The four ingredient recipe for AI

- Data
- Model Architecture
- Learning Algorithm
- Compute Resources

The 5 modalities of data

Images/Videos (or Computer Vision)

3D Vision

Speech/Audio

Text – natural language

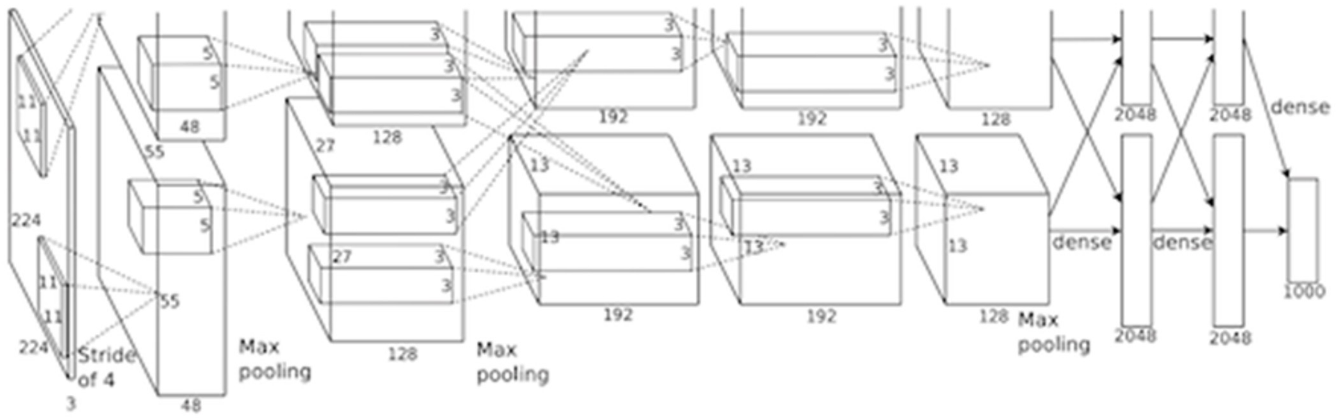
Sensor Data

Key architectures

AlexNet (2012)

The network was made up of

- 5 Conv layers
- Max-Pooling la
- Dropout layers
- 3 Fully Connec layers.



AlexNet architecture (May look weird because there are two different "streams". This is because the training process was so computationally expensive that they had to split the training onto 2 GPUs)

Main Points

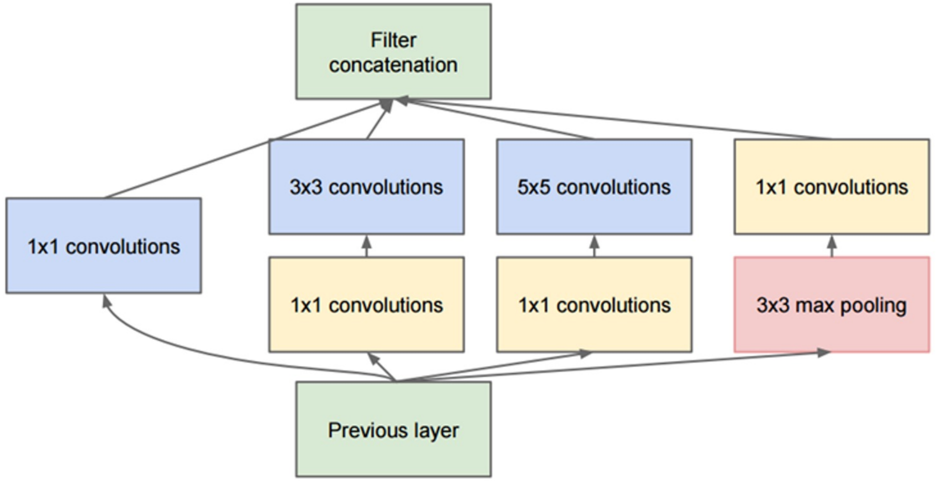
- Trained the network on ImageNet data, which contained over 15 million annotated images from a total of over 22,000 categories.
- Used ReLU for the nonlinearity functions (Found to decrease training time as ReLUs are several times faster than the conventional tanh function).
- Used data augmentation techniques that consisted of image translations, horizontal reflections, and patch extractions.
- Implemented dropout layers in order to combat the problem of overfitting to the training data.
- Trained the model using batch stochastic gradient descent, with specific values for momentum and weight decay.
- Trained on two GTX 580 GPUs for **five to six days**.

Why It's Important

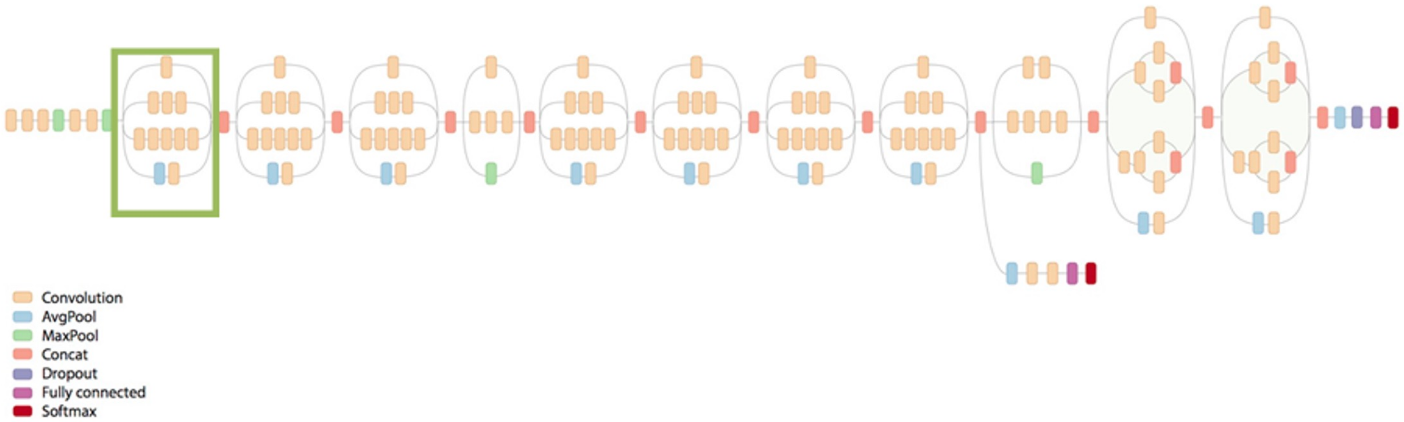
The neural network developed by Krizhevsky, Sutskever, and Hinton in 2012 was the coming out party for CNNs in the computer vision community. This was the first time a model performed so well on a historically difficult ImageNet dataset. Utilizing techniques that are still used today, such as data augmentation and dropout, this paper really illustrated the benefits of CNNs and backed them up with record breaking performance in the competition.

GoogLeNet (2015)

- Used 9 Inception modules in the whole architecture, with over 100 layers in total! Now that is deep…
- No use of fully connected layers! They use an average pool instead, to go from a 7x7x1024 volume to a 1x1x1024 volume. This saves a huge number of parameters.
- Uses 12x fewer parameters than AlexNet.



Full Inception module



Green box shows parallel region of GoogLeNet

Well said Leo, well said

Microsoft ResNet (2015)

- 152 layers…
- Interesting note that after only the *first 2* layers, the spatial size gets compressed from an input volume of 224x224 to a 56x56 volume.
- Authors claim that a naïve increase of layers in plain nets result in higher training and test error (Figure 1 in the paper).
- The group tried a 1202-layer network, but got a lower test accuracy, presumably due to overfitting.

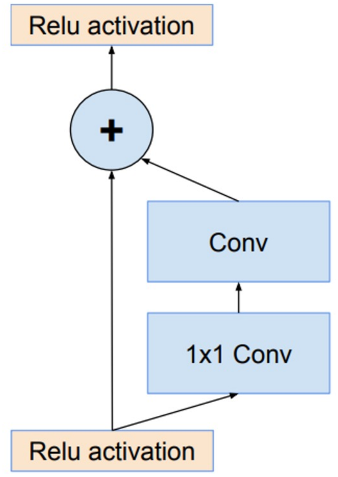
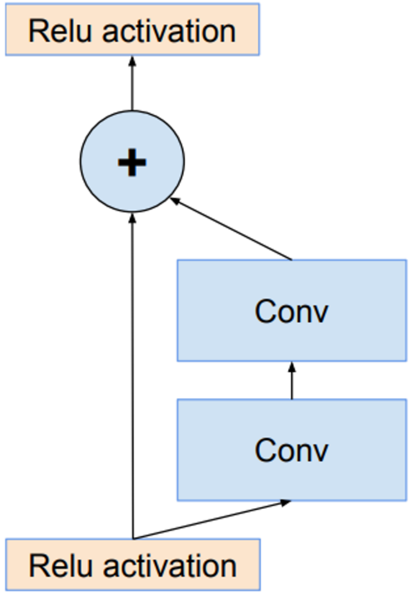
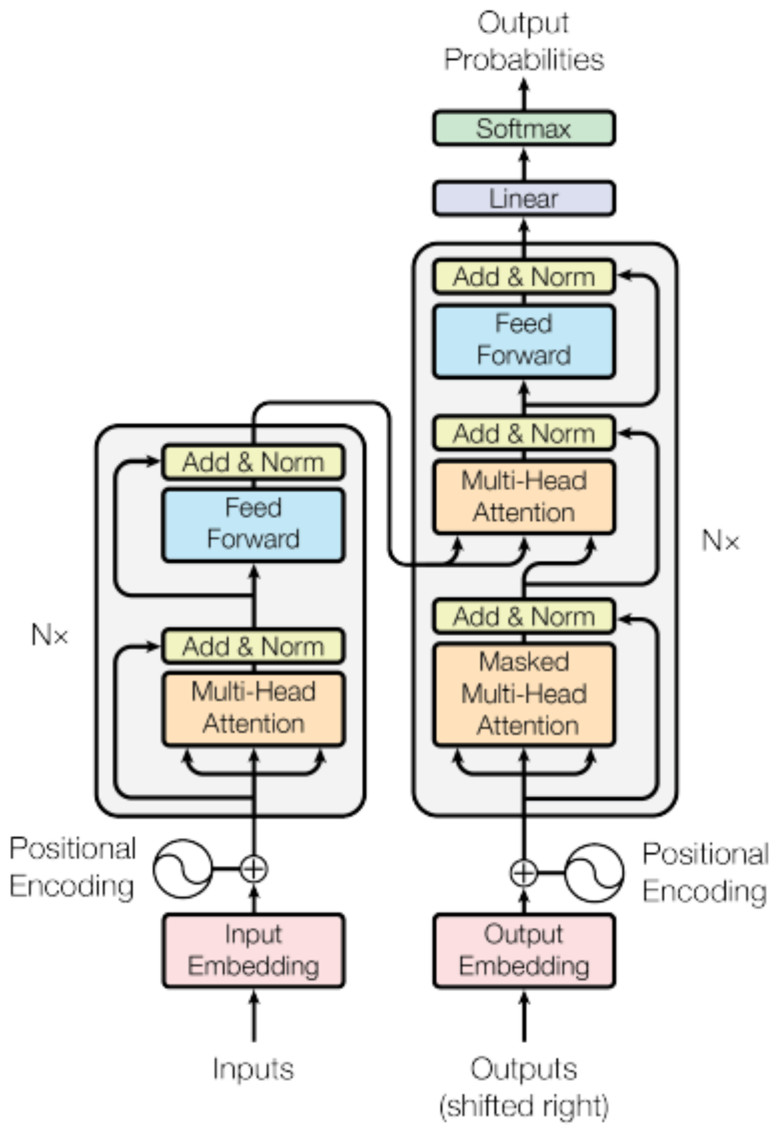


Figure 2. Optimized version of ResNet connections by [5] to shield computation.



Transformer architecture

- *Attention is All You Need* (Vaswani, 2017)
- The GPT-*n* series of models uses this architecture with only minor adjustments
  - PreNorm layers – move normalization layers before attention/feed forward blocks
- Encoder – Self Attention
- Decoder – Cross Attention

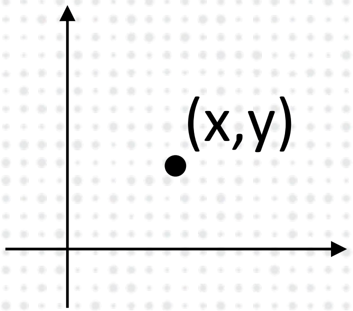


https://arxiv.org/abs/1706.03762

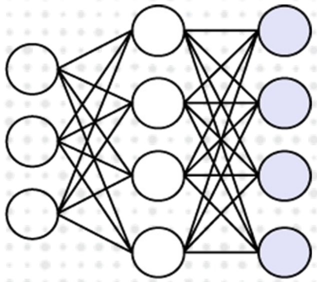
Concluding Remarks

Neural Fields (or INRs)

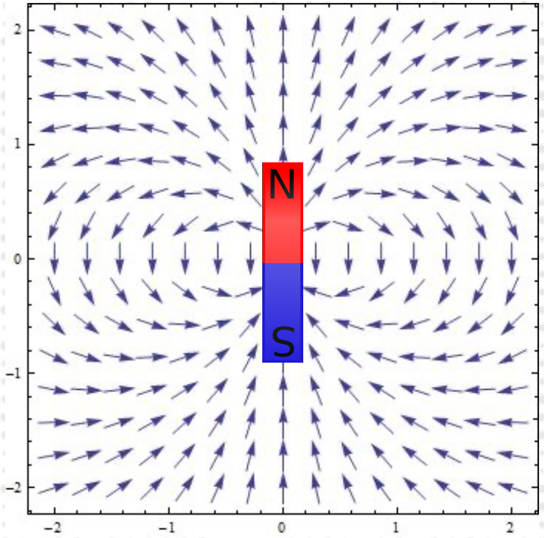
What are neural fields?



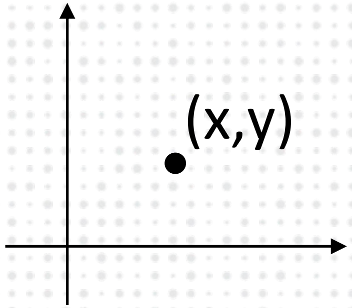
$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^2$



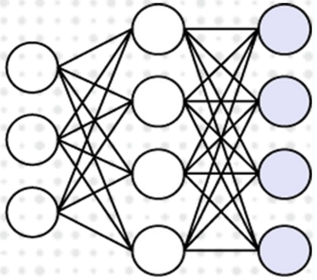
Neural Network (Φ)



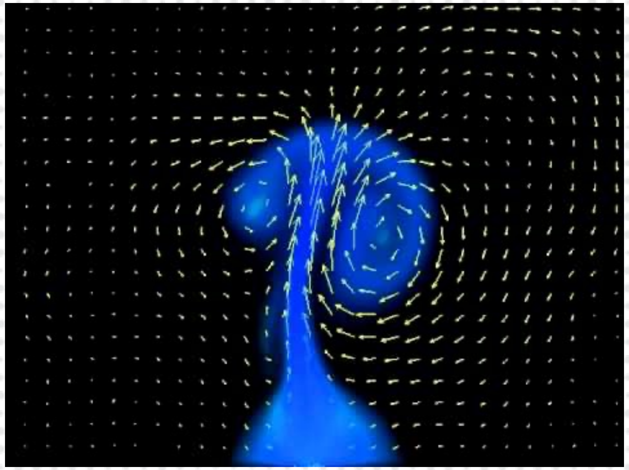
Magnetic Field



$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^2$



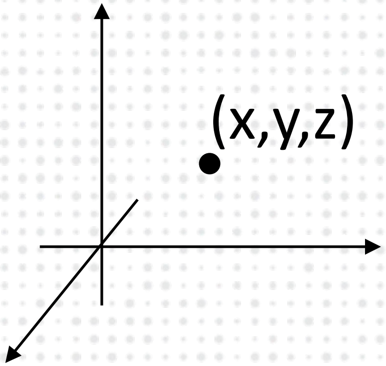
Neural Network (Φ)



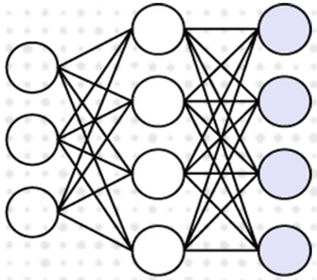
Eulerian Flow Field of a Fluid

[Koldora CC]

What are neural fields?

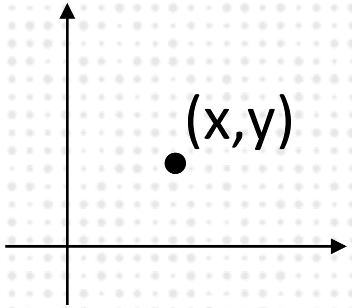


$\Phi: \mathbb{R}^n \rightarrow \mathbb{R}$

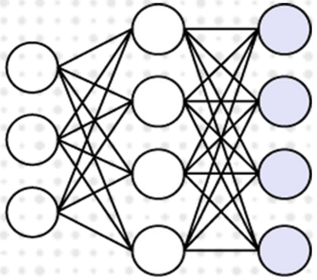


Neural Network
(Φ)

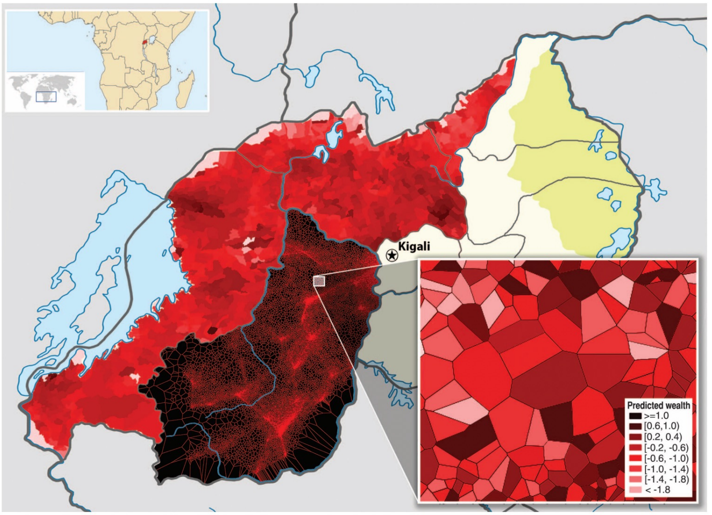
Signed Distance Function (SDF)



$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^n$



Neural Network
(Φ)



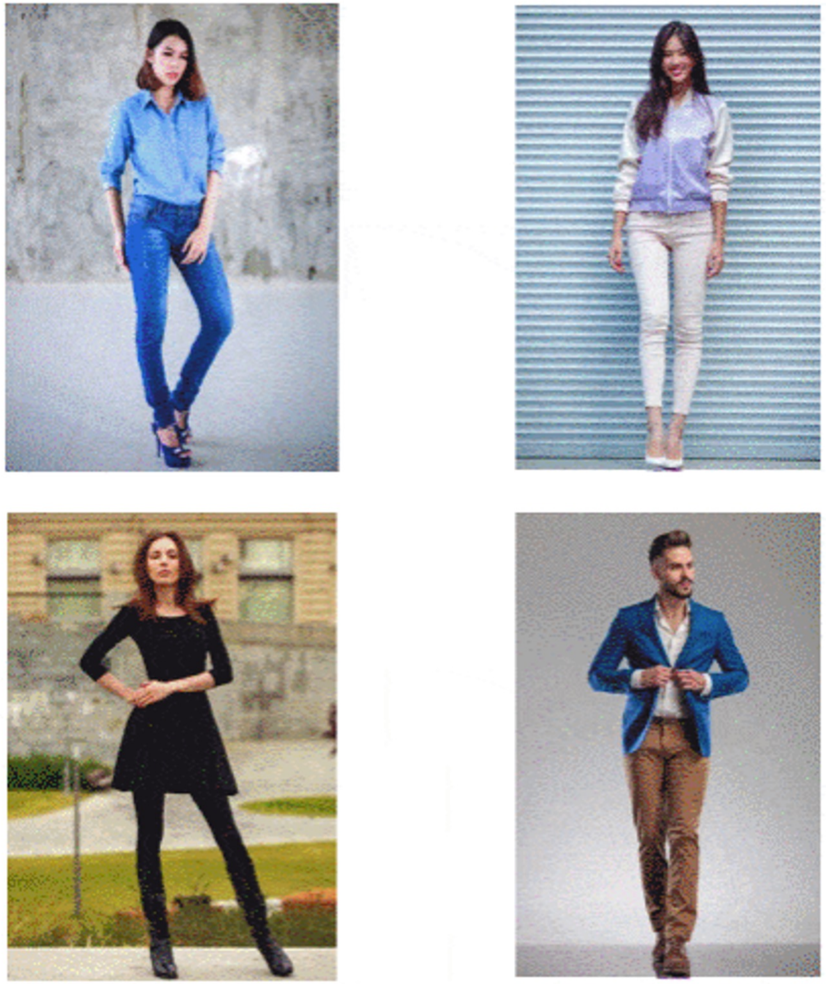
Geospatial Data

[Blumenstock et al. 2015]

Neural Fields for Vision and Graphics

[@smallfly, Müller et al. 2022]

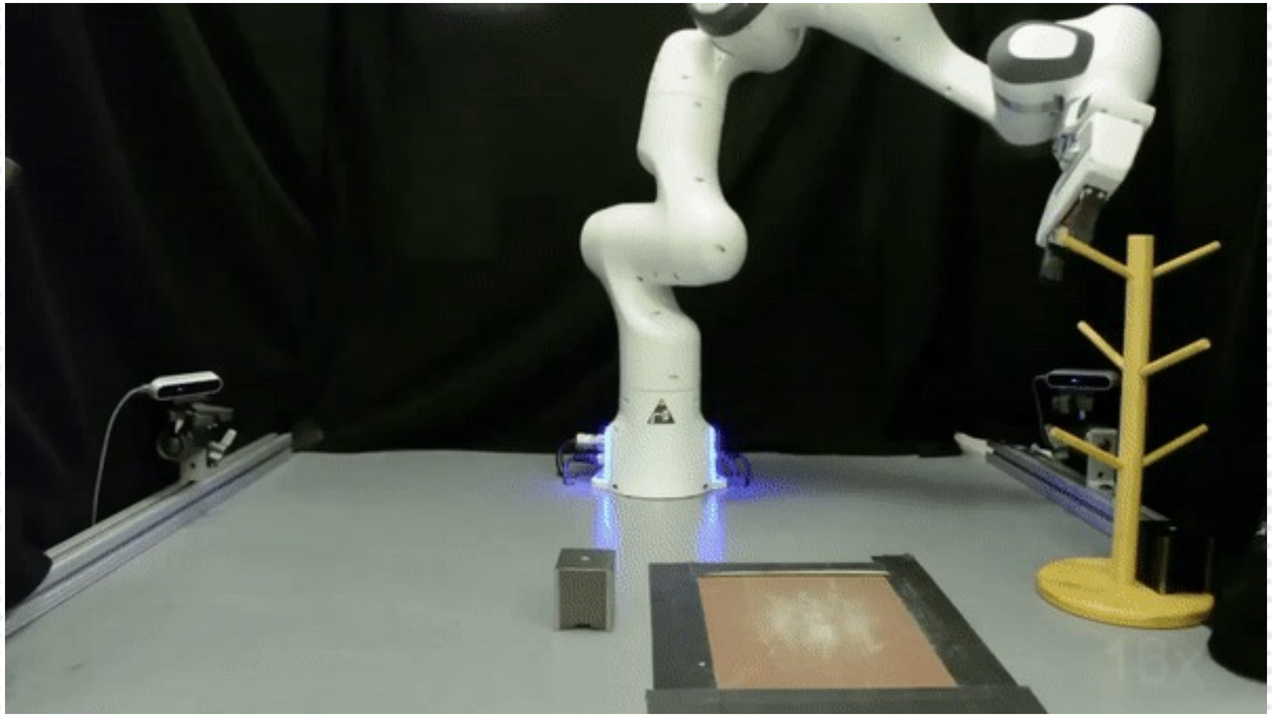
Neural Fields for Vision and Graphics



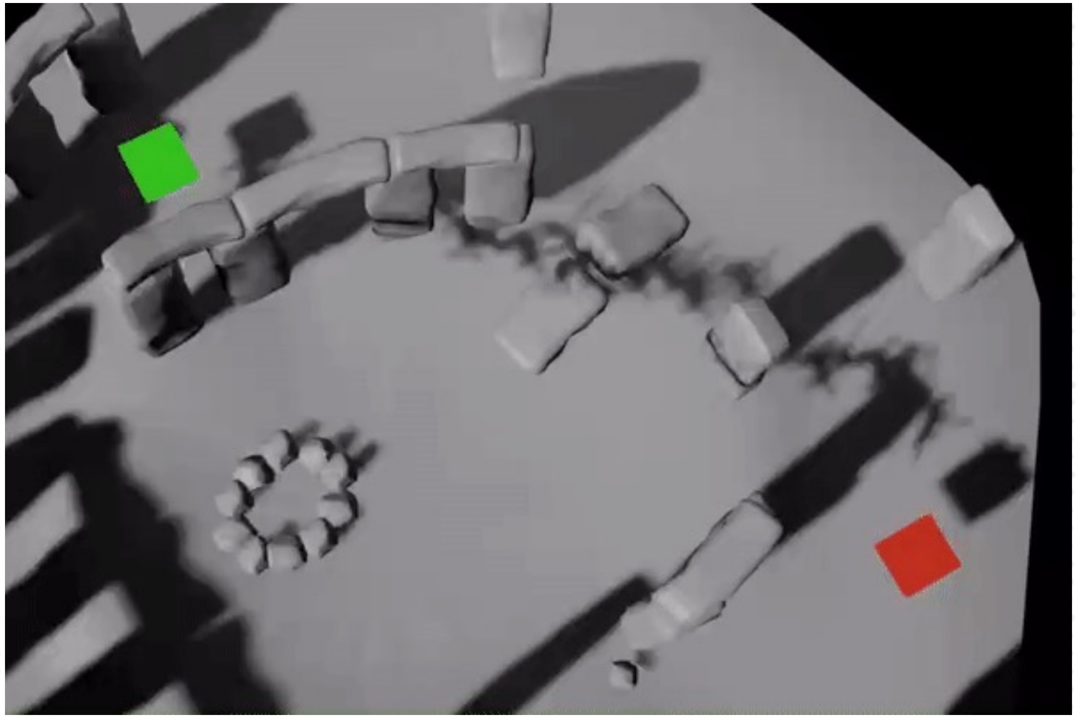
[Saito et al. 2020 (PIFu)]

[Chan et al. 2021]

Neural Fields for Robotics

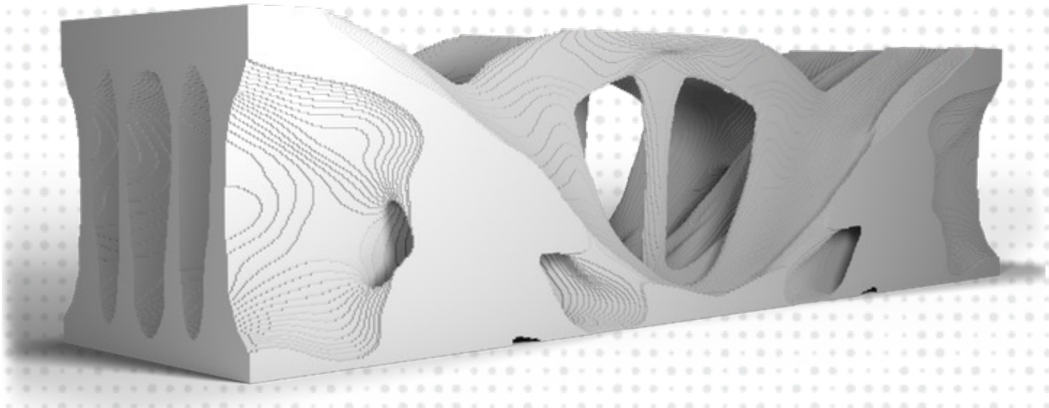


[Simeonov et al. 2021 (Neural Descriptor Fields)]

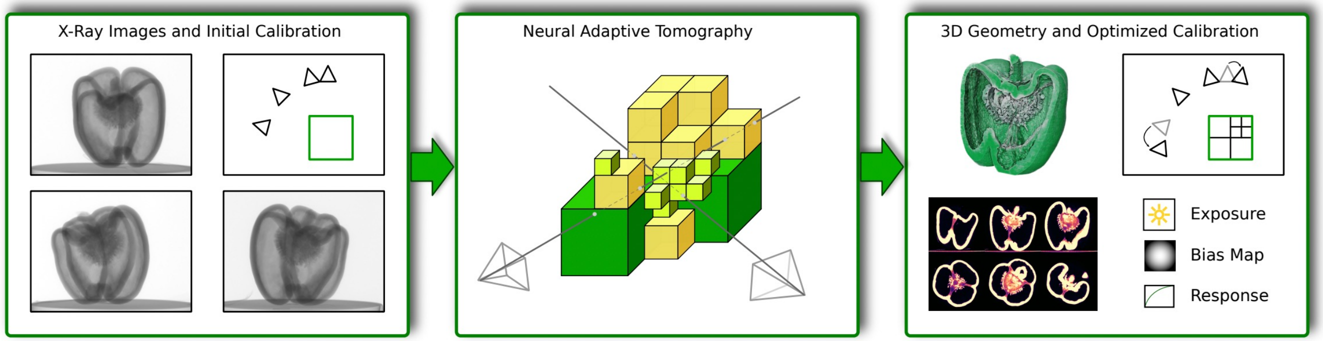


[Adamkiewicz, Chen et al. 2021]

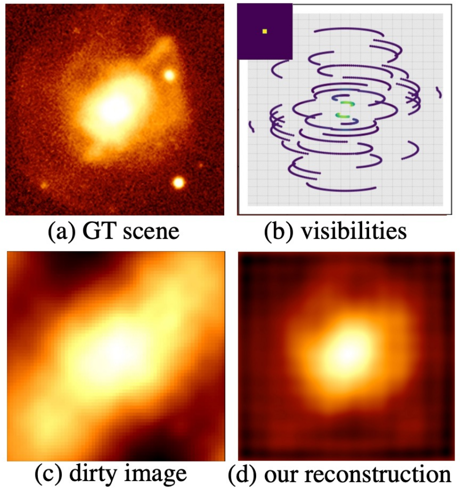
Neural Fields for Science and Engineering



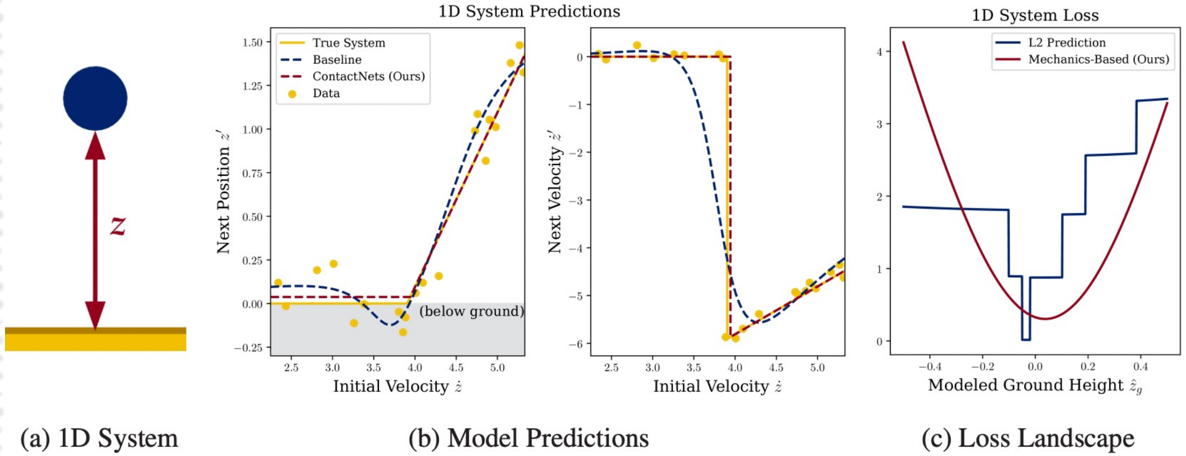
Topology Optimization [Doosti et al. 2021]



Tomographic Reconstruction [Ruckert et al. 2022]



Astronomical Interferometry [Wu et al. 2021]

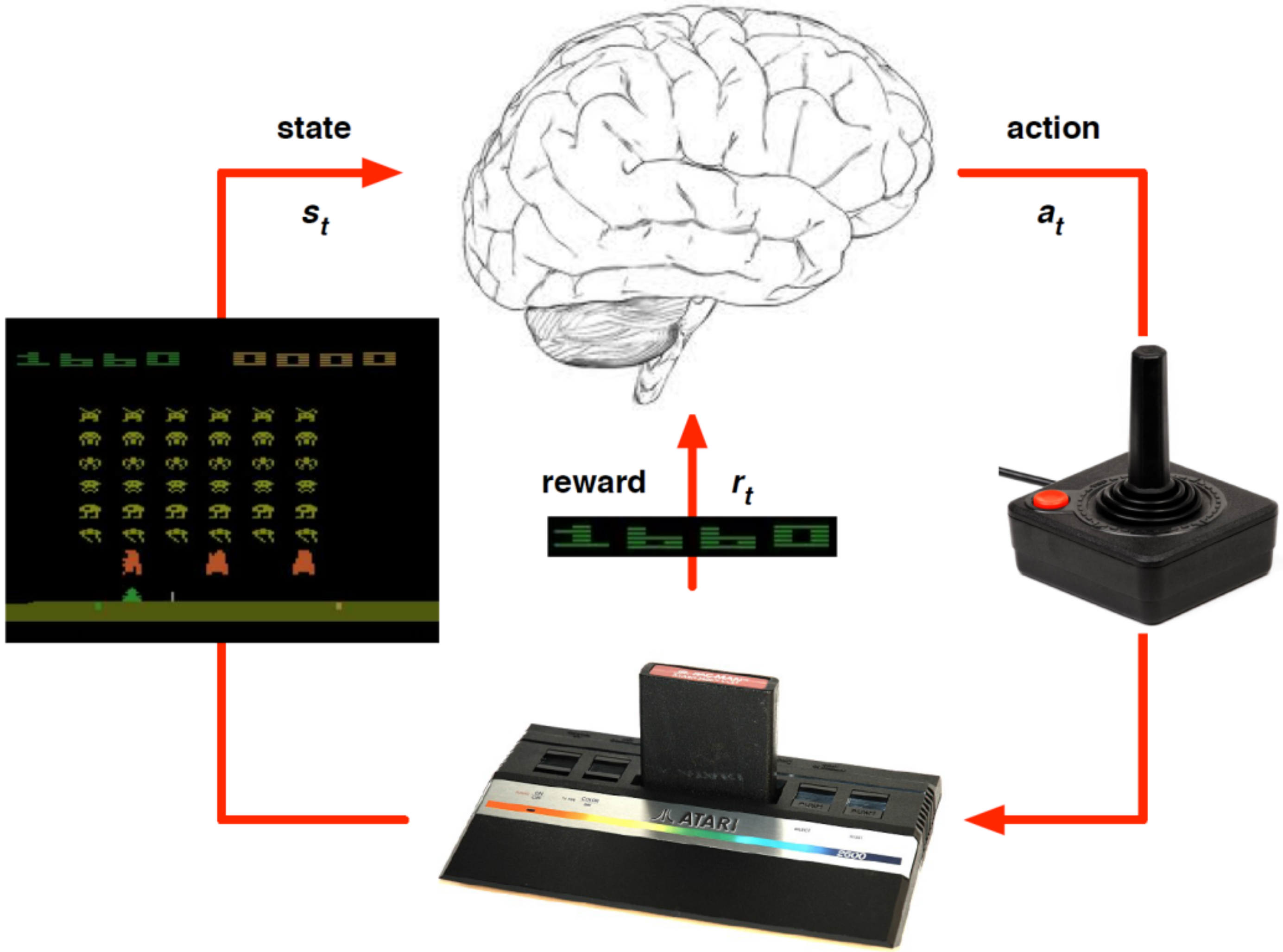


Contact Dynamics [Pfrommer, Halm et al. 2022]

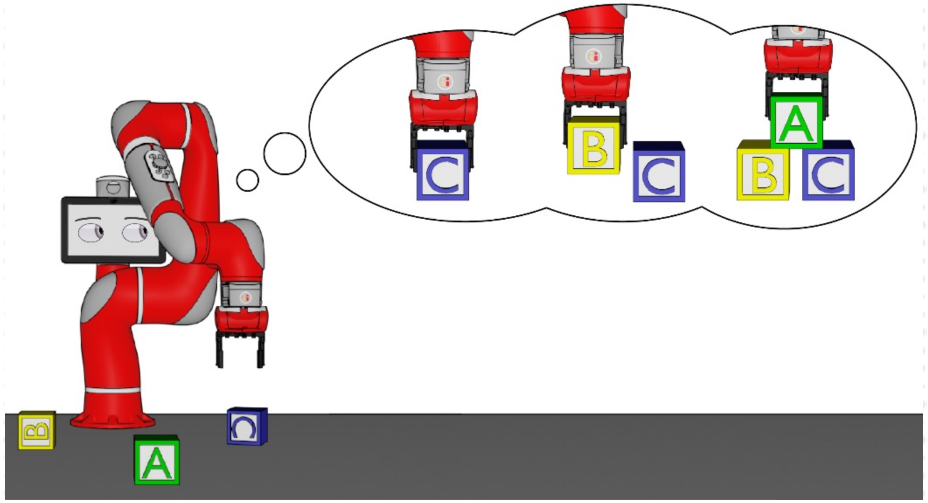
Translational AI Center

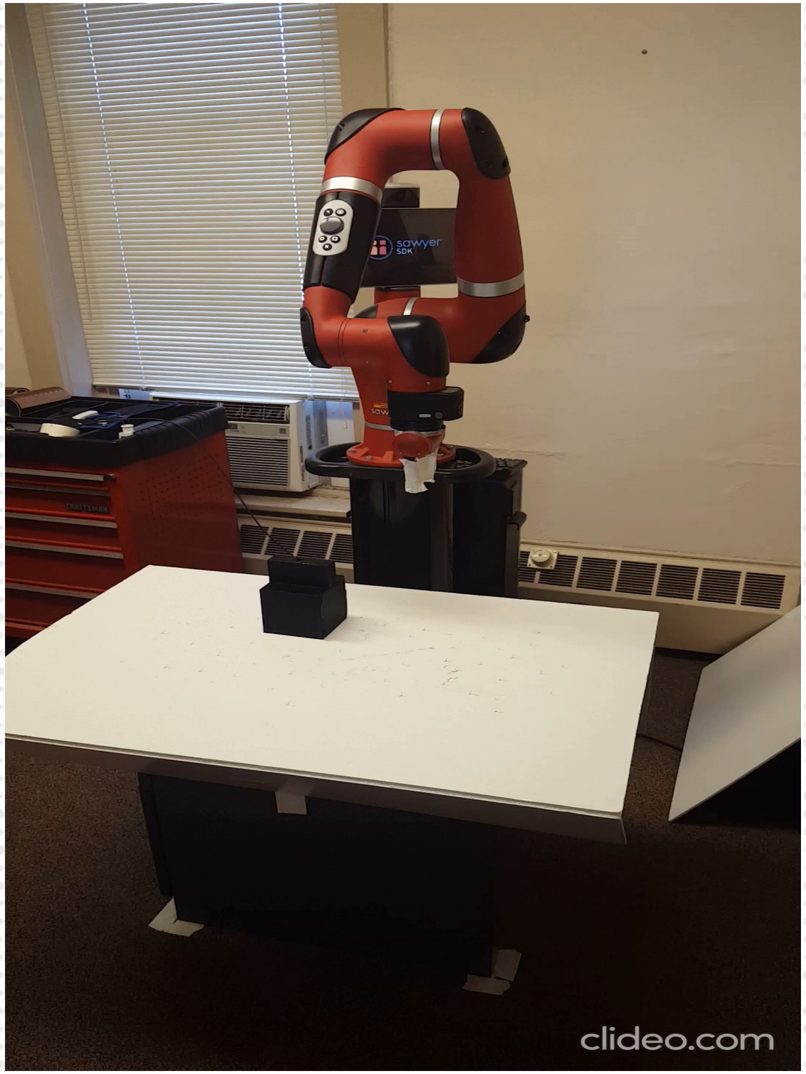
Reinforcement Learning

Learning from feedback



From playing games to robotic manufacturing



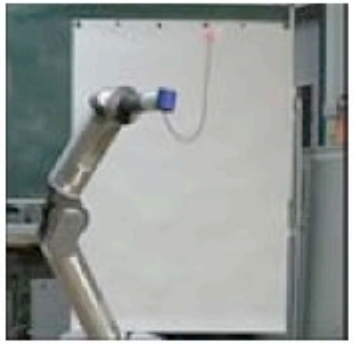


Some initial success stories

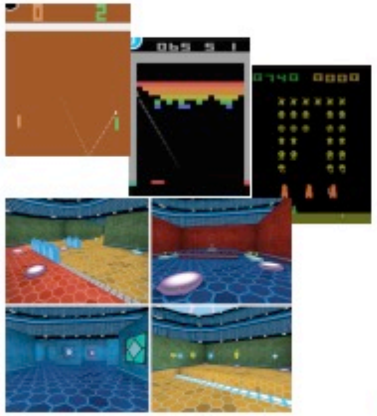
Kohl and Stone, 2004

Ng et al, 2004

Tedrake et al, 2005



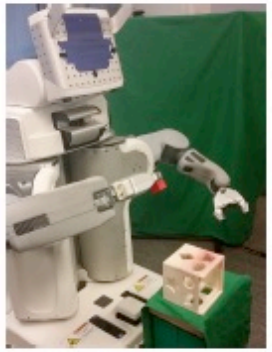
Kober and Peters, 2009



Mnih et al 2013 (DQN)
Mnih et al, 2015 (A3C)

Silver et al, 2014 (DPG)
Lillicrap et al, 2015 (DDPG)

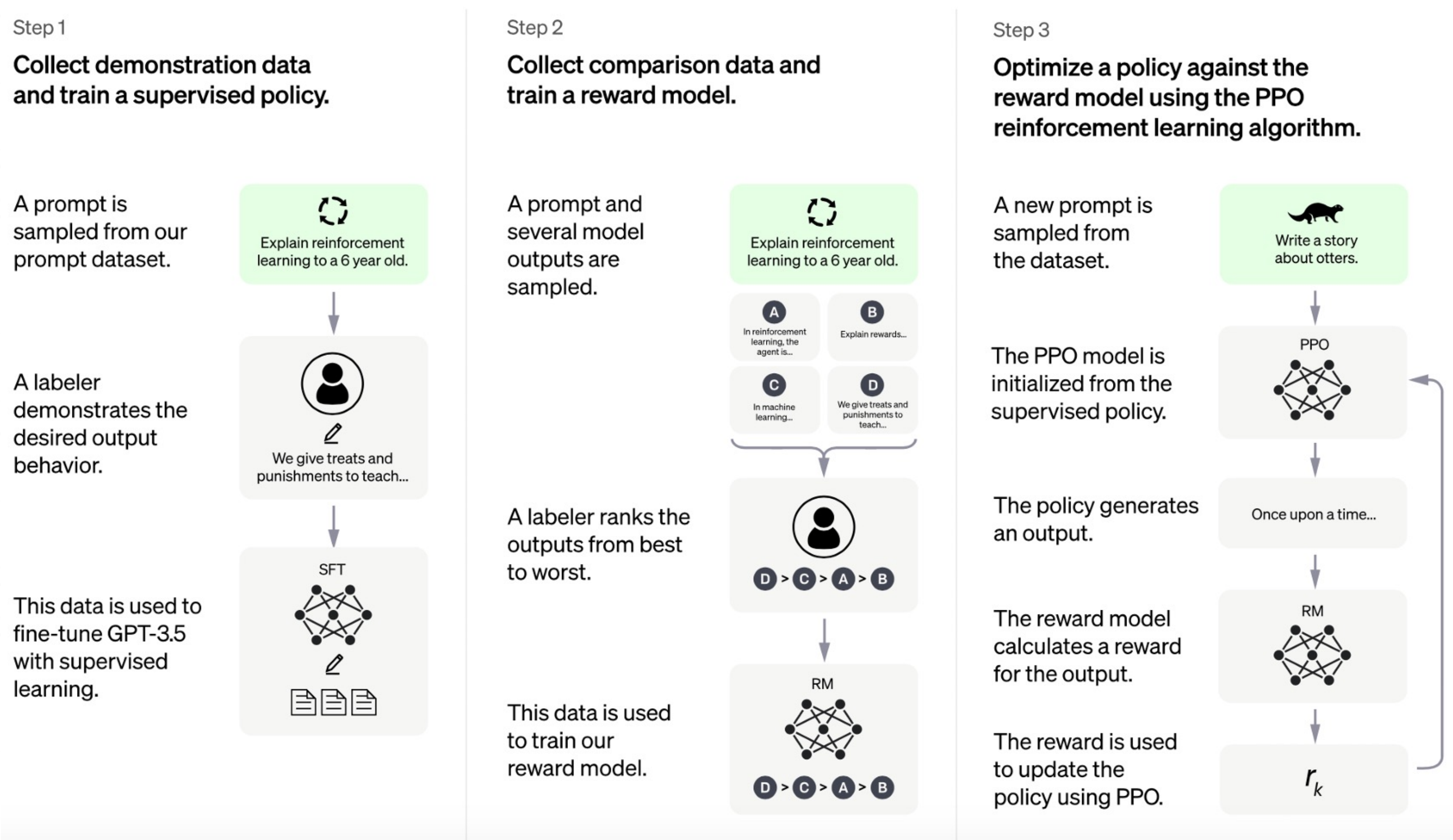
Schulman et al, 2016 (TRPO + GAE)



Levine*, Finn*, et al, 2016 (GPS)

Silver*, Huang*, et al, 2016 (AlphaGo)

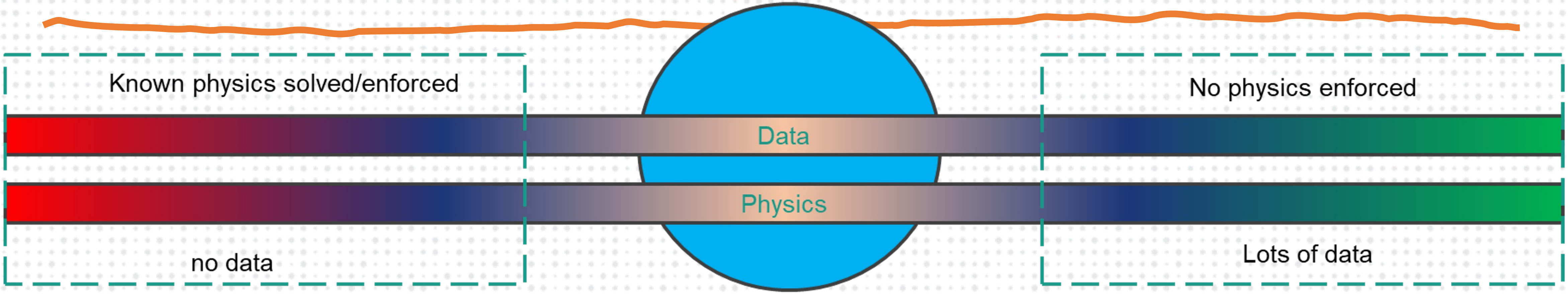
ChatGPT – RL with human feedback



Translational AI Center

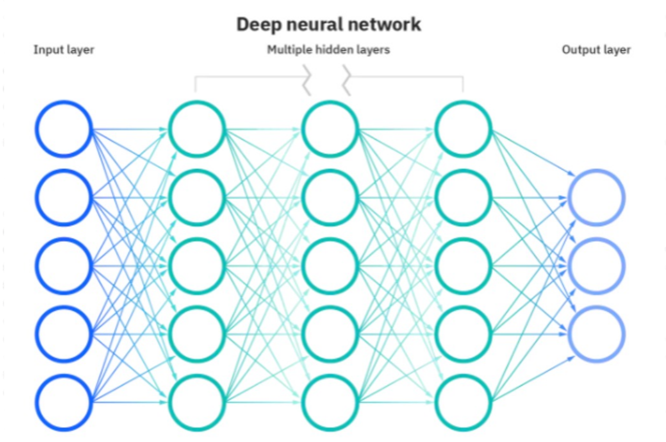
Scientific Machine Learning

SciML Overview

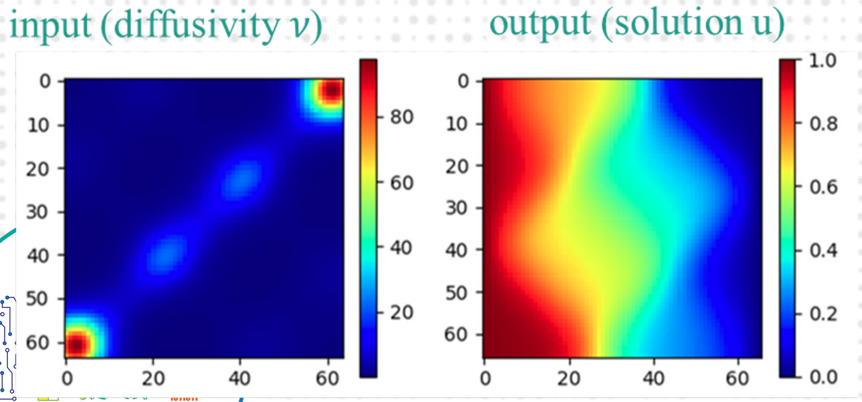


- Look at ML through the lens of amount of data vs amount of information
- Information (or domain knowledge) considered in terms of constraints
- Spectrum of methods being developed by the community
- Several recent works on 'neural PDE' solvers
- Significant implication for multi-scale simulations, real time control, design, inverse problems

SciML Overview



https://www.ibm.com/cloud/learn/neural-networks



Basic ML: Input output data pairs

Know some underlying physics

- constraints, invariances
- Perhaps know the equations themselves

Can integrate these 'domain knowledge' into the network

AI in Molecular Synthesis

- AI allows scientists to determine what potential drugs are worth evaluating in the lab and the effective ways of synthesizing them

CHEMICAL SYNTHESIS PLANS BENCHMARK: TOP-1 TEST ACCURACY

Source: Schwaller, 2020 | Chart: 2021 AI Index Report

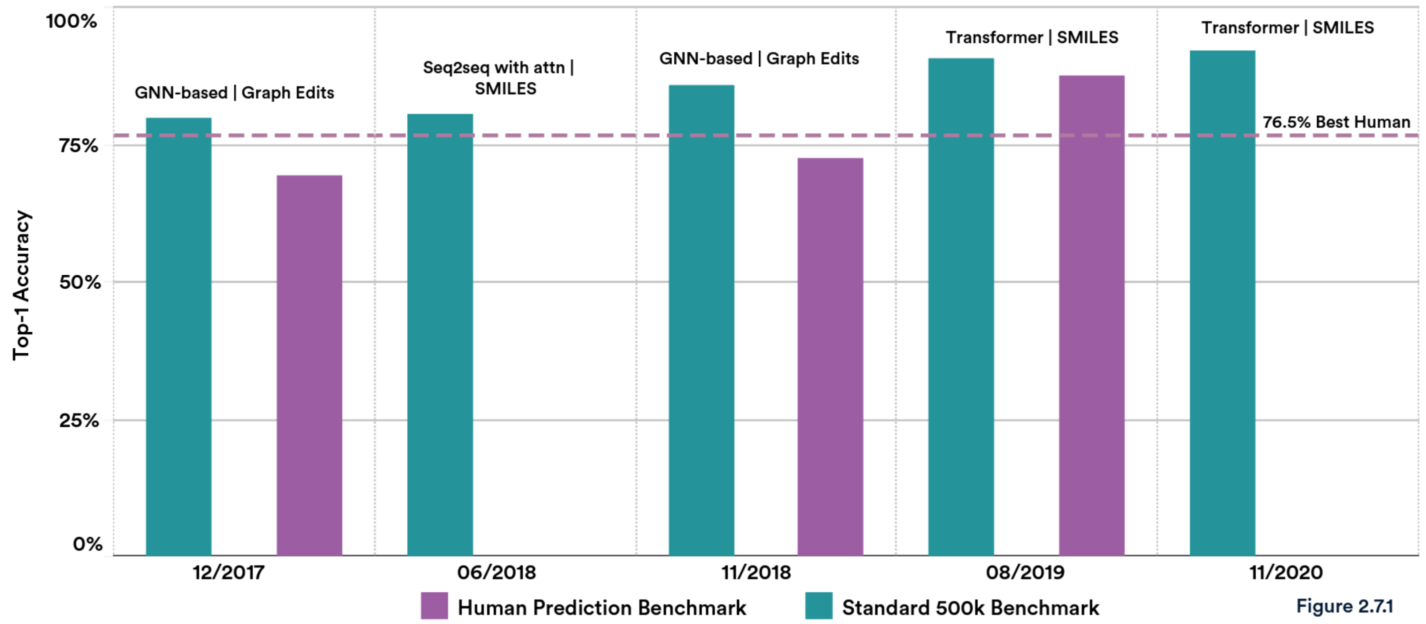
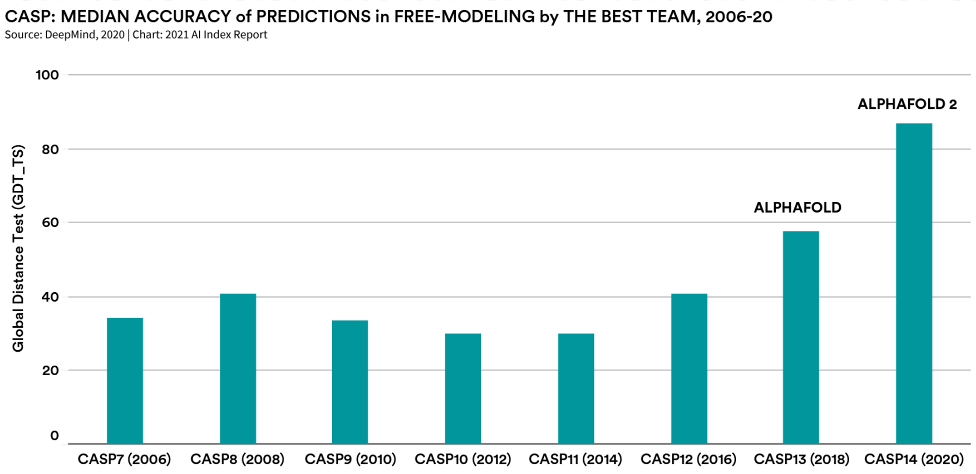


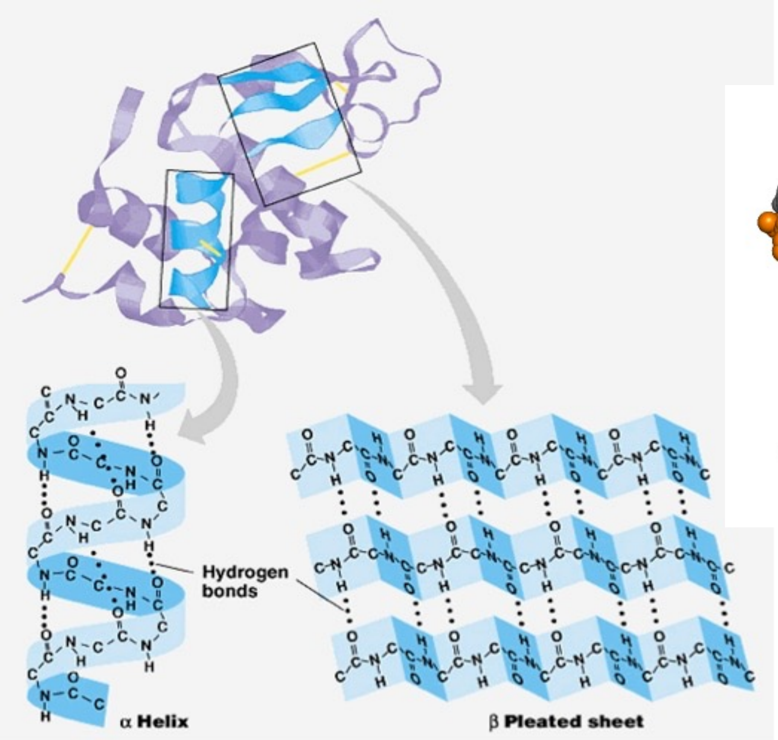
Figure 2.7.1

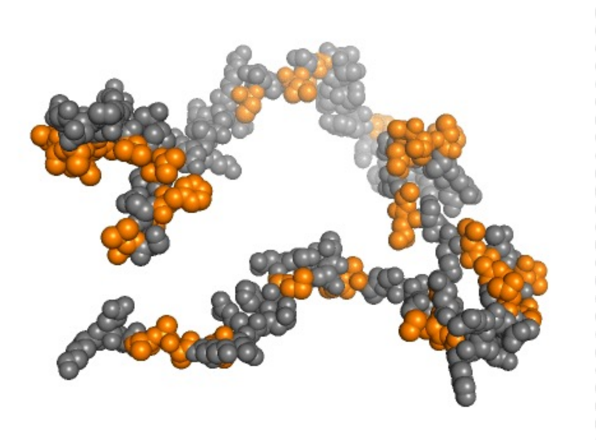
6 Acknowledgment: Philippe Schwaller at IBM Research–Europe and the University of Bern provided instructions and resources for gathering and analyzing the data.

Alphafold and Protein Folding

- The protein folding problem is to determine the 3D structures of proteins from sequences of amino acids





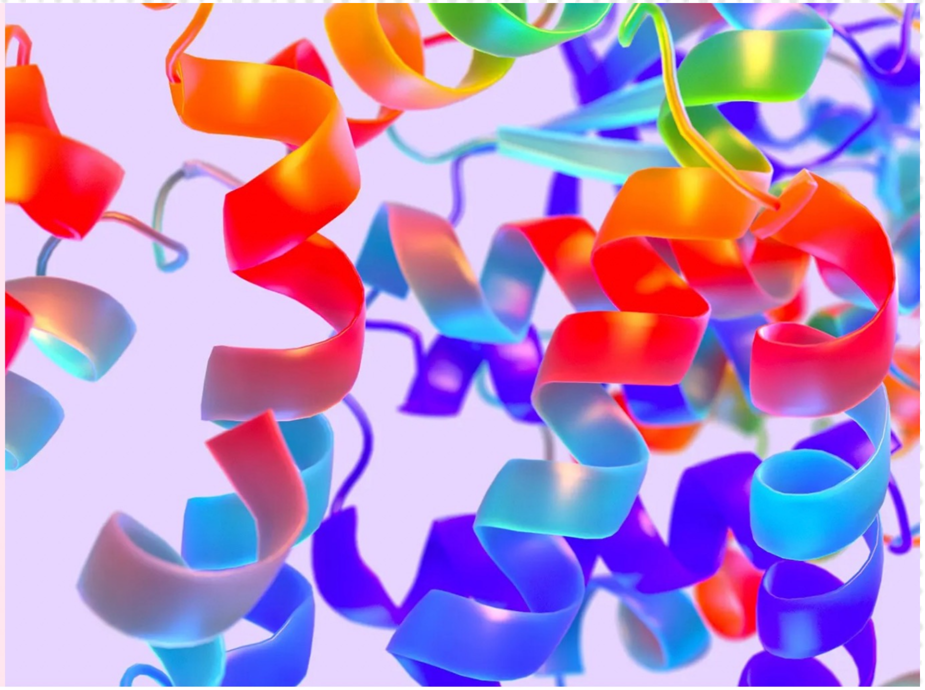


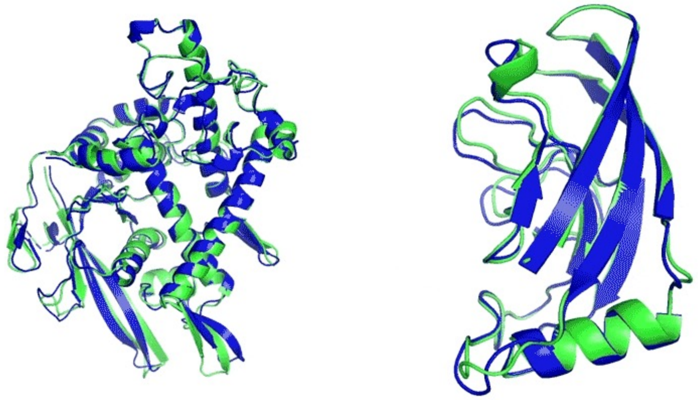
AlphaFold 1

Research

AlphaFold: Using AI for scientific discovery

January 15, 2020





T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

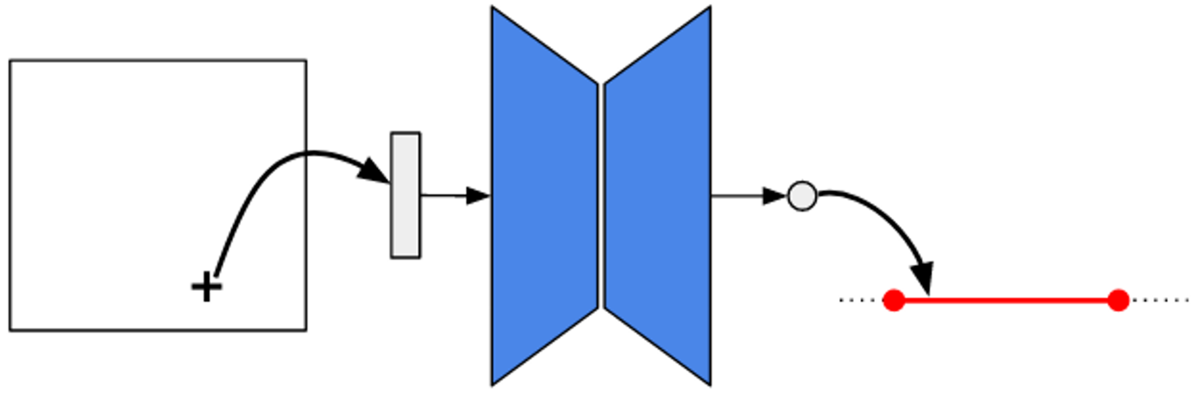
T1049 / 6y4f
93.3 GDT
(adhesin tip)

- Experimental result
- Computational prediction

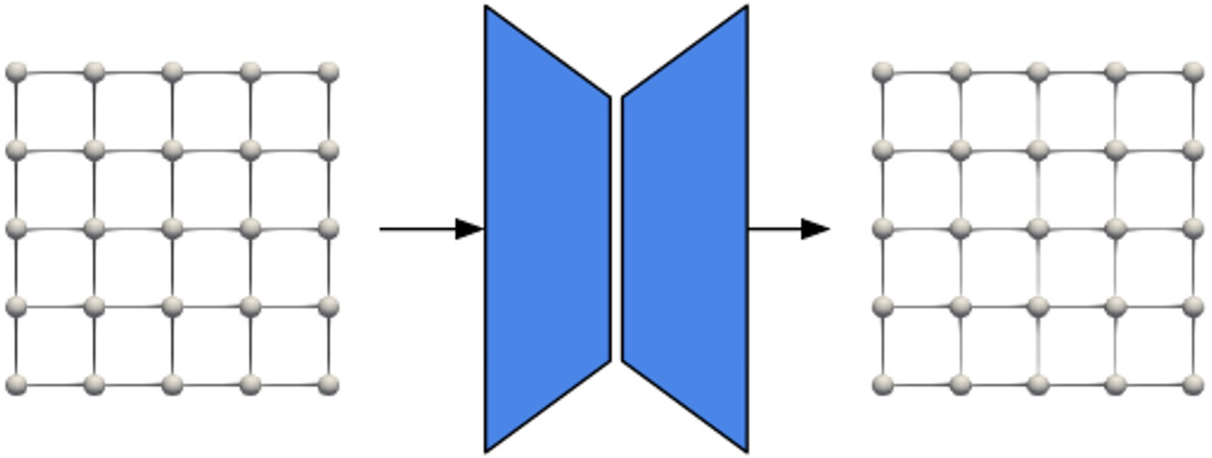
SciML: Architecture and loss functions

Pointwise architectures:

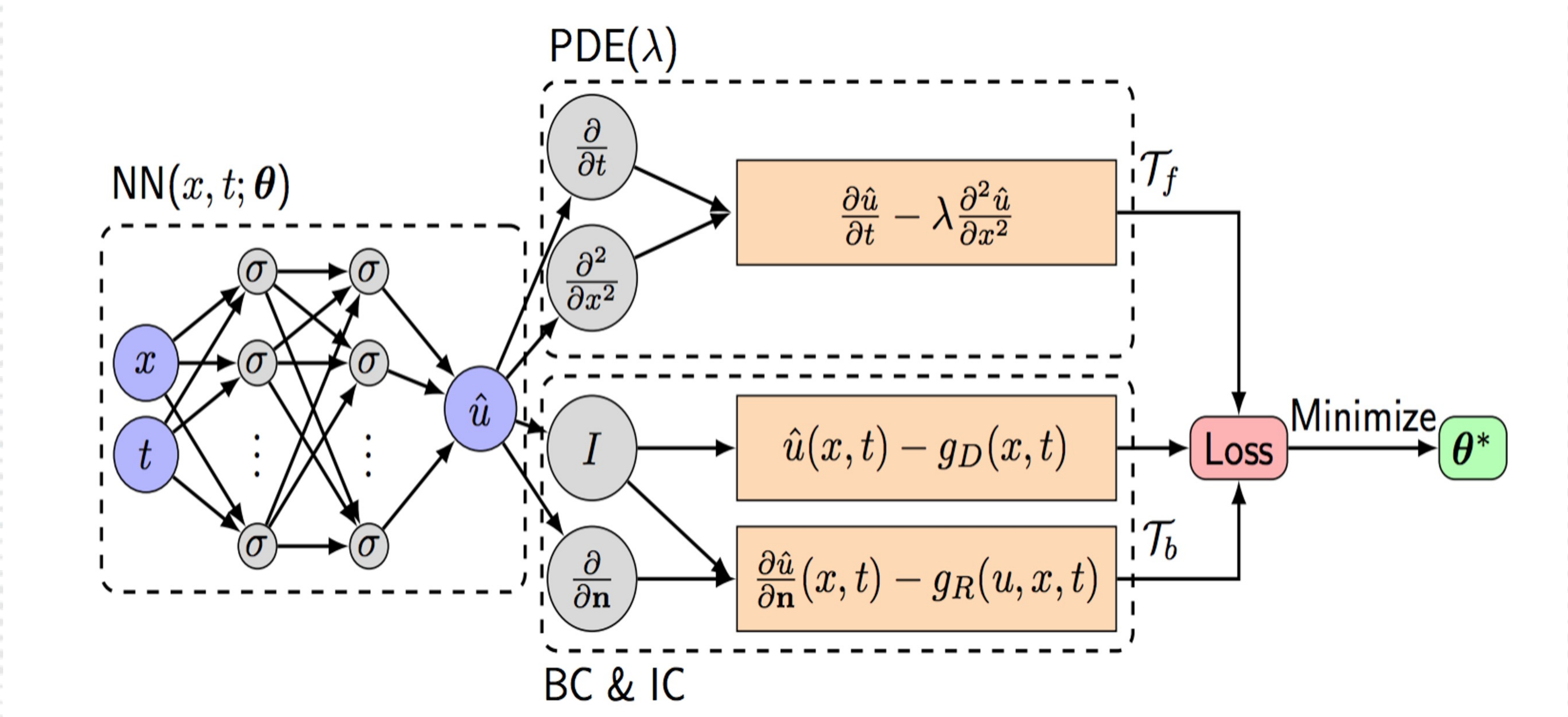
PINNs Physics Informed Neural Networks



Field based architectures

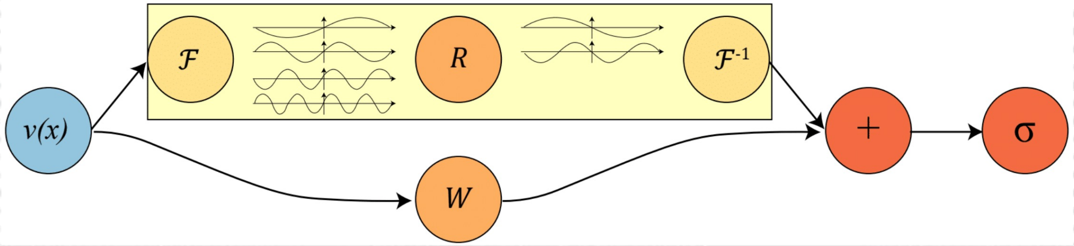


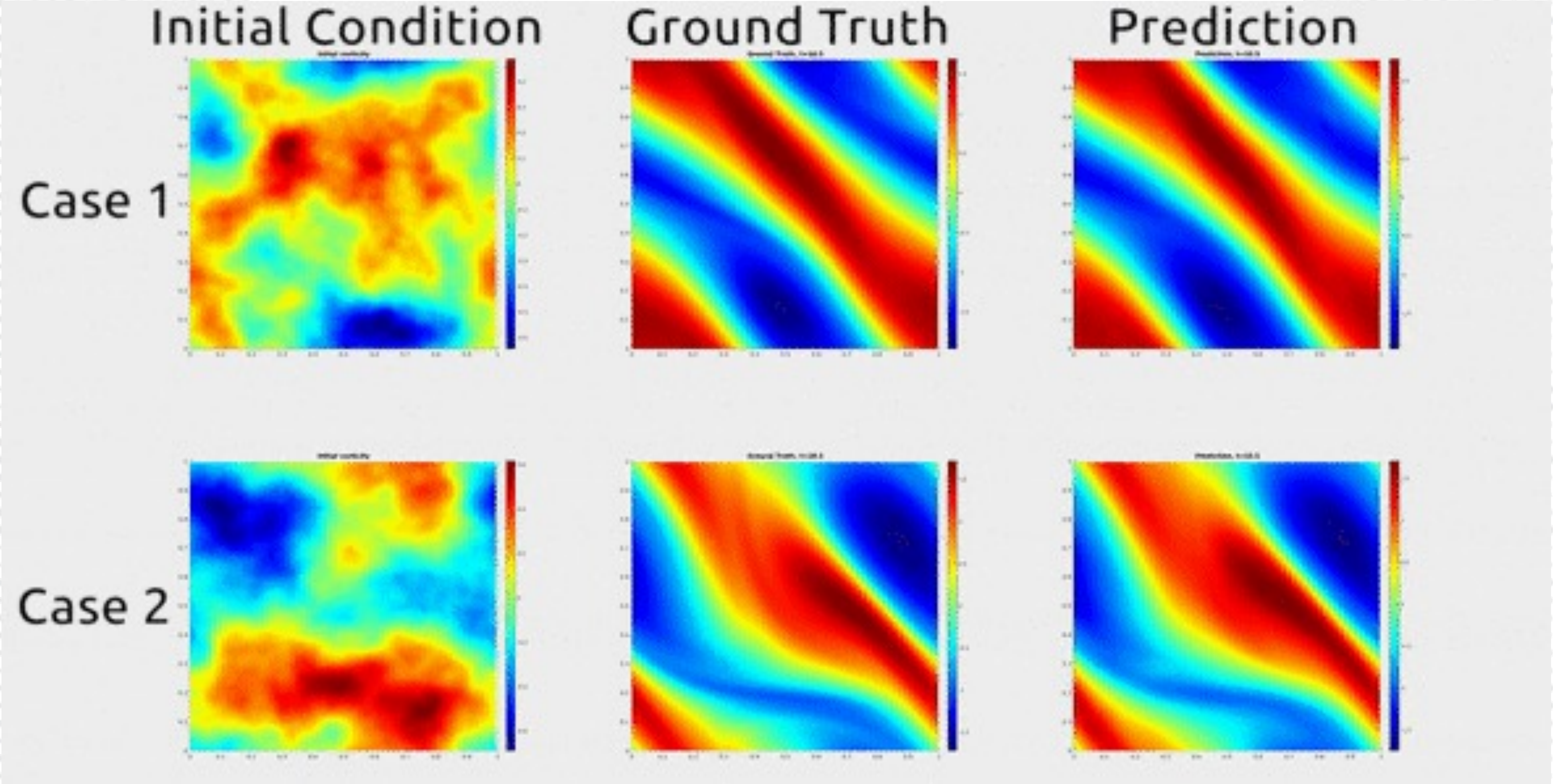
Physics Informed Neural Networks for point-wise solutions



Fourier Neural Operators

Operator learning can be taken as an image-to-image problem. The Fourier layer can be viewed as a substitute for the convolution layer.



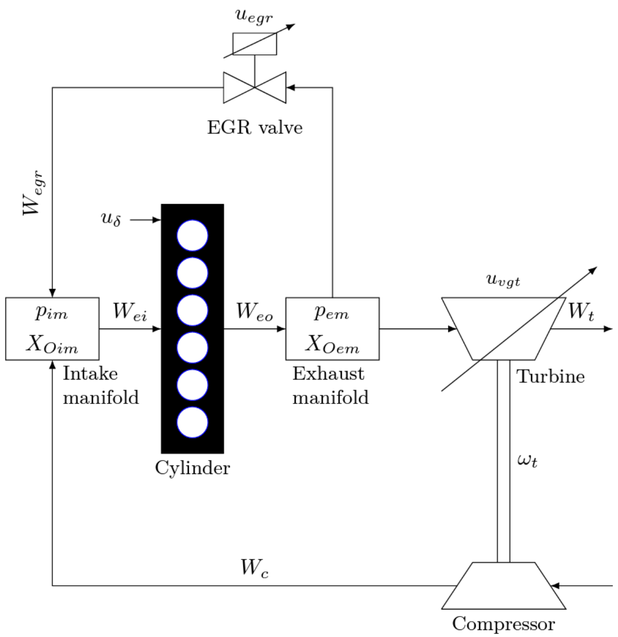


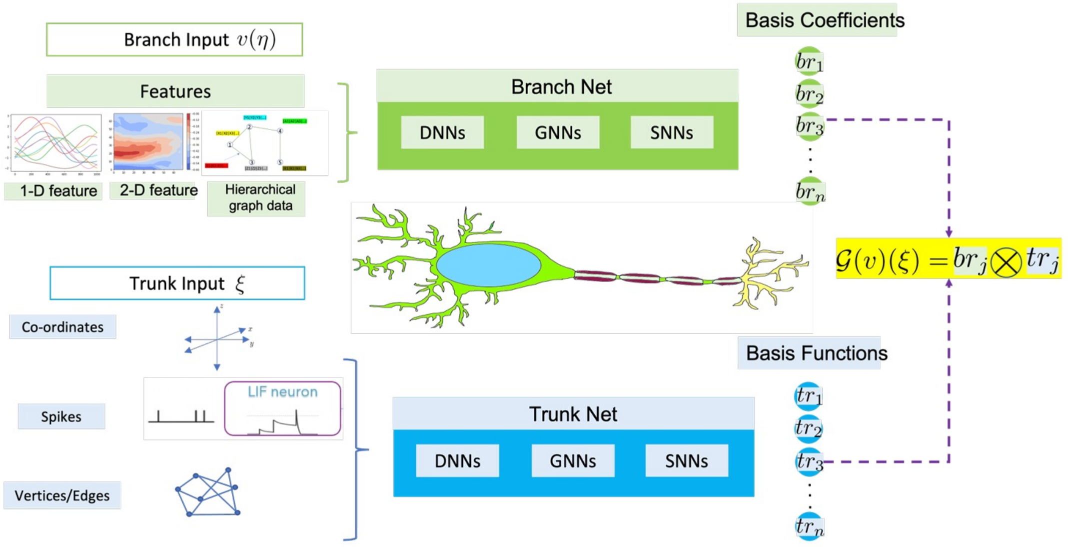
The Fourier layer just consists of three steps:

1. Fourier transform $\mathcal{F}$
2. Linear transform on the lower Fourier modes R
3. Inverse Fourier transform $\mathcal{F}^{-1}$

DeepONets

- Extends PINNs to arbitrary geometries and parametric datasets etc.

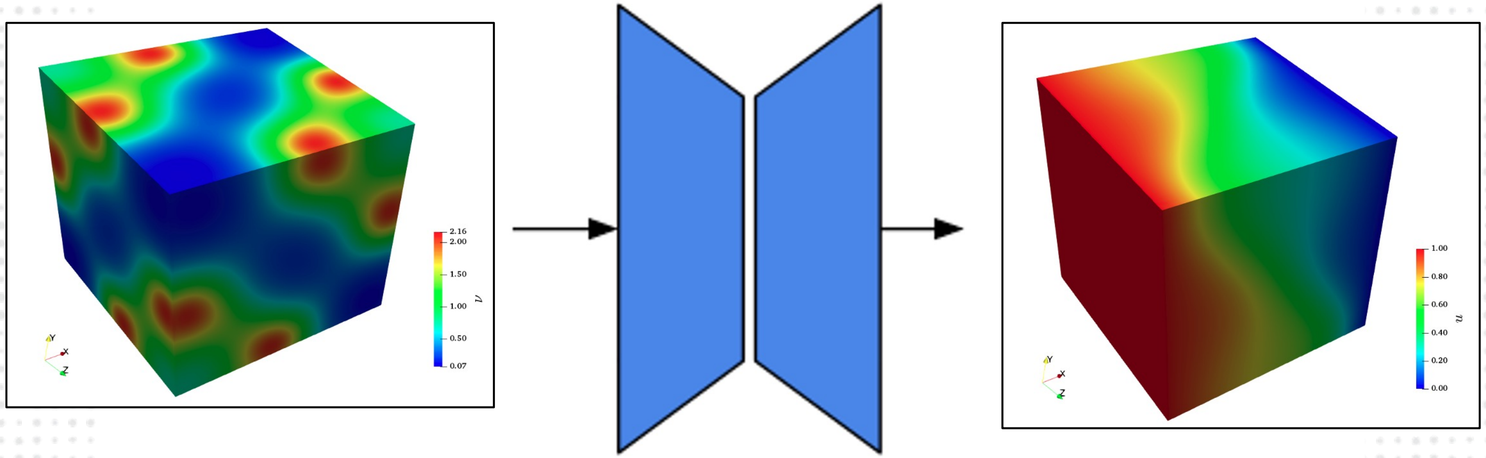




SciML: Forward problems vs Inverse problems

Forward solvers:

Neural PDE solvers



SciML for inverse and design