Research Area:

The success of deep learning comes from three aspects: efficient algorithm, powerful hardware, and large-scale dataset. Our lab targets the first two aspects. MIT HAN Lab is looking for motivated students in the area of deep learning and computer architecture to solve impactful AI problems with light-weight models and high computational efficiency. We have internship openings.

Introduction:

Song Han is an assistant professor at MIT's EECS. He received his PhD degree from Stanford University. His research focuses on efficient deep learning computing. He proposed "deep compression" technique that can reduce neural network size by an order of magnitude without losing accuracy, and the hardware implementation "efficient inference engine" that first exploited pruning and weight sparsity in deep learning accelerators. His team's work on hardware-aware neural architecture search (ProxylessNAS, Once-for-All Network (OFA), MCUNet) was highlighted by MIT News, Wired, Qualcomm News, VentureBeat, IEEE Spectrum, adopted by PyTorch and AutoGluon, received six low-power computer vision contest awards in flagship AI conferences. Song received Best Paper awards at ICLR'16 and FPGA'17, Amazon Machine Learning Research Award, SONY Faculty Award, Facebook Faculty Award, NVIDIA Academic Partnership Award. Song was named "35 Innovators Under 35" by MIT Technology Review; he received the NSF CAREER Award and the IEEE "AIs 10 to Watch: The Future of AI" award.

Desired Skills:

We are looking for exceptional researchers to help develop next-generation machine learning model optimization techniques.

- Familiar with Python and PyTorch
- Experience of developing and training state-of-art deep learning models
- Experience with developing large scale distributed machine learning systems and deployment on edge devices.
- Deep knowledge in computer vision and/or time-series analysis
- (Bonus) Experience with machine learning model deployment with ONNX, TFLite, TensorRT, Qualcomm SNPE, etc.
- (Bonus) Experience with embedded Linux, Android, iOS development
- (Bonus) Experience with AWS or other cloud services

Contact:

Email CV (with class ranking) and a short (1-page) research statement to songhan@mit.edu