

Machine Learning @ Tsinghua AI Institute

Jun Zhu

**Institute for Artificial Intelligence
Department of Computer Science and Technology
Tsinghua University**

Basic Theory Research Center @ AI Institute

◆ Perform fundamental research towards the 3rd generation of AI

◆ Research focus on

- Bayesian methods
- Deep learning
- Adversarial machine learning
- Reinforcement learning
- Brain-inspired AI
- Continual learning

◆ International Advisor Board



Bo Zhang
Member CAS



Yue Hao
Member CAS



Manuela Veloso
ACM/IEEE/AAAI Fellow

Research Focus

◆ Fundamentals of AI and machine learning (ML)

- Bayesian theory
- Probabilistic machine learning
- Probabilistic programming

◆ Reinforcement learning

- Decision-making in uncertain domains

◆ Robust and Interpretable AI/ML

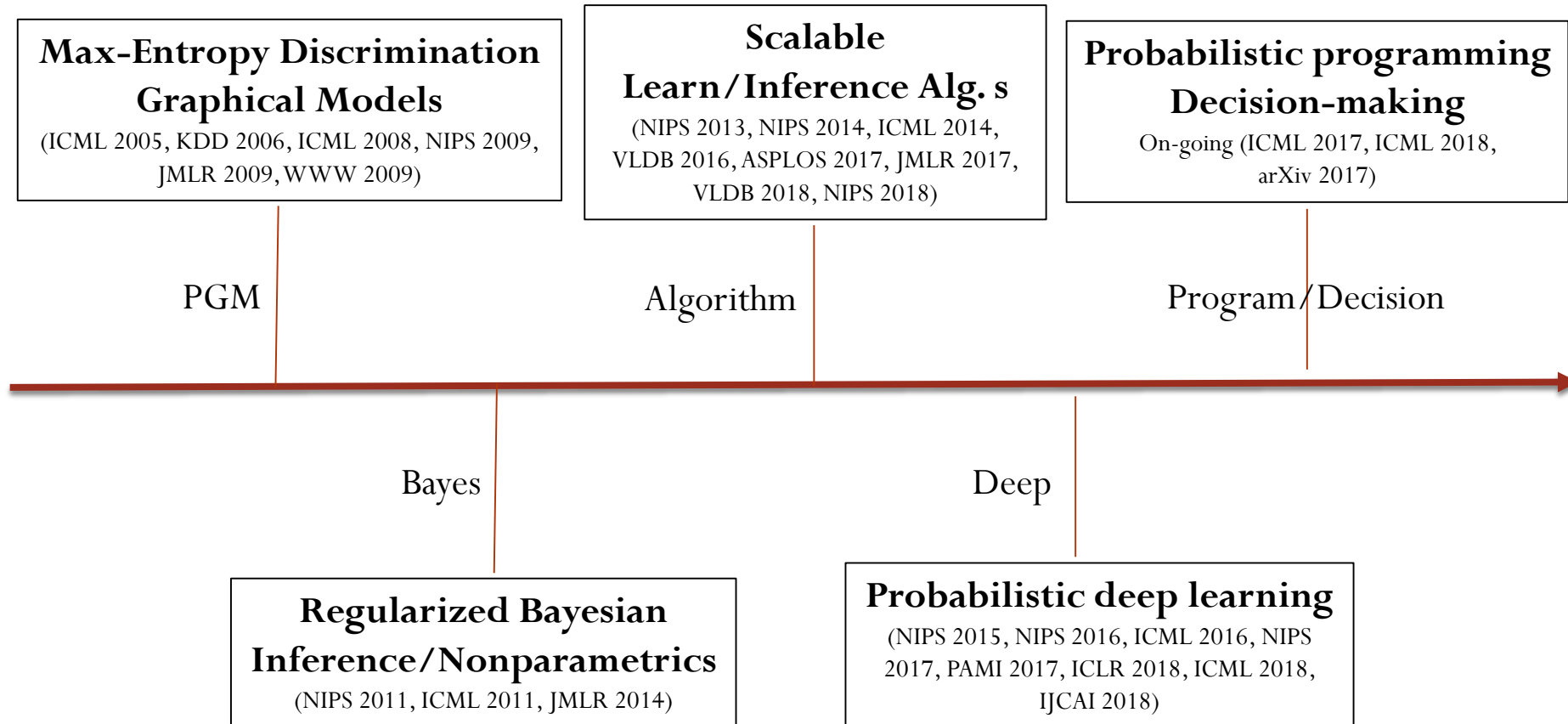
- Adversarial attack and defense
- Interpretability of deep learning models

◆ Low-energy ML

- ANNs on spiking networks
- Bayesian inference on spiking networks

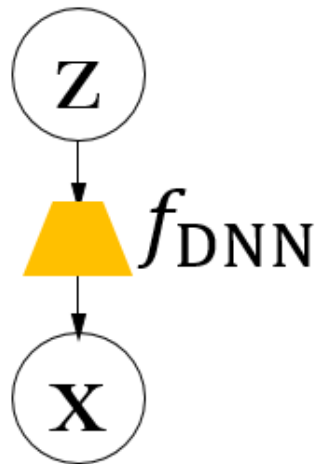
◆ Applications: Vision, text, network data analysis

Our journey with probabilistic machine learning

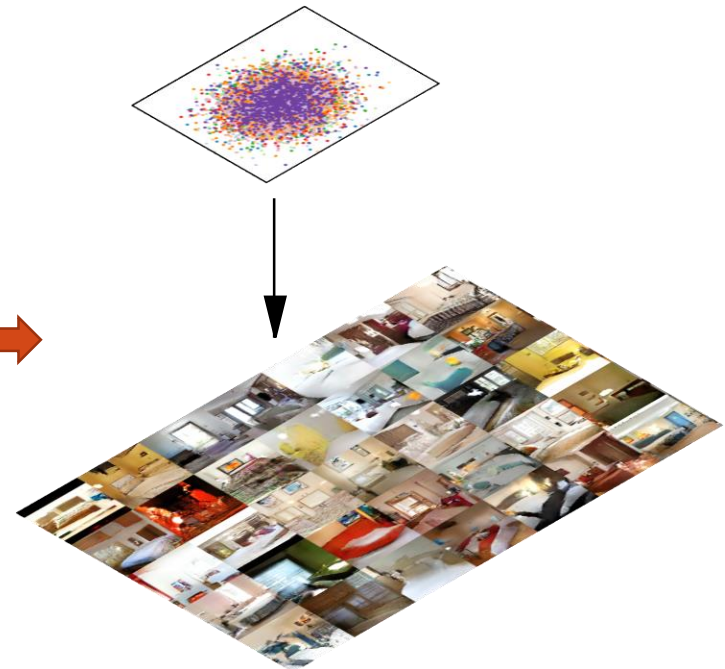


Bayesian deep learning

- ◆ Conjoin complimentary advantages of deep learning and Bayesian methods
 - Pioneering work by R. Neal and Sir D. MacKay in 1990's
 - Much recent progress on learning/inference algorithms as model structure goes deeper! E.g.: GANs, VAEs, Moment-matching networks...



Unsupervised training
on natural images



ZhuSuan: a GPU library for probabilistic deep learning

◆ Open-sourced in GitHub:

<https://github.com/thu-ml/zhusuan>



ZhuSuan: A Library for Bayesian Deep Learning

[J. Shi](#), [J. Chen](#), [J. Zhu](#), [S. Sun](#), [Y. Luo](#), [Y. Gu](#), [Y. Zhou](#)

arXiv preprint, arXiv:1709.05870 , 2017

Online Documents:

• <http://zhusuan.readthedocs.io/>

◆ Related publications on Algorithms & Applications

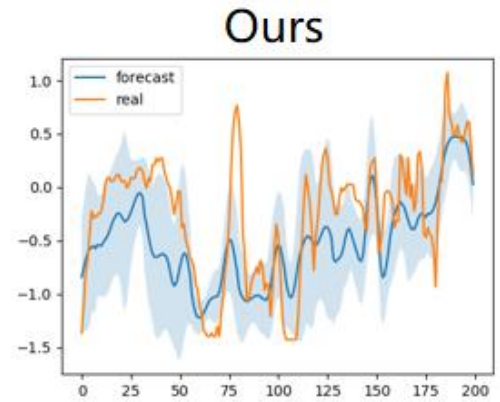
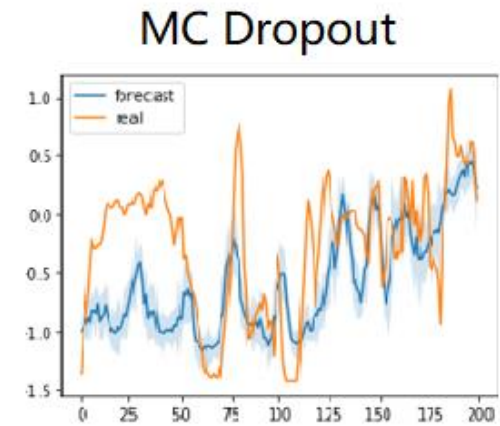
- Implicit variational inference (Shi et al., ICLR 2018; ICML 2018)
- Stein variational inference on graphs or manifolds (Zhuo et al., ICML 2018; Liu & Zhu, AAAI 2018)
- Stochastic gradient MCMC (Du et al., TNNLS 2017)
- Semi-supervised learning (Li et al., NIPS 2015, PAMI 2017; Luo et al., CVPR 2018, NIPS 2018)
- Triple GANs (Li et al., NIPS 2017)
- Textbook QA (Li et al., CVPR 2018)
- Style-transfer of handwritten characters (Sun et al., IJCAI 2018)

Some examples – Air Quality Prediction

◆ Bayesian inference on a LSTM with attention to calculate uncertainty

- Decrease the error vs. baseline model
- Provide uncertainty calculation

MSE (NO ₂)	BNN	Baseline NN
+1 hr	0.145	0.160
+7 hr	0.371	0.423
+16 hr	0.389	0.508

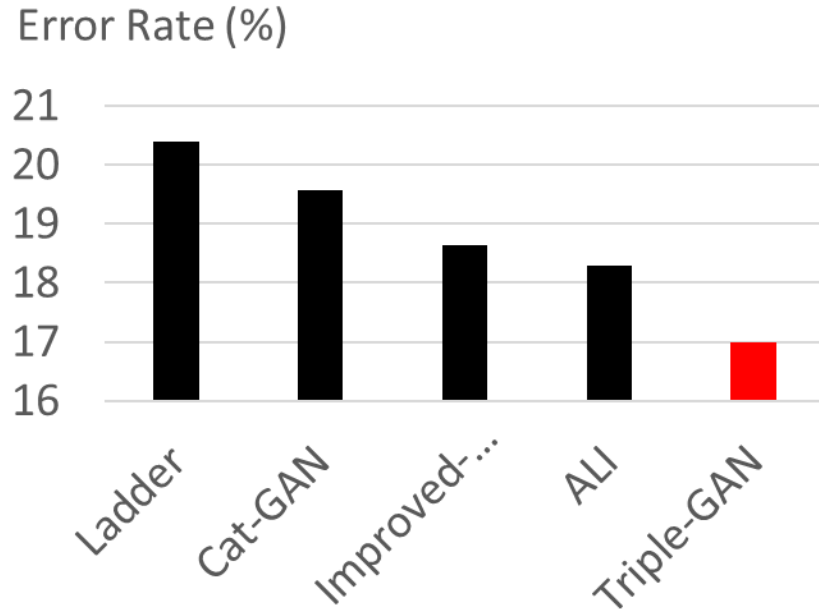


[Function space particle optimization for Bayesian neural networks. Wang et al., ICLR 2019]

Some examples – SSL

Advance previous state-of-the-art results on natural images (CIFAR10) substantially

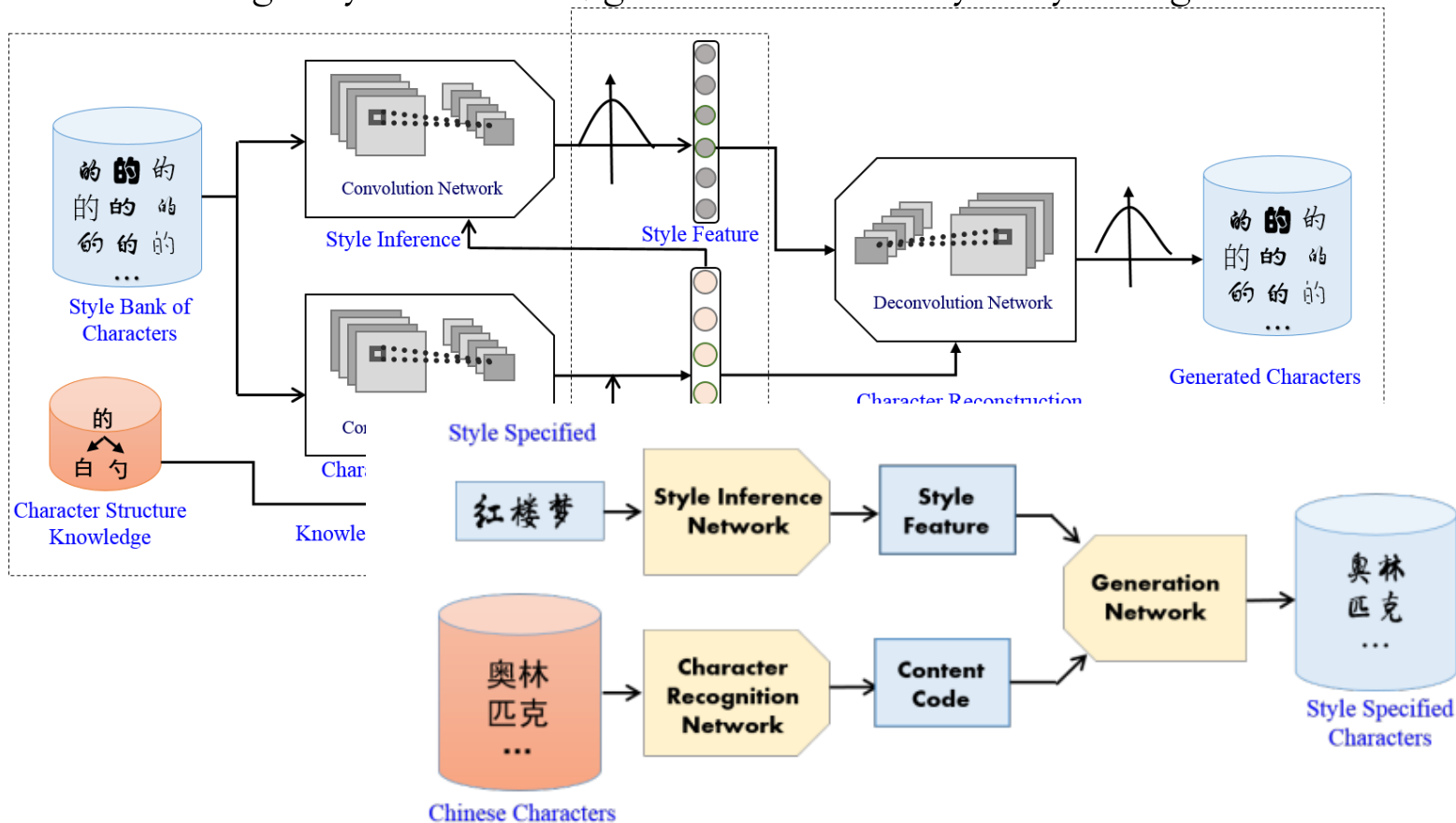
First GAN-based model to generate data in a specific class in SSL



Triple Generative Adversarial Nets (Li et al., NIPS 2017)

Some examples – Style transfer via few-shot learning

◆ Disentangle style & content, generalize to new styles by seeing one or a few examples



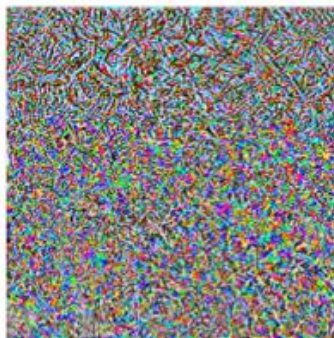
Learning to Write Stylized Chinese Characters by Reading a Handful of Examples
(Sun et al., IJCAI 2018)

Adversarial noise for deep learning

- ◆ Almost all popular networks suffer



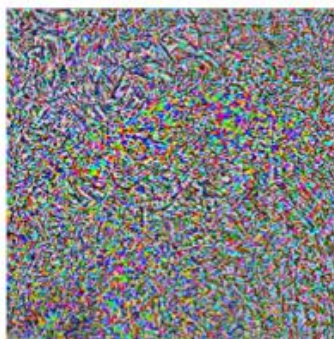
Alps: 94.39%



Dog: 99.99%



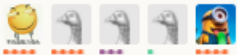
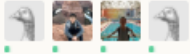

Puffer: 97.99%



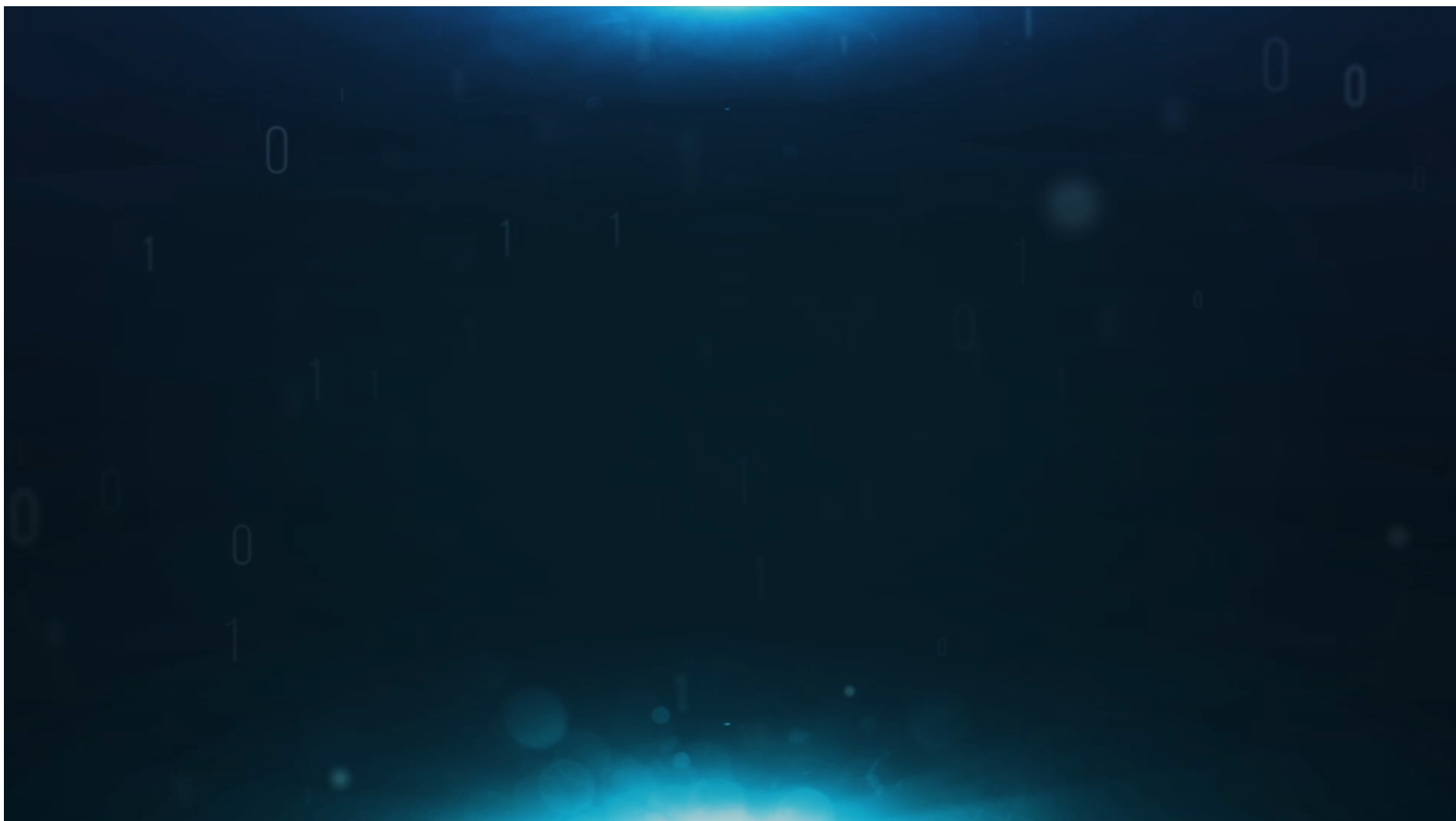
Crab: 100.00%

Adversarial attack and defense

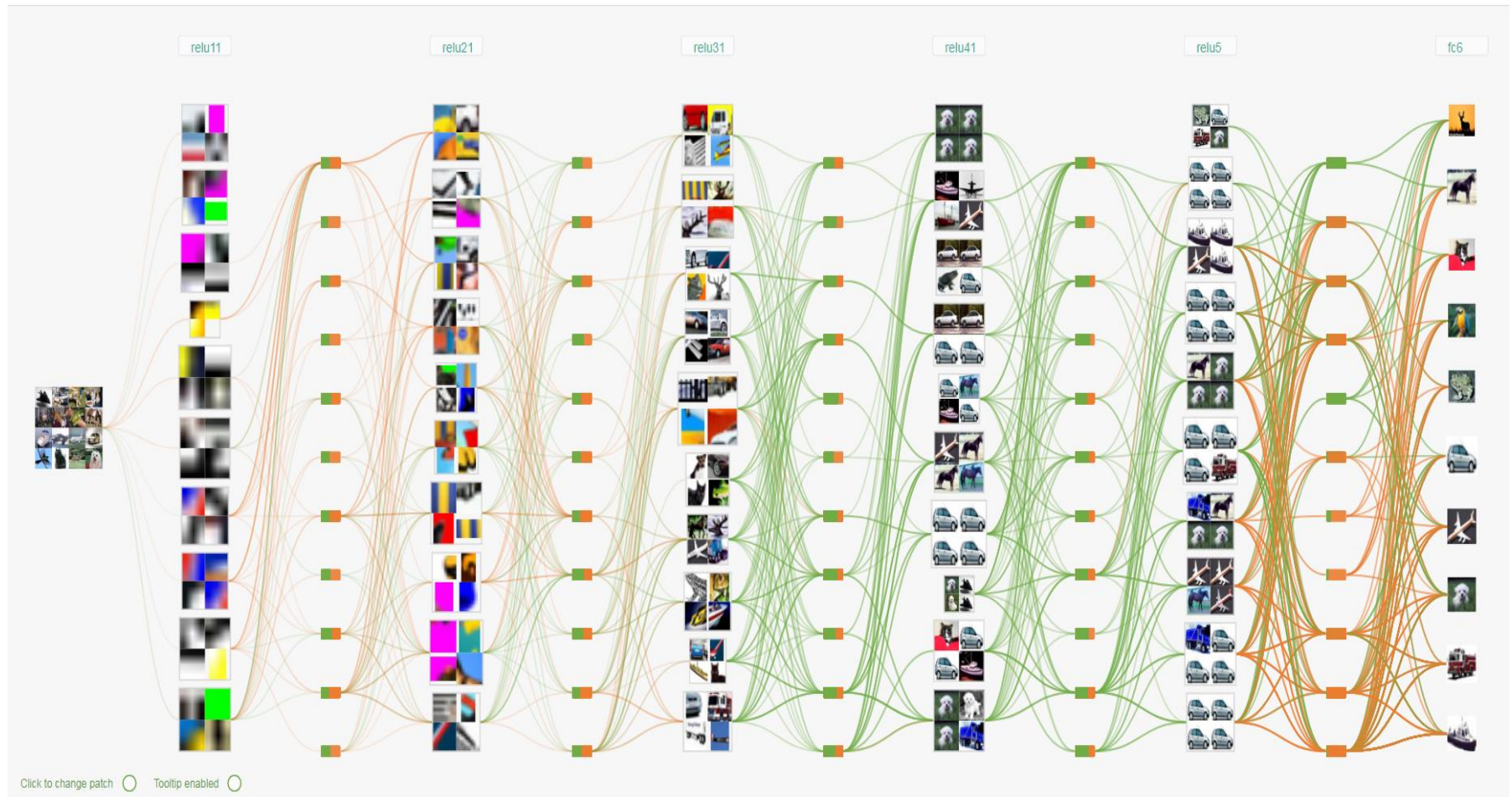
- ◆ Google Brain organized a competition on adversarial attack & defense
- ◆ Three tasks ([black-box](#))
 - Non-targeted adversarial attack (91 teams)
 - Targeted adversarial attack (65 teams)
 - Defense against adversarial attack (107 teams)
- ◆ We won all three tasks with a large margin ([2 papers at CVPR 2018](#))

#	▲pub	Team Name	Kernel	Team Members	Score ?	Entries	Last
1	—	TsAIL			0.95316	1	6mo
2	—	iyswim			0.92352	1	6mo
3	—	Anil Thomas			0.91484	1	6mo

- ◆ Related publications: CVPR (x4), ICML (x2), NIPS, AAAI



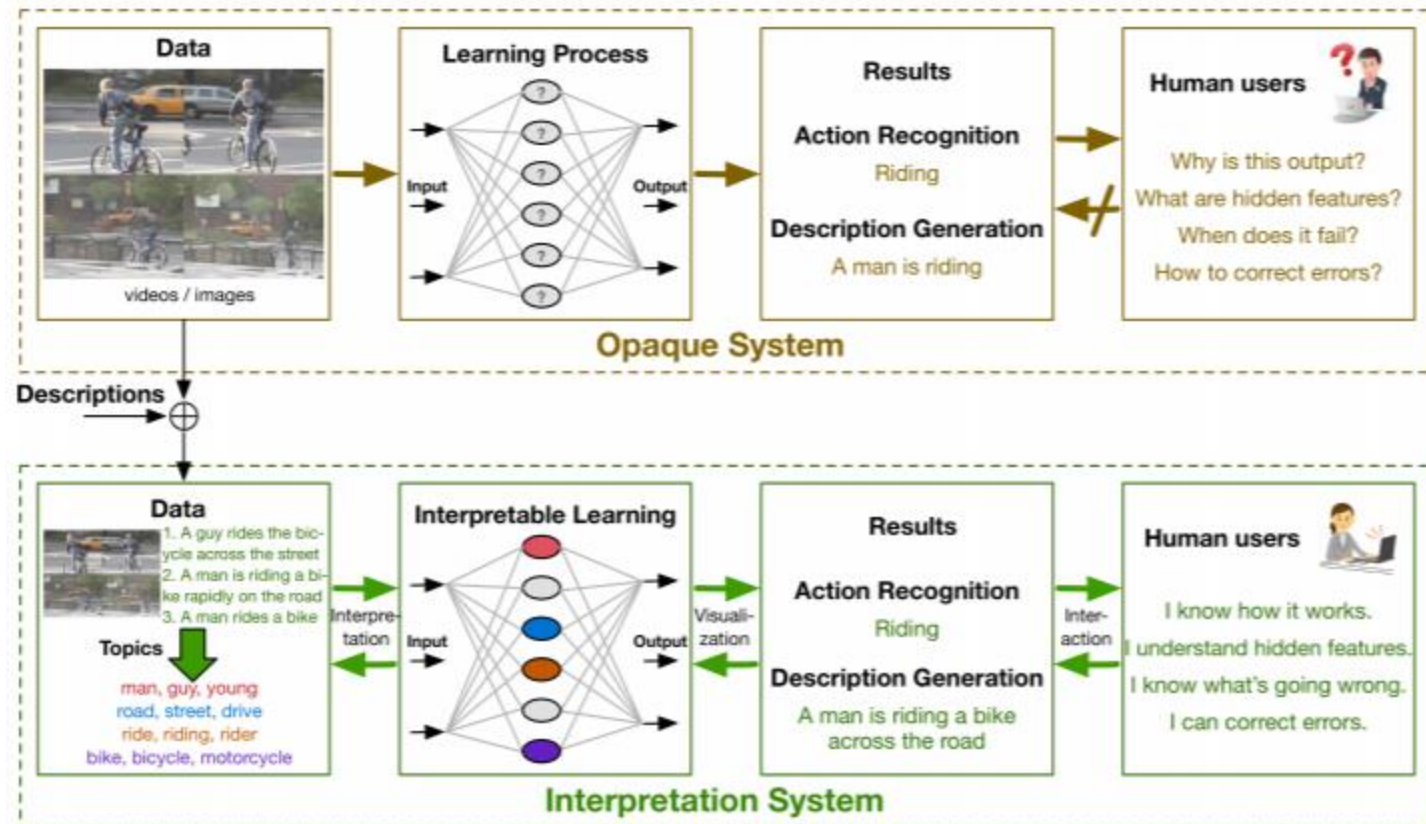
Towards interpretable ML



- Liu, M., Shi, J., Li, Z., Li, C., Zhu, J. and Liu, S., 2017. Towards better analysis of deep convolutional neural networks. *IEEE TVCG*, 23(1), pp.91-100. (Top-3 Popular Articles; Highly cited)

Improve interpretability with knowledge

- ◆ Bridge the gap between data-driven learning and interpretable knowledge



- Y. Dong, H. Su, J. Zhu, B. Zhang. "Improving Interpretability of Deep Neural Networks with Semantic Information," in *IEEE CVPR*, 2017

Human-in-loop Learning



Human-in-the-loop Learning

A woman is singing → A woman is dancing

A man is doing → A man is riding a motorcycle

This dog → The dog is swimming

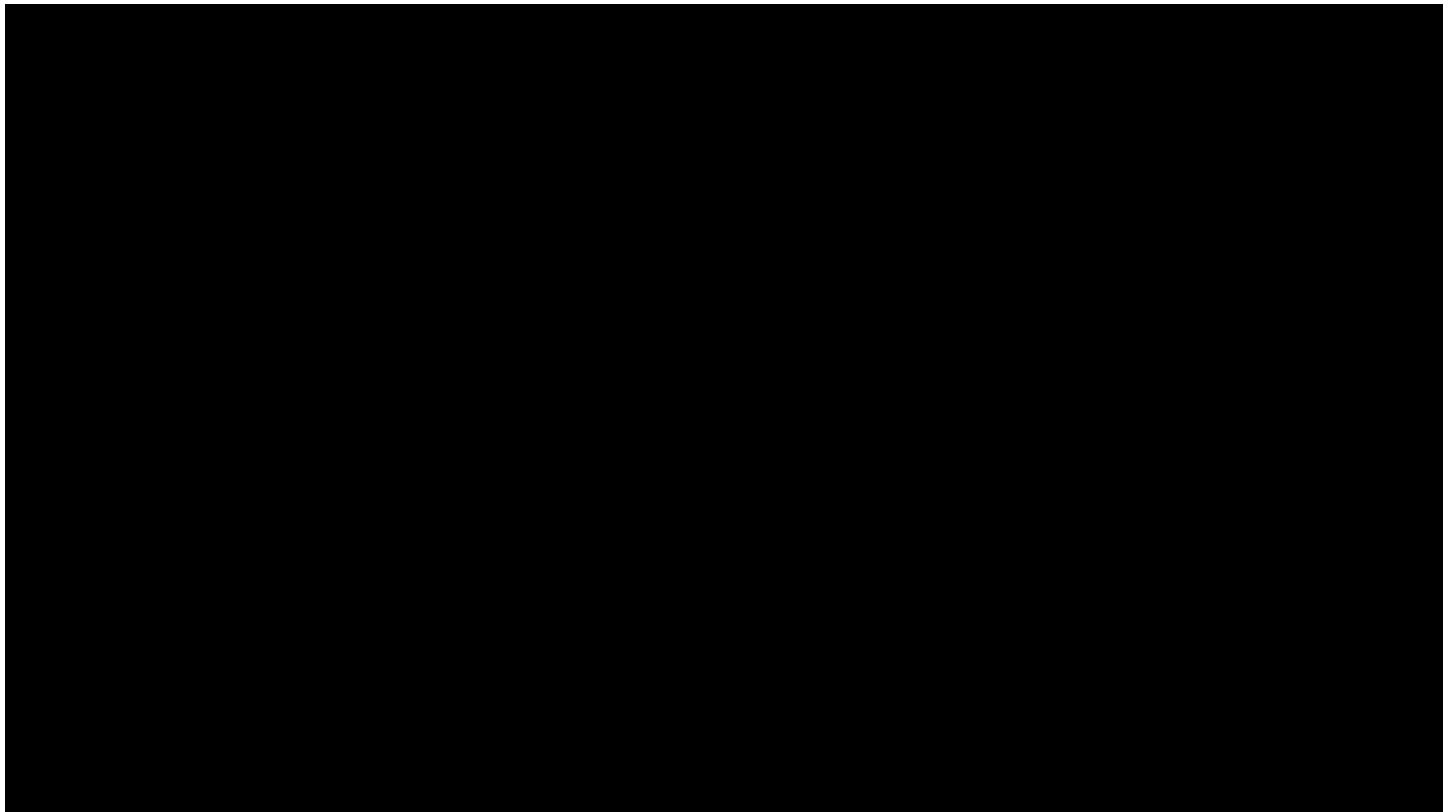
Someone is playing → Polar bears are playing

- ◆ providing the missing topics (“dance”, “motorcycle”, “swim” and “polar bear”) for the first half of these four videos and refining the model, the predictions for the second half are more accurate.

Decision-Making

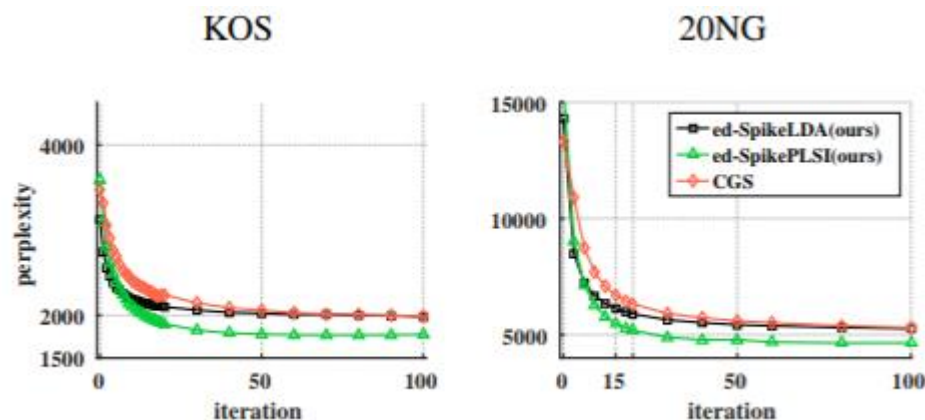
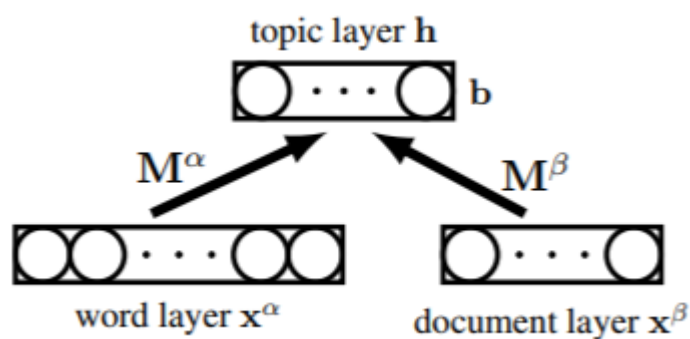
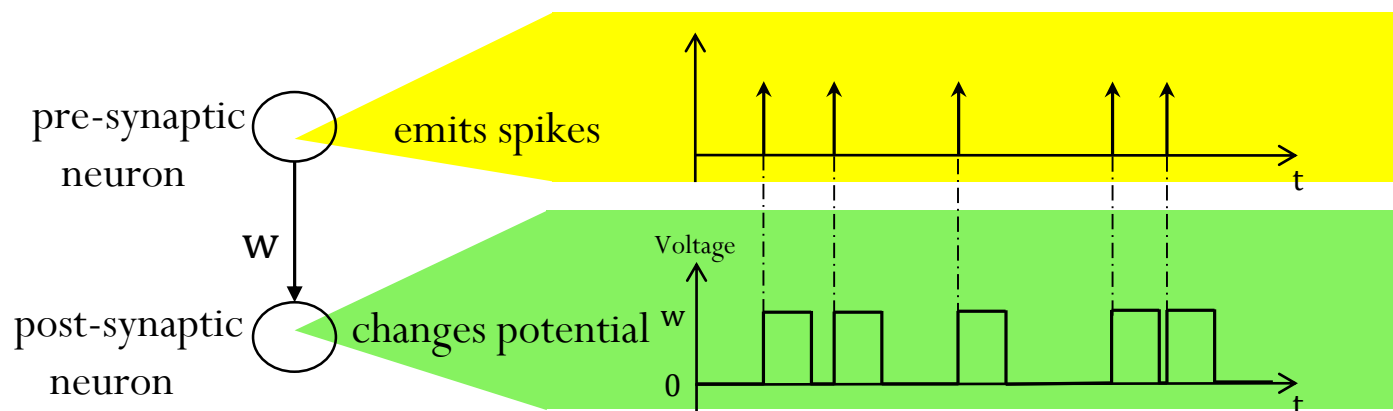
◆ Decision-making in uncertain environments

- Bayesian bandits (ICML 2017, ICML 2018); Imitation learning (AAAI 2018)
- ViZDoom competition (2nd place in 2017; 1st place + 2nd place in 2018)



Towards Low-Energy ML

◆ Probabilistic inference on spiking networks



Xiao, et al. Towards Training Probabilistic Topic Models on Neuromorphic Multi-chip Systems, *IJCAI*, 2018

Yu, et al., Direct Training for Spiking Neural Networks: Faster, Larger, Better, *AAAI*; *Frontiers in Neuroscience*, 2018

Summary

◆ Our research focus:

- Probabilistic ML: theory, algorithms, probabilistic programming library
- Adversarially robust and interpretable machine learning
- Decision making in uncertain environments
- Low-energy ML methods
- Applications in image, text, network analysis

Thank you!

J. Zhu. Probabilistic Machine Learning: Models, Algorithms and a Programming Library. *Proc. of the 27th International Joint Conference on Artificial Intelligence*, Early Career. Pages 5754-5759.