GENERALIZABLE AI: A NEW FOUNDATION

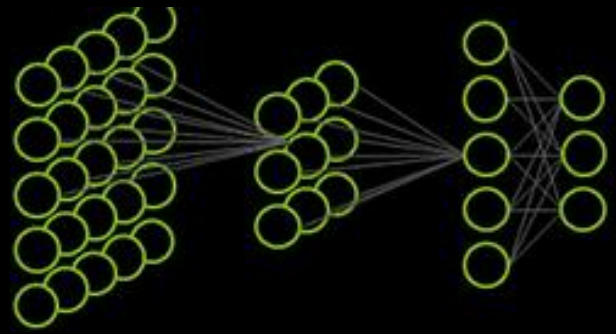
NVIDIA

Caltech

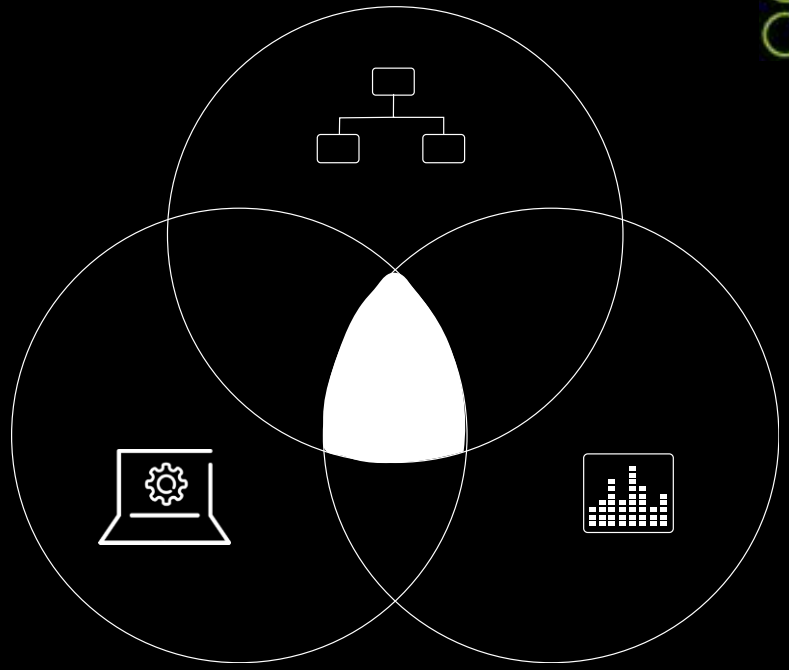
ANIMA ANANDKUMAR

TRINITY OF AI

ALGORITHMS



COMPUTE



DATA

IMPRESSIVE GROWTH OF AI

Wide range of domains

Deep reinforcement learning beats human champion

NVIDIA GAN generates photo-realistic images, passes Turing test

BUT NOTABLE FAILURES

AI is not living up to its hype

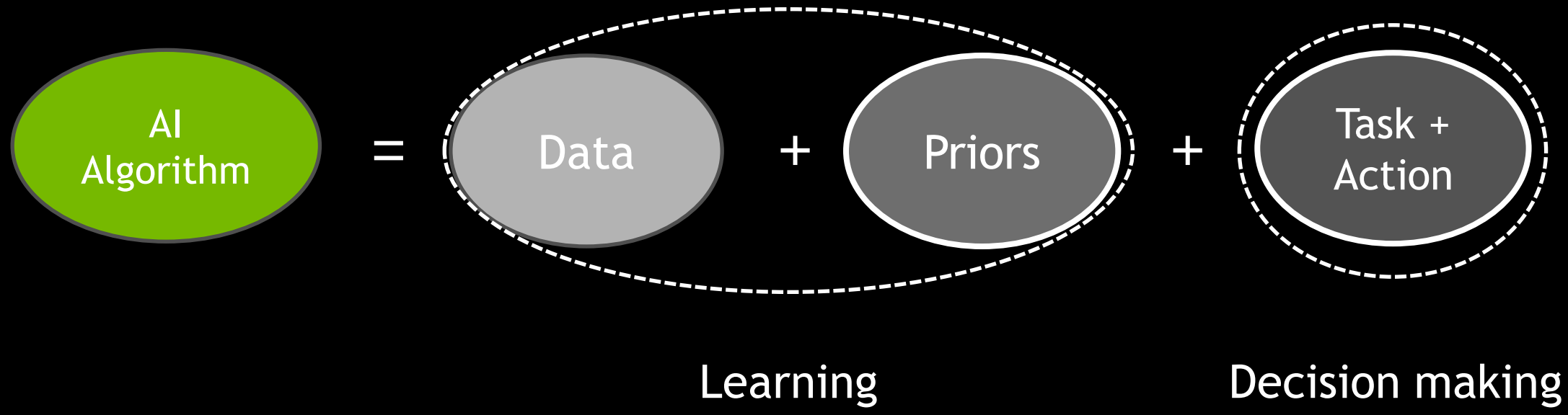
Safety-critical applications

Language understanding

How do we fix these gaps in deep learning?

PATH TO GENERALIZABLE AI

INGREDIENTS OF AN AI ALGORITHM



DEEP LEARNING STATUS QUO

AI Algorithm

=

Data

+

Priors

+

Task

Data hungry

- Massive datasets
- Expensive human labeling

DEEP LEARNING STATUS QUO

AI Algorithm

=

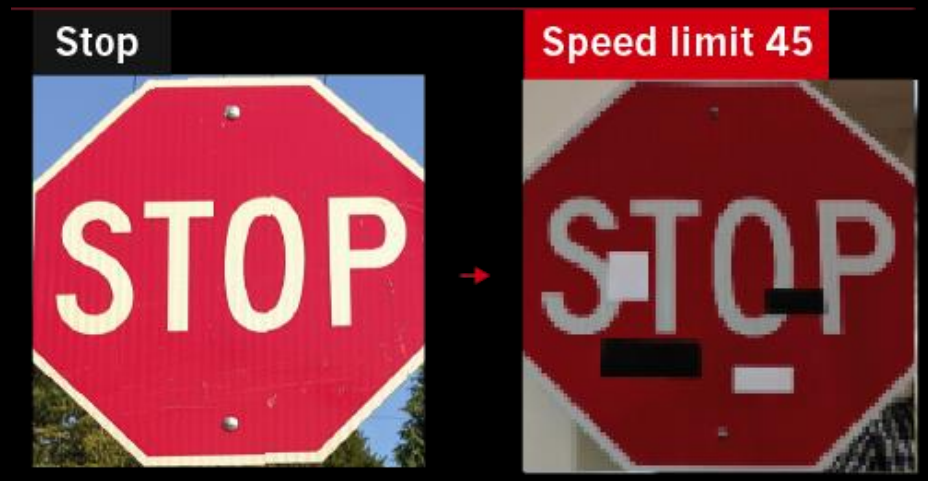
Data

+

Priors

+

Task



Not robust

- Easy to fool current models
- Not domain specific

DEEP LEARNING STATUS QUO

AI Algorithm

=

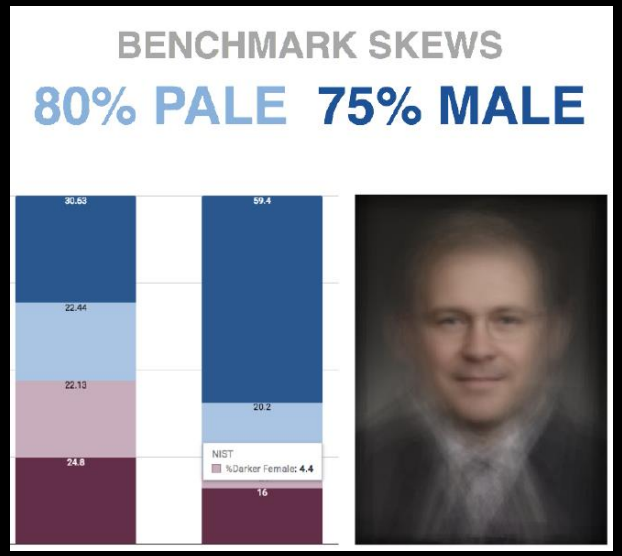
Data

+

Priors

+

Task



Simplistic

- Fixed tasks
- Limited benchmarks

NEXT FRONTIER IN AI

AI Algorithm

=

Data

+

Priors

+

Task

Unsupervised

- Disentanglement learning
- Domain adaptation

Robust

- Recurrent feedback
- Domain knowledge
- Compositionality

Adaptive

- Multi-task & domains
- Online and continual learning

NEXT FRONTIER IN AI

AI Algorithm

=

Data

+

Priors

+

Task

Unsupervised

- Disentanglement learning
- Domain adaptation

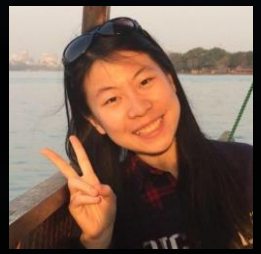
Robust

- Recurrent feedback
- Domain knowledge
- Compositionality

Adaptive

- Multi-task & domains
- Online and continual learning

BRAIN-INSPIRED ARCHITECTURES WITH RECURRENT FEEDBACK

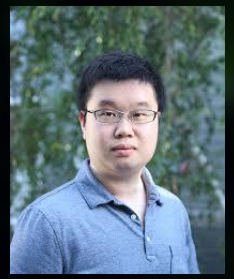


Yujia Huang

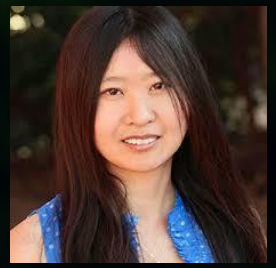
Sihui Dai

James Gornet

Tan Nyugen



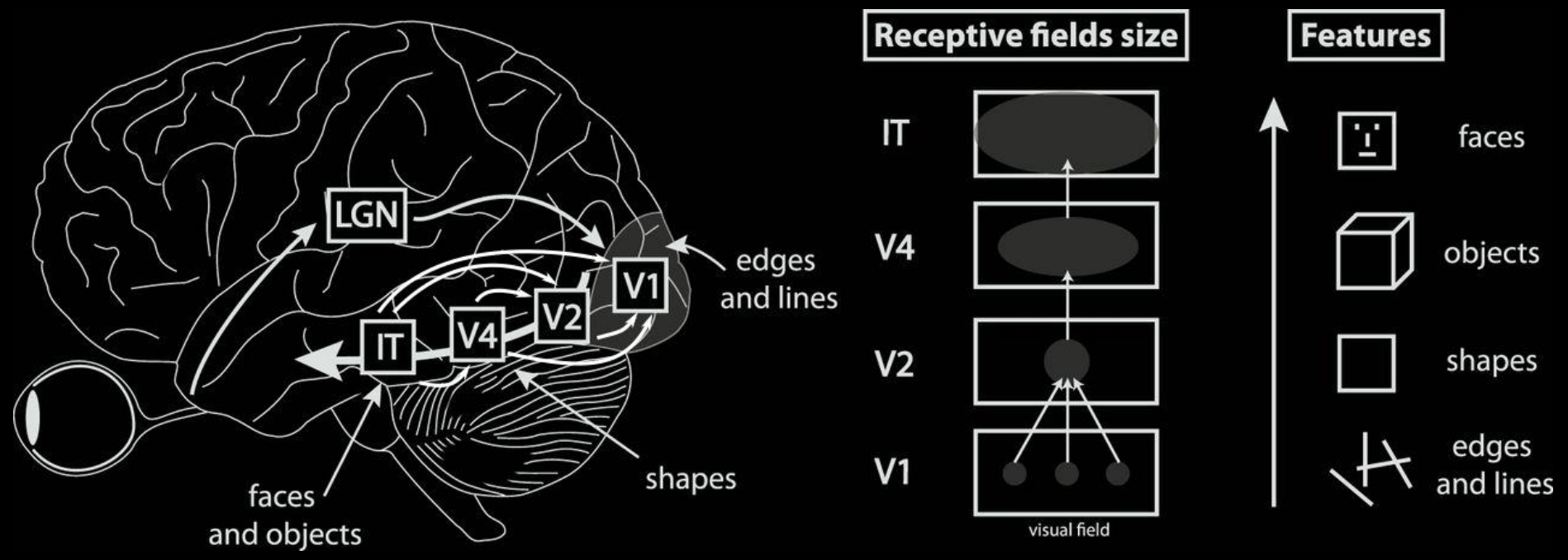
Zhiding Yu



Doris Tsao

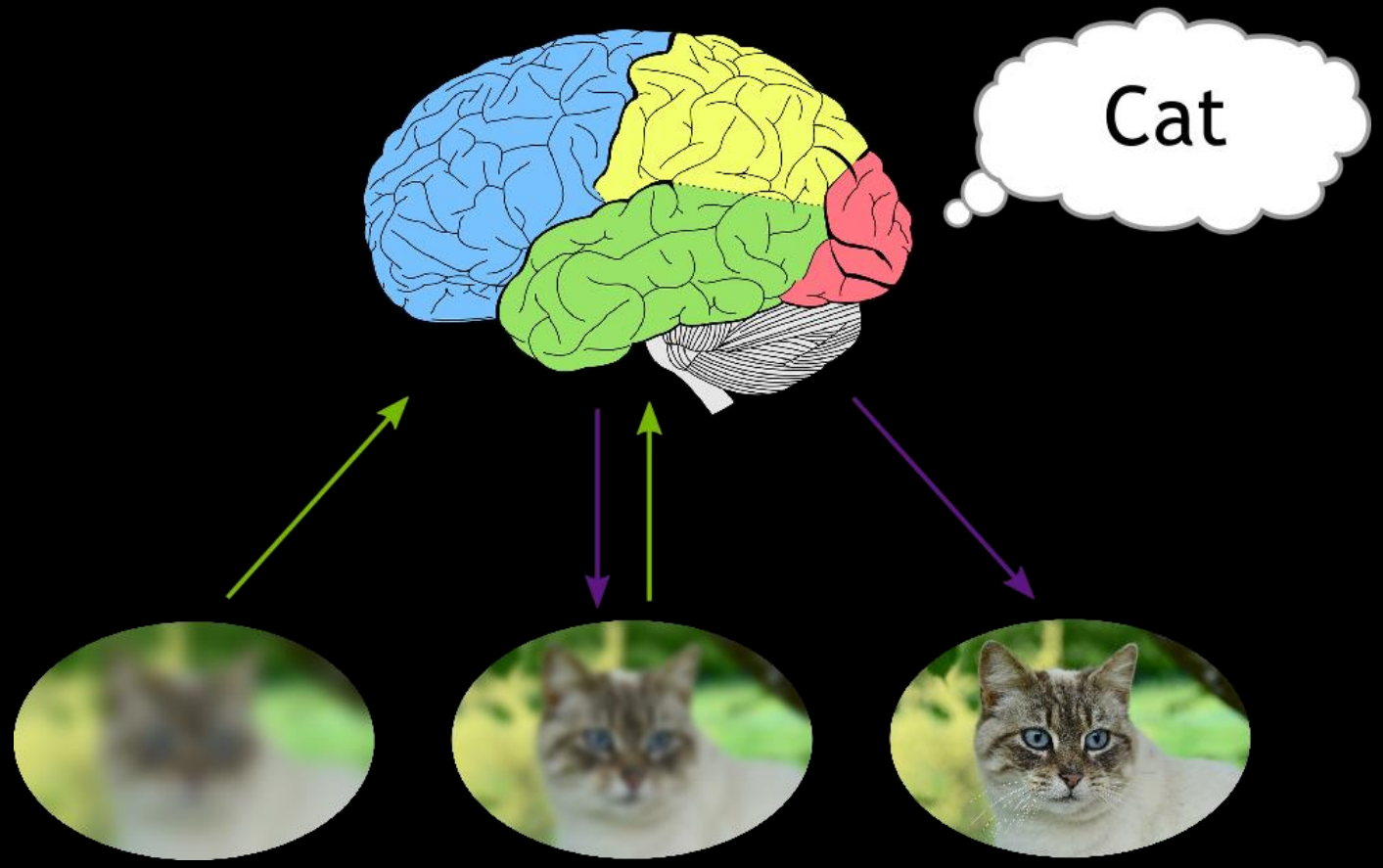
A

THE HUMAN BRAIN IS HIERARCHICAL



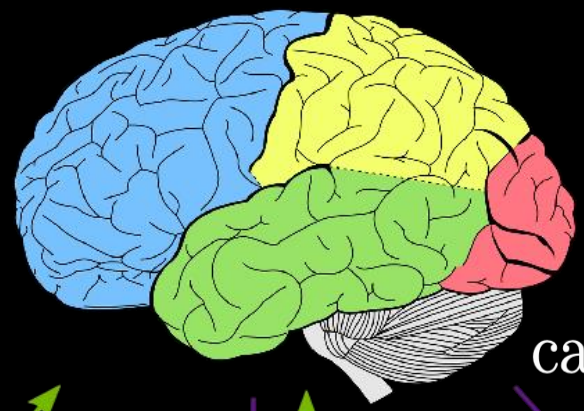
Adapted from *Journal of Vision* (2013), **13**, 10

HUMAN VISION IS ROBUST



THE BRAIN IS BAYESIAN

$$\underbrace{p(\text{image} \mid \text{objects})}_{\text{likelihood}} \underbrace{p(\text{objects})}_{prior}$$

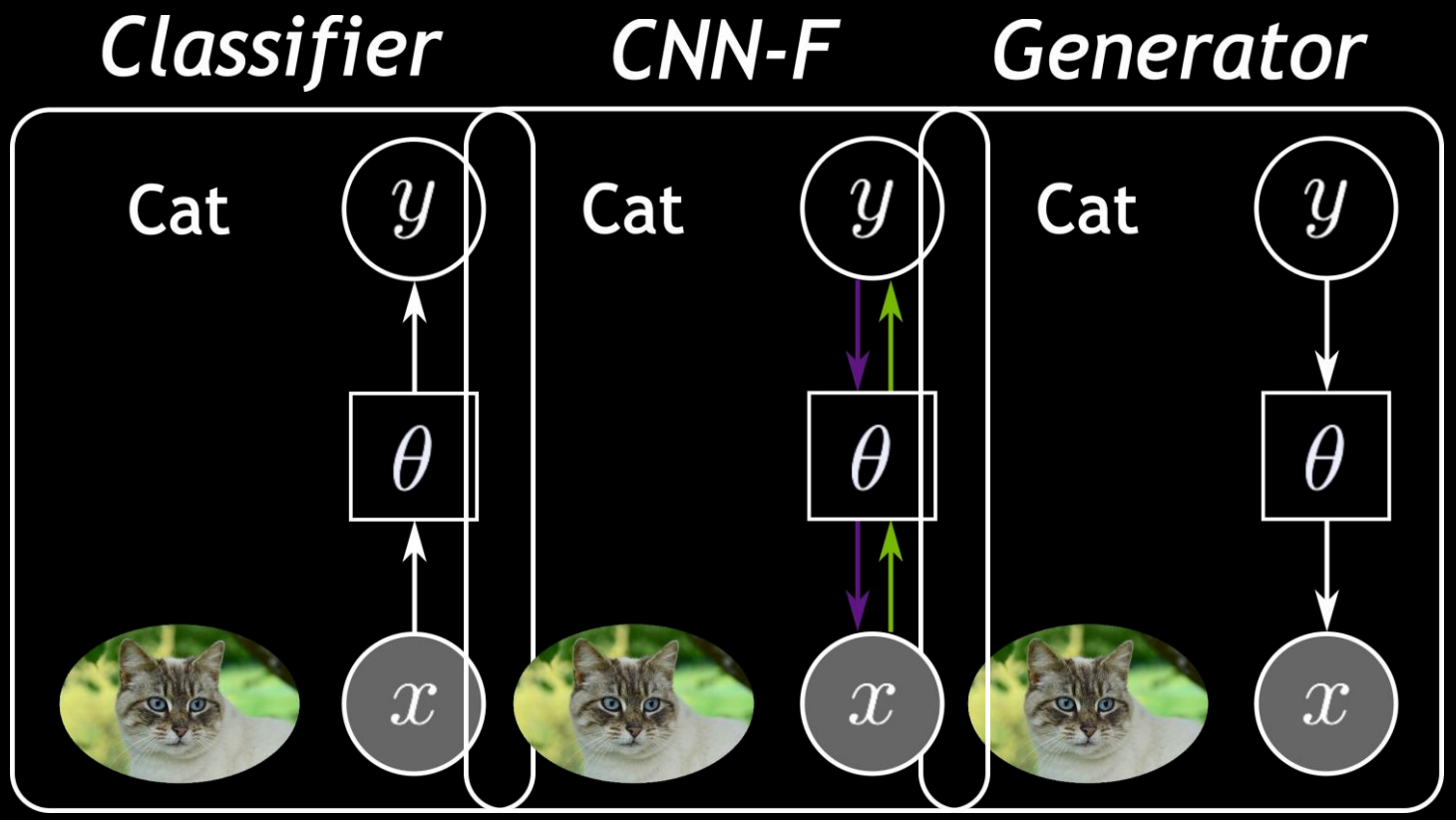


$$\text{cat} = \arg\max_{\text{object}} p(\text{object} \mid \text{image})$$

image

image of a cat

COMBINING CLASSIFIER AND GENERATOR THROUGH FEEDBACK CONNECTIONS

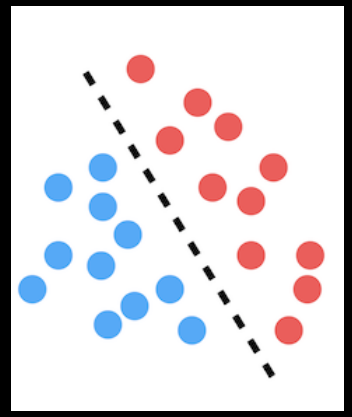


GENERATIVE VS DISCRIMINATIVE CLASSIFIER

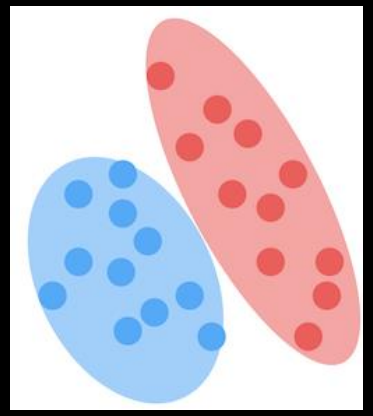
$$p(x, y) \rightarrow p(y|x)$$

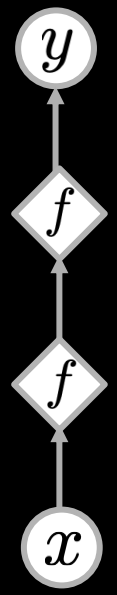
$$p(x, y, z) \rightarrow p(y|x, z)$$

Logistic Regression

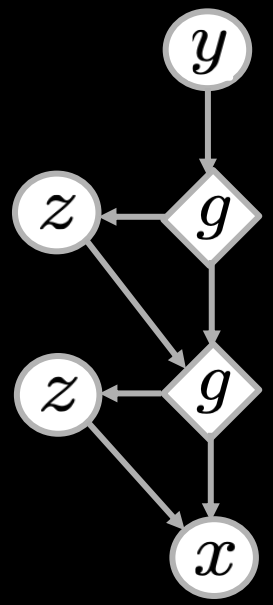


Gaussian Mixture



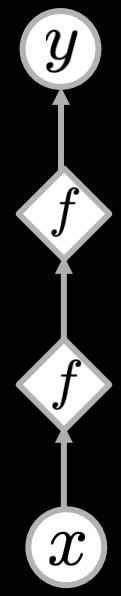


CNN

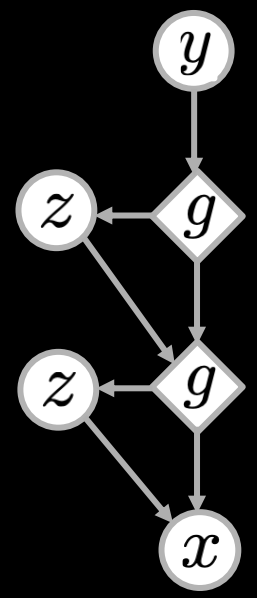


Deconvolutional Generative Model

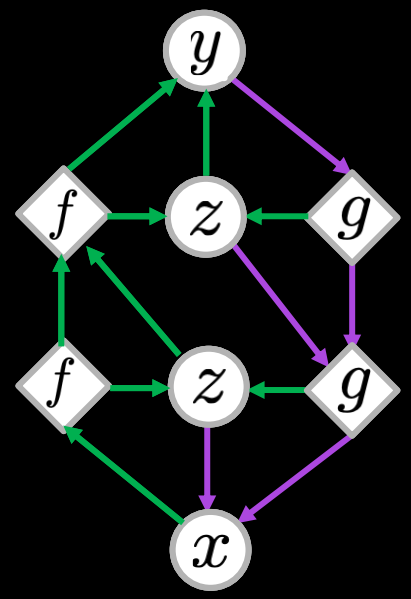
MESSAGE PASSING NETWORK



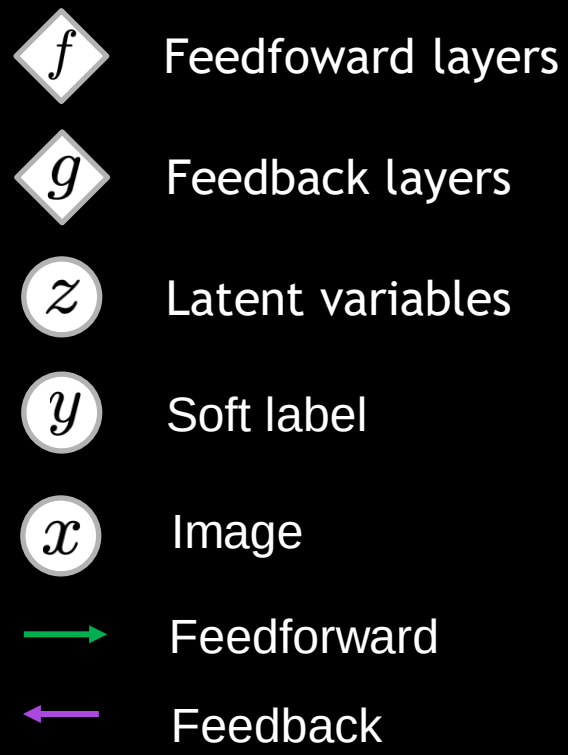
CNN



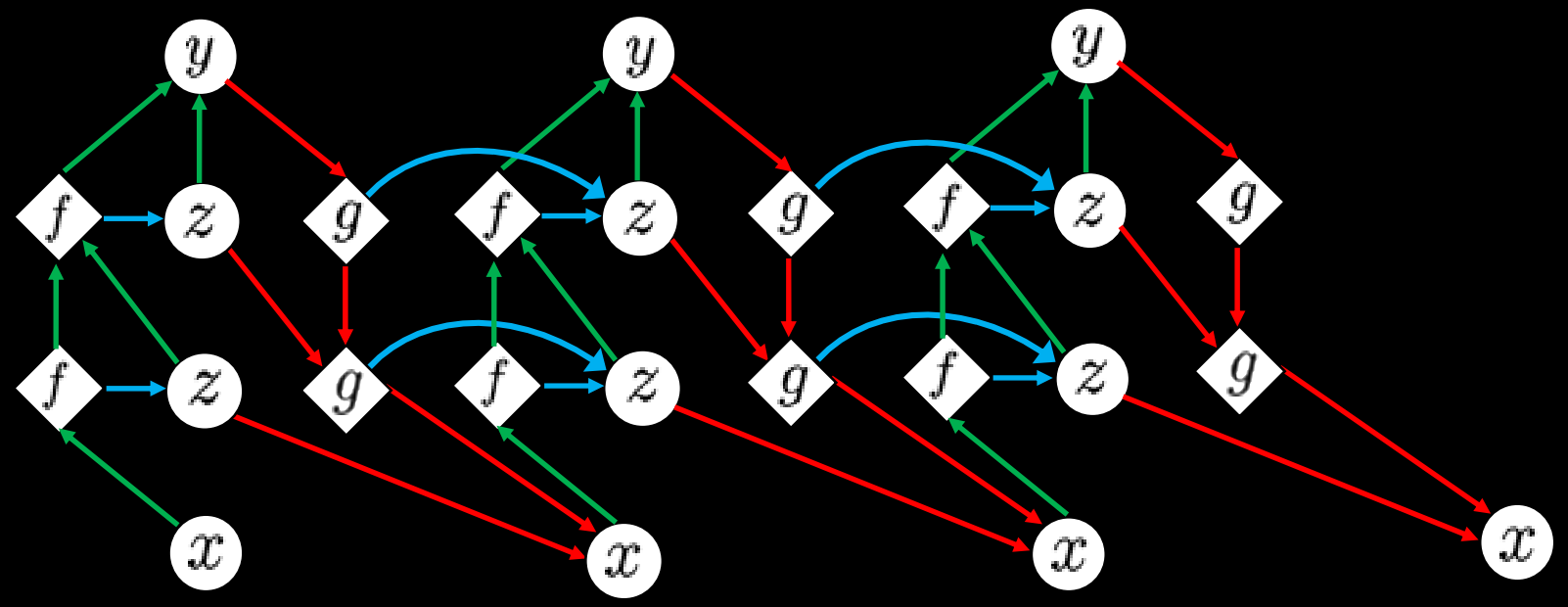
Generative feedback



CNN-F



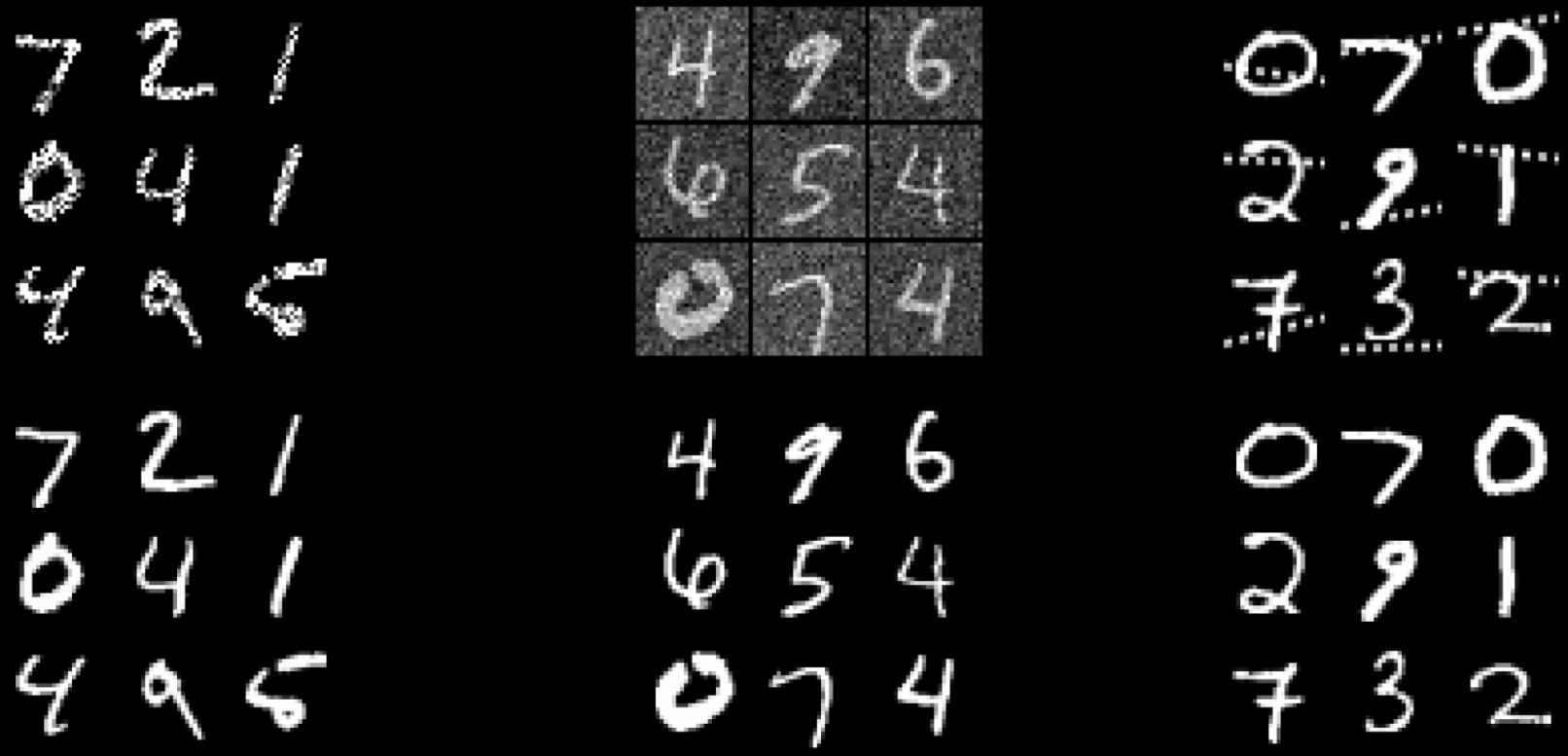
SELF-CONSISTENCY THROUGH RECURRENT FEEDBACK



CNN-F CAN REPAIR DISTORTED IMAGES WITHOUT SUPERVISION

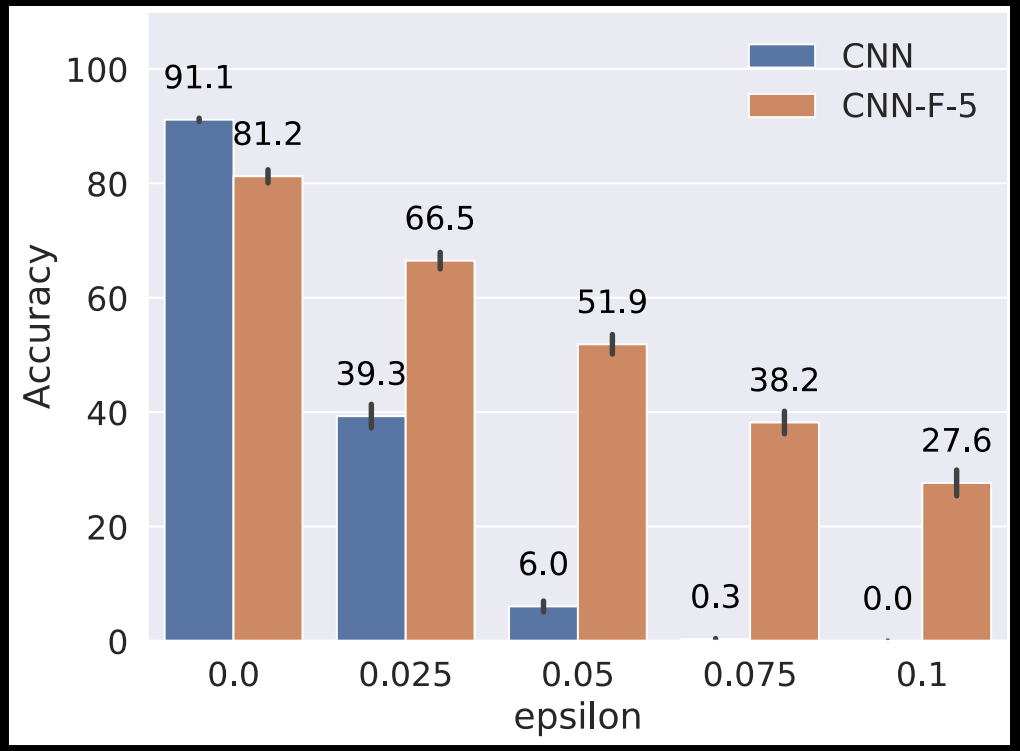
Shot Noise | *Gaussian Noise* | *Dotted Line*

Corrupted images



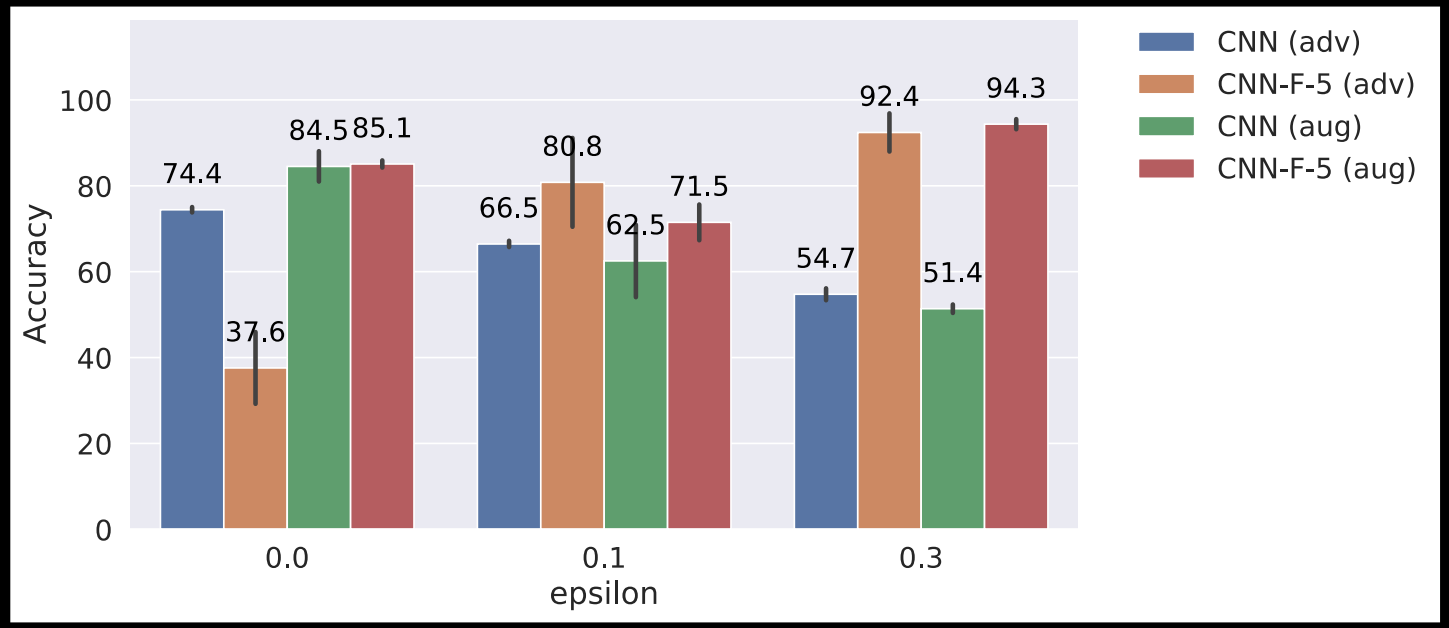
Ground-truth

CNN-F IMPROVES ADVERSARIAL ROBUSTNESS



- Standard training on Fashion-MNIST.
- Attack with PGD-40.
- CNN-F has higher adversarial robustness than CNN.

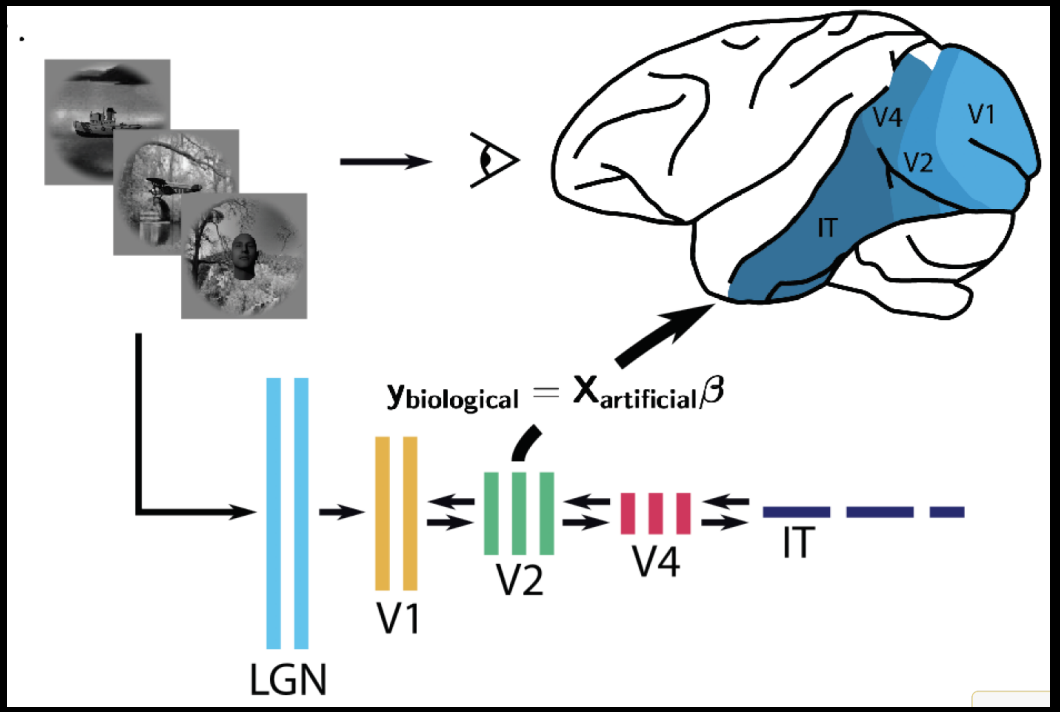
CNN-F COMBINED WITH ADVERSARIAL TRAINING

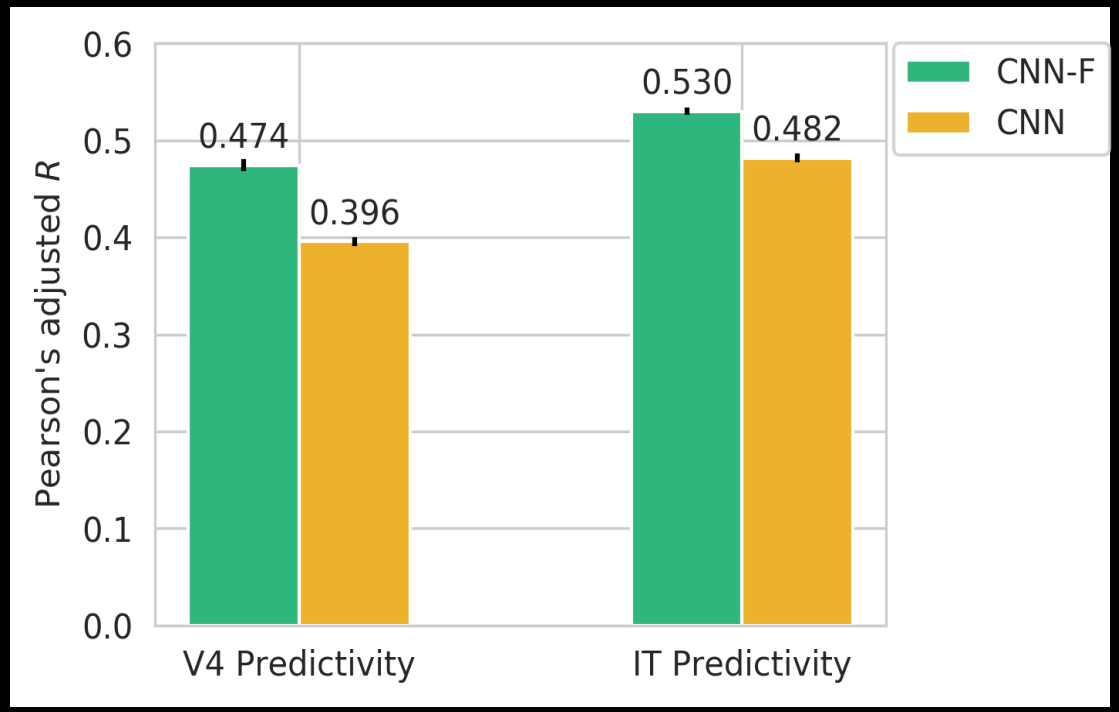


- Adversarial training on Fashion-MNIST.
- Trained with PGD-40 (eps=0.3). Attack with PGD-40.
- CNN-F augmented with adversarial images achieves high accuracy for both clean and adversarial data.

CNN-F HAS HIGHER BRAIN SCORE

Feedback is biologically more plausible





TAKE-AWAYS

Recurrent generative feedback for robust learning

- Human brain has feedback pathways for top-down inference
- Internal generative model of the world
- Bayesian brain: bottom up feedforward + top down feedback
- Robustness is inherent in CNN-F
- Biological plausibility in CNN-F

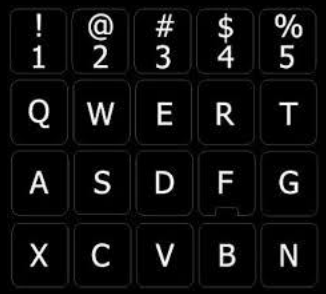
NEURO-SYMBOLIC SYSTEMS FOR COMPOSITIONAL REASONING

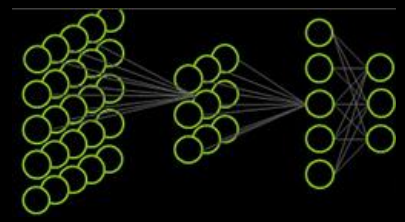
Forough Arabshahi

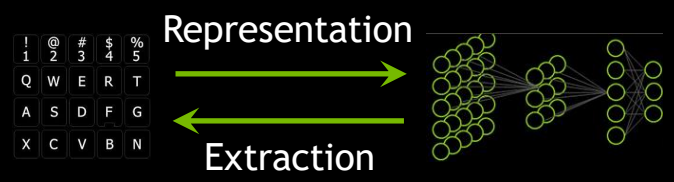
Sameer Singh

A

SYMBOLISTS VS. CONNECTIONISTS

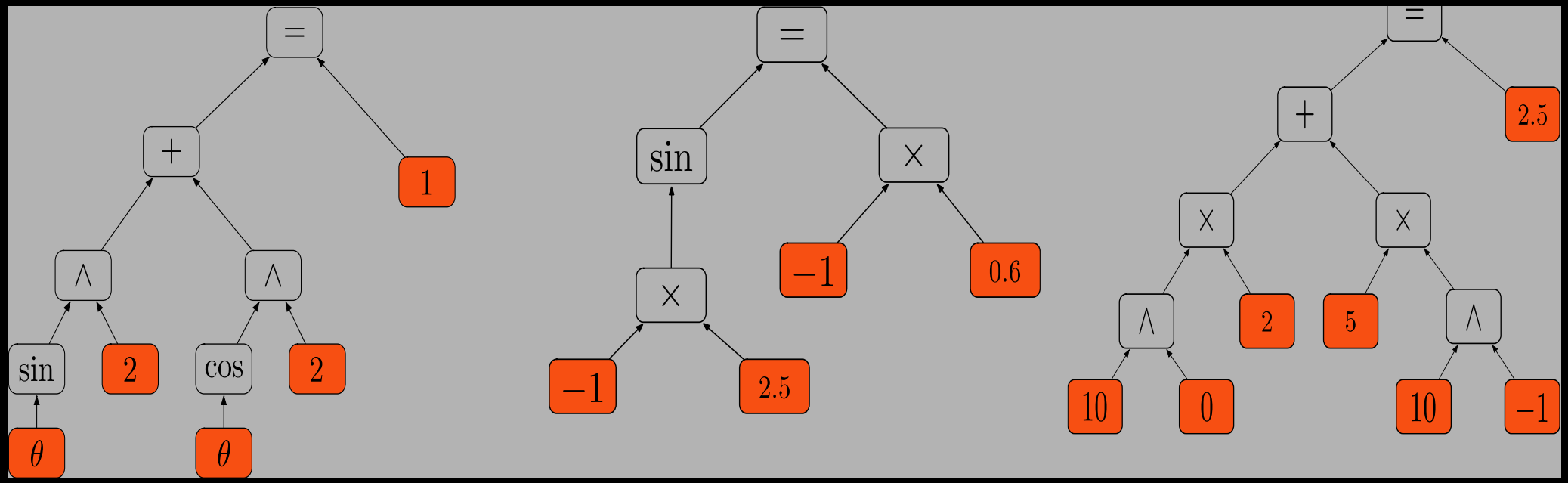






	Symbolic	Connectionist	Neuro-Symbolic
Explainability	✓	✗	✓
Generalization & knowledge coverage	✗	✓	✓
Extrapolation	✗	✗	✓

TYPES OF TRAINING EXAMPLES



$\sin^2(\theta) + \cos^2(\theta) = 1$

Symbolic Expressions

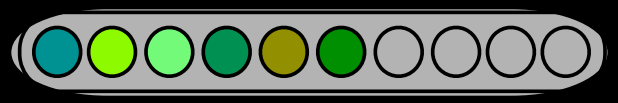
$\sin(-2.5) = -0.6$

Function Evaluation

Decimal Tree for 2.5

Number Encoding

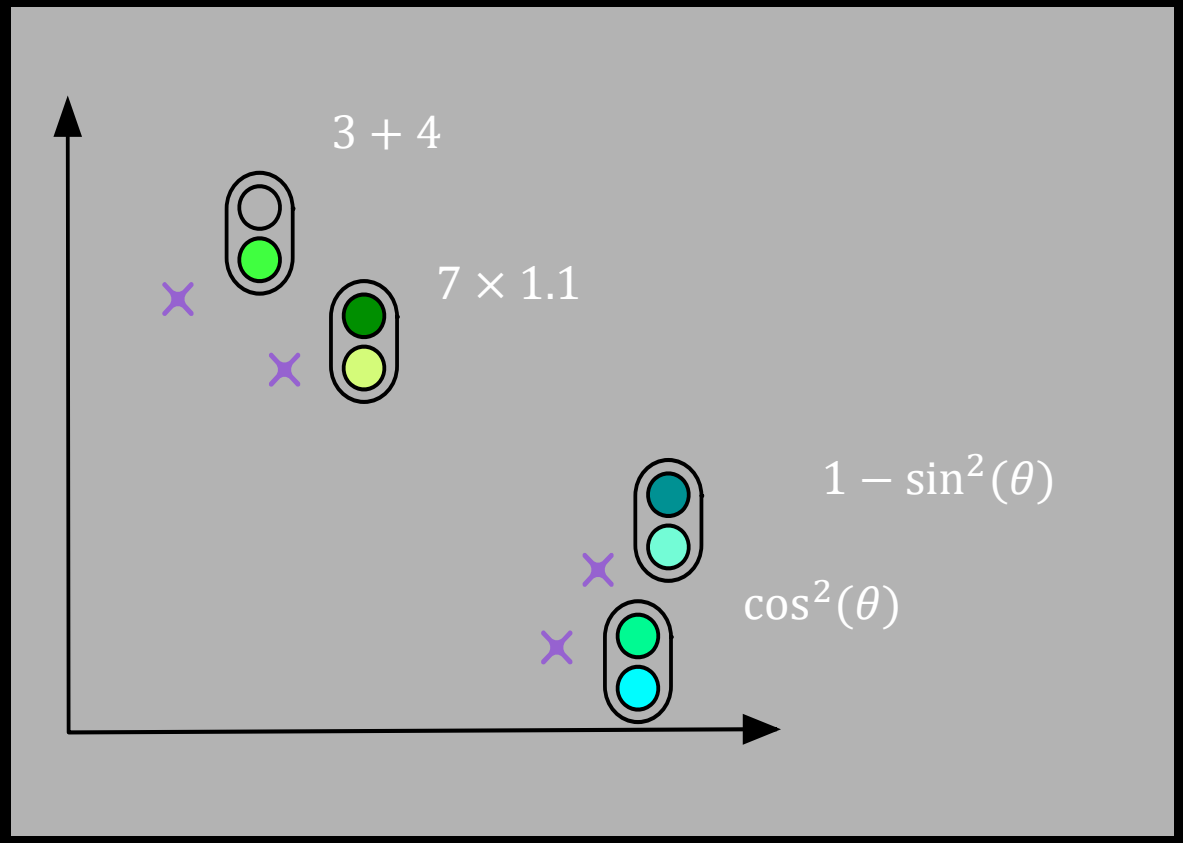
CONTINUOUS REPRESENTATIONS FOR REASONING



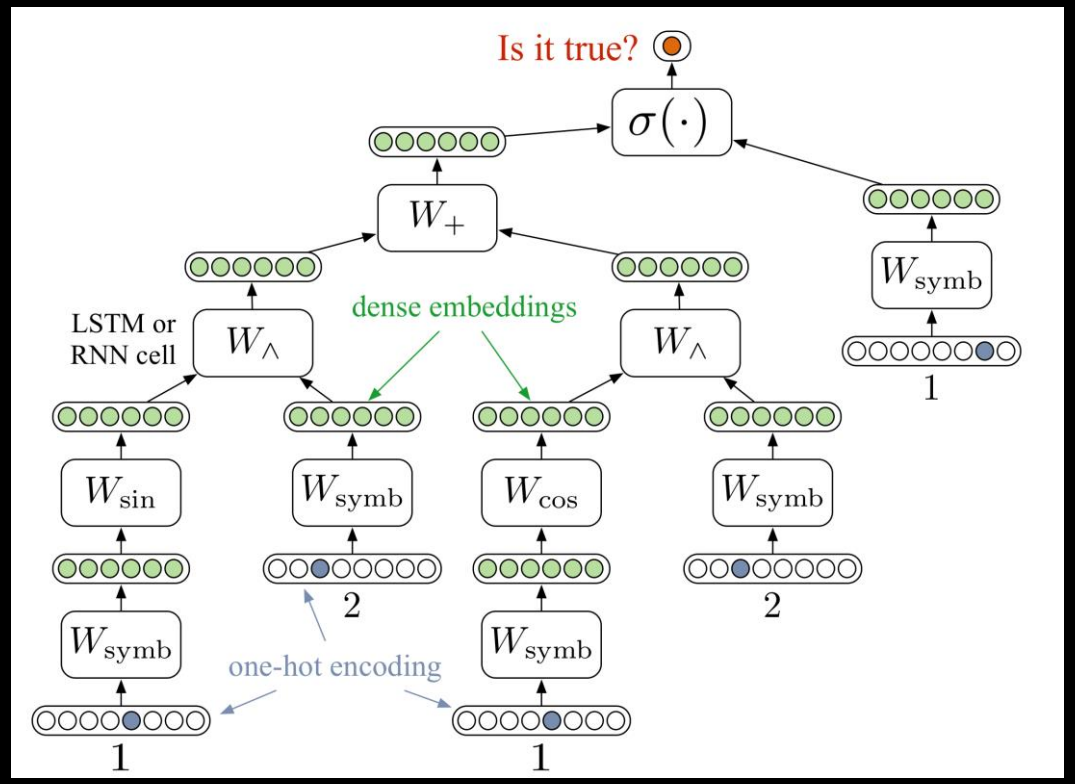
Representations of symbols, numbers and functions in common embedding space

2.45 θ 2 ...

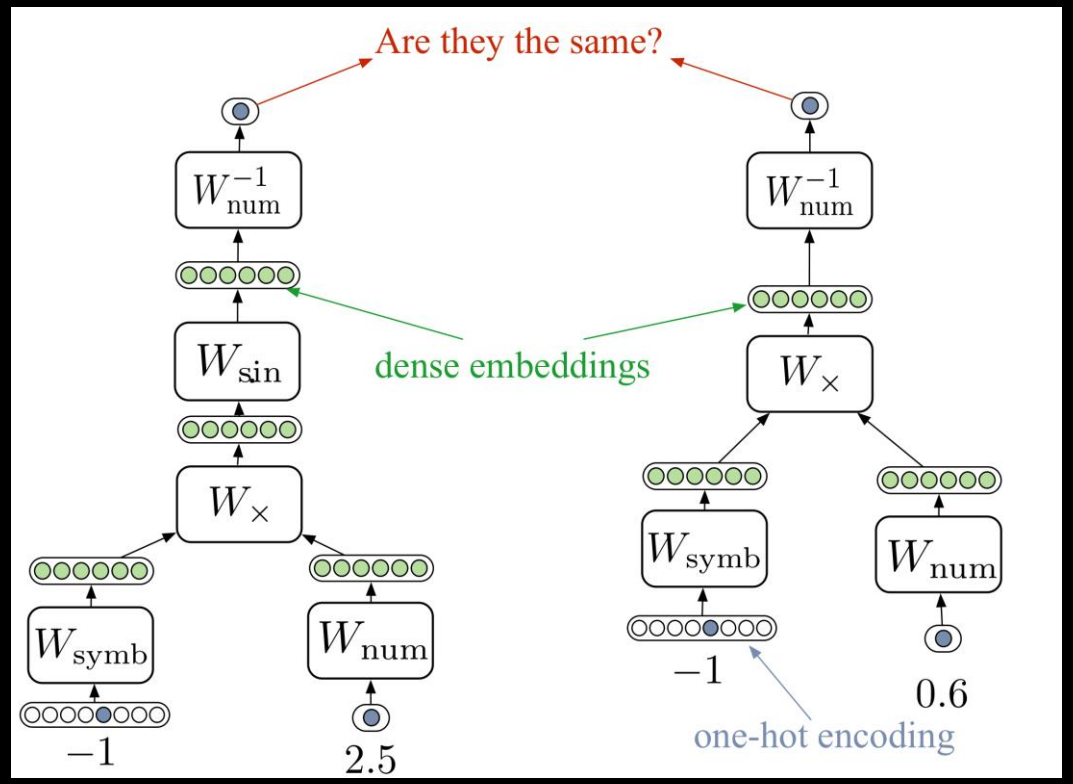
sin cos $\times$ ...



TREE-LSTM FOR COMPOSITIONALITY

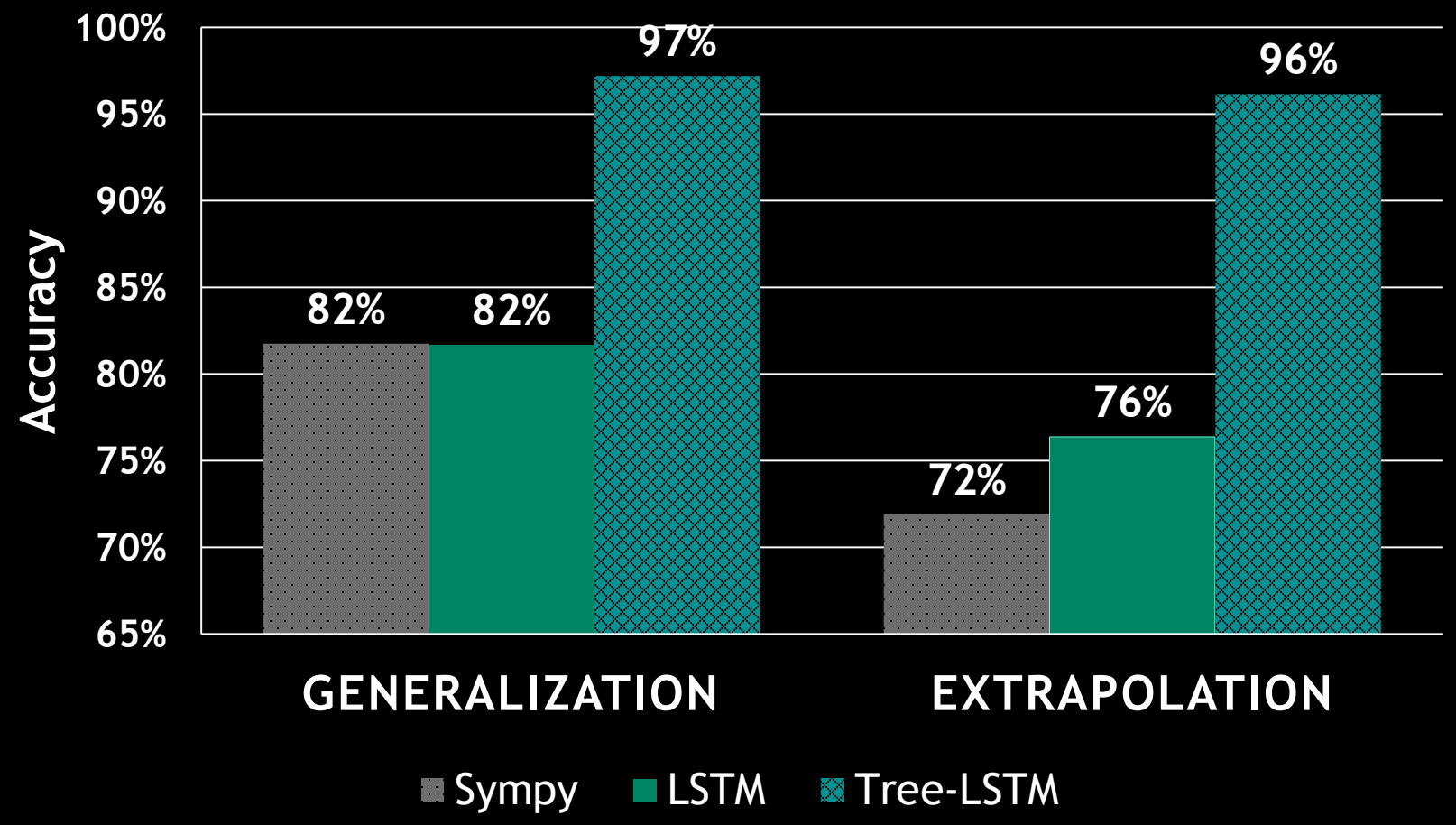


$$\sin^2(\theta) + \cos^2(\theta) = 1$$



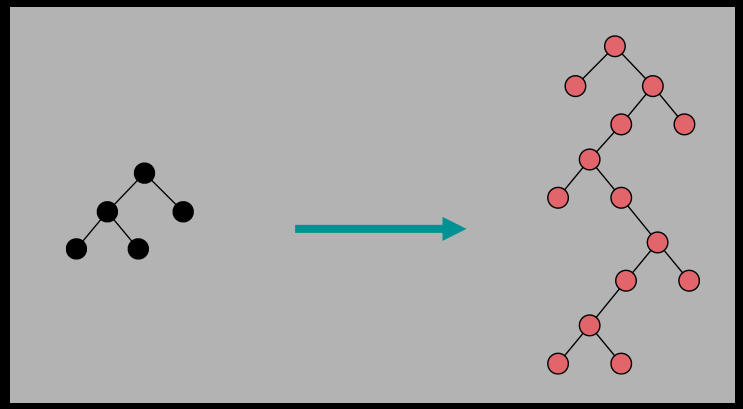
$$\sin(-2.5) = -0.6$$

EQUATION VERIFICATION



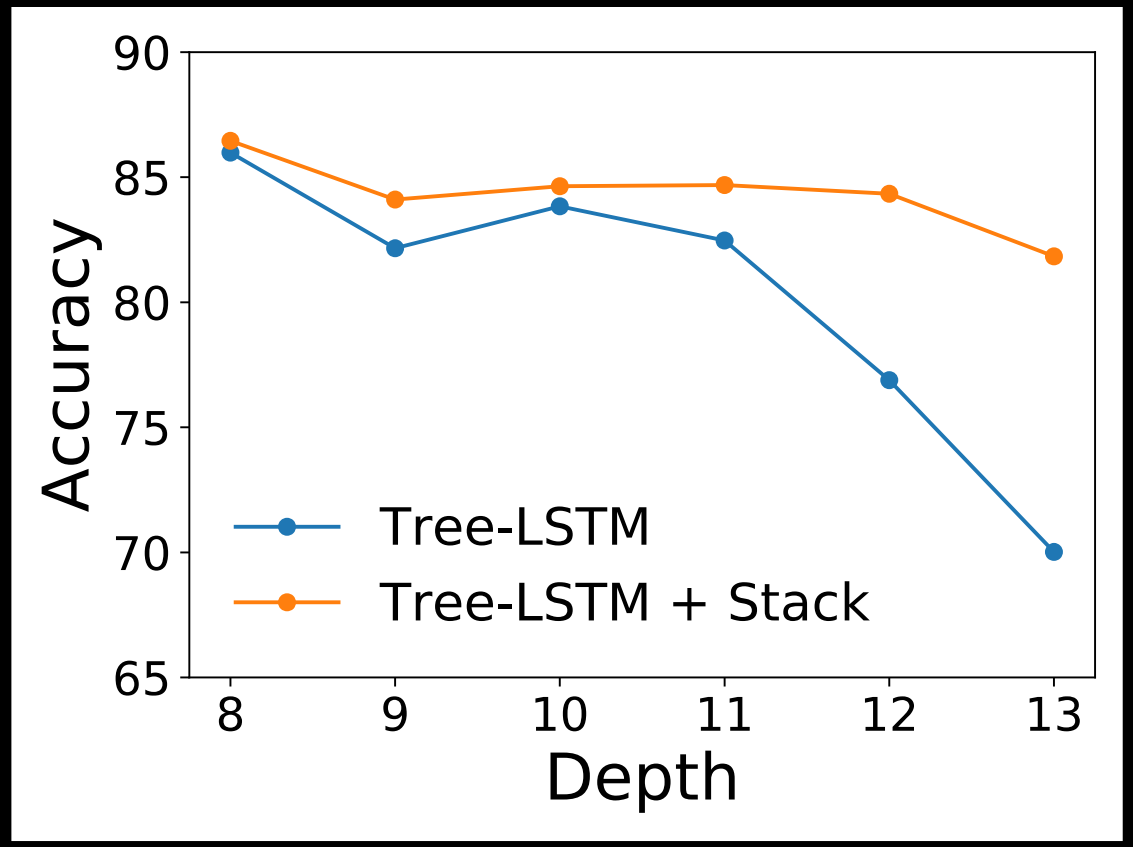
AUGMENTING WITH STACK MEMORY

Differentiable memory for extrapolation to harder examples



Train:
Depth 1-7

Test:
Depth 8-13



TAKE-AWAYS

Neuro-symbolic systems for compositional learning

- Math reasoning tasks
- Combine symbolic expressions and numerical data
- Generalizable and composable representation of functions
- Differentiable memory stack for extrapolation to harder examples

AI4SCIENCE

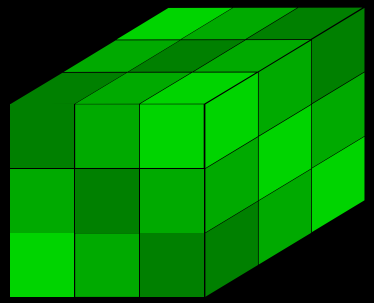
ROLE OF PRIORS

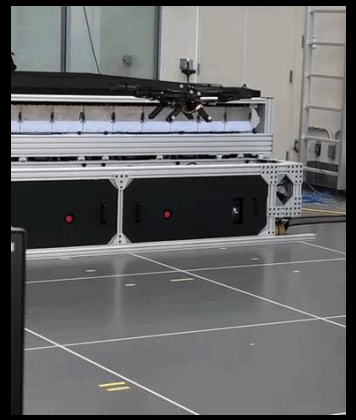
Learning = Data + Priors

How to use structure and domain knowledge to design Priors?

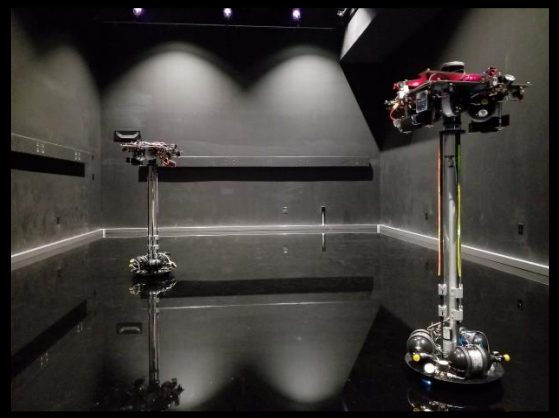
Examples of Priors

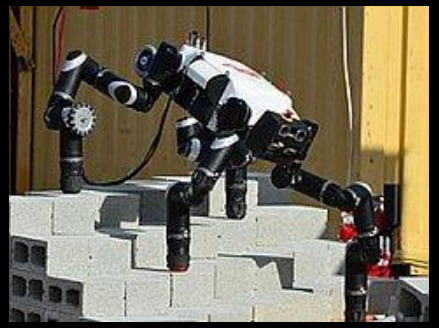
- Tensors and graphs
- Laws of nature
- Simulations

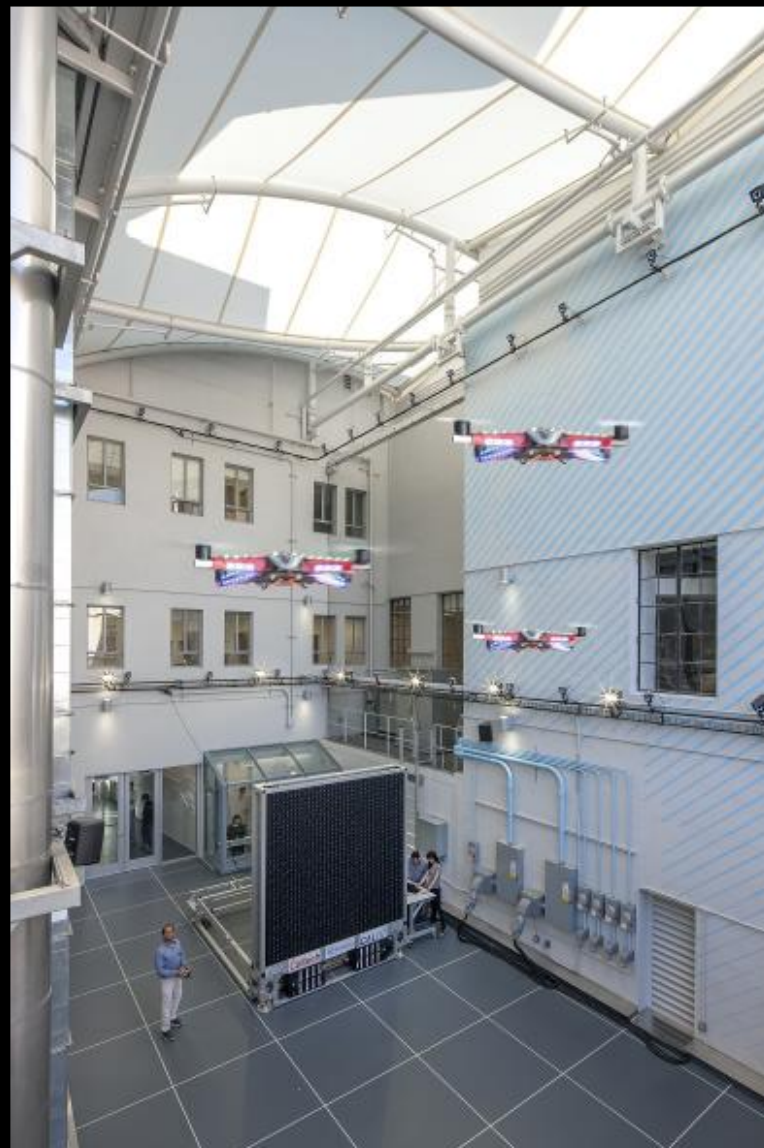




AUTONOMOUS DYNAMIC ROBOTS AT CAST, CALTECH







LEARNING RESIDUAL DYNAMICS FOR DRONE LANDING

f = nominal dynamics
$\tilde{f}$ = learned dynamics

New State

Current Action (aka control input)

$$s_{t+1} = f(s_t, a_t) + \tilde{f}(s_t, a_t) + \epsilon$$

Current State

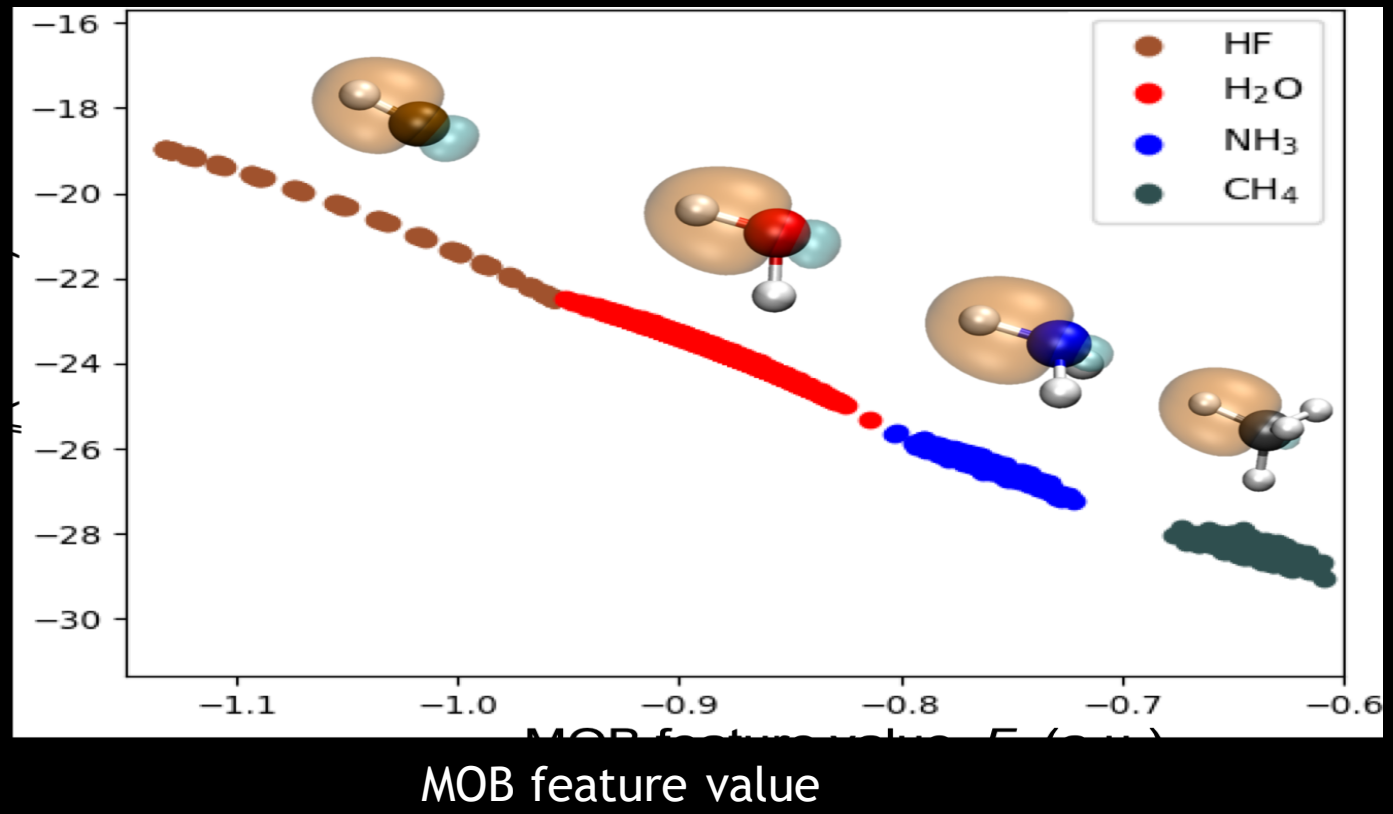
Unmodeled Disturbance

Our method is

- Provably robust and safe
- Generalizes to higher landing speeds

3D Landing Performance

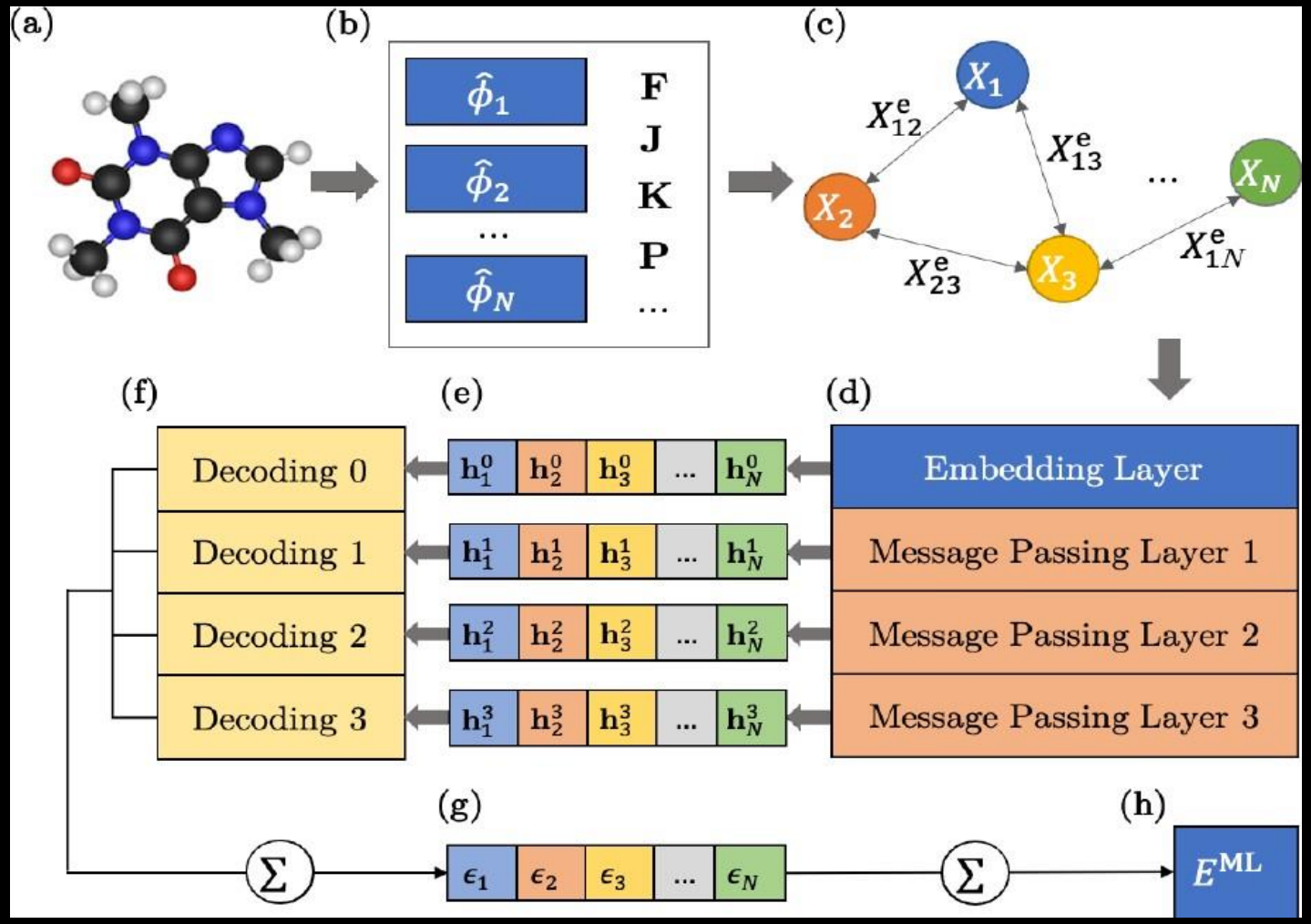
QUANTUM FEATURES IN CHEMISTRY



MOB-ML features: universal mapping in chemical space

M. Welborn, Z. Qiao, F. R. Manby, A. Anandkumar, T. F. Miller III, arXiv:2007.08026.

ORBNET: MOB + GRAPH NEURAL NETWORKS

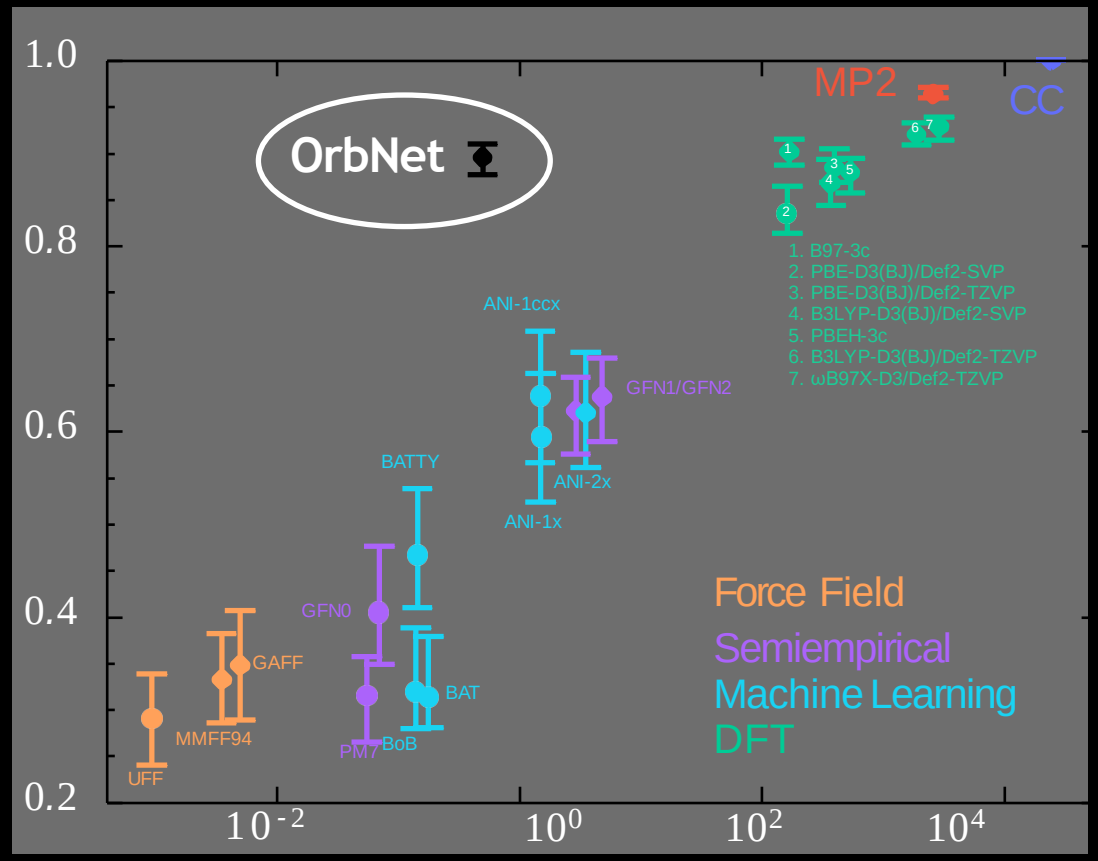


M. Welborn, Z. Qiao, F. R. Manby, A. Anandkumar, T. F. Miller III, arXiv:2007.08026.

ORBNET: 1000X SIMULATION SPEED-UP

Quantum-mechanical accuracy at semi-empirical cost

Drug-molecule conformer stability rankings (R^2)



Time-to-solution (s)

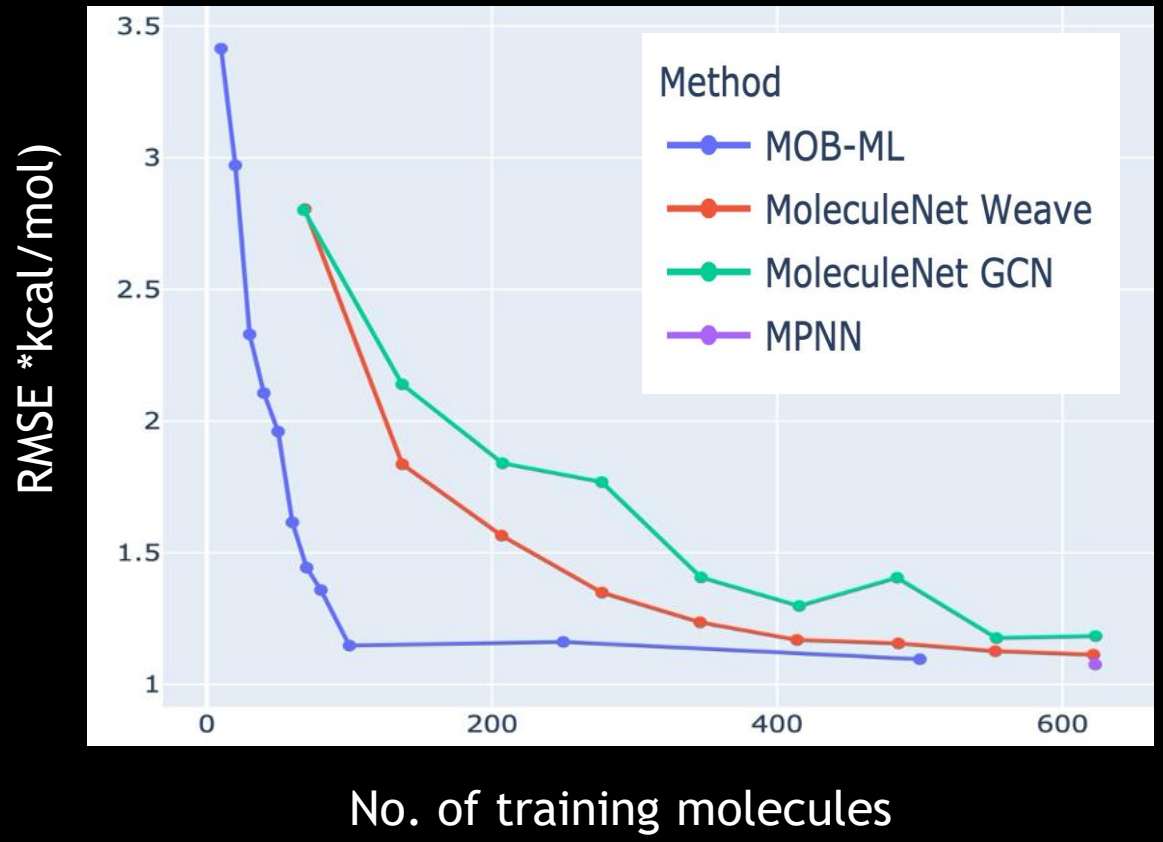
Test data: Drug-like molecules with 10-50 heavy atoms

Zero shot generalization: testing on molecules ~10x larger

Welborn, Z. Qiao, F. R. Manby, A. Anandkumar, T. F. Miller III, arXiv:2007.08026.

STATE-OF-ART DATA EFFICIENCY

MOB-ML vs. others for water



MOB-ML works across solvents

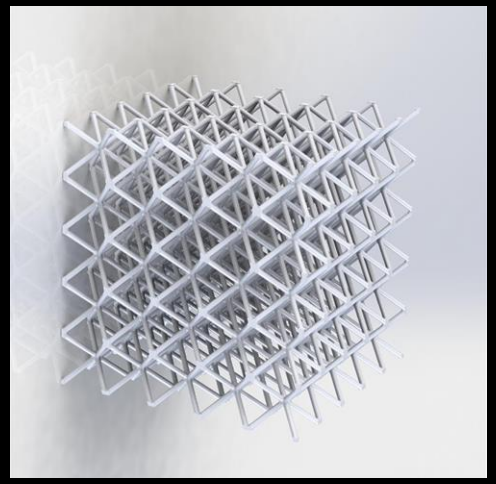
Solvent	Training Data	MAE (kcal/mol)
Benzene	50	0.57
Carbon tetrachloride	50	0.40
Chloroform	40	0.84
Cyclohexane	30	0.44
Diethylether	40	0.58
Hexadecane	40	0.57
Octanol	50	0.84

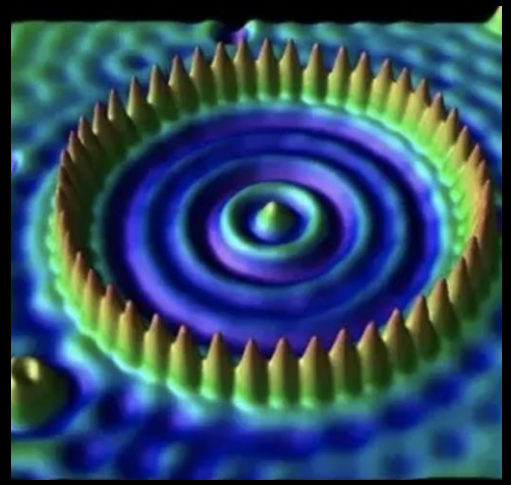
MoleculeNet: *Chem. Sci.*, 2018, **9**, 513; MPNN: *J. Chem. Inf. Model.* 2019, **59**, 3370

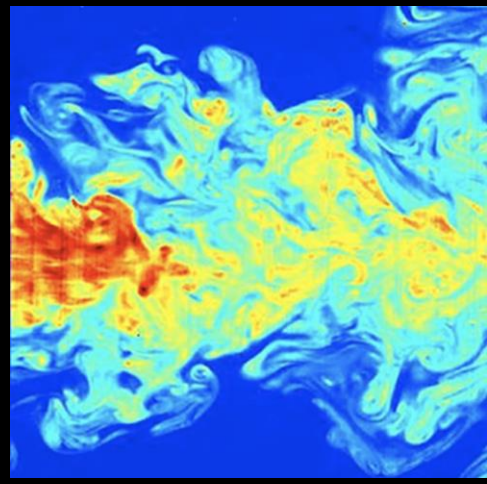
LEARNING FAMILY OF PDE

Problems in science and engineering reduce to PDEs.

Learning mapping from parameters to output through operator learning

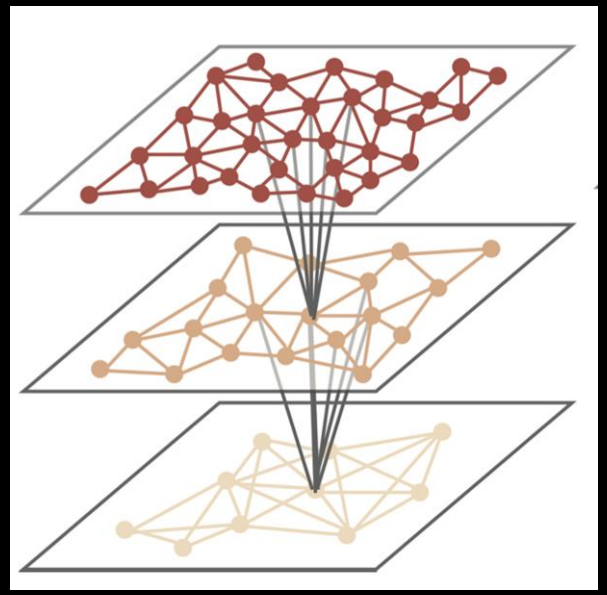


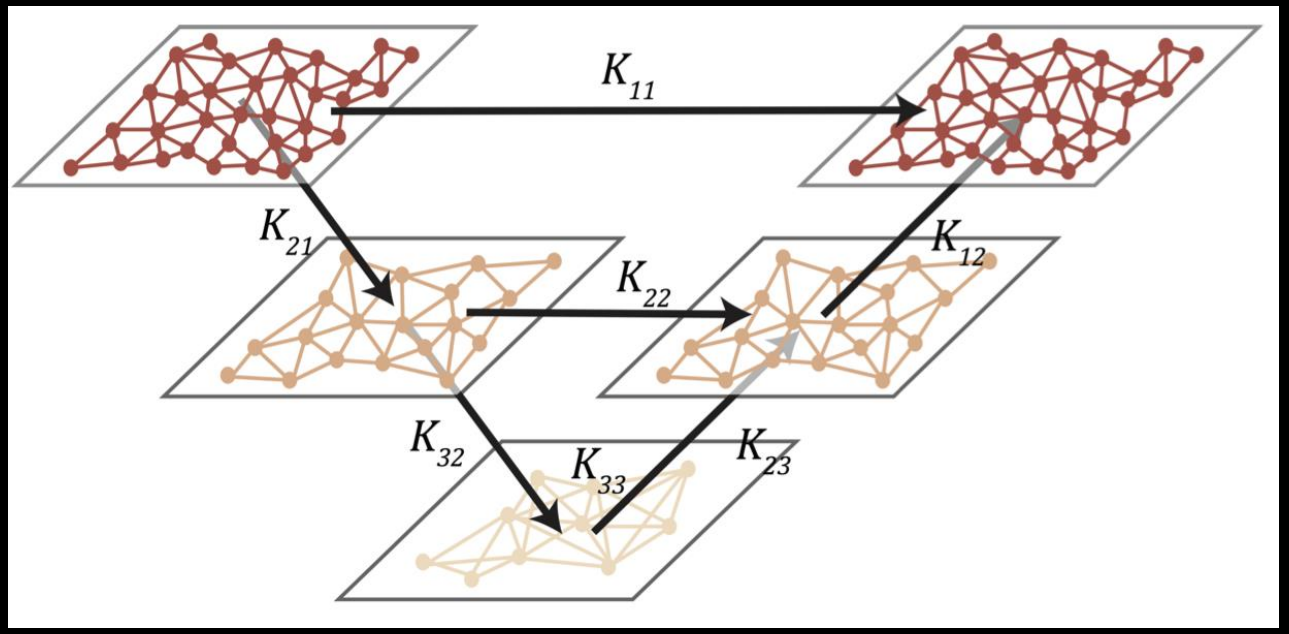




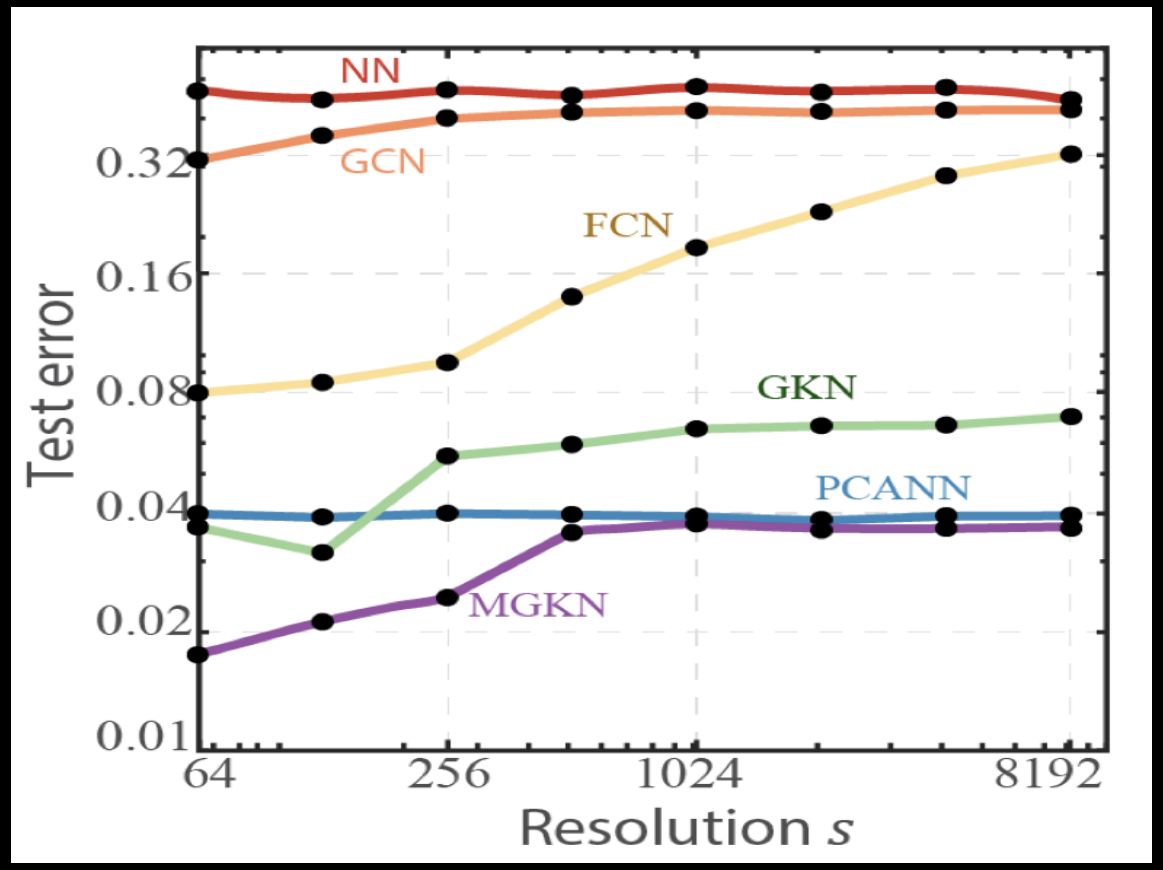
MULTIPOLE GRAPHS

- Multi-scale graphs to capture different ranges of interaction
- Linear complexity





EXPERIMENTAL RESULTS



Burgers equation

Graph neural networks for operator learning

Super-resolution and generalization within family of PDEs

TAKE-AWAYS

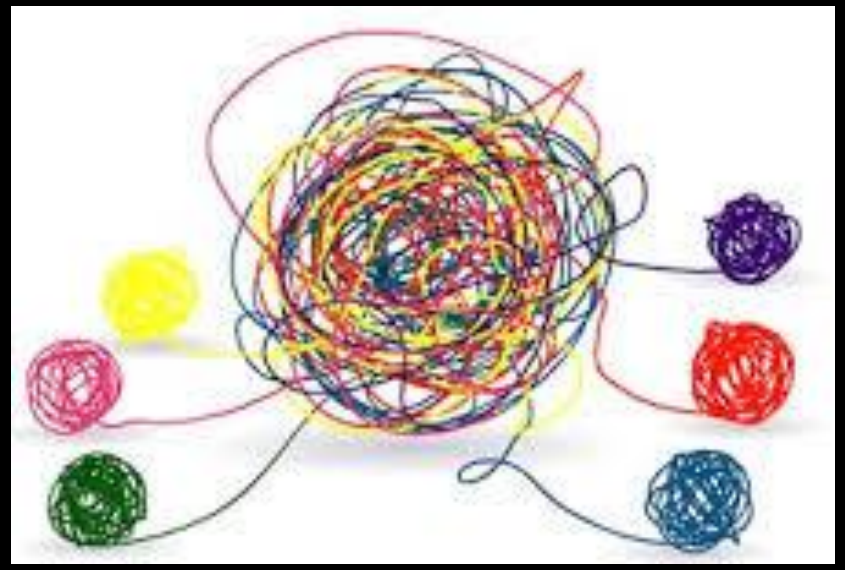
Domain knowledge augments deep learning

- Black-box deep learning is unsuitable for scientific domains
- Lack of labeled data and robustness
- Domain knowledge can tailor learning to the problem
- What is right mix of priors + deep learning?

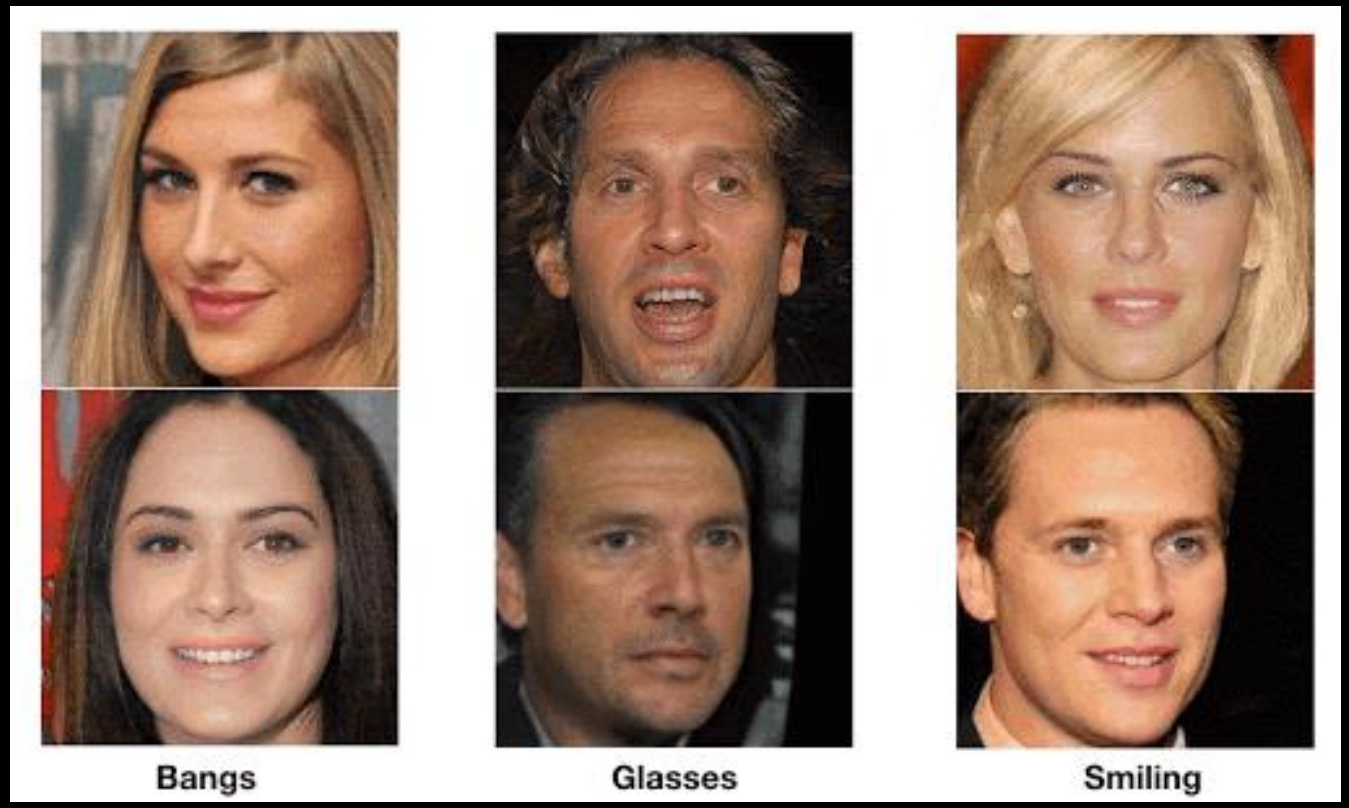
UNSUPERVISED LEARNING

DISENTANGLEMENT LEARNING

Learning latent variables that disentangle data



DISENTANGLED GENERATION



- Semi-supervised learning with very little labeled data
- Both unsupervised and supervised auxiliary losses help in disentanglement

Semi-supervised learning on 0.5% Labeled data

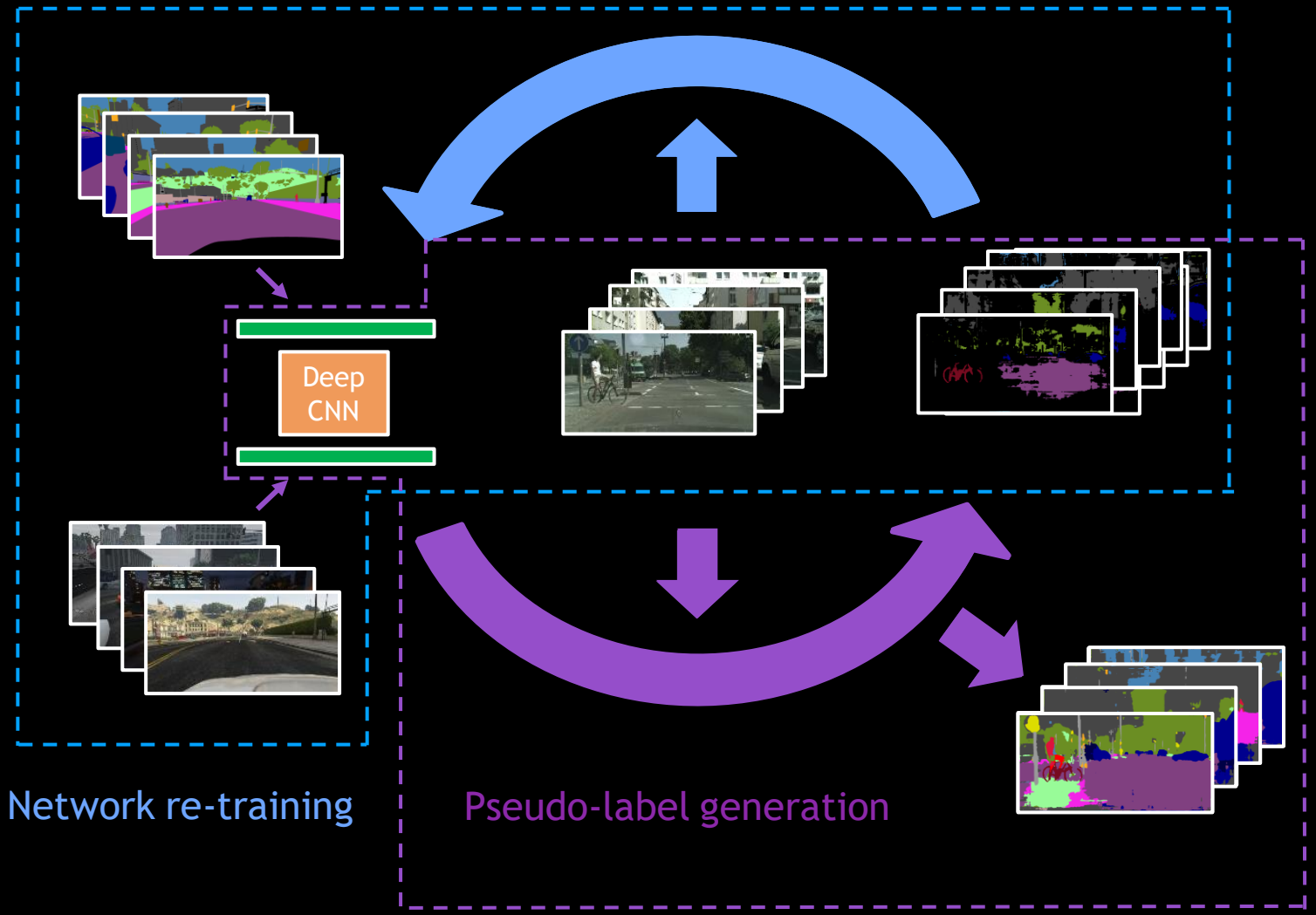
https://sites.google.com/nvidia.com/semi-stylegan

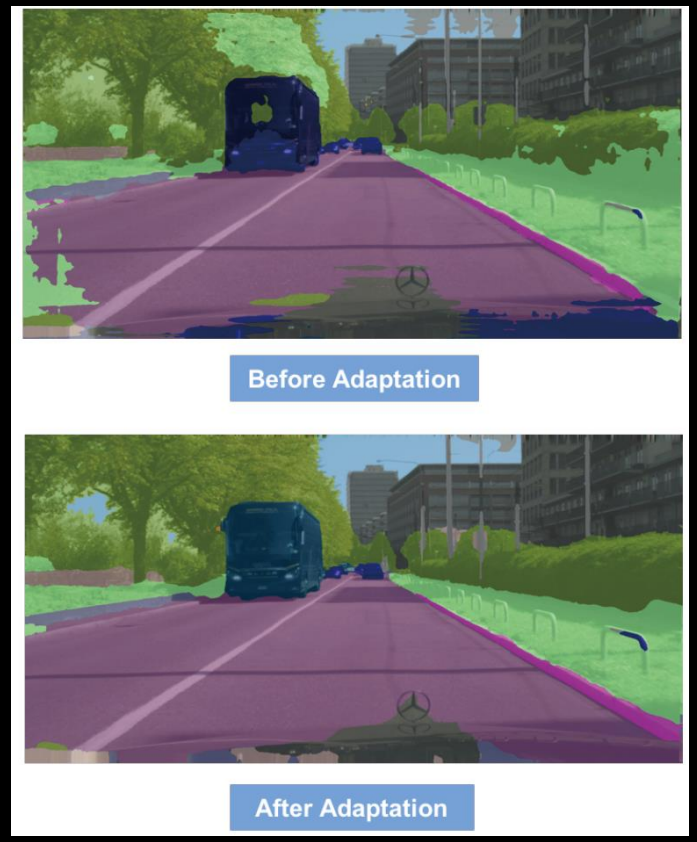
SELF-SUPERVISED LEARNING

Robust measures of confidence

- Data invariances provide supervision
- Self training with pseudo-labels
- Need confidence measure to select pseudo-labels

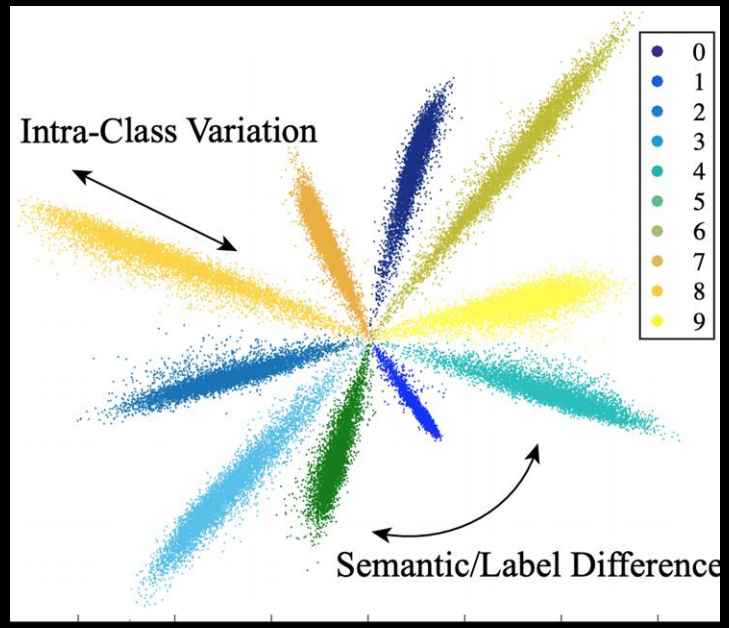
DOMAIN ADAPTATION THROUGH SELF-TRAINING

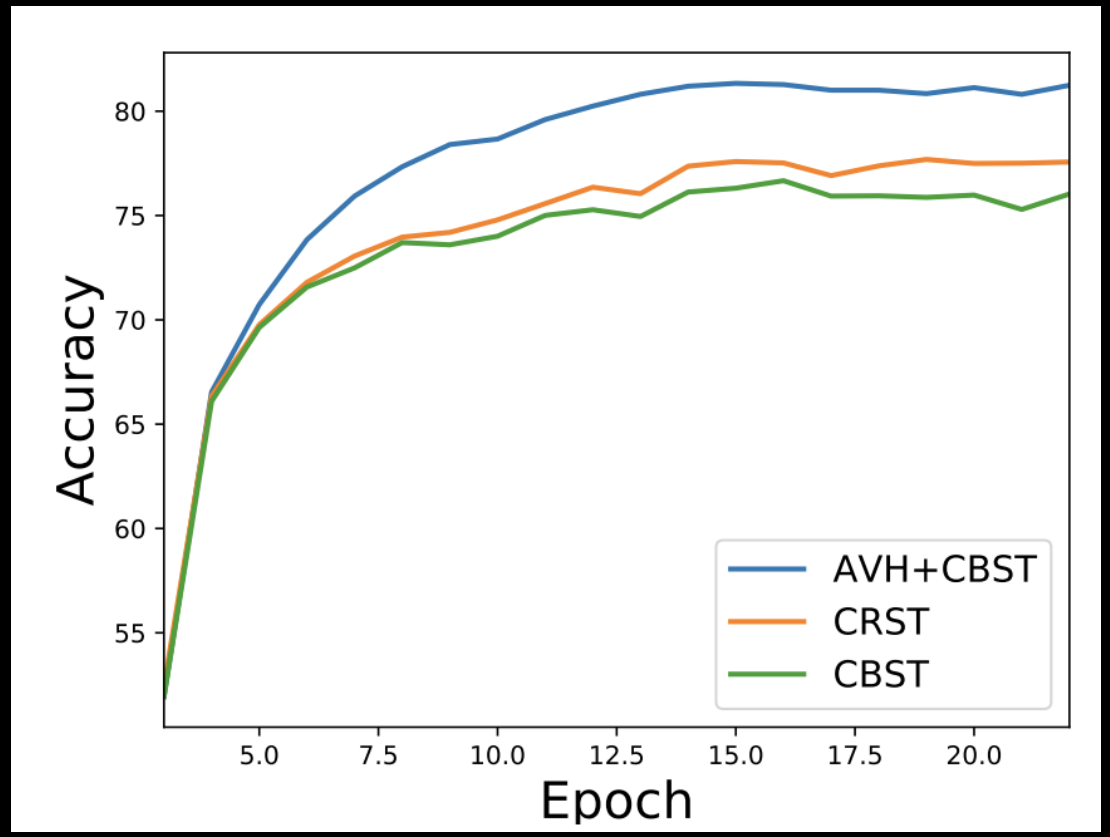




ANGULAR MEASURE IMPROVES SELF-TRAINING

Angular measure of hardness for sample selection for self training





https://sites.google.com/nvidia.com/avh

TASK ADAPTATION AND GENERALIZATION

Hongyu Ren

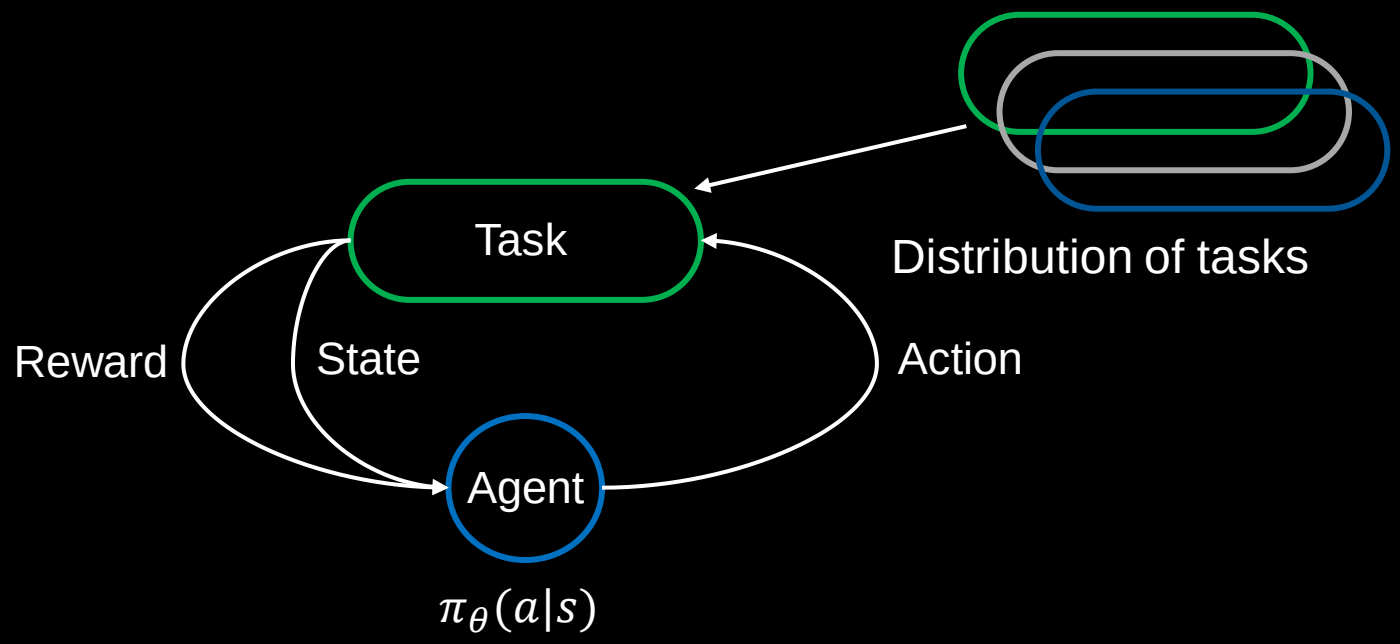
Yuke Zhu

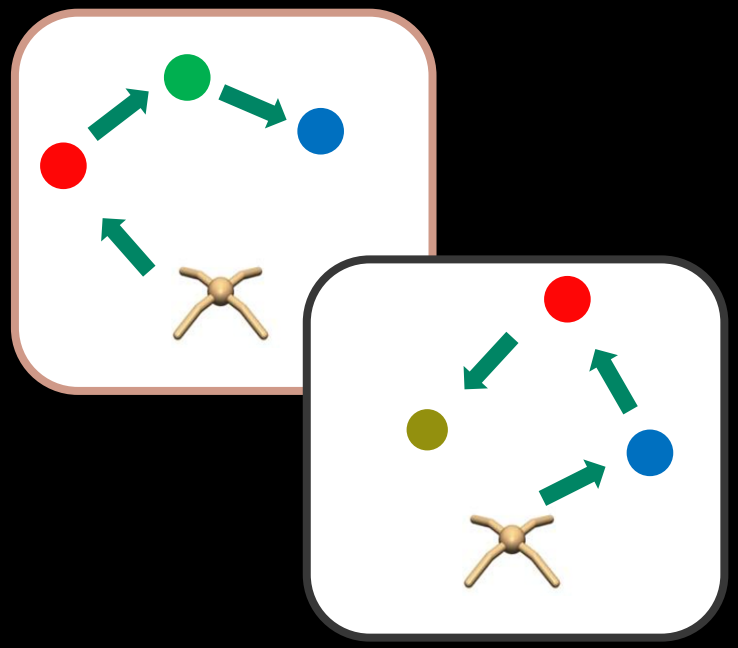
Animesh Garg

A

META-REINFORCEMENT LEARNING

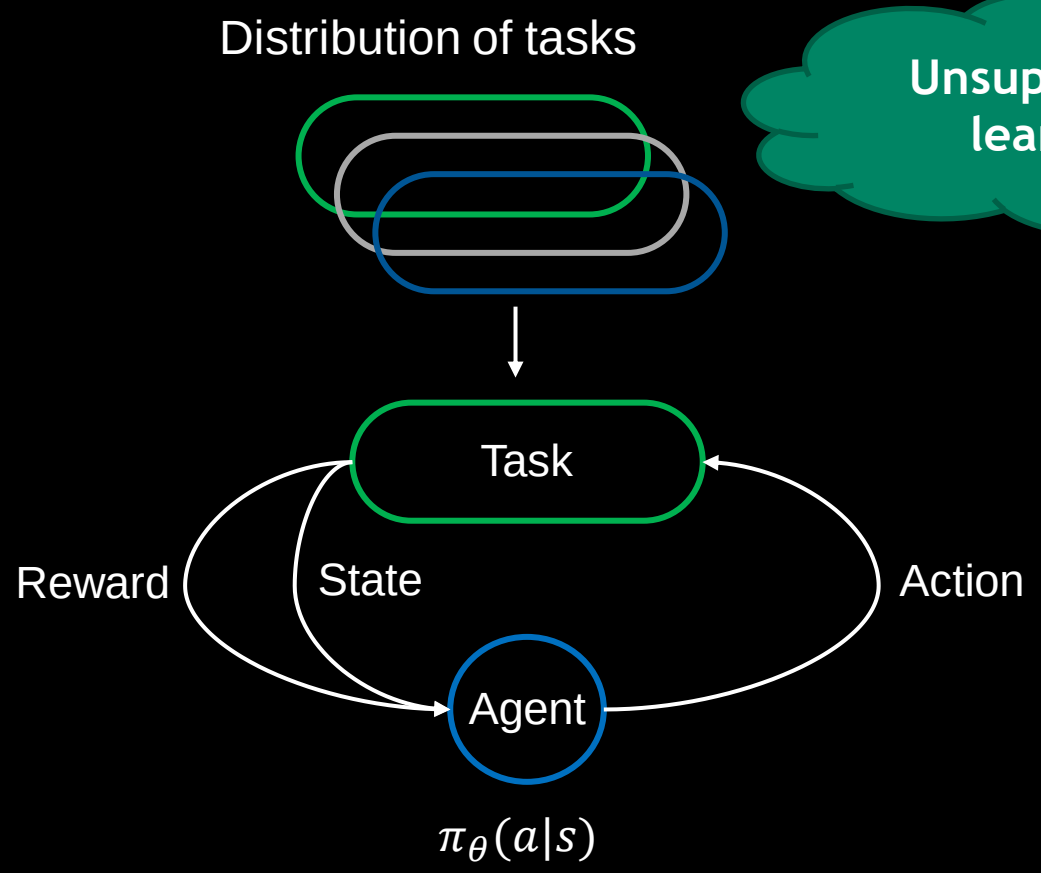
- Agents should be versatile! Given a new task, should quickly adapt
- Each task is complex and requires finishing a sequence of sub-tasks.

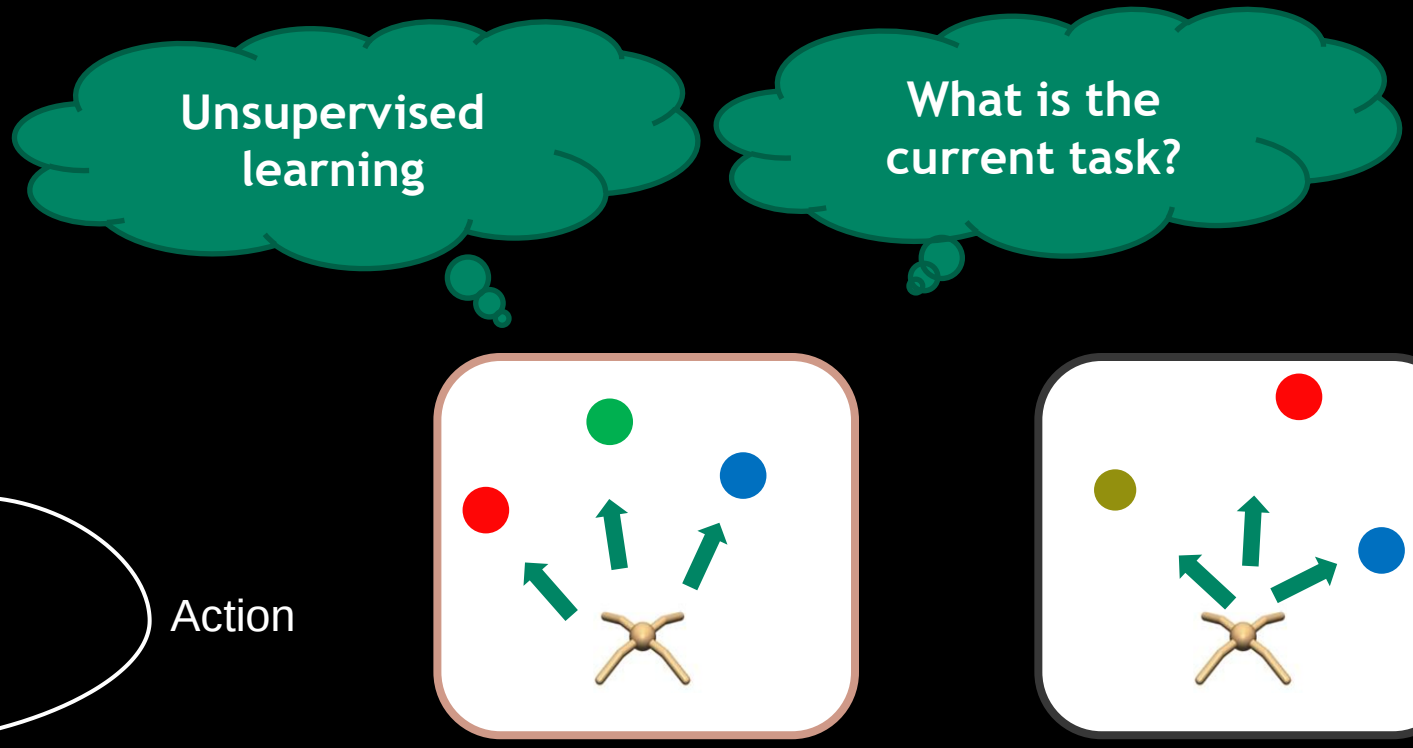




META-REINFORCEMENT LEARNING

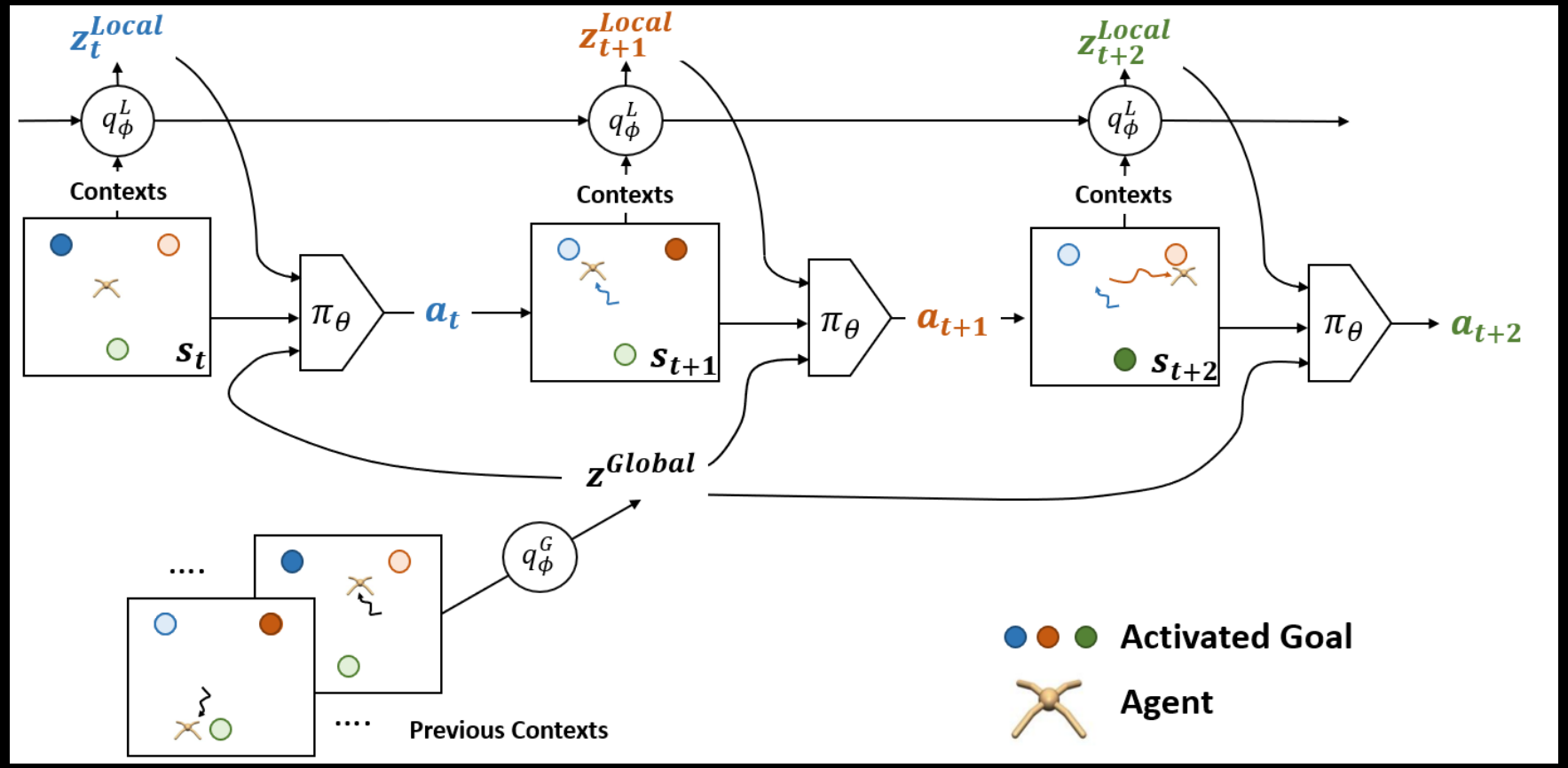
Task inference is key!





OCEAN: ONLINE CONTEXT ADAPTATION FOR TASK INFERENCE

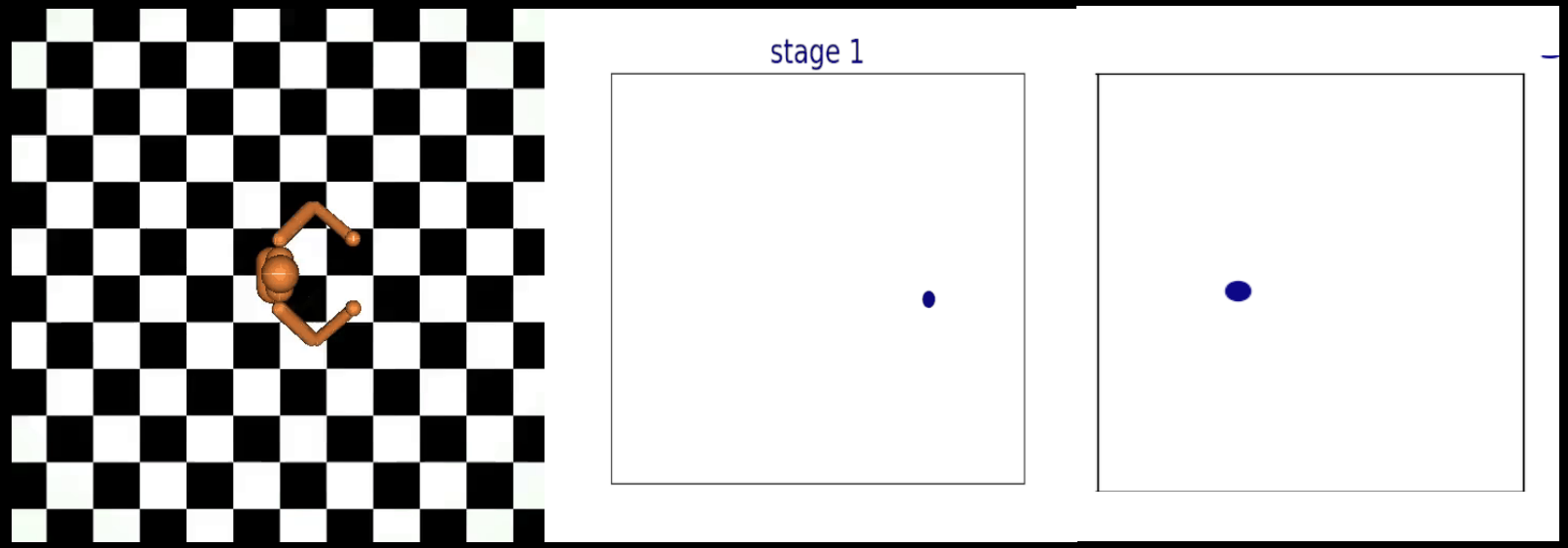
OCEAN combines both global and local task inference.



http://snap.stanford.edu/ocean/

DEMONSTRATION

Our algorithm finishes each subtask by accurately



Our algorithm Baseline

- Accurate task inference for Meta-RL.
- Real-world tasks are sequential and compositional.
- Tasks should be inferred and updated online.

CONCLUSION

- Generalizable AI needs rethinking of deep learning
- **Unsupervised learning** is the key
- Brain-inspired NN with recurrent feedback for robustness
- Neuro-symbolic AI enables compositionality
- Domain knowledge is critical for AI4science