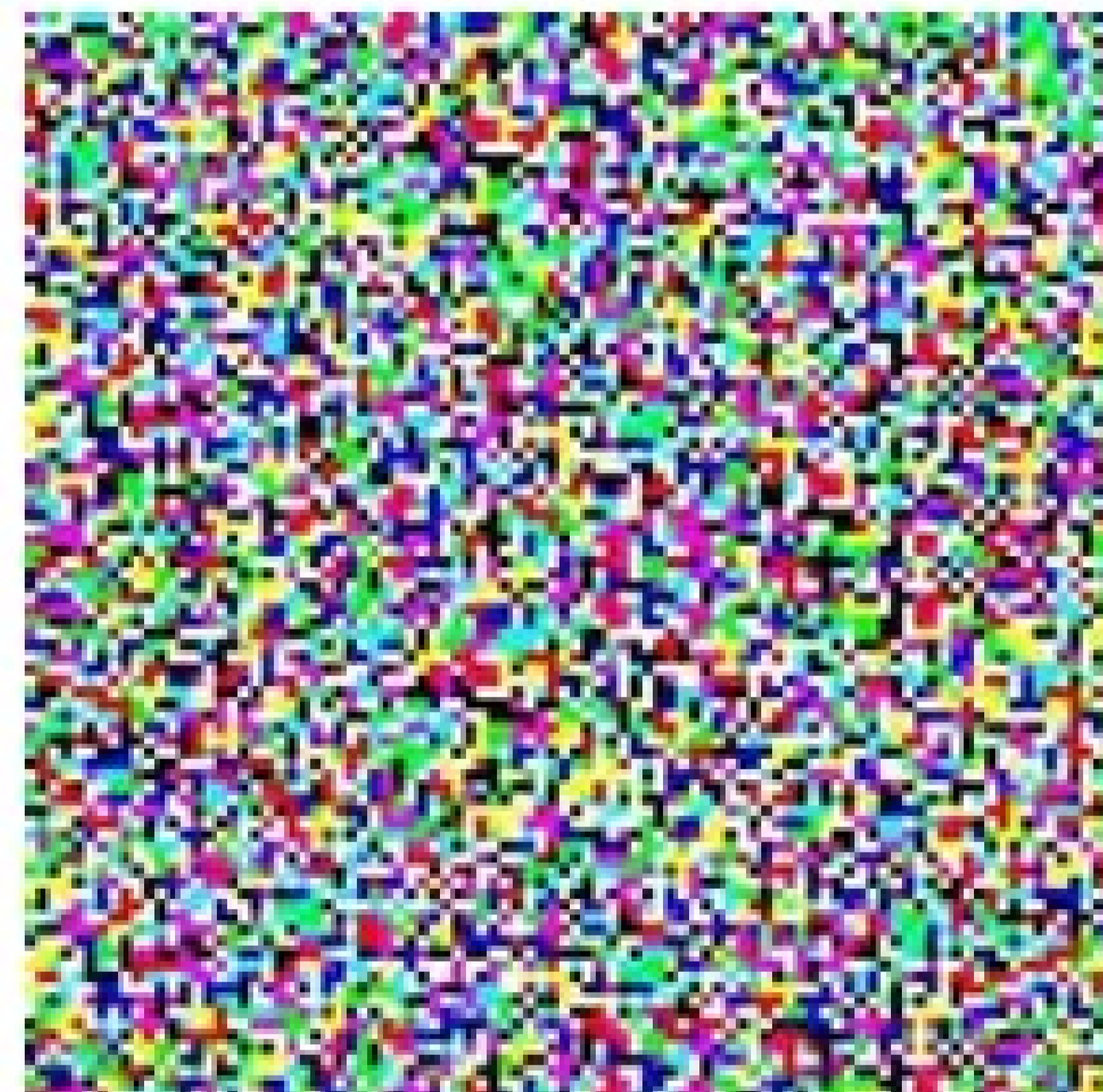
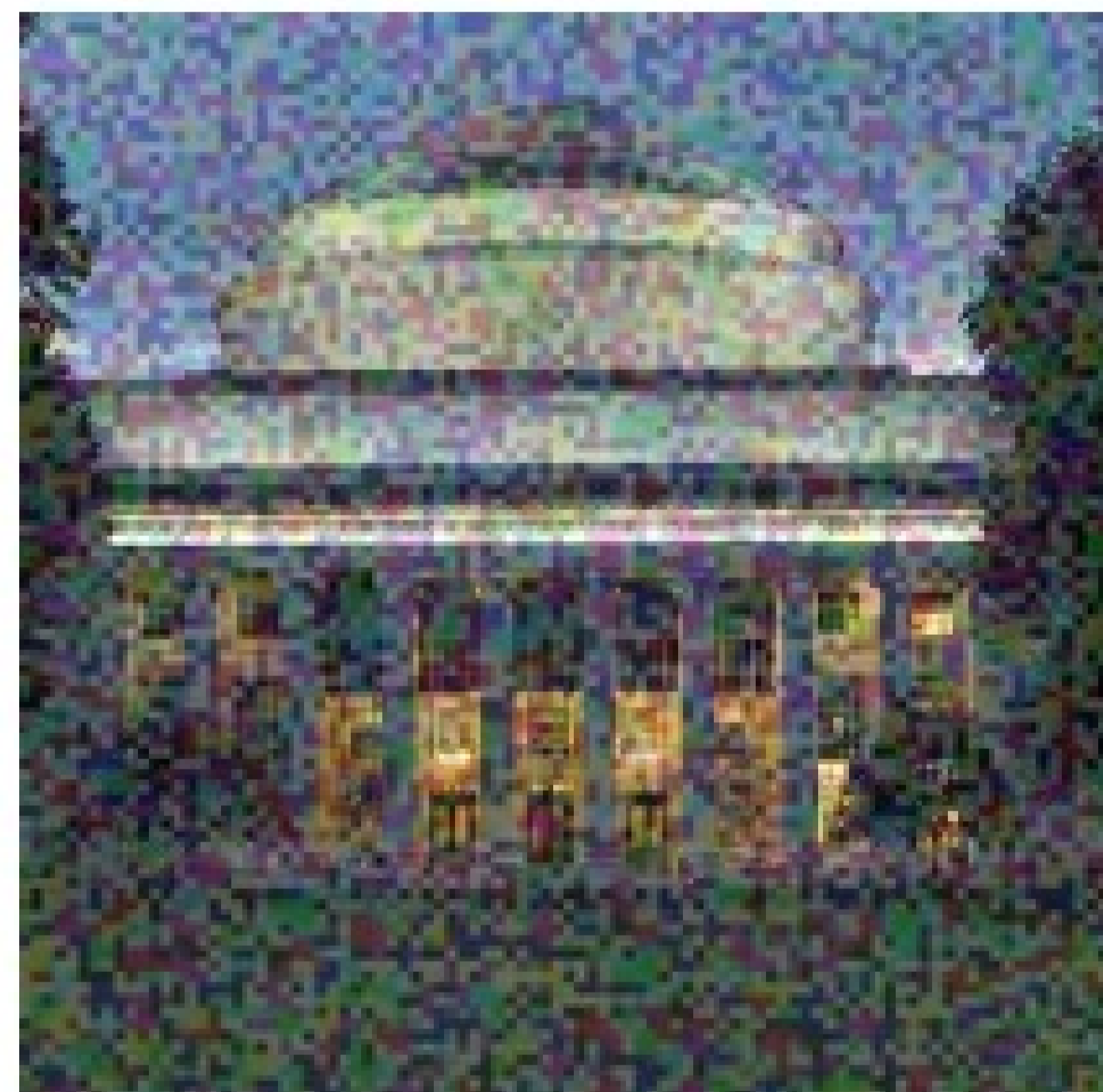
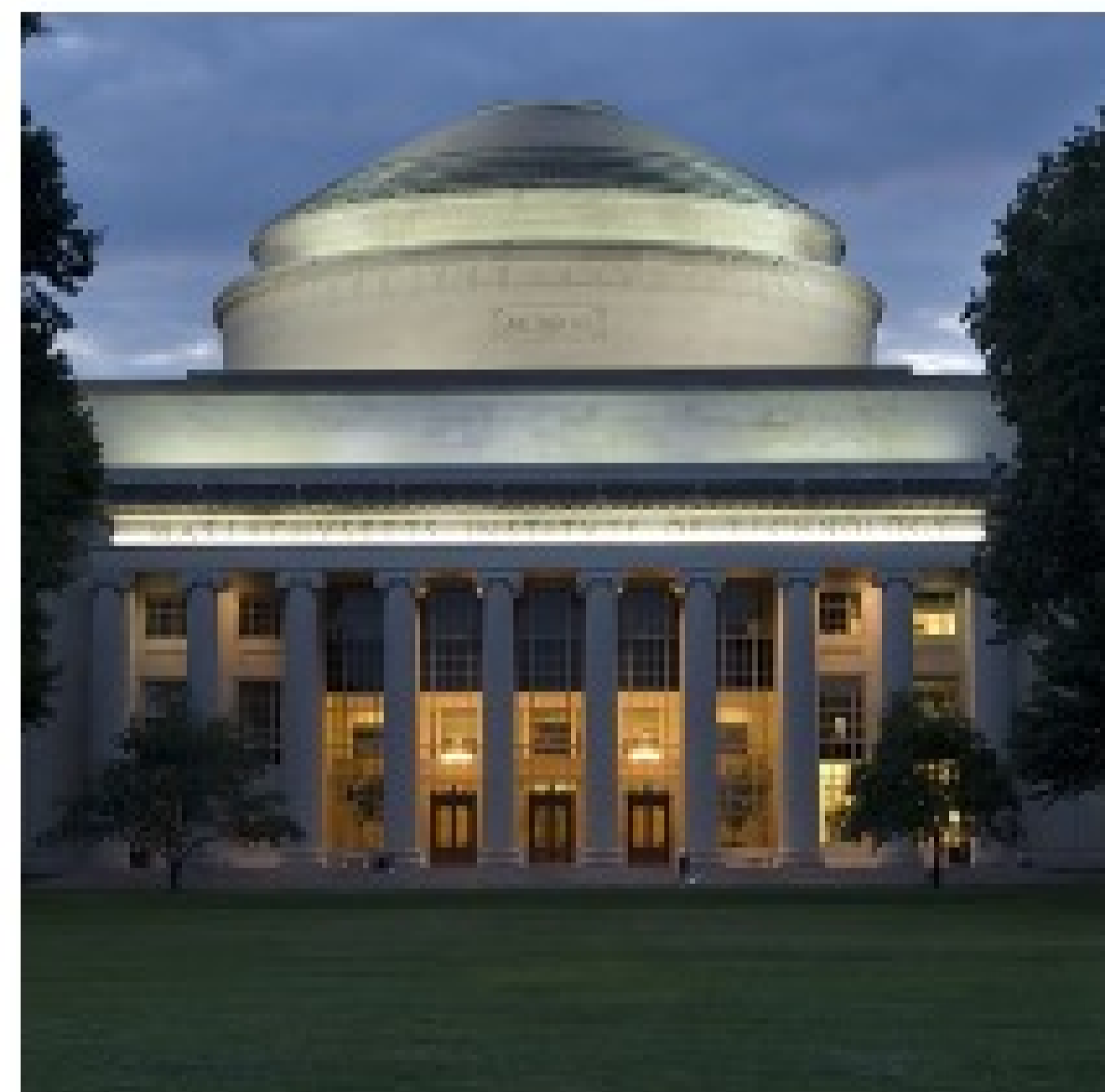


Diffusion: Generating images iteratively by repeatedly refining and removing noise



The Diffusion Process

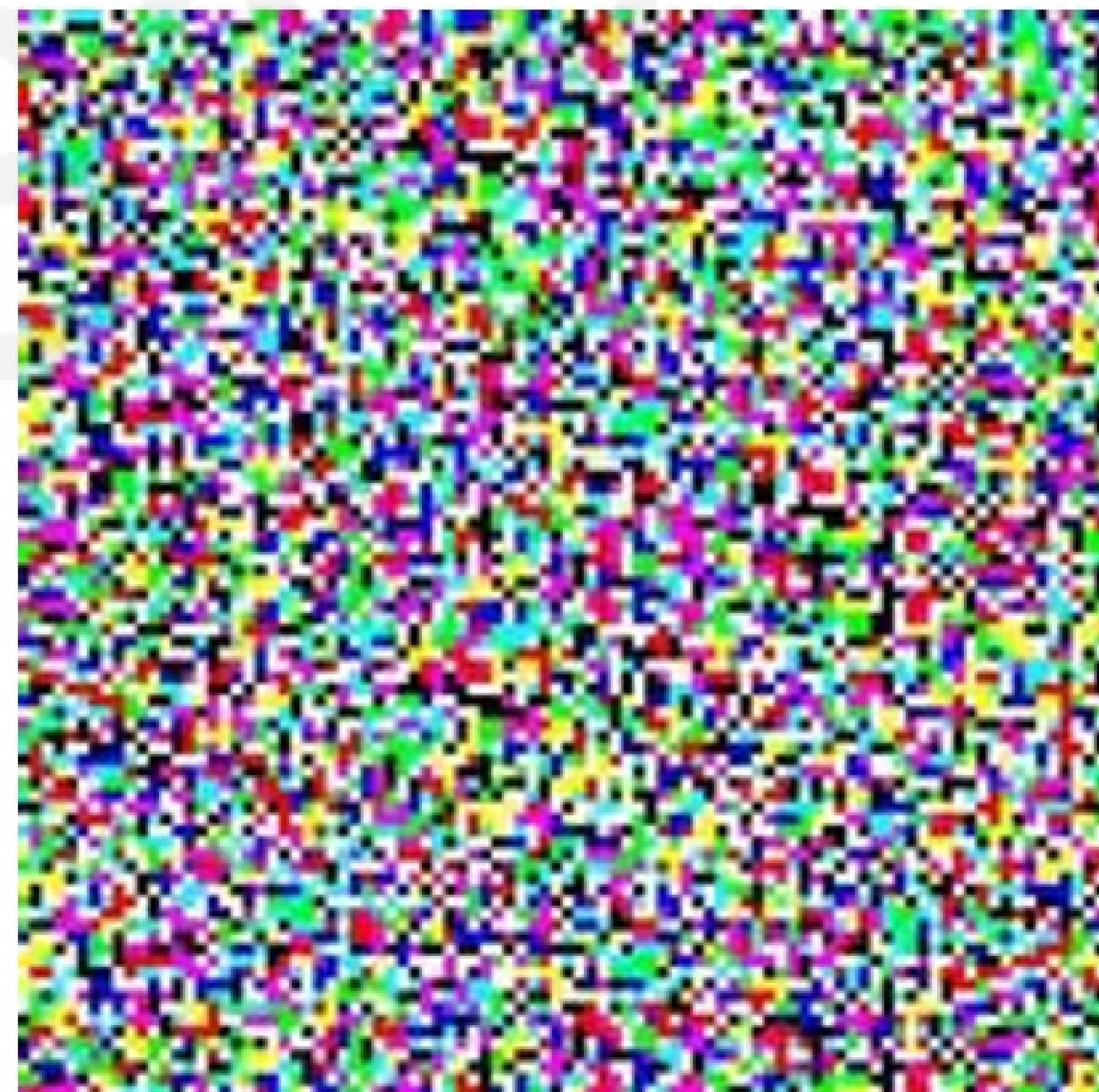
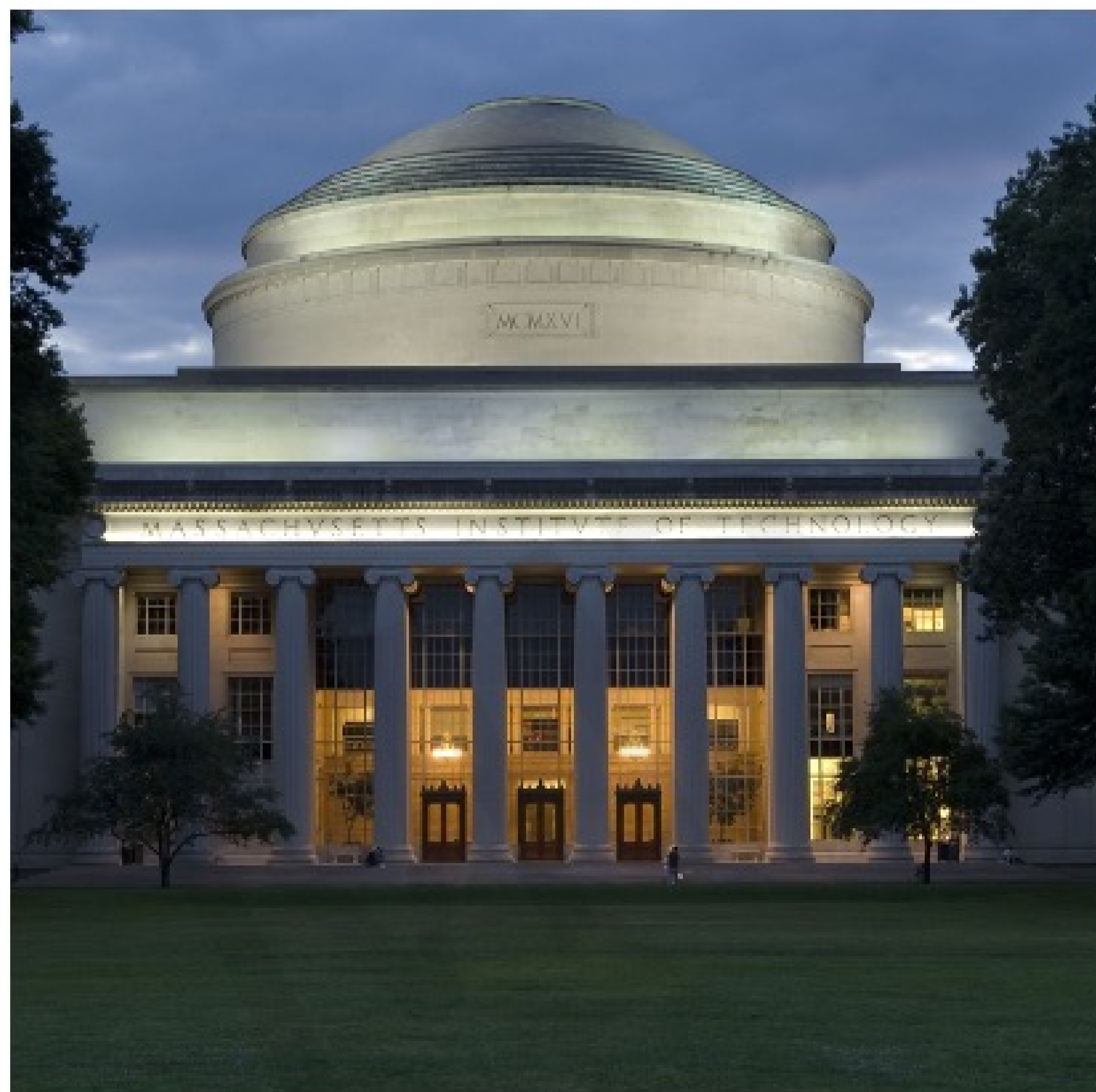
Forward noising
(data-to-noise)



Reverse denoising
(noise-to-data)

Forward Noising

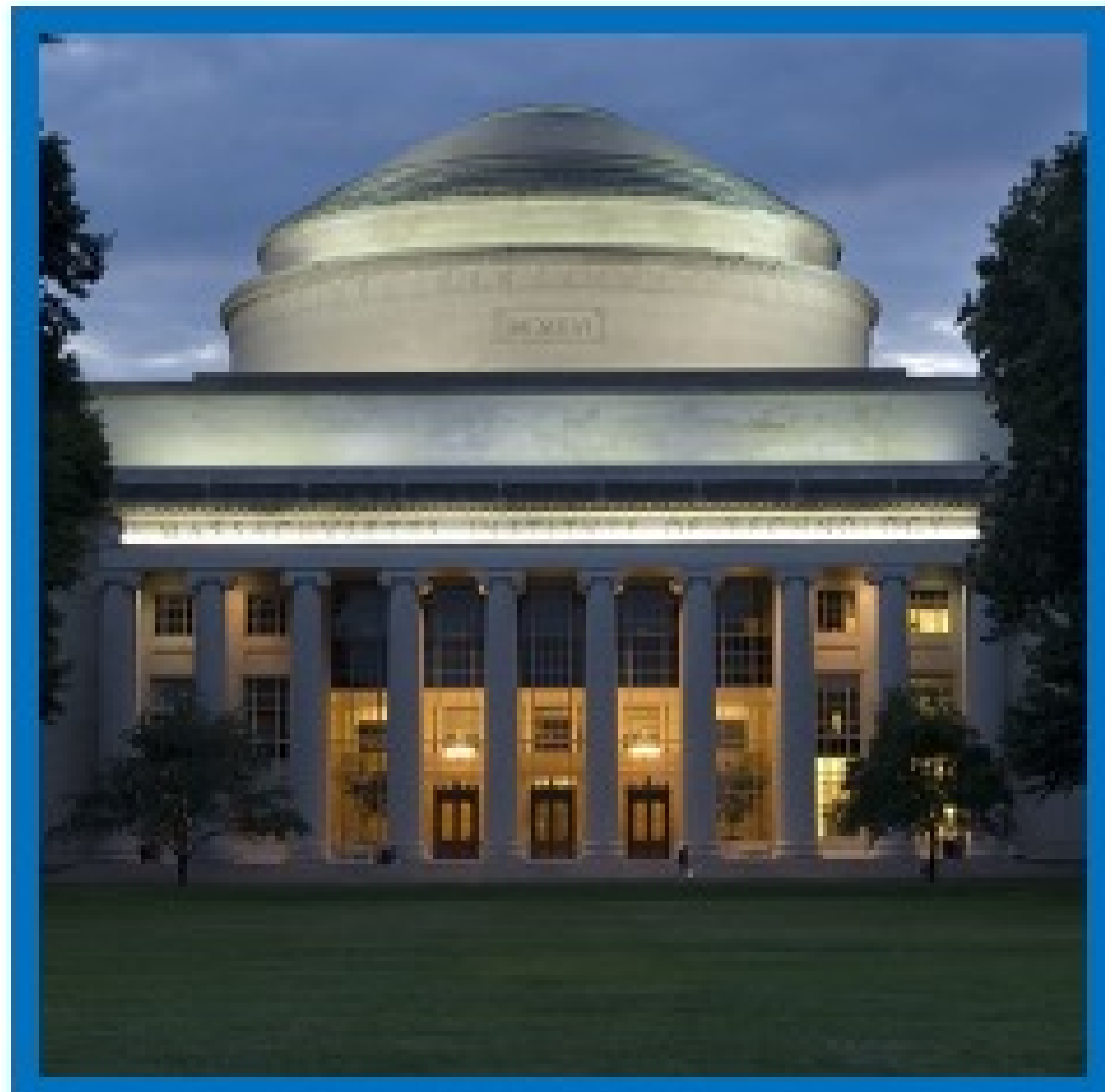
Step 1: Given an image (left), randomly sample a random noise pattern (right)



Forward Noising

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

T = 0



100% image
0% noise

T = 1



75% image
25% noise

T = 2



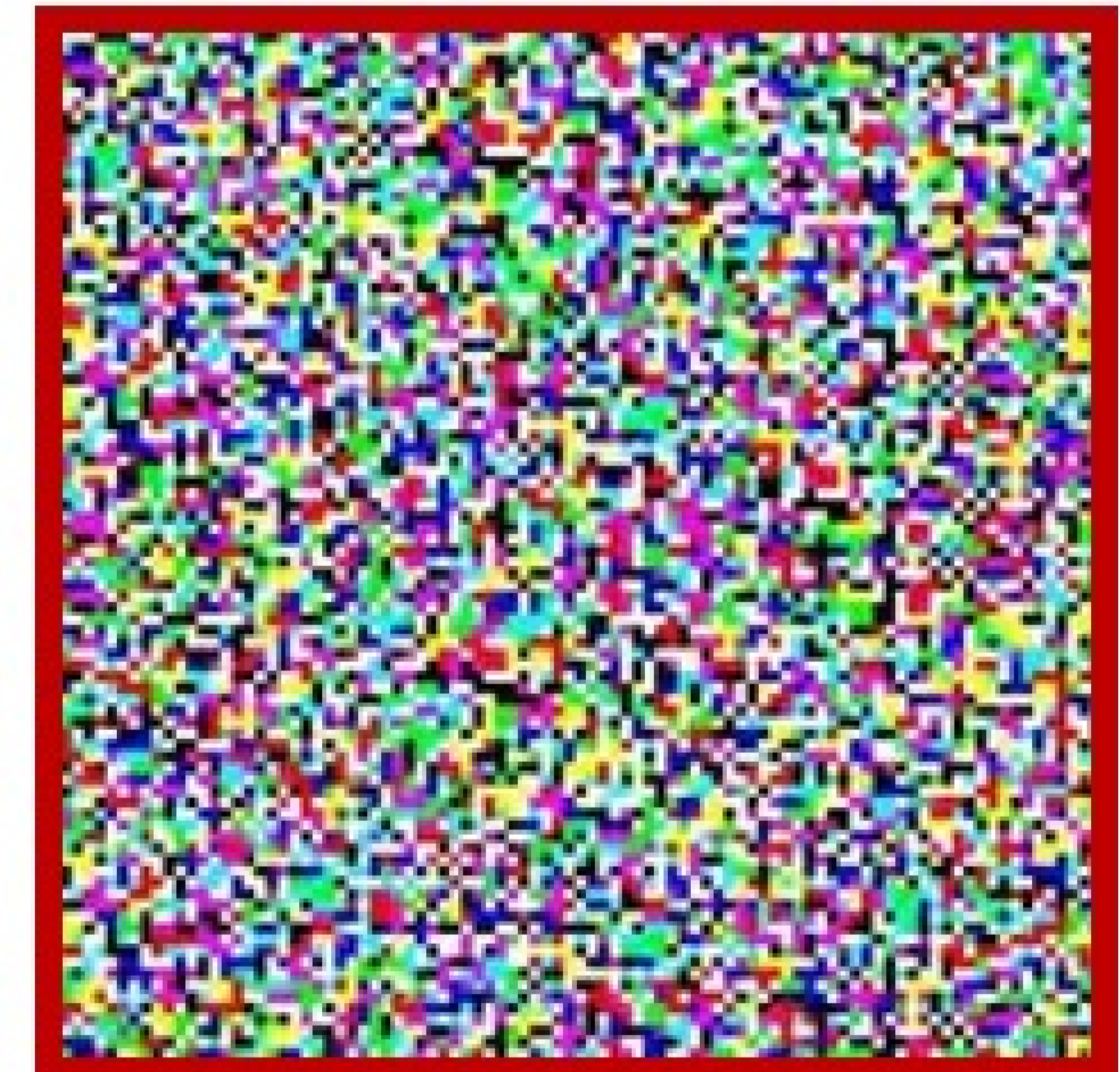
50% image
50% noise

T = 3



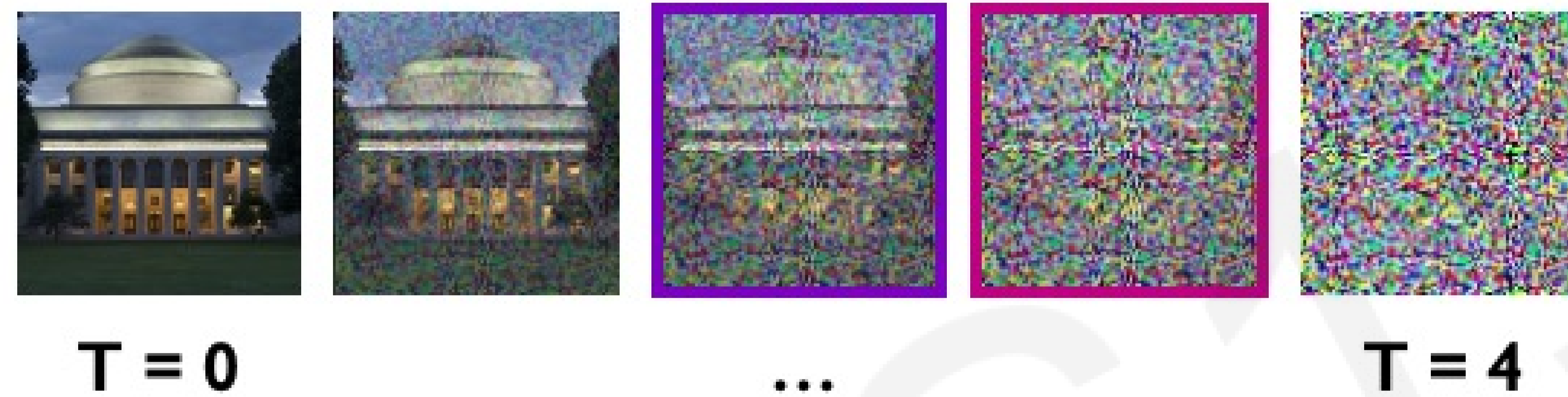
25% image
75% noise

T = 4

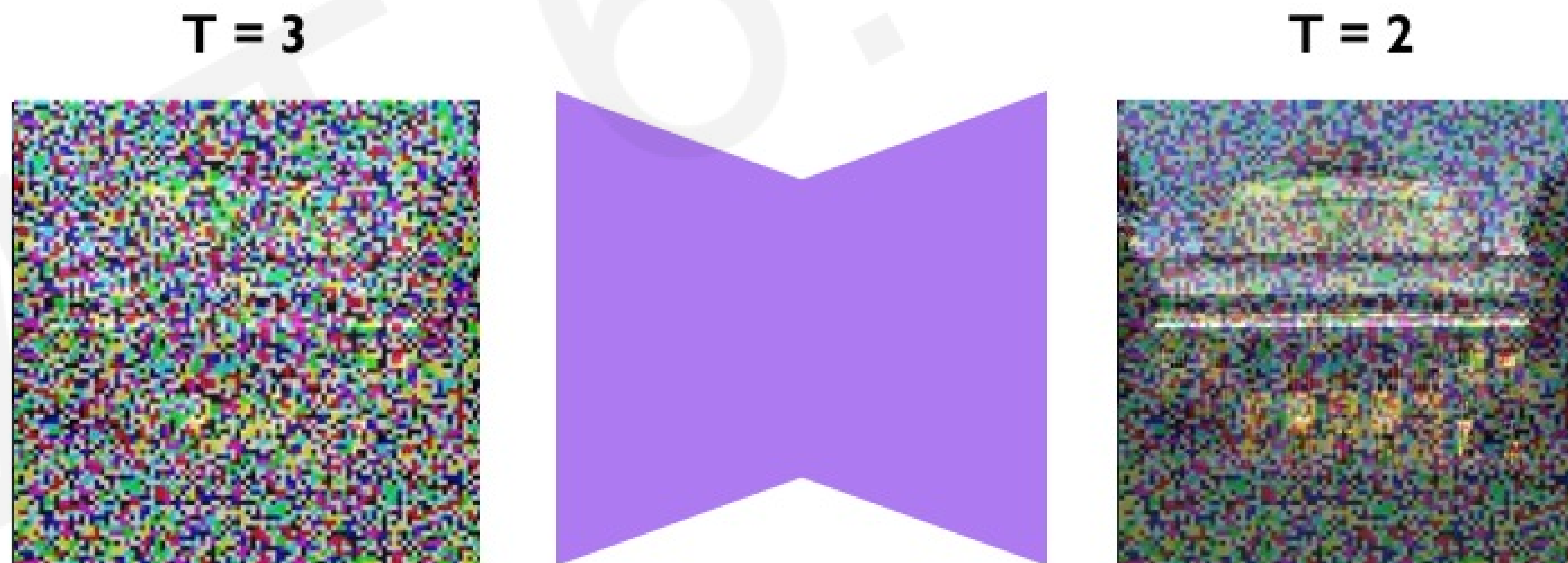


0% image
100% noise

Reverse Denoising



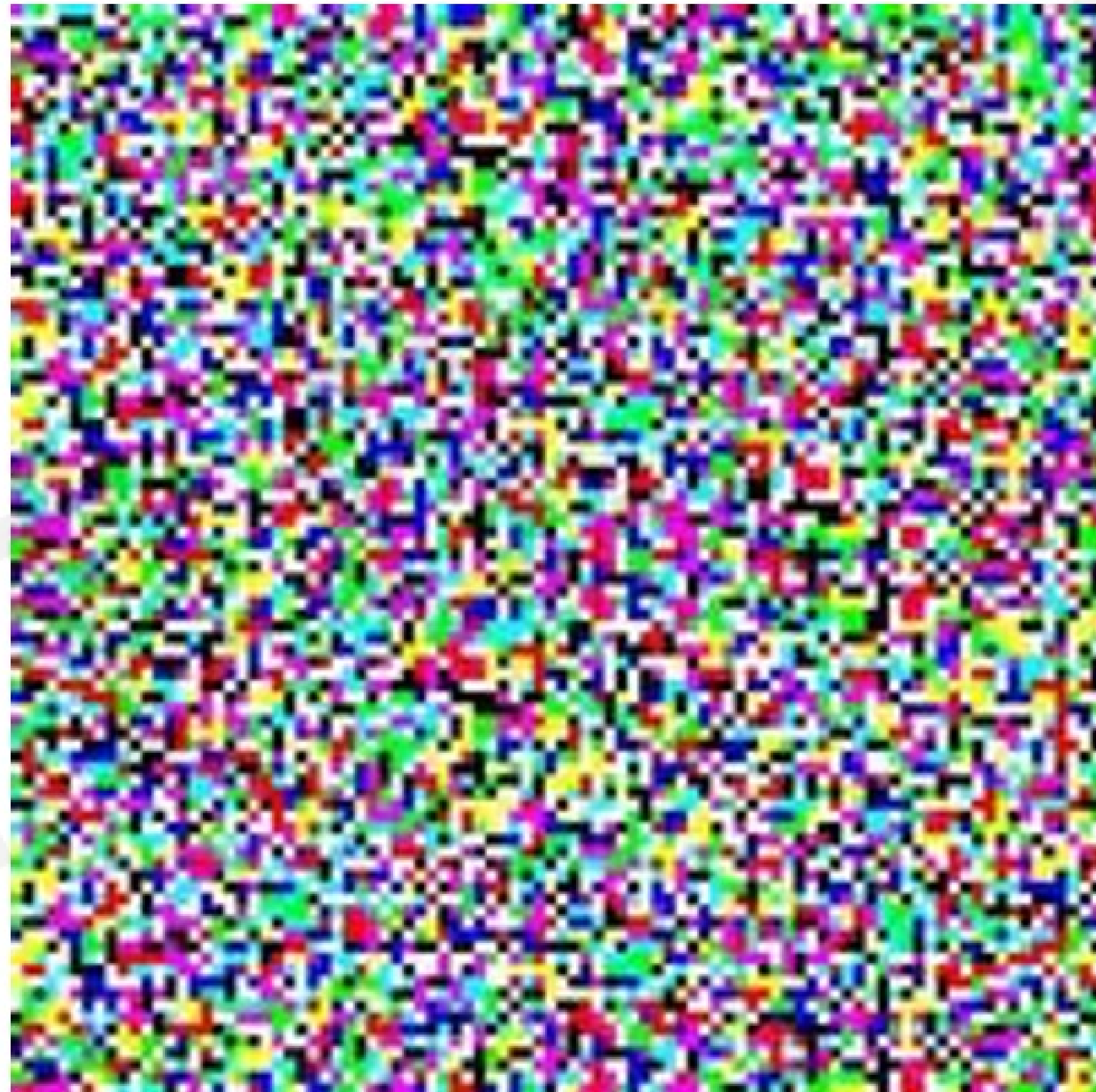
Goal: Given image at T, can we learn to estimate image at T-1?



How can we train this network?

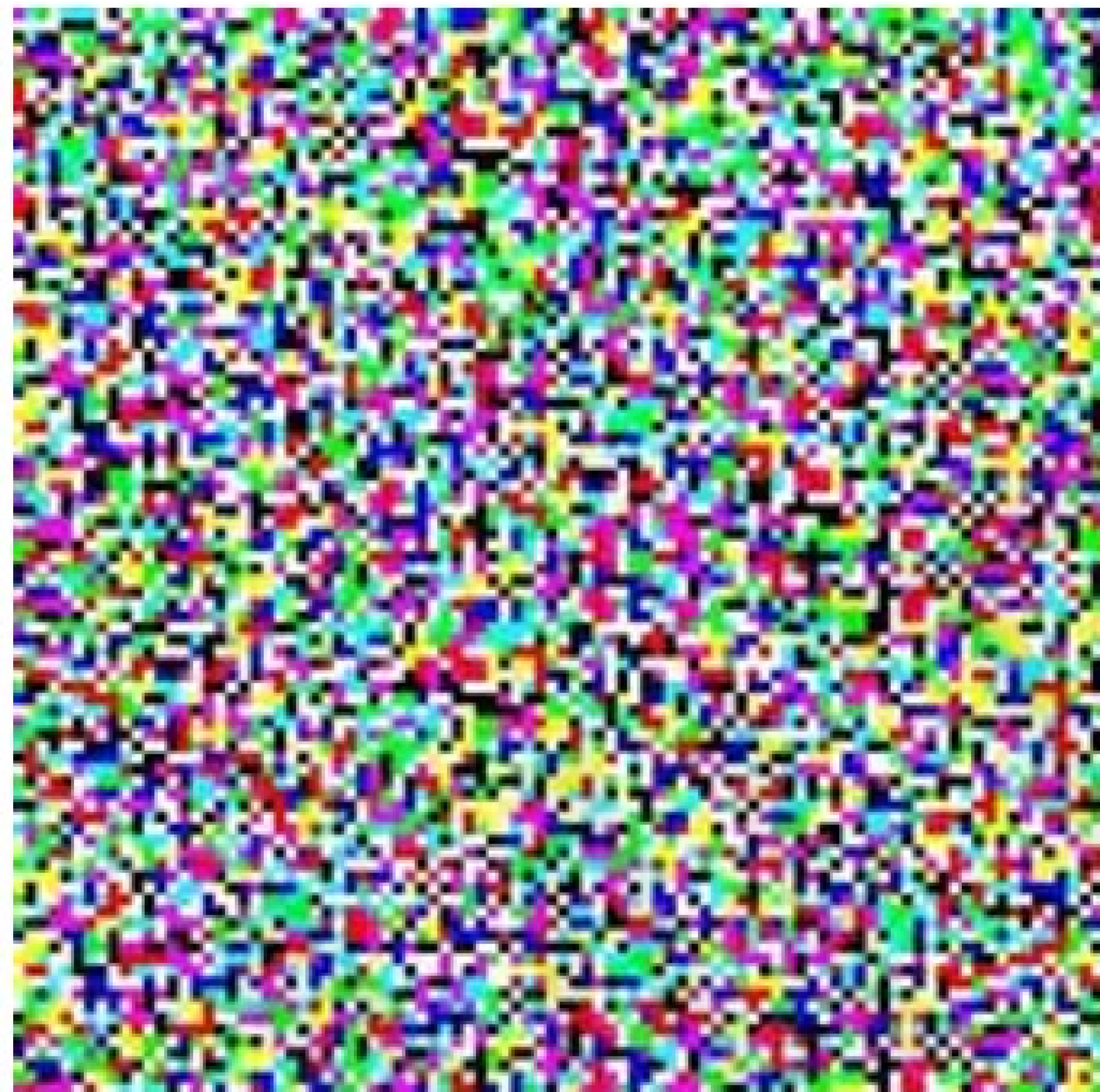
Sampling Brand New Generations

T

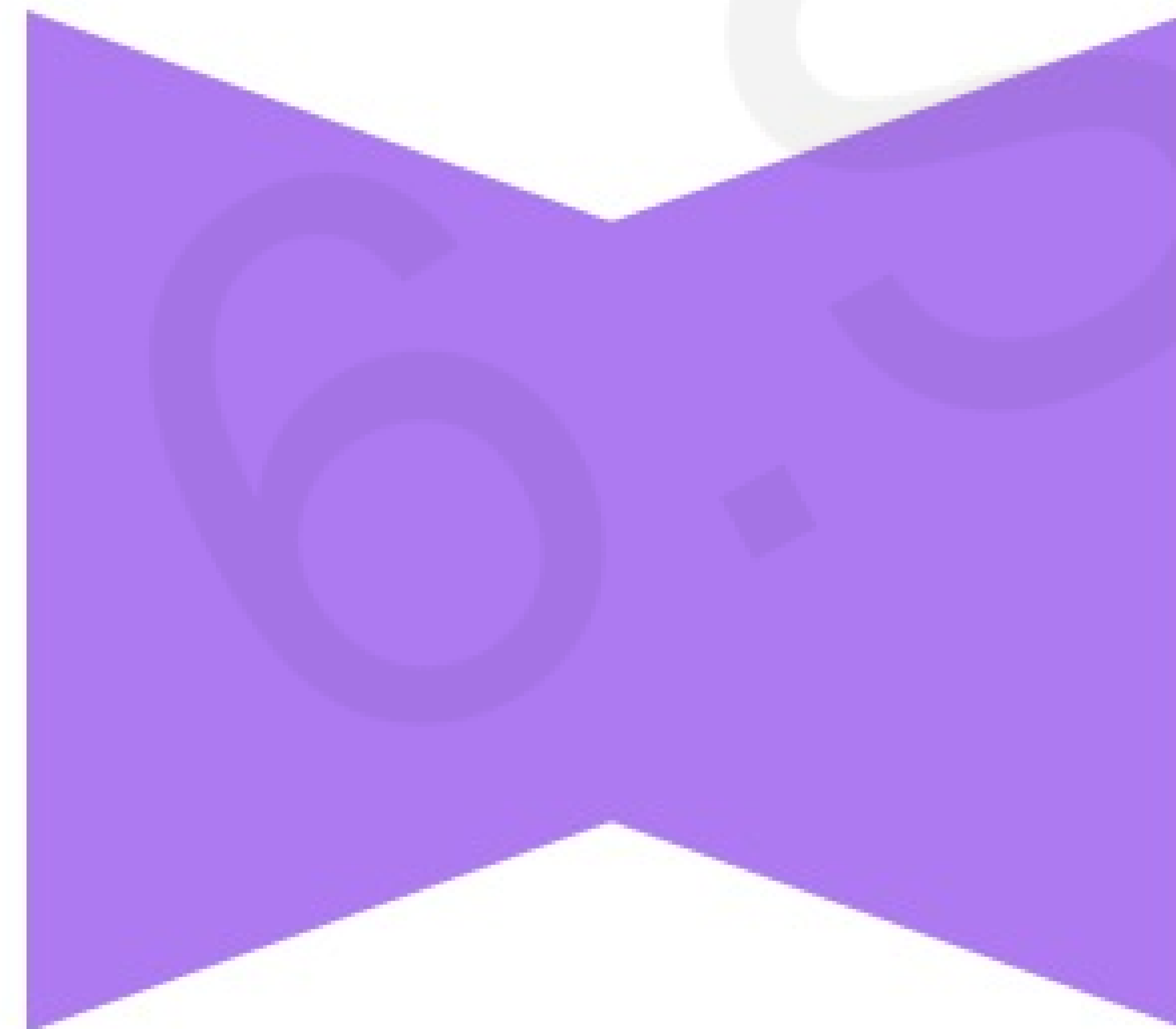


Sampling Brand New Generations

T

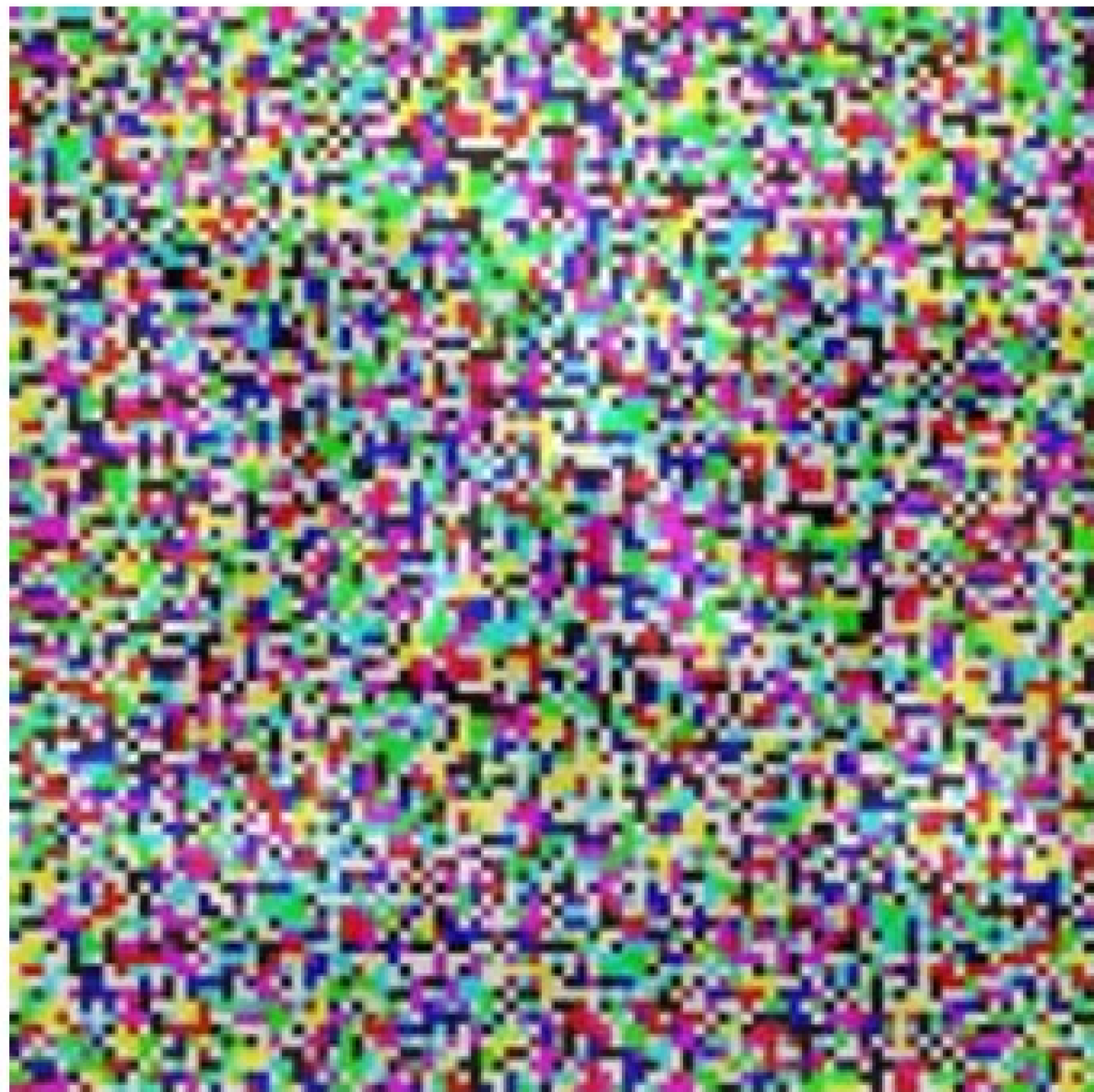


T-1



Sampling Brand New Generations

T-1

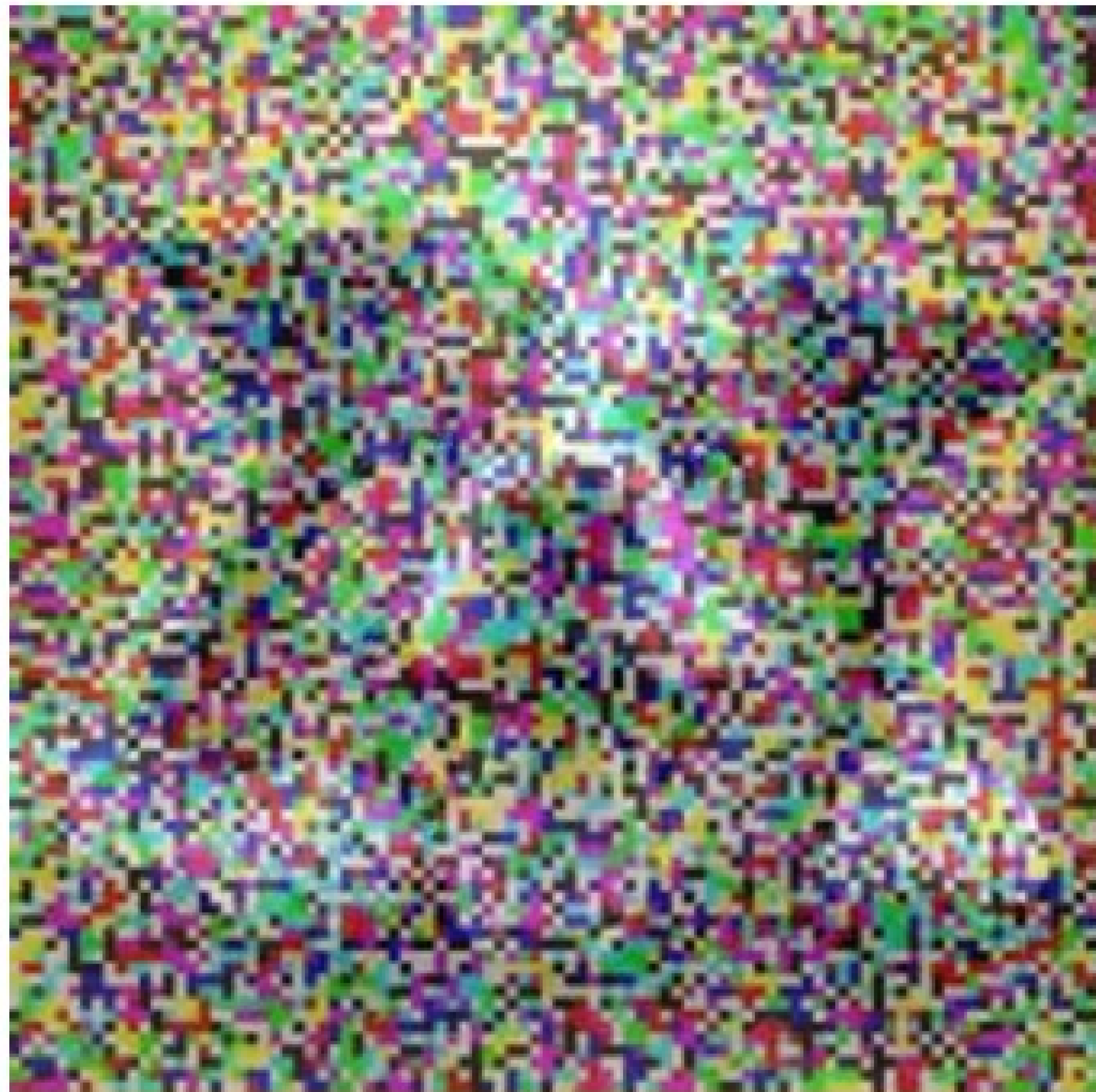


T-2



Sampling Brand New Generations

T-2



T-3



Sampling Brand New Generations

T-3



T-4



Sampling Brand New Generations

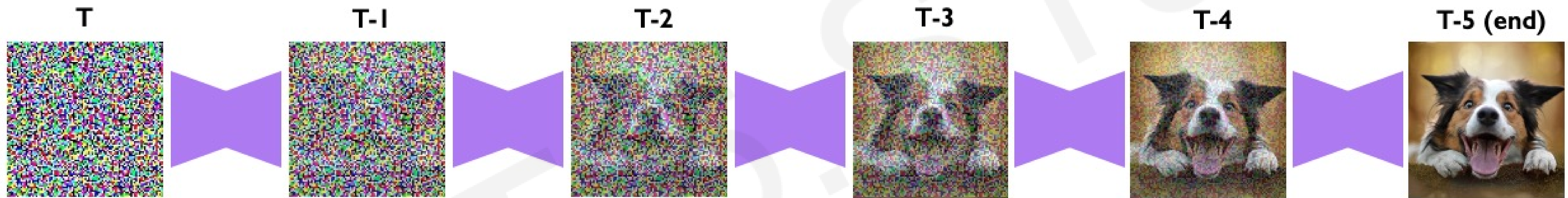
T-4



T-5 (end)



Sampling Brand New Generations

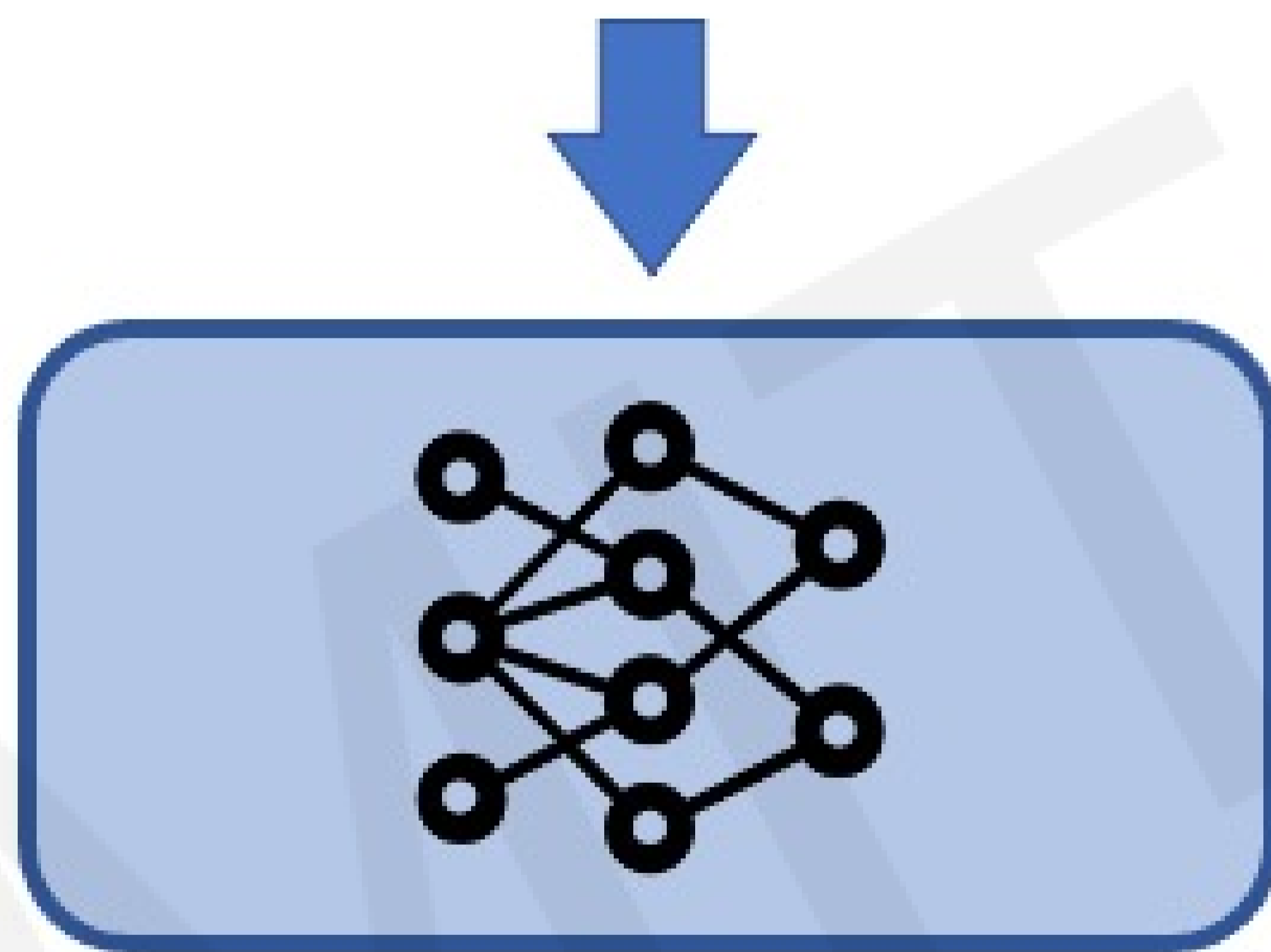






Generating Images from Natural Language

“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”

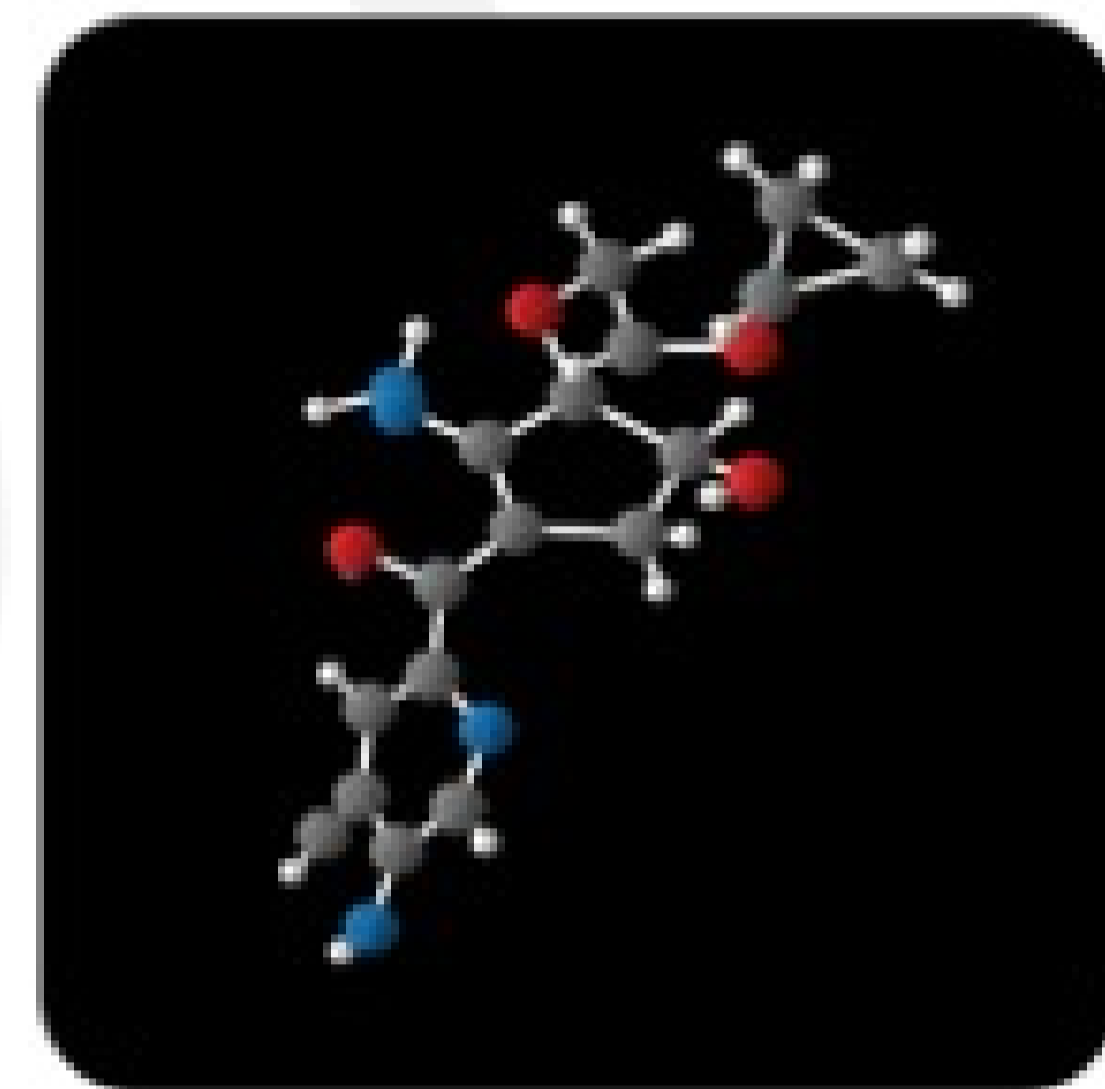
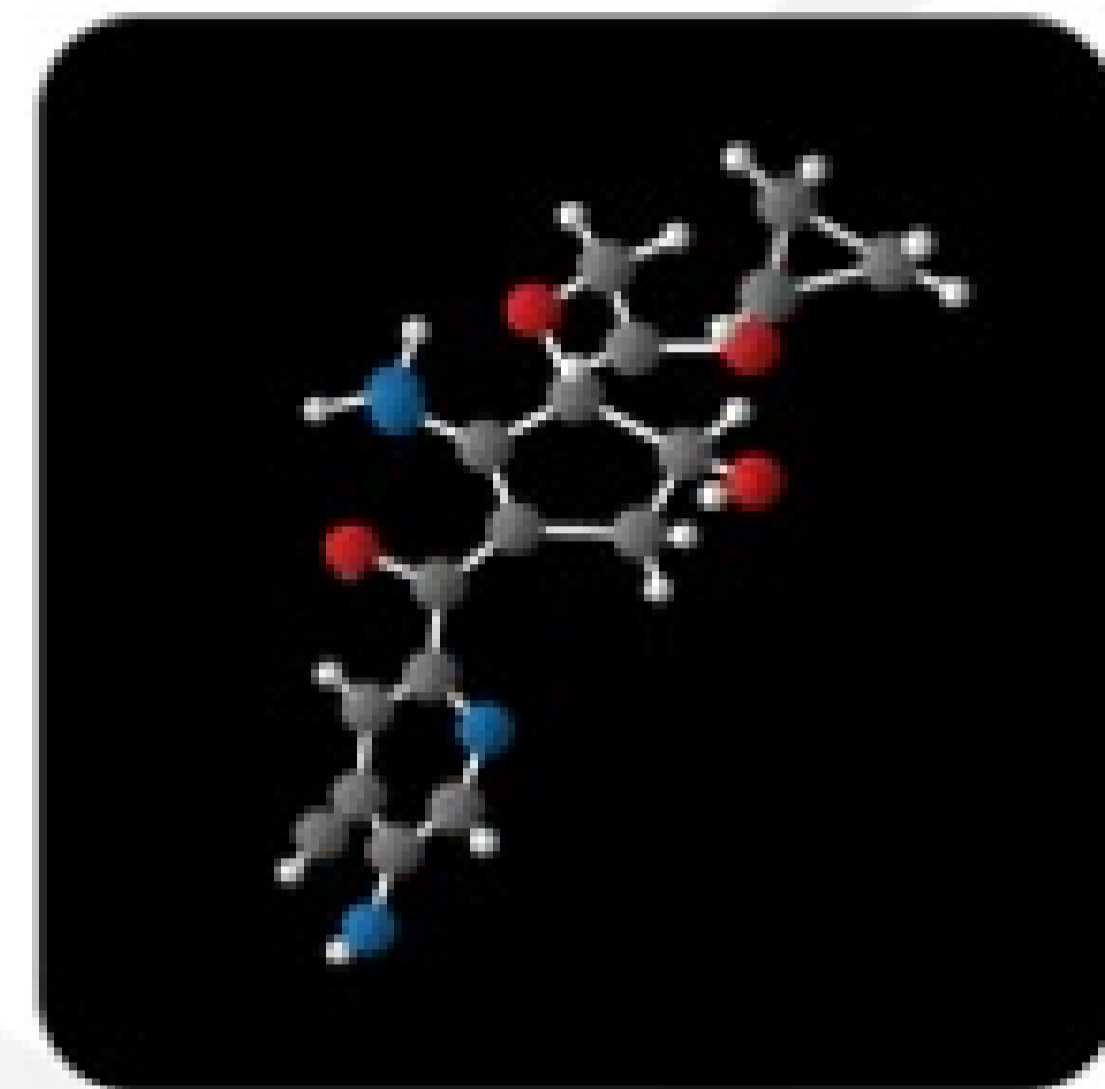
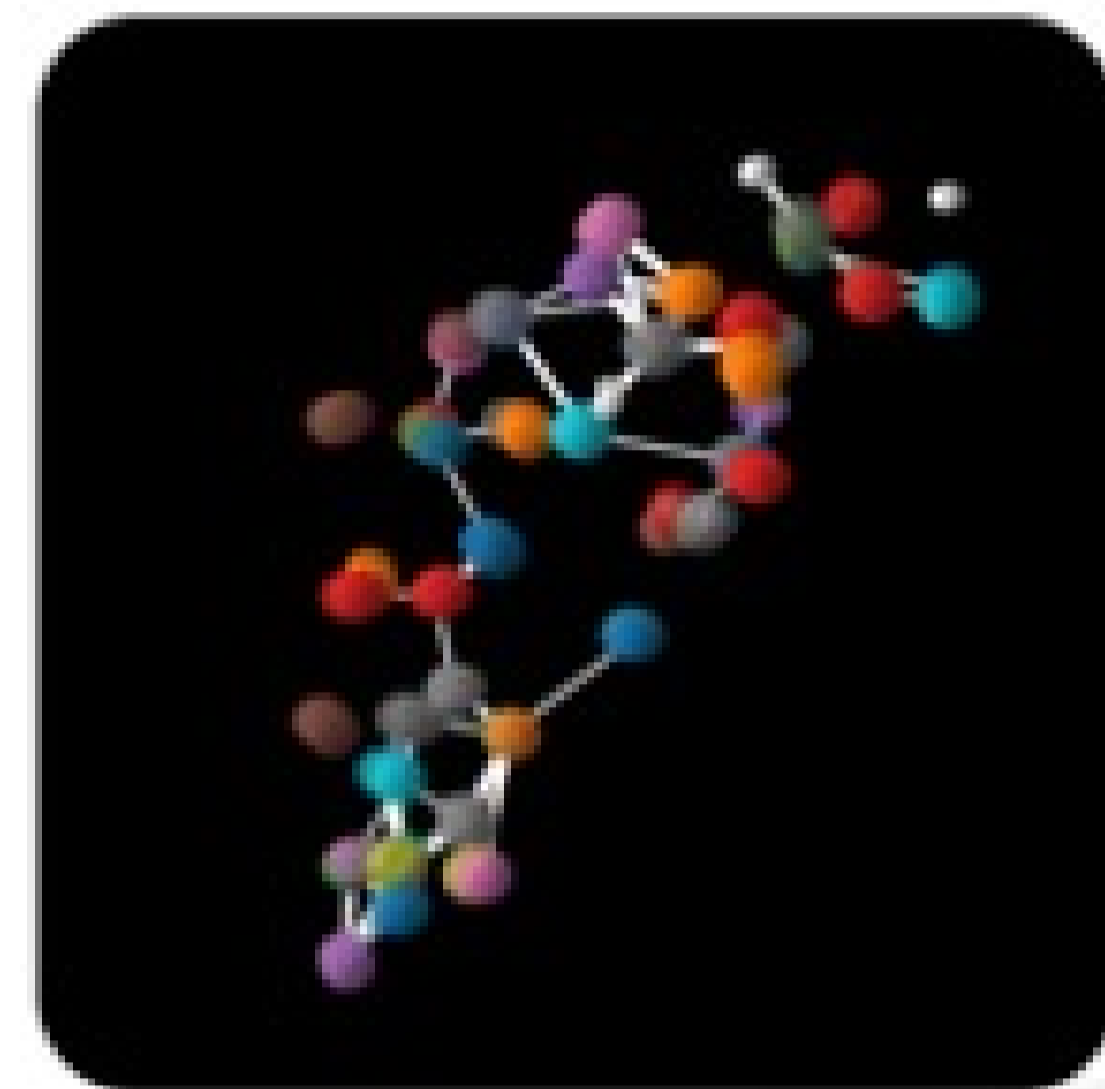
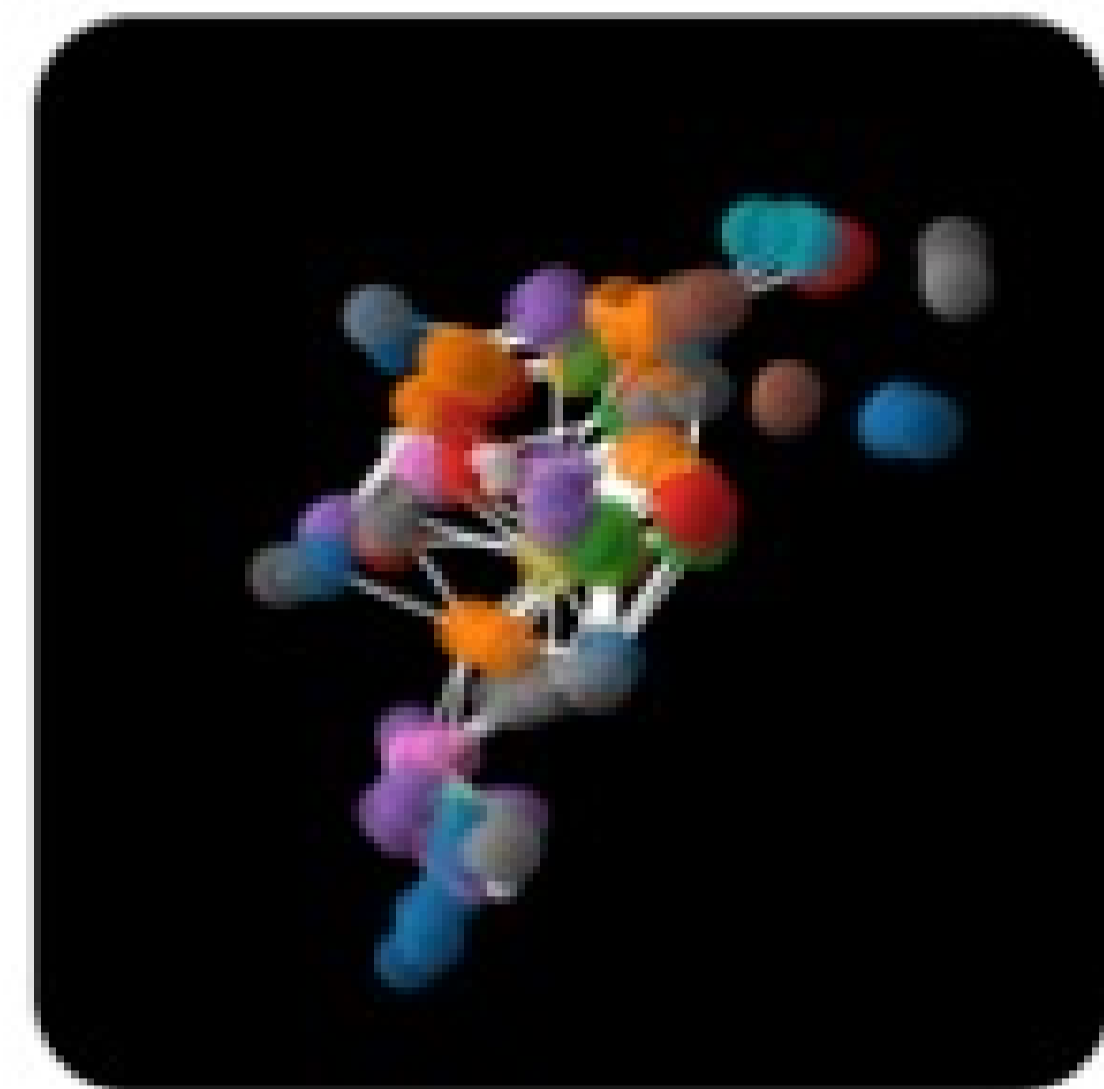


“close-up of a snow leopard in the snow hunting, rack focus, nature photography”



Beyond Images: Molecular Design

Chemistry: Generating Molecules in 3D

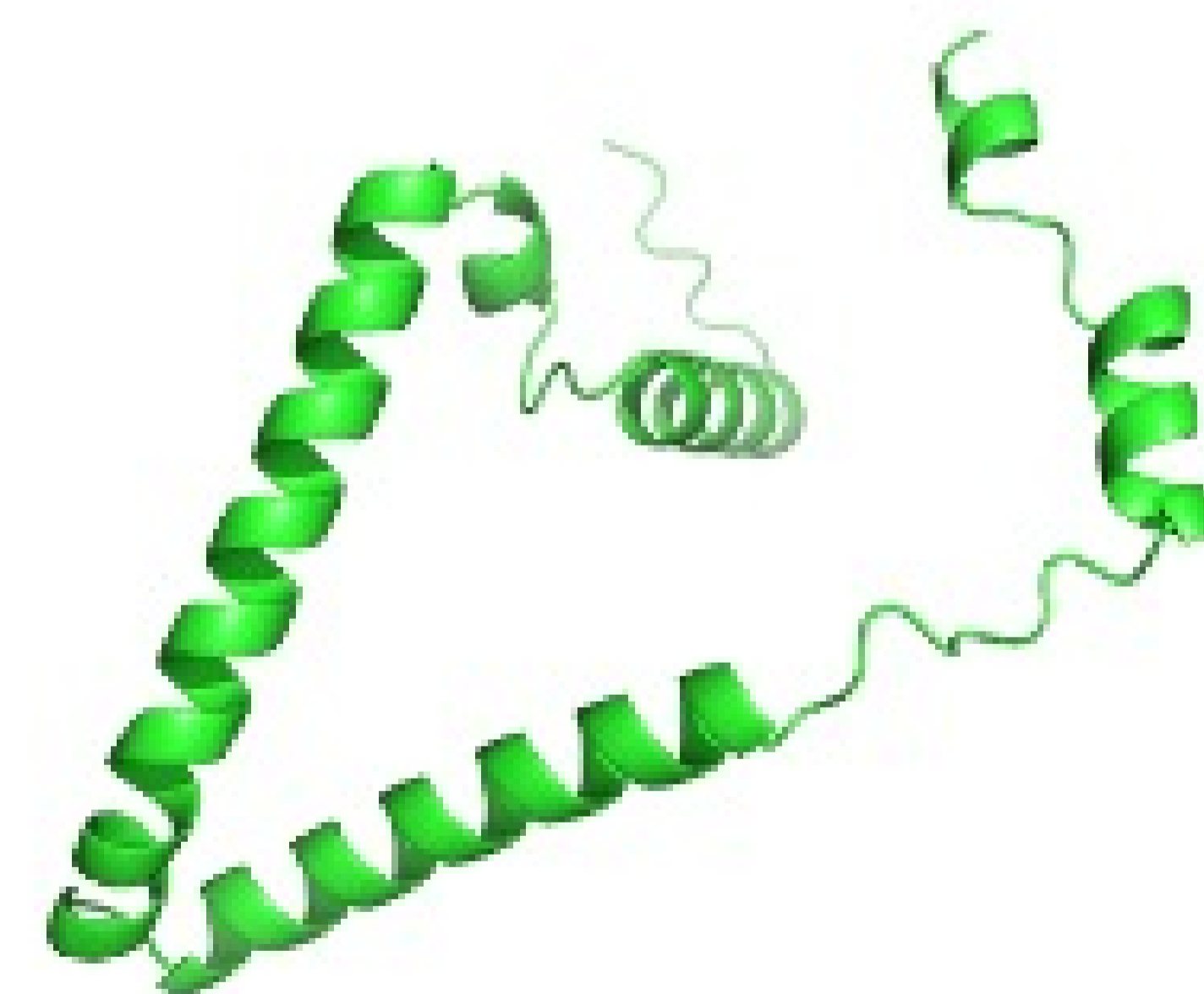
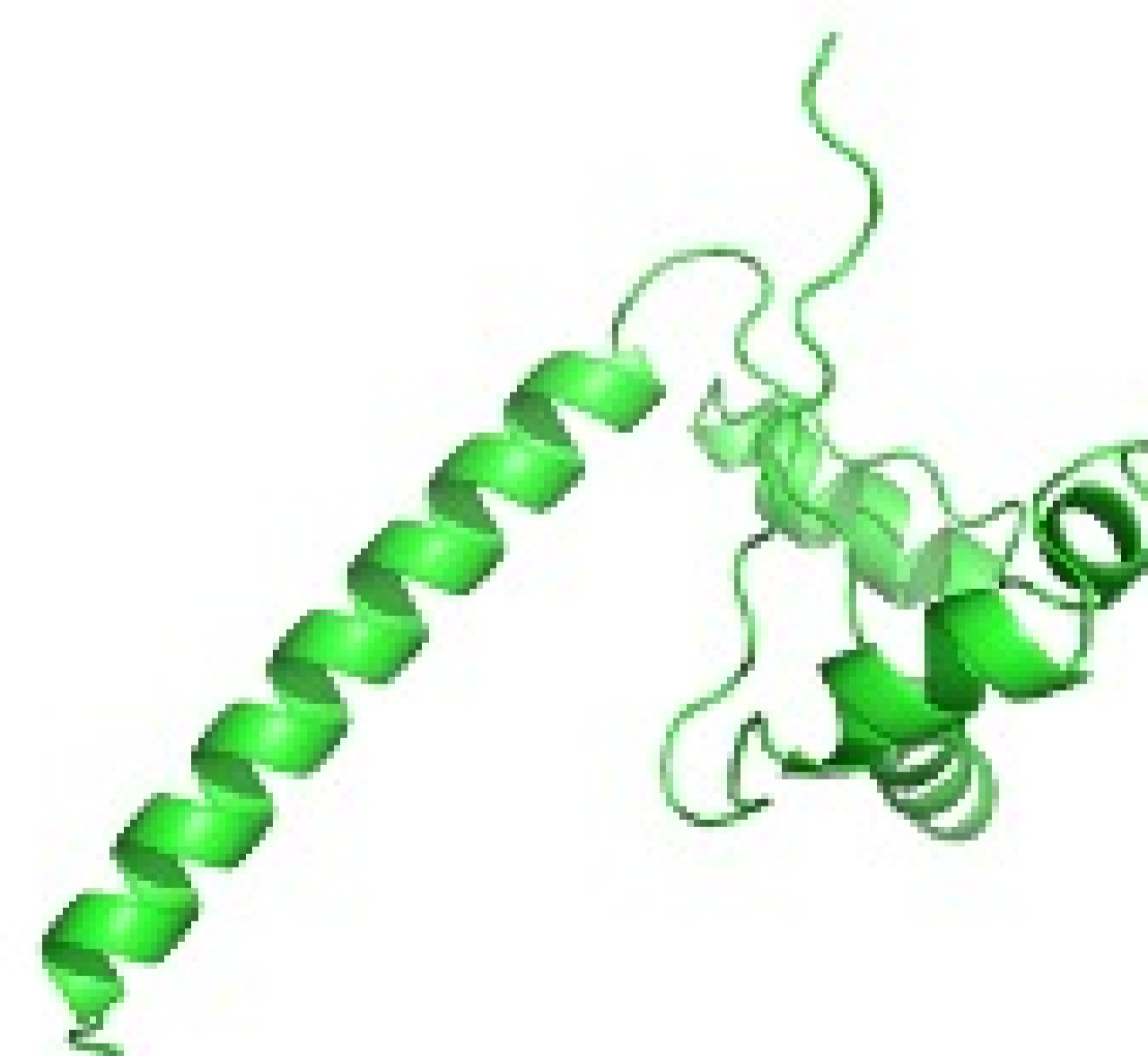
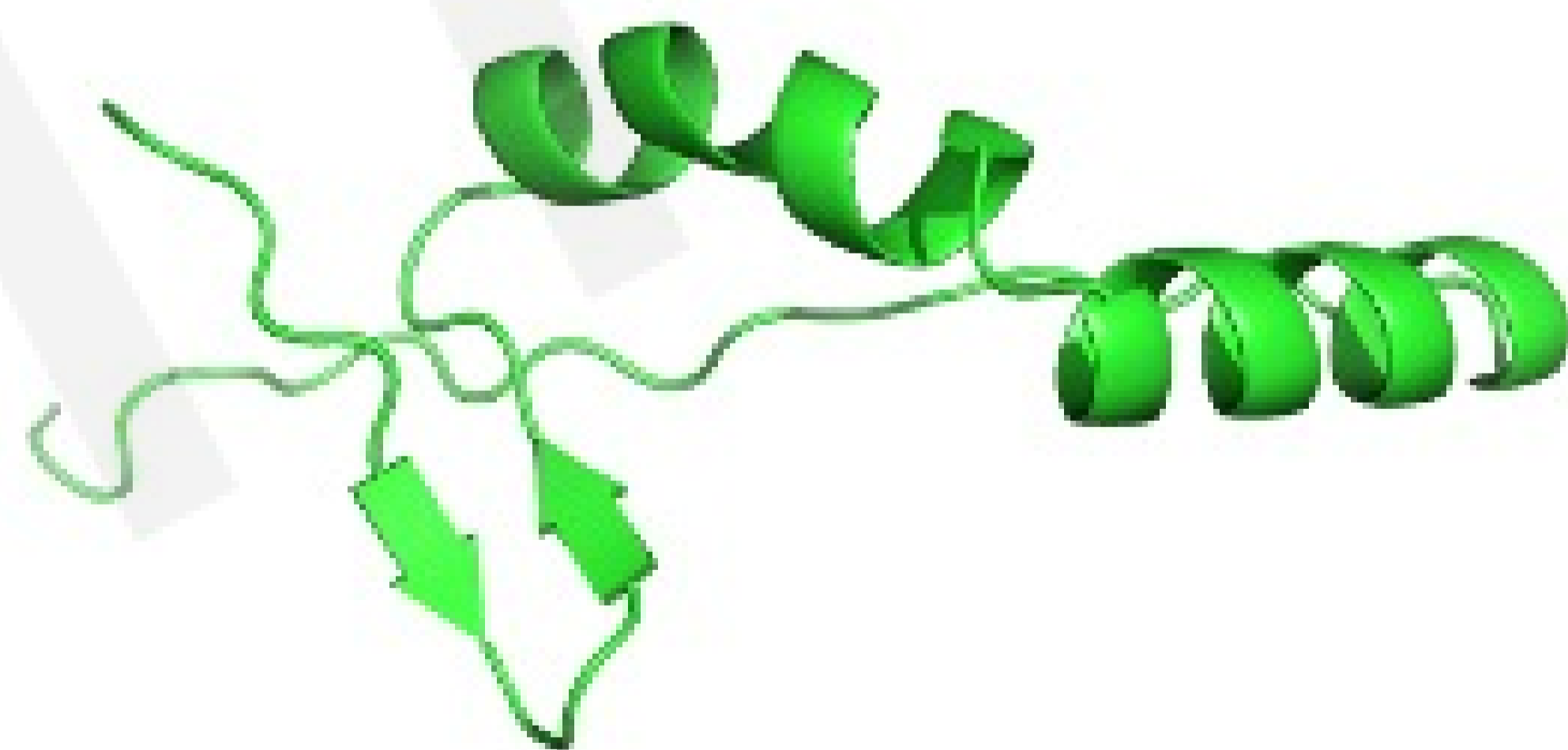
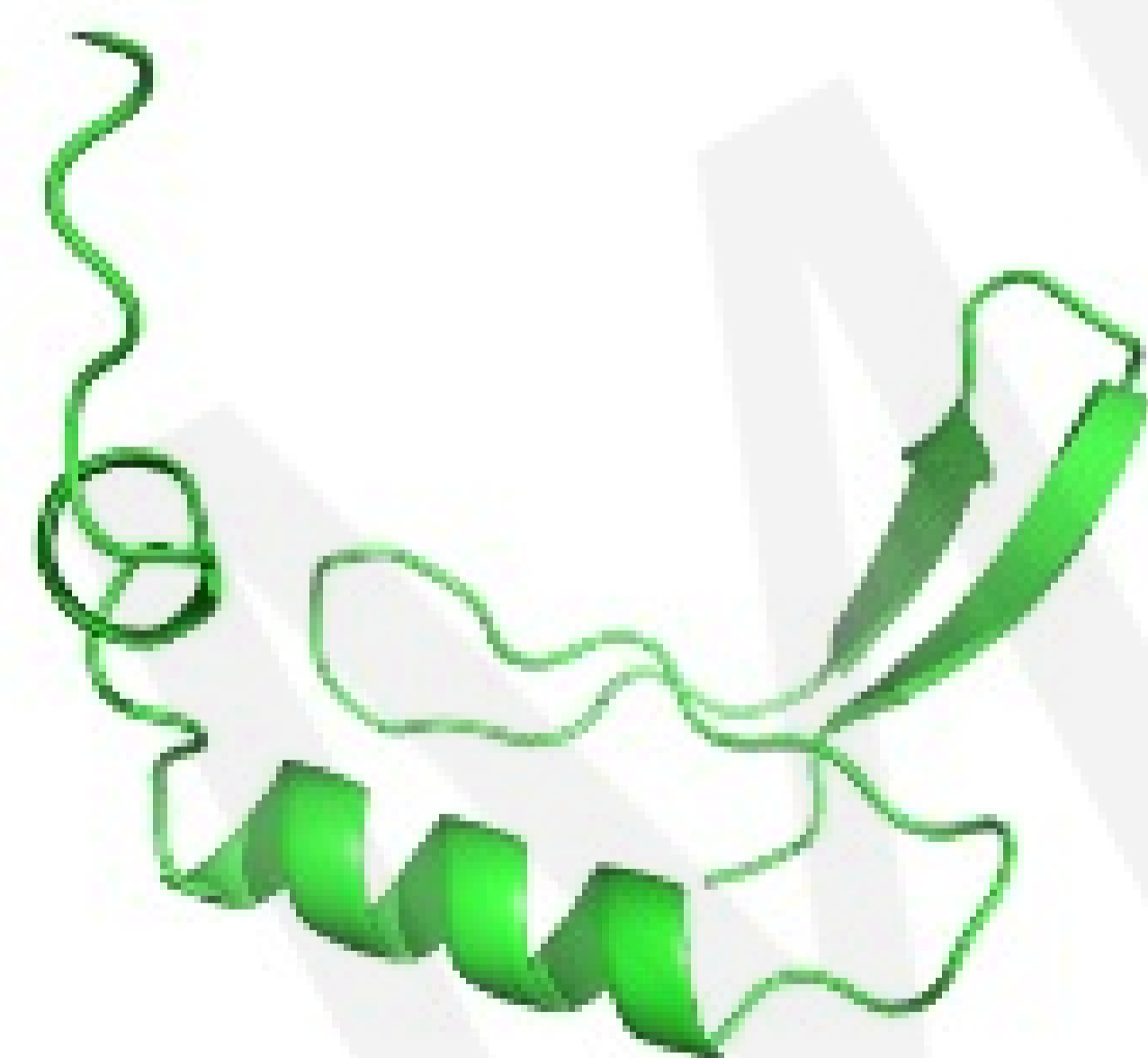


Noise

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?

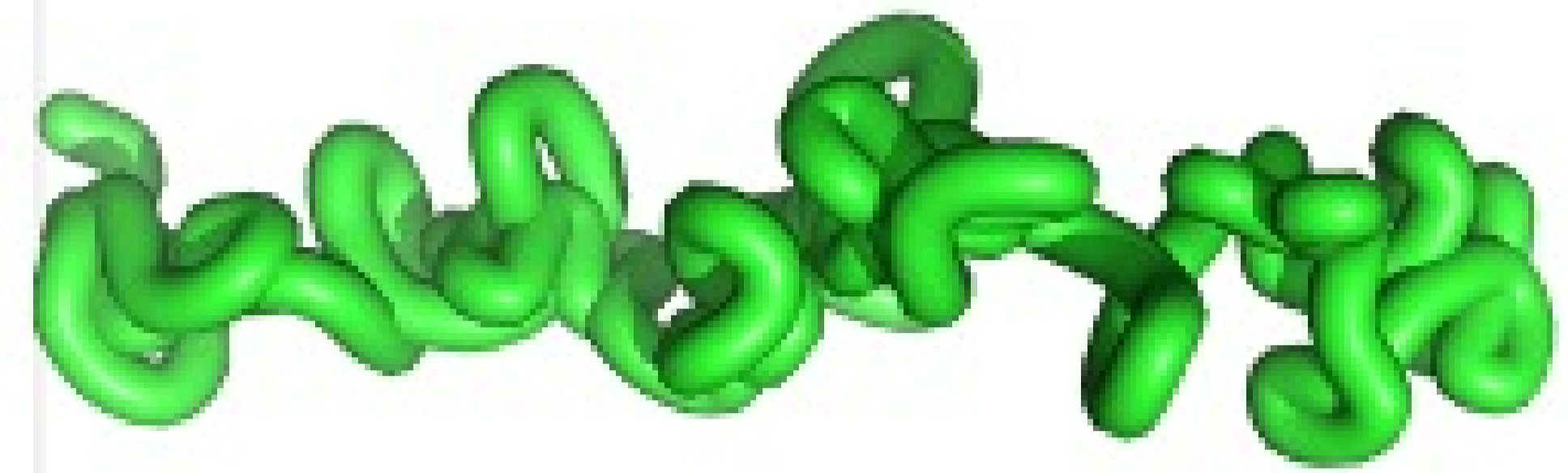


Protein function is defined by structure

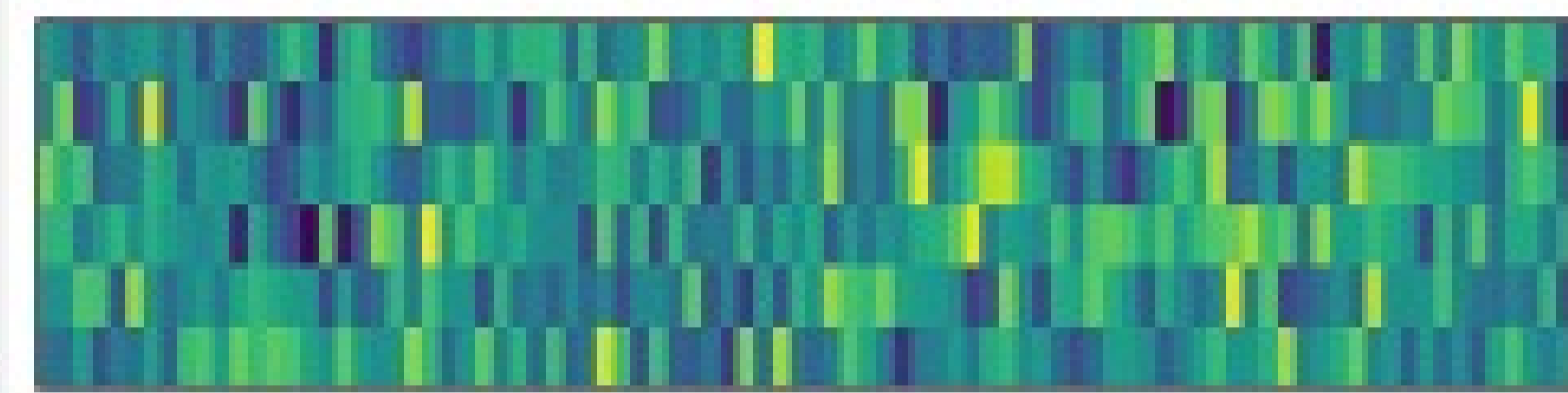
Protein folding leads to structure

Protein Structure Generation via Folding Diffusion

Structure, unfolded

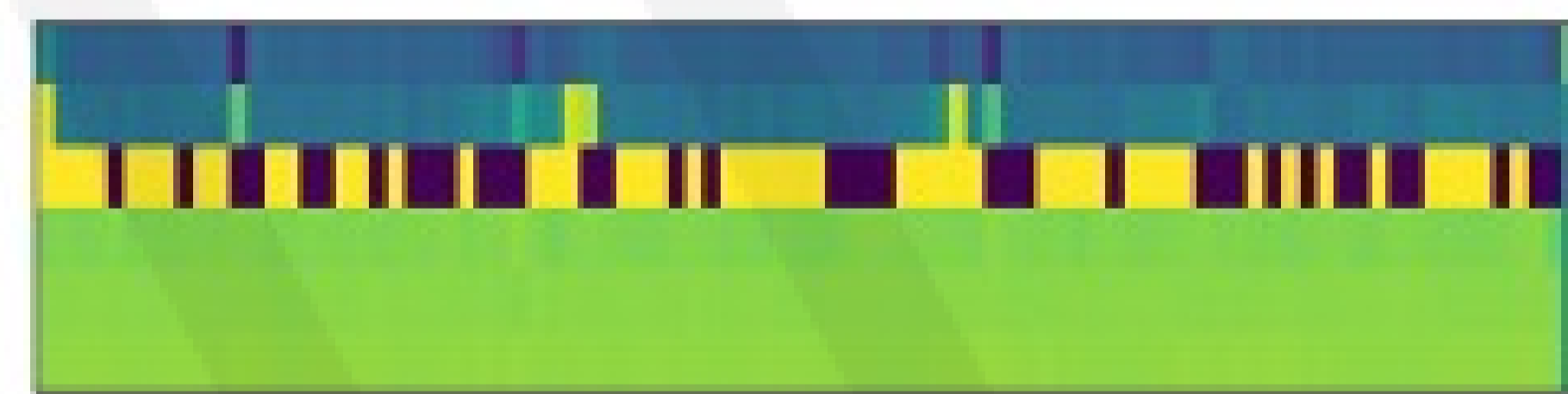


Angles, unfolded

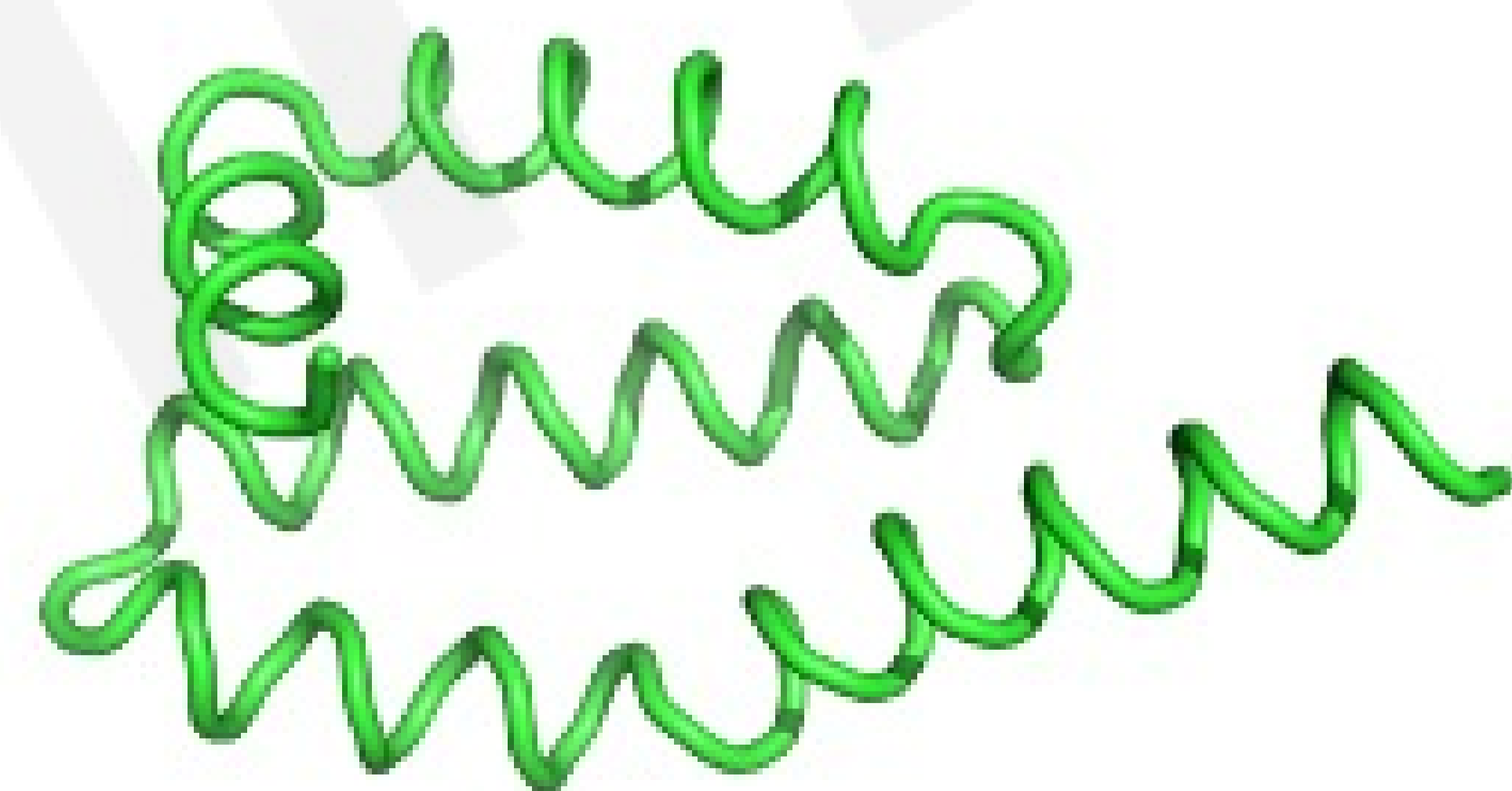


Diffusion Model

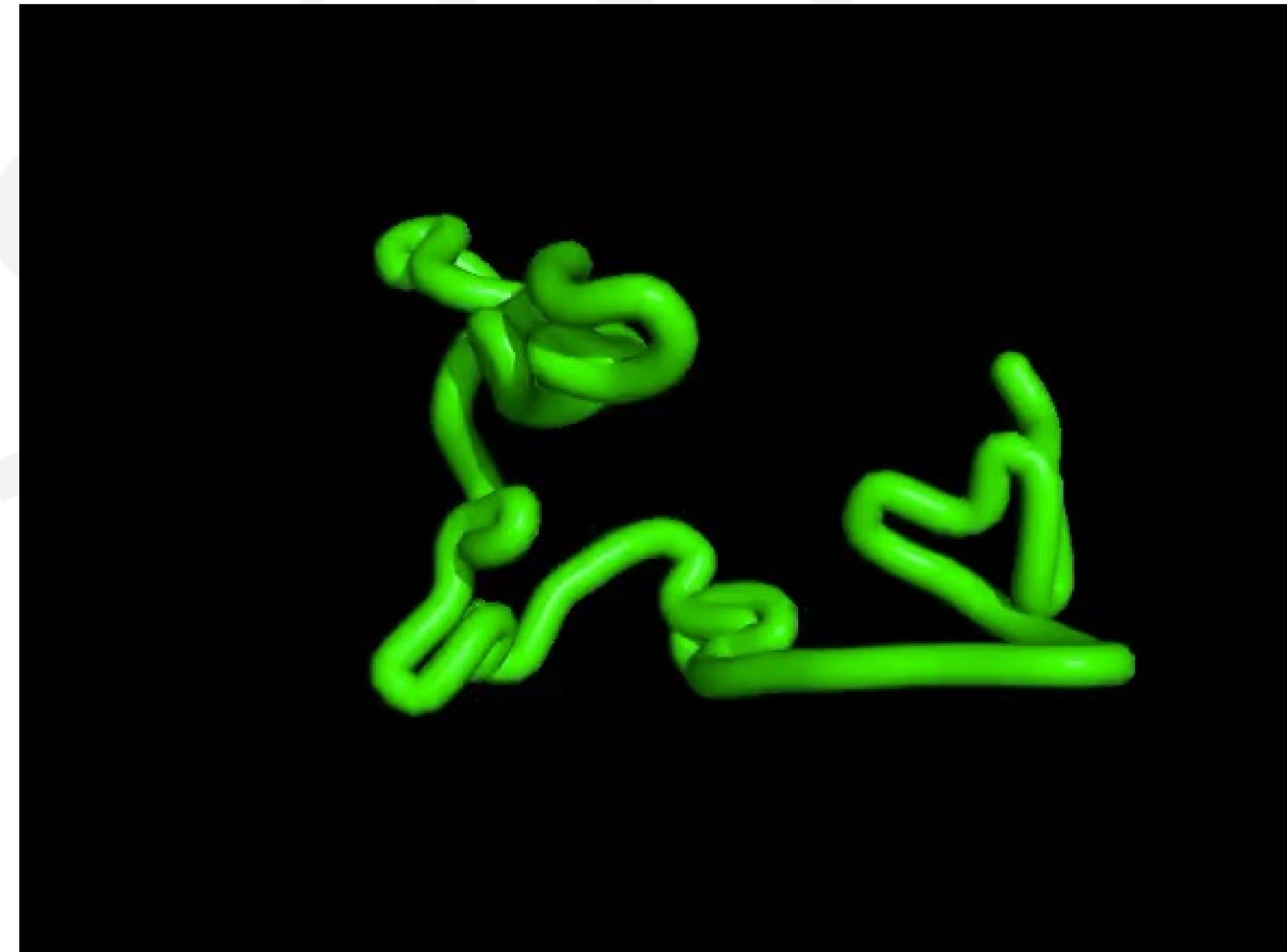
Angles, generated



Structure, generated

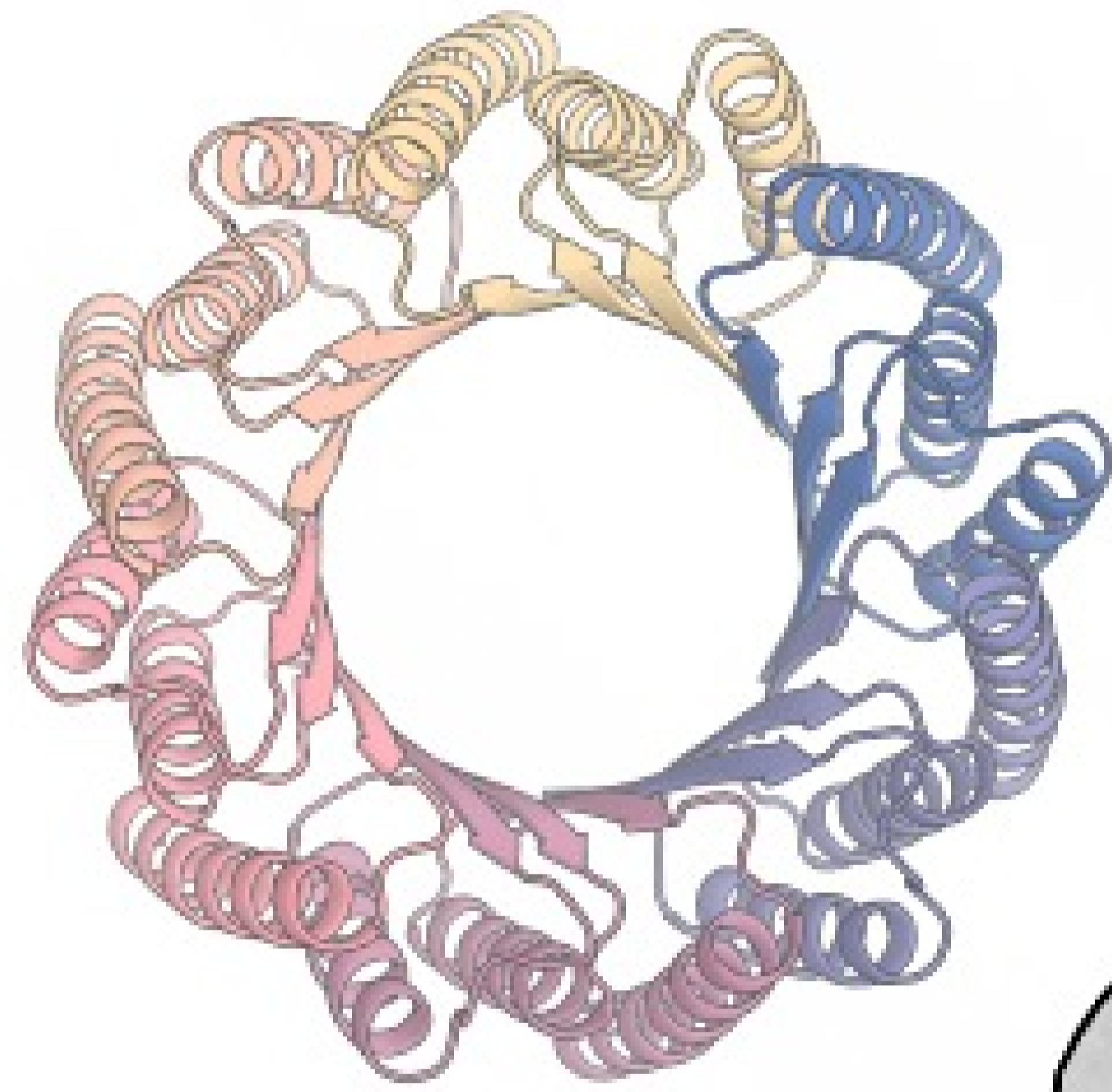


Denoising to generate a structure!



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

Diffusion Models: Foundation for Programmable Protein Design

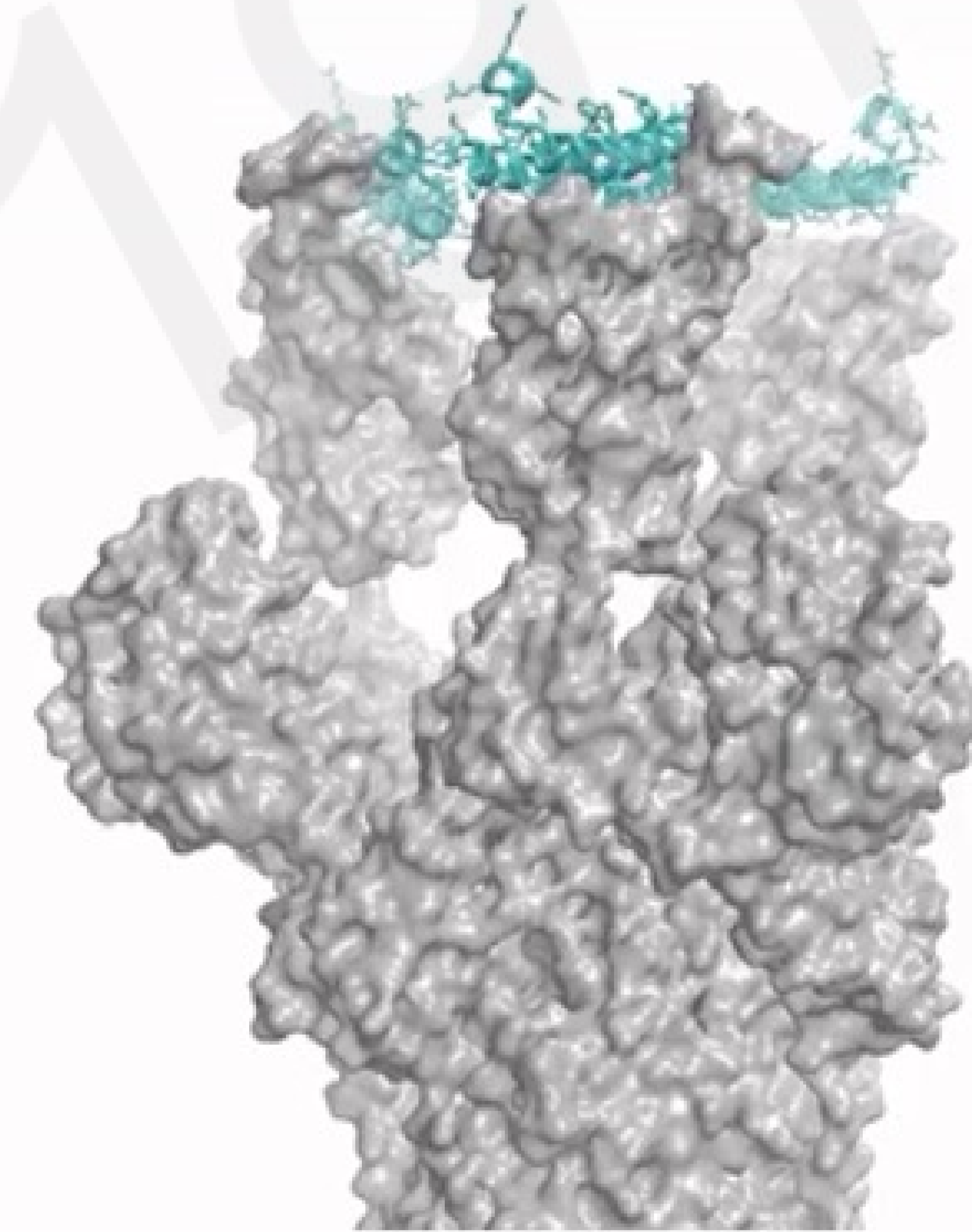


**Real-world
structure**

**AI-generated
protein design**



**Generating a novel
binder to COVID spike**





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