

# Deep Learning for the Internet of Things

Shuochao Yao, Yiran Zhao, Aston Zhang,  
Shaohan Hu, Huajie Shao, Chao Zhang,  
Lu Su, Tarek Abdelzaher



UNIVERSITY OF MINNESOTA

Driven to Discover<sup>SM</sup>

# Introduction

- Internetnetworked mobile & embedded devices -> Internet of Things (Sensor-rich world) -> Revolutionize the interactions
- Build smarter and more user-friendly applications
- Deep learning has greatly changed the way that computing devices process human-centric content such as images, video, speech, and audio

# Four Key Research Questions

1. What deep neural network structures can **effectively process and fuse sensory input data** for diverse IoT applications?
2. How to **reduce the resource consumption** of deep learning models such that they can be efficiently deployed on resource-constrained IoT devices?
3. How to **compute confidence measurements** in the correctness of deep learning predictions for IoT applications?
4. How to **minimize the need for labeled data** in learning?

# Deep Learning Models For Sensor Data

- IoT applications often depend on collaboration among multiple sensors;
- The tasks on IoT devices can be generally categorized as: **estimation tasks** & **classification tasks**.

# Deep Learning Models For Sensor Data

- For **estimation-oriented problems** (tracking/localization), sensors generate measurements of physical quantities.
- **Challenges:** Noisy (nonlinear & correlated over time); hard to separate signal from noise; lead to estimation errors and bias.

# Deep Learning Models For Sensor Data

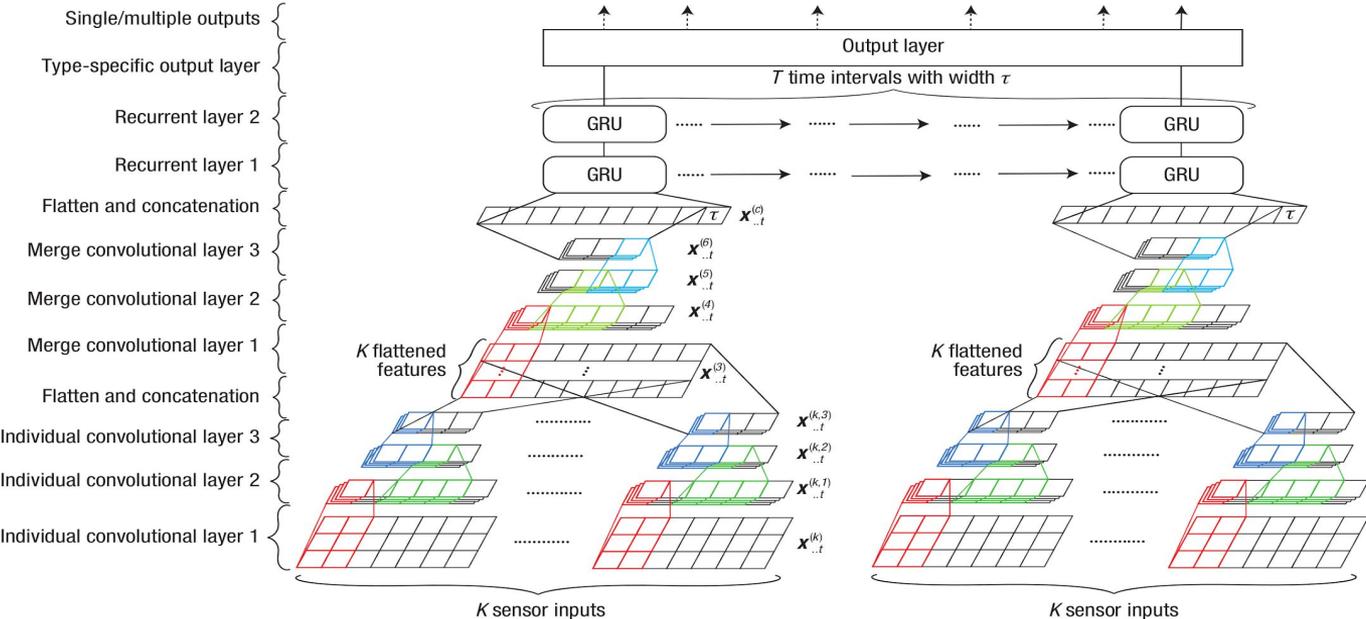
- For **classification-oriented problems** (activity/context recognition), a typical approach is to hand-crafted features derived from raw sensor data.
- **Challenges:** time-consuming; requires extensive experiments to generalize well.

# Deep Learning Models For Sensor Data

- Design novel neural network structures for multisensor data fusion:
  1. Model complex interactions among multiple sensory inputs;
  2. Encode features of sensory inputs effectively.

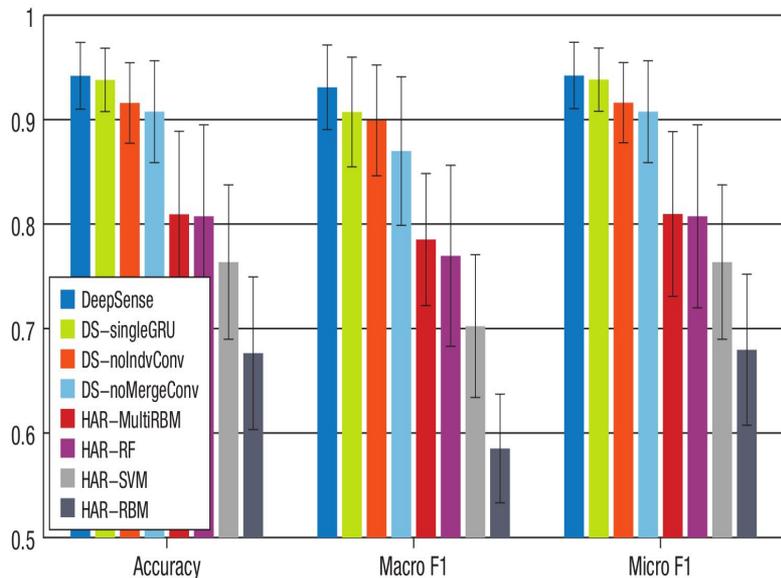
DeepSense!

# Deep Learning Models For Sensor Data

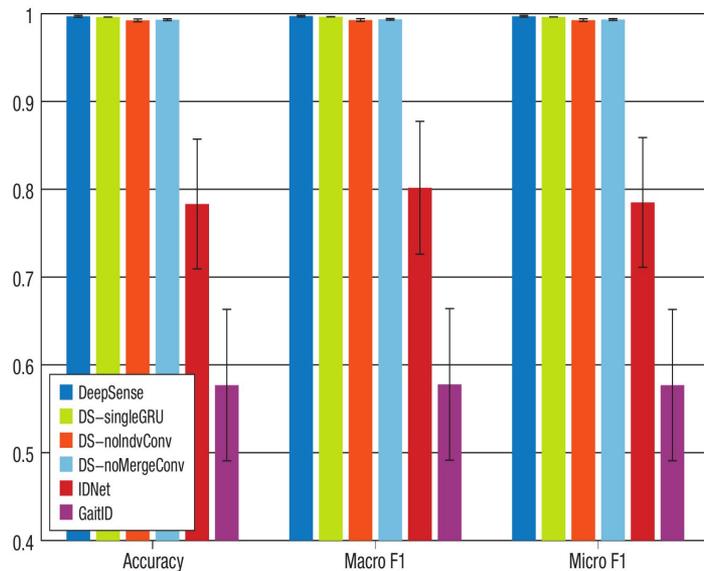


**FIGURE 1.** Main architecture of the DeepSense framework.

# Deep Learning Models For Sensor Data



**FIGURE 2.** Performance metrics of heterogeneous human activity recognition (HHAR) task with the DeepSense framework.

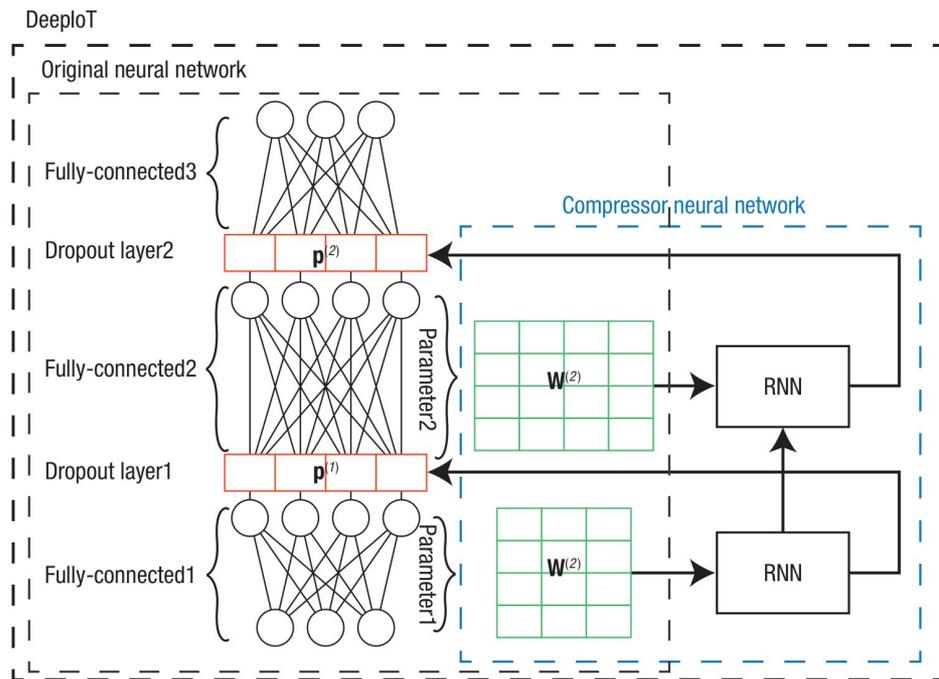


**FIGURE 3.** Performance metrics of UserID task with the DeepSense framework.

# Compressing Neural Networks Structures

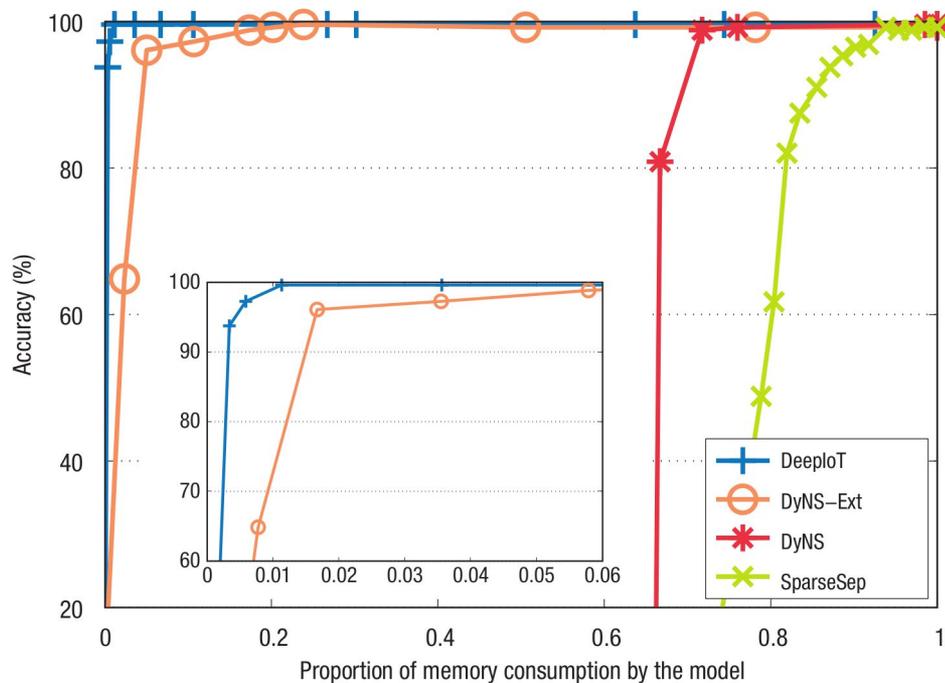
- Resource constraints.
  - A key question is whether it is possible to compress deep neural networks.
1. Can a unified approach compress commonly used deep learning structures, including fully connected, convolutional, and recurrent neural networks, as well as their combinations?
  2. To what degree does the resulting compression reduce energy, execution time, and memory needs in practice?

# Compressing Neural Networks Structures



**FIGURE 4.** Overall DeeploT system framework. Orange boxes represent dropout operations. Green boxes represent parameters of the original neural network.

# Compressing Neural Networks Structures



**FIGURE 5.** The tradeoff between testing accuracy and memory consumption by models.

# Compressing Neural Networks Structures

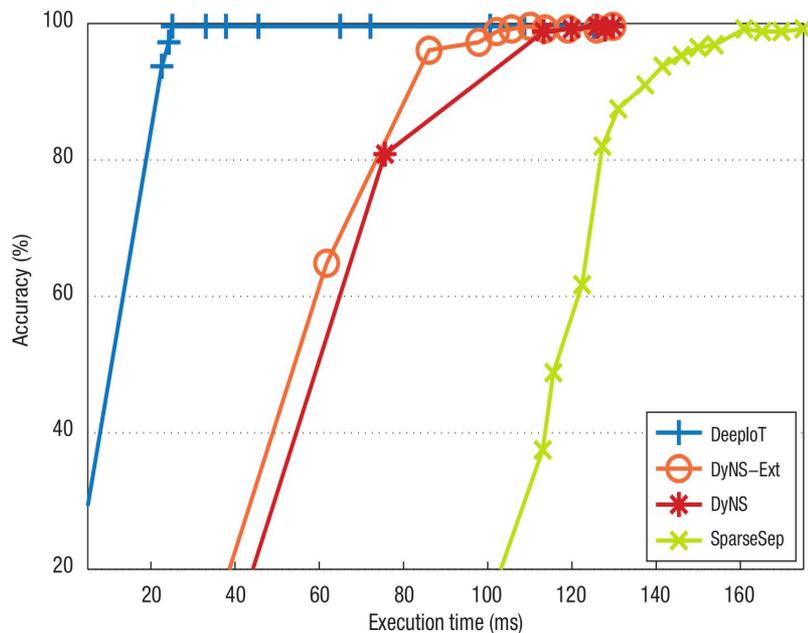


FIGURE 6. The tradeoff between testing accuracy and execution time.

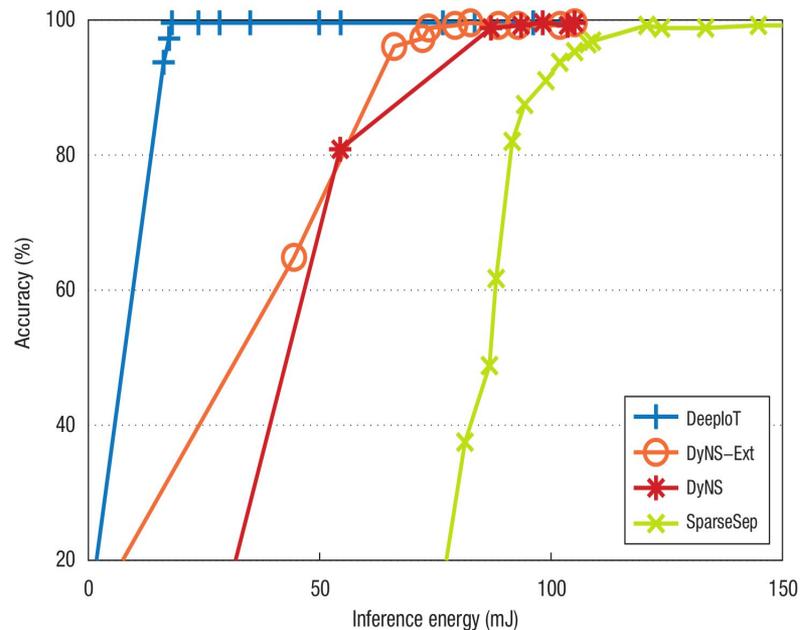


FIGURE 7. The tradeoff between testing accuracy and energy consumption.

# Estimating Uncertainty

- How to offer principled uncertainty estimates that can faithfully reflect the correctness of model predictions?
  1. How to develop methods that provide accurate uncertainty estimates in prediction results obtained from deep learning models?
  2. How to develop resource-efficient solutions for the uncertainty estimation problem, such that they can be implemented on resource-limited IoT devices?

# Estimating Uncertainty

## RDeepSense:

1. A simple, well-calibrated, and efficient uncertainty estimation algorithm for a multilayer perceptron (MLP);
2. Apply a tunable function, based on a weighted sum of negative log-likelihood and mean square error, as the loss function.

# Estimating Uncertainty

**TABLE 1.** Mean absolute error (MAE) and negative log-likelihood (NLL) for the **NYCommute** task.

Deep learning algorithm	MAE	NLL
RDeepSense	5.64	7.7
SSP-1	8.15	4.86
SSP-3	7.90	4.67
SSP-5	7.51	4.84
SSP-10	7.03	4.81
MCDrop-3	5.69	19,995.6
MCDrop-5	5.64	1,335.73
MCDrop-10	5.61	640.35
MCDrop-20	5.61	640.35
Gaussian Process	11.84	7.46

# Estimating Uncertainty

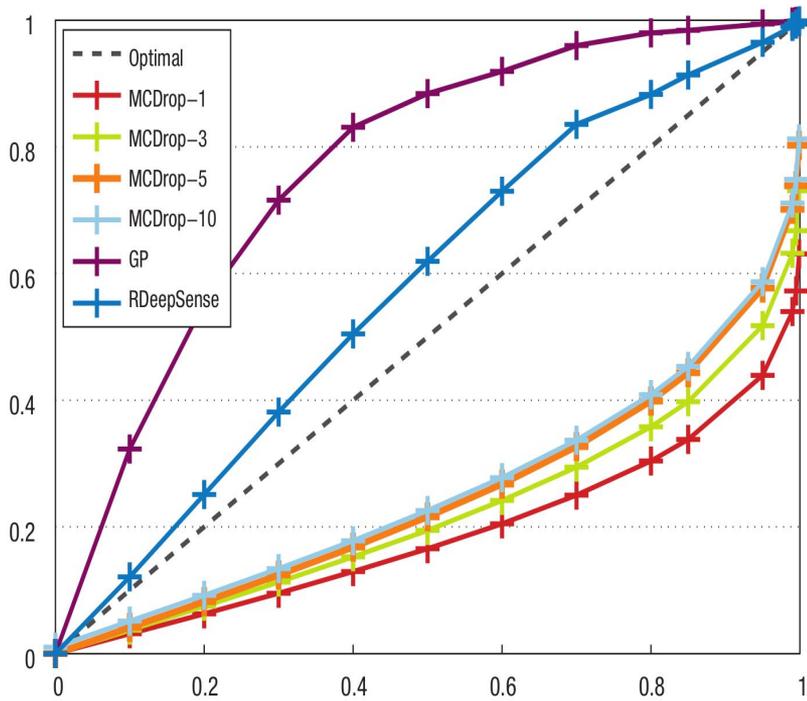


FIGURE 8. The calibration curves of RDeepSense, GP, and MCDrop-k.

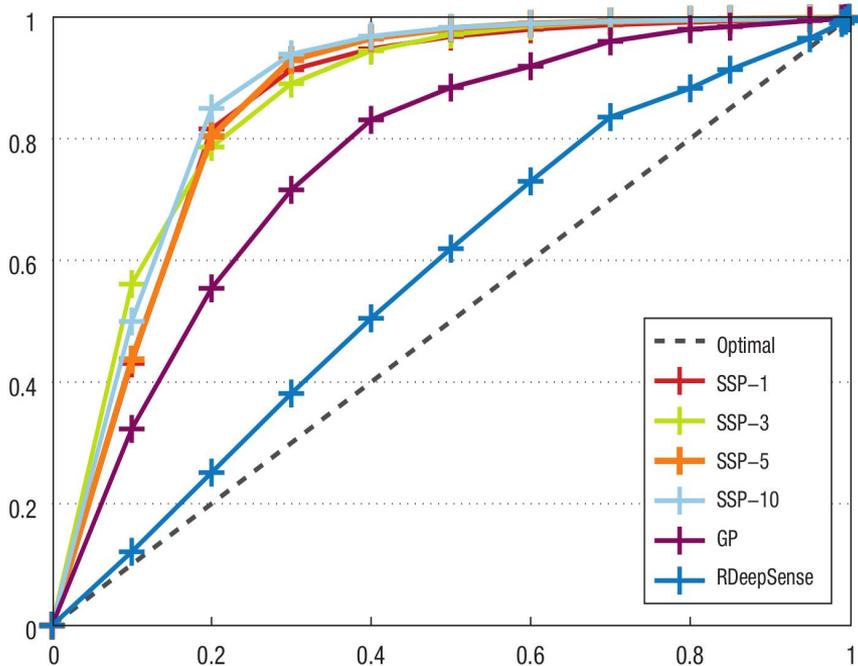
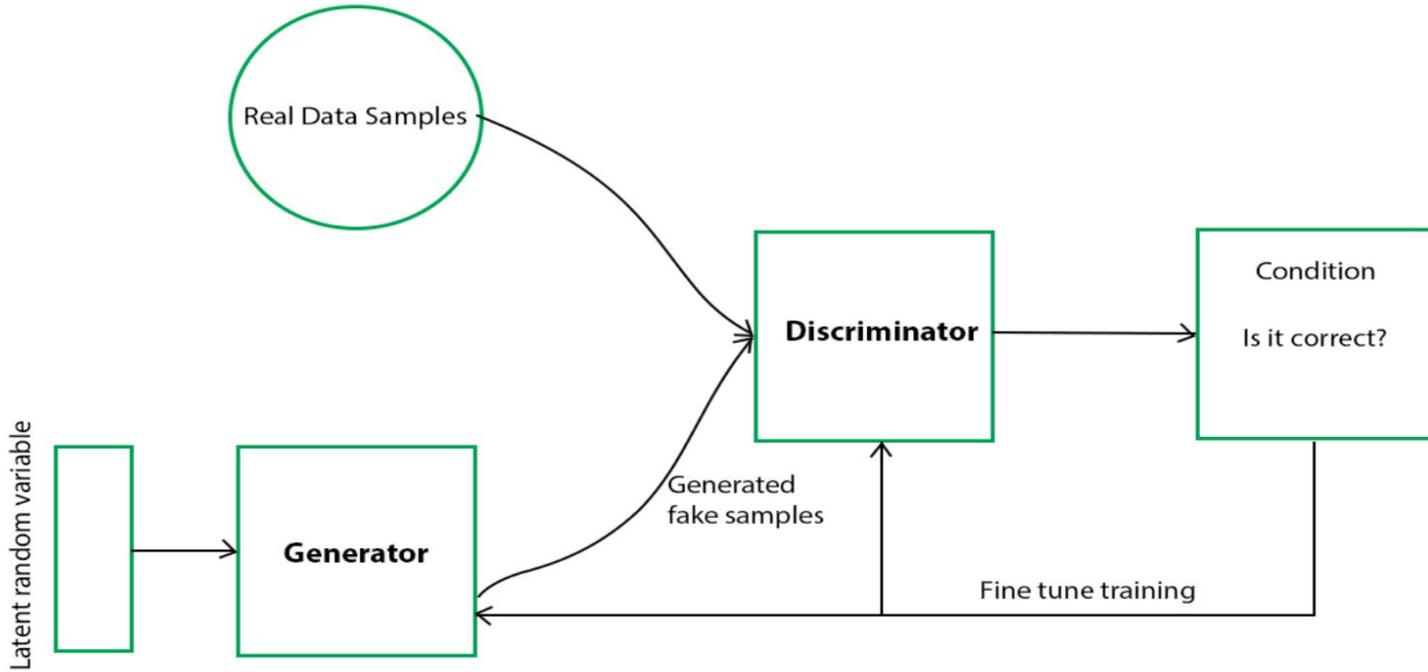


FIGURE 9. The calibration curves of RDeepSense, GP, and SSP-k.

# Minimizing Labeled Data

- The need for labeling offers a significant practical impediment to the use of deep learning in IoT contexts, where labeling cannot be easily done.
- **Generative adversarial networks (GAN)** has been proposed as a promising deep learning technique for unsupervised and semi-supervised learning.

# Minimizing Labeled Data



# Minimizing Labeled Data

**TABLE 2.** Semisupervised training of HHAR with DeepSense framework.

<i>p</i> %	10%	5%	3%	2%	1%
Sense-GAN	94.8%	92.5%	91.4%	90.4%	88.3%
DeepSense	92.0%	89.3%	85.3%	83.6%	79.1%

# Future Work

1. Can one build a unified deep learning framework for largely heterogeneous sensory inputs, such as audio signals, Wi-Fi signals, and motion inputs?
2. What are the impact of neural network compression on system performance, such as execution time and energy consumption?
3. Can one extend uncertainty measurements to other deep learning models besides MLPs?
4. How does one learn in highly dynamic environments where it is impossible to collect a large number of data samples?