Virginia Tech
Hume Center for National Security and Technology

Learning Signal Processing and Communications Systems from Data
IEEE-CCAA Workshop, OAI, OH 2017

Tim O'Shea, oshea@vt.edu
Research Faculty, Virginia Tech
Founder/Chief Scientist , DEEPSIG Inc

DEEPSIG

hume@vt.edu
www.hume.vt.edu

Overview

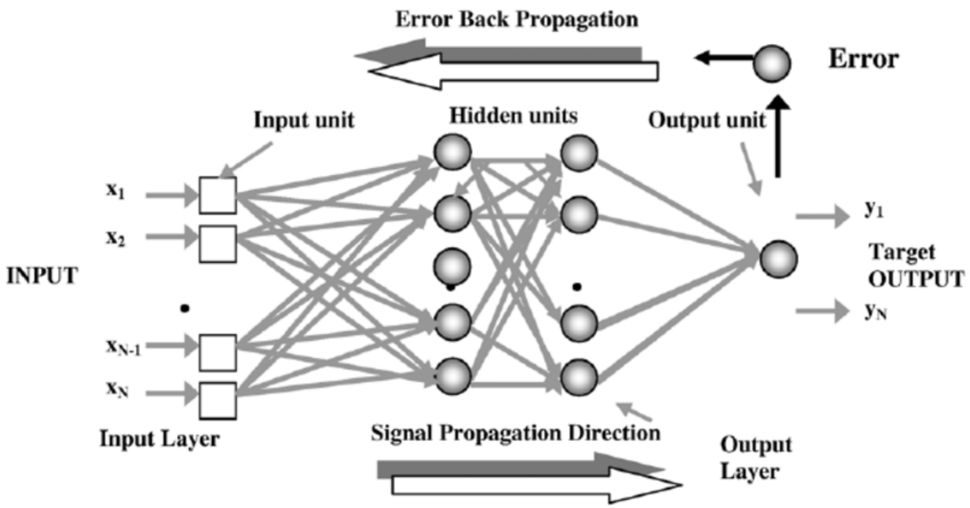
- Background
- Deep Learning
  - Enablers
  - Impacts
  - Implications
- Learning Radio Systems
  - Estimation Problems
  - Communications Systems
  - Radio Domain Knowledge
  - Signal Detection
  - Signal Recognition
- Future Work & Remaining Challenges

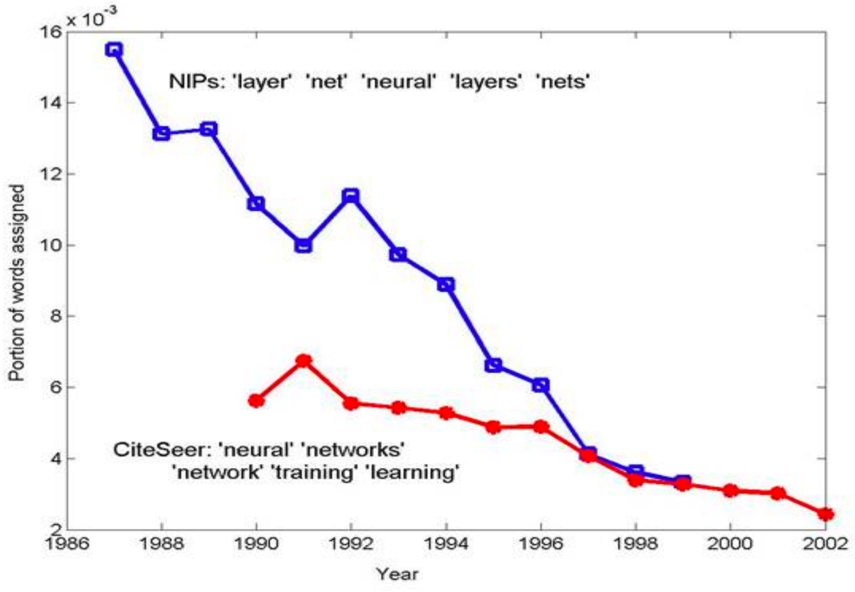
Deep Learning

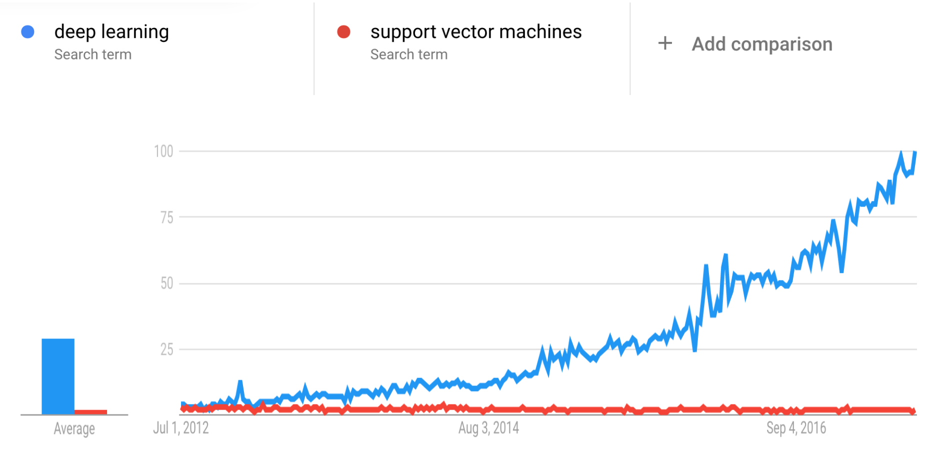
Enablers, Impacts & Implications

Deep Learning Enablers and Implications

- Neural Networks have been a long time coming
  - Rosenblatt's perceptron ~1957
  - NN Backpropagation of Loss ~1960



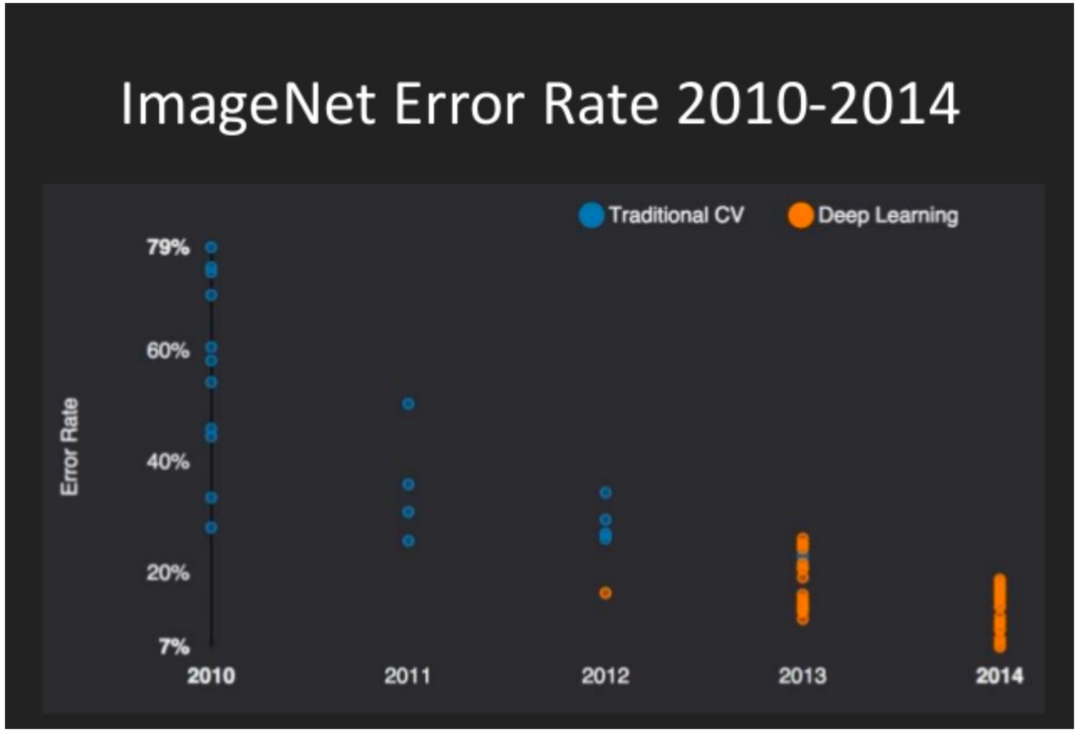


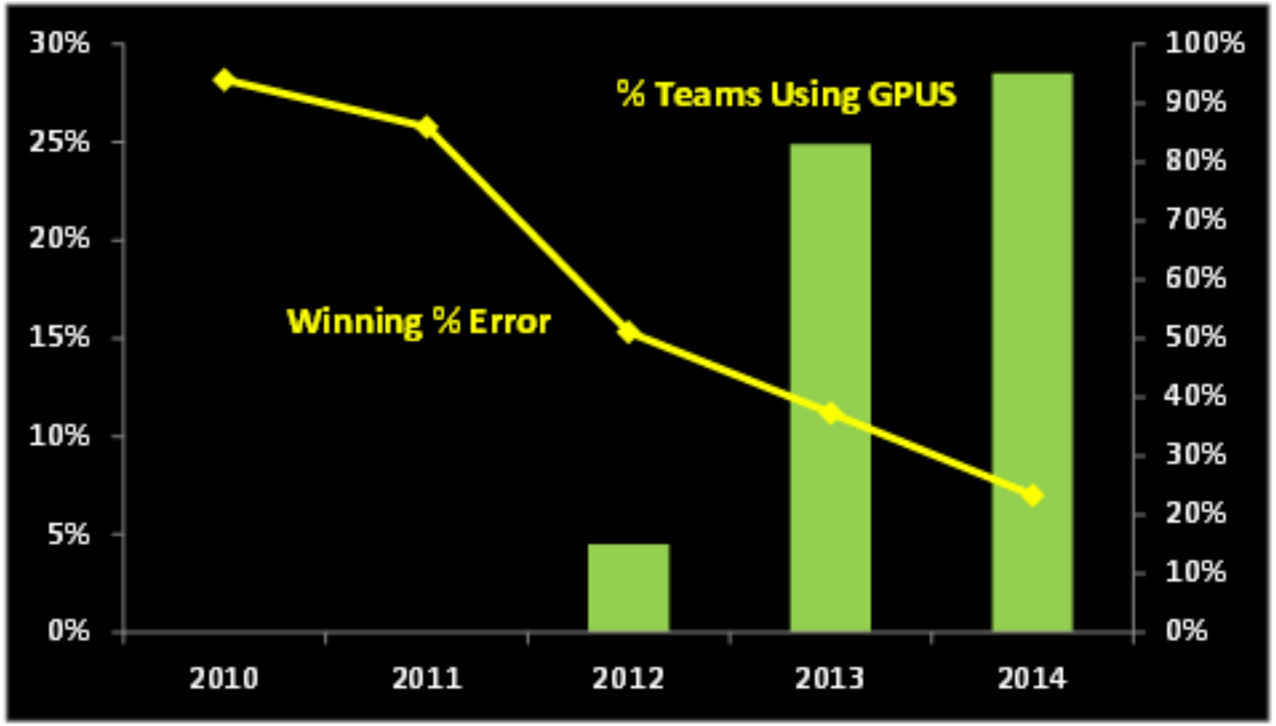


  - AI Winters: 74-80, 87-93, death of NNs
    - Max-margin methods, compact models
  - Cognitive Radio Winter ~(2010-2014)

Enter Deep Learning

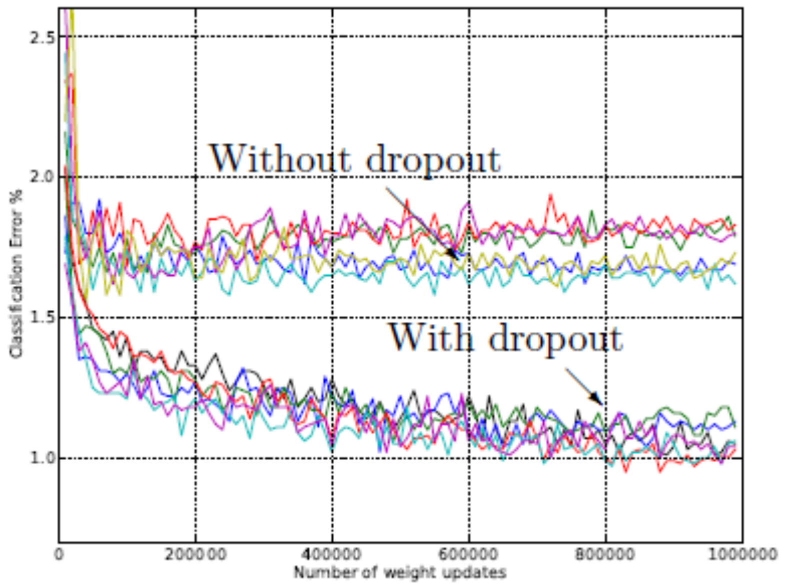
- 2012-2014 Completely up-ended the Computer Vision Domain
  - Complete shift from expert feature extraction: SIFT, Gabor, etc
  - Major decrease in error rates
  - Major adoption of Deep Learning methods & concurrent architectures





Enter Deep Learning

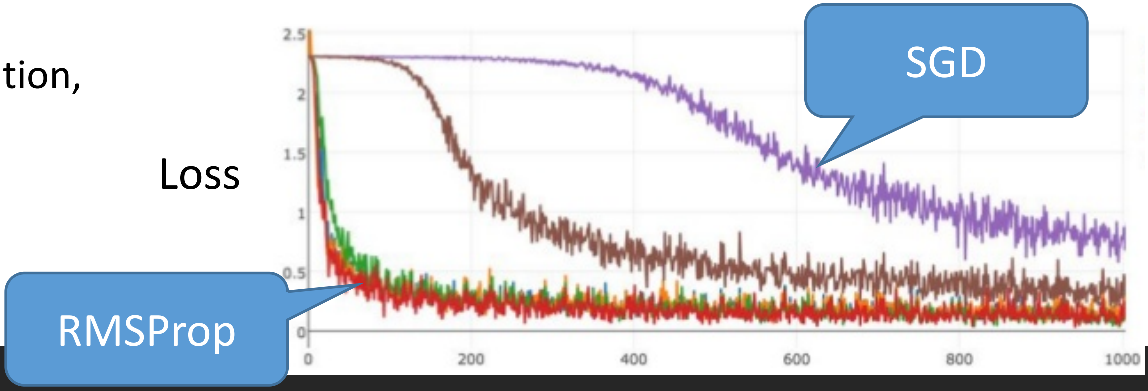
- Alexnet (2012): first modern deep learning network
  - Many convolutional layers, followed by FC, softmax (VGG architecture)
  - Improvements in gradient behavior
    - Normalization & rectified linear units
  - Strong regularization using Dropout
  - Mini-batch gradient descent /w momentum
  - Efficient concurrent distribution
- ML/DL Momentum & results have only increased since then
  - Pre-publication process (arxiv), rapid iteration
  - Commercial GPU compute widely available (low $/flop)
  - Massive improvements in
    - Attention, Sequence models, adversarial models
  - Currently massive interest and proven effective in:
    - Computer Vision, Natural Language Processing, Voice Recognition, Finance,
    - Applications to *new* domains is surprisingly slow given the advances we are seeing!
  - None of enablers is really specific to these fields!



Regularization Advances

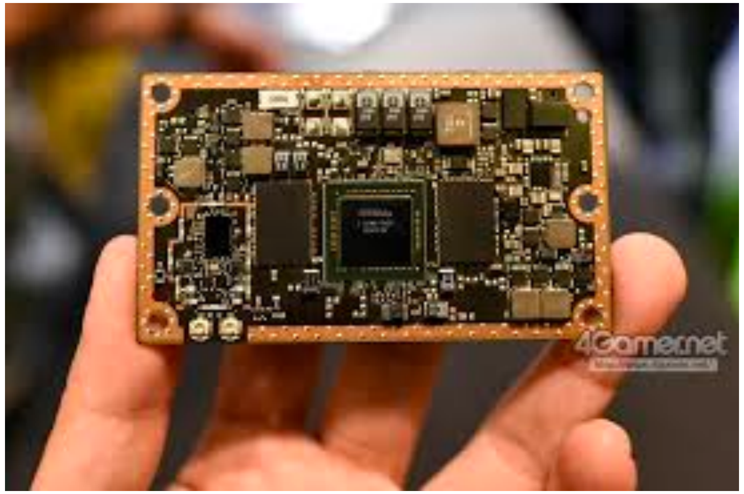
Solver Advances

Loss

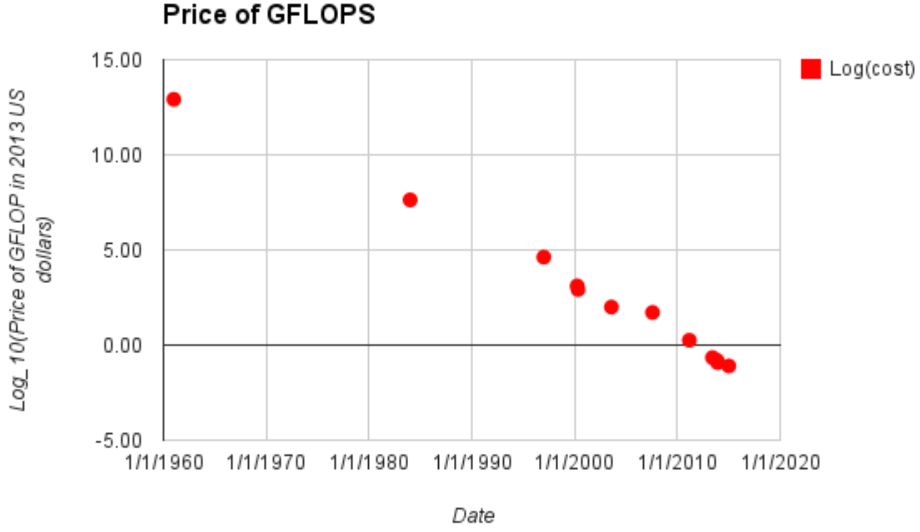


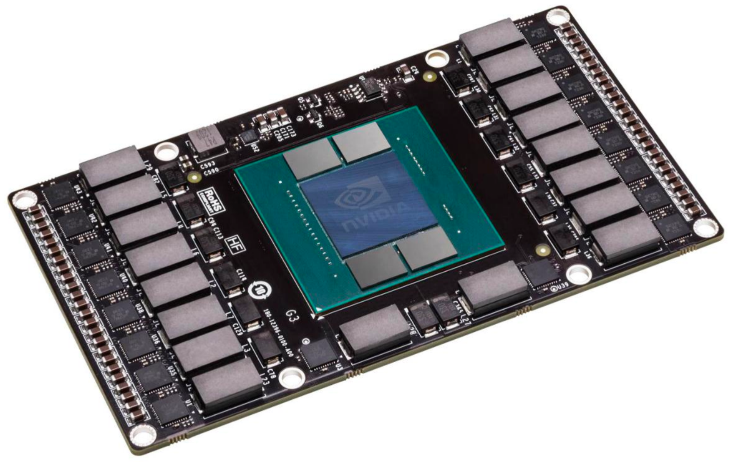
Computational Economies of Scale

- **Processing architecture economy of scale**
  - Many tasks centralizing around similar NN / tensor structure
  - Very well suited for concurrent architectures / GPUs
  - GPU compute architectures/software finally maturing
  - Programming models and cost of devices making major impact
- **Platforms for computer vision & autonomous vehicle navigation**
  - Numerous solutions competing for this space
  - Many of them very well suited for low-SWaP signal processing as well



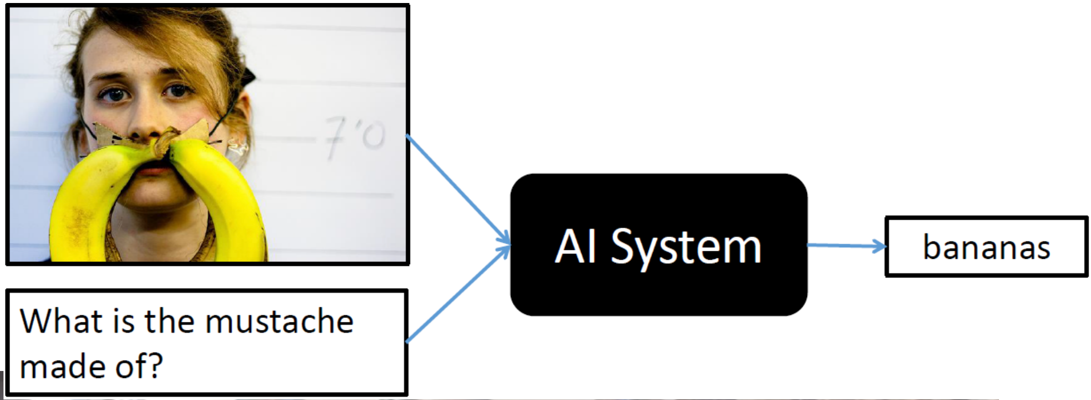
~20 dB Flops/$ Improvement!

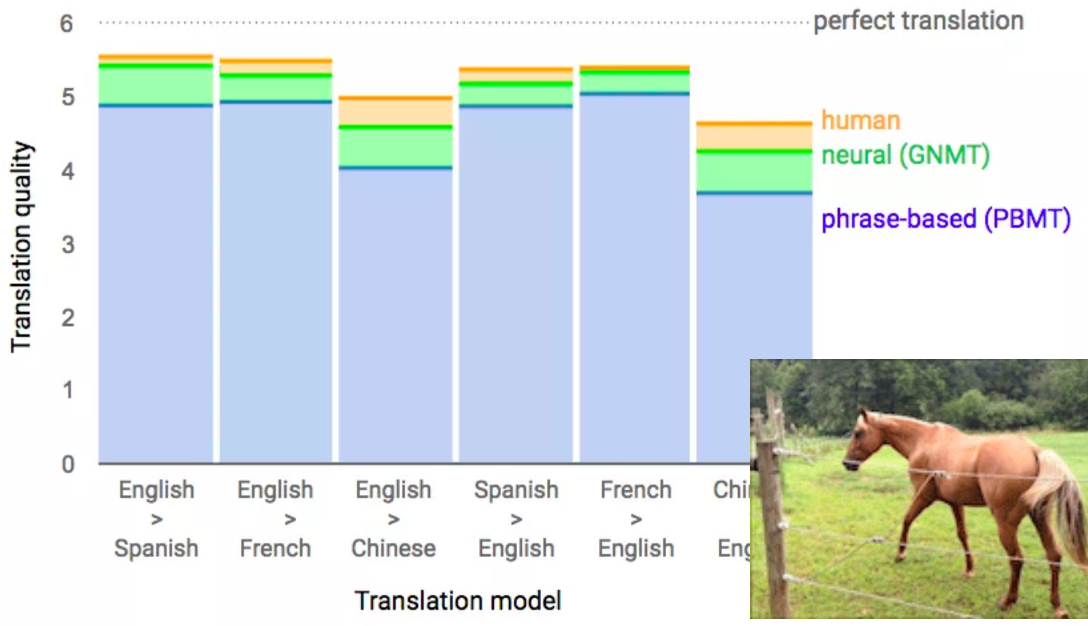


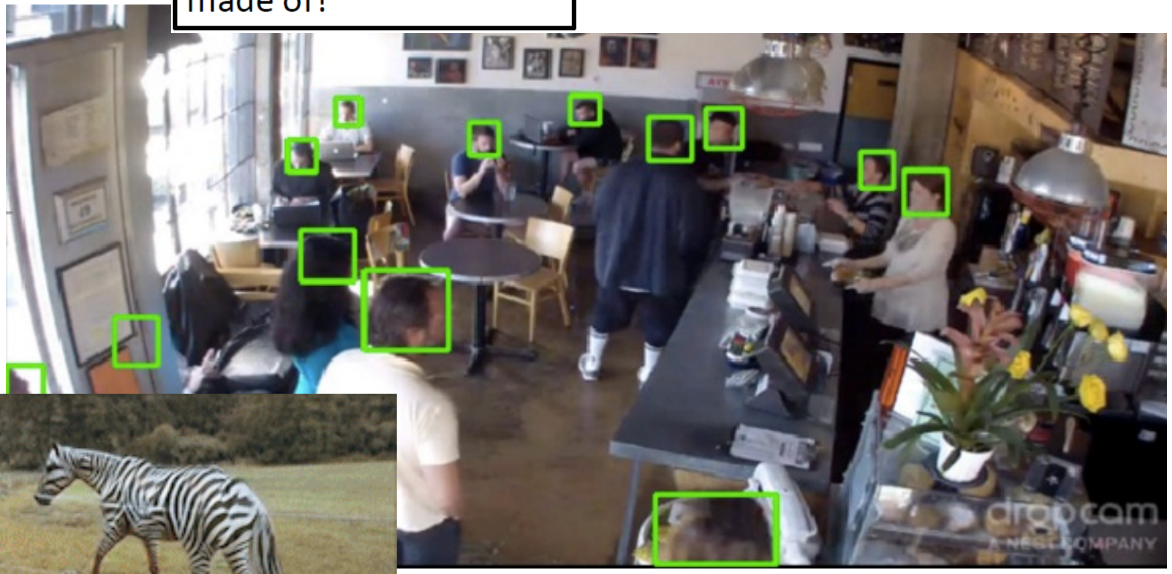


Deep Learning Vision in 2016 & 2017

- DL Machine translation extremely close to human level performance
- Large scale object recognition used at scale
  - Distributed embedded production systems
  - Self driving cars trusted with these approaches
- Powerful Visual Question Answering
- Powerful unsupervised style transfer for video

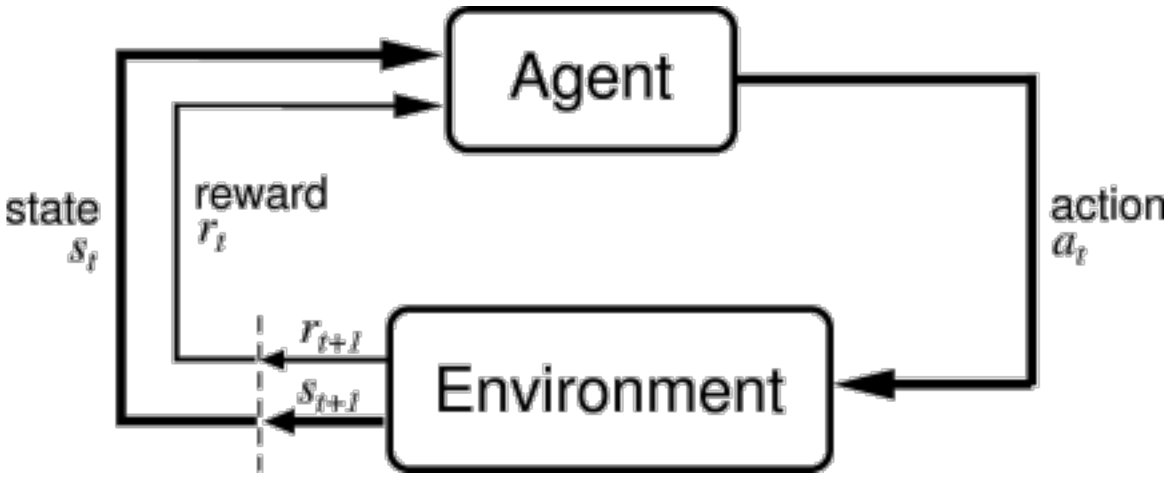






Deep Reinforcement Learning (aside)

- Sutton / Deepmind pioneer Deep-Q Learning
  - Function approximation of complex reward functions using experience and deep neural networks
- 2015: Deep Q Networks (in Nature)
  - Outperformed humans on a majority of Atari games
  - AlphaGO beats a top Go player (4-1)
- 2016:
  - A3C major improvements to DQN
    - Advantage and critic methods for adversarial learning)
  - Datacenter optimization (40% reduction cooling)
- 2017:
  - AlphaGO beats world #1 (Jie) and team China
  - ATARI feats becoming ridiculous
    - Score overflow in Ms Pacman
  - Improvements in learning, regularization, experience replay, policy function representation all major



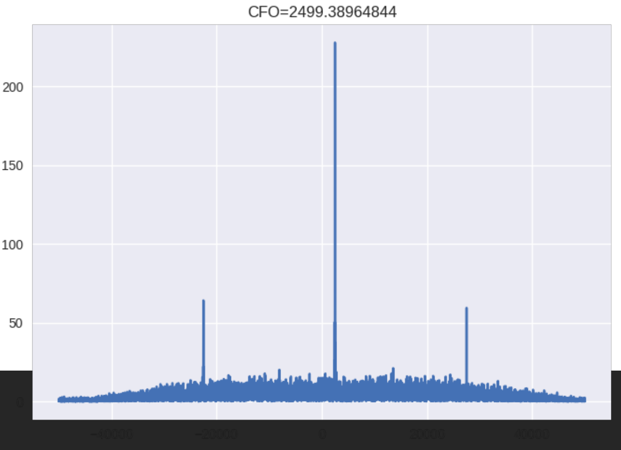
$$Q^*(s,a) = \mathbb{E}_{s' \sim \mathcal{E}}\left[r + \gamma \max_{a'} Q^*(s',a') \middle| s, a \right]$$

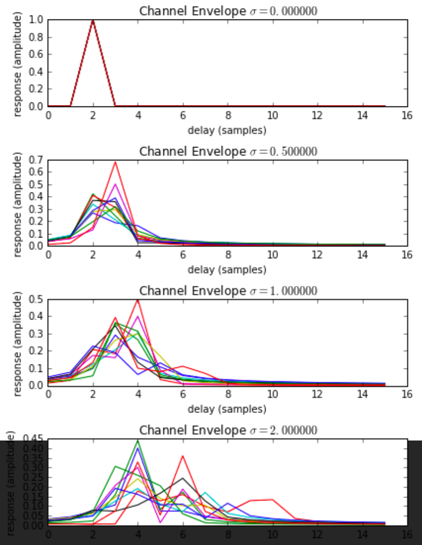
Rethinking Radio Signal Processing

Taking Advantage of Data-Driven Large-Scale Optimization in Radio

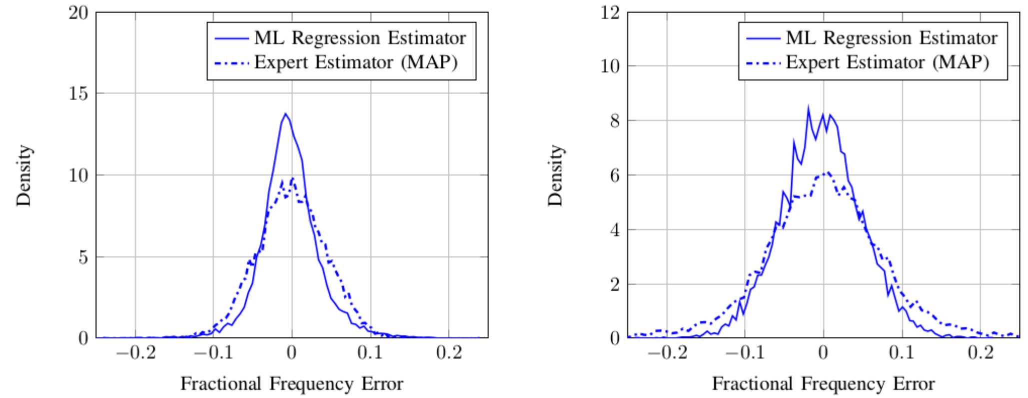
Resilient Communications System Learning

- Estimation is central to signal processing
- Traditionally MSE/MLD/MAP optimization on closed form analytic signal expressions
  - Often makes AWGN channel assumption
  - Analytic cyclo-stationary features widely used in radio communications system today
    - Works well for long integration and channels which matches analytic model (AWGN)
  - We can do better for short time problems!
    - Single high order moment does not capture all available information!
- Wide ranging implications to signal processing estimation problems!

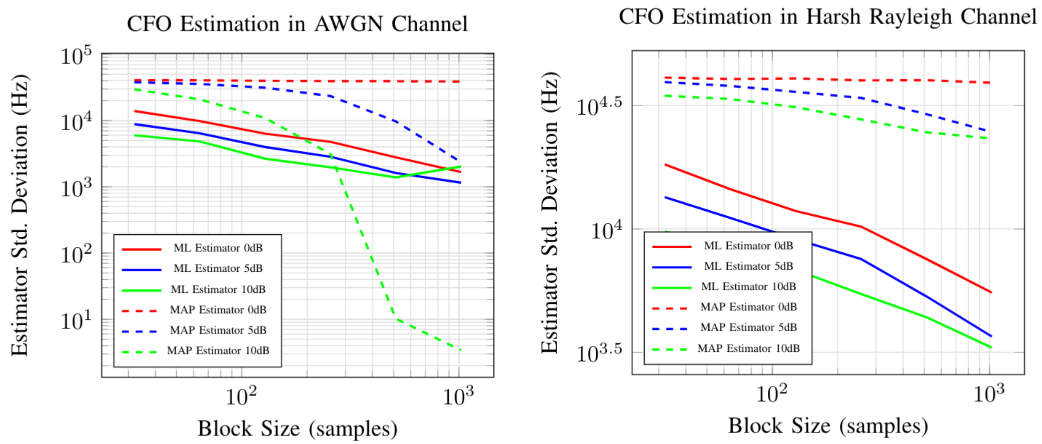




Short time CFO regression error distribution

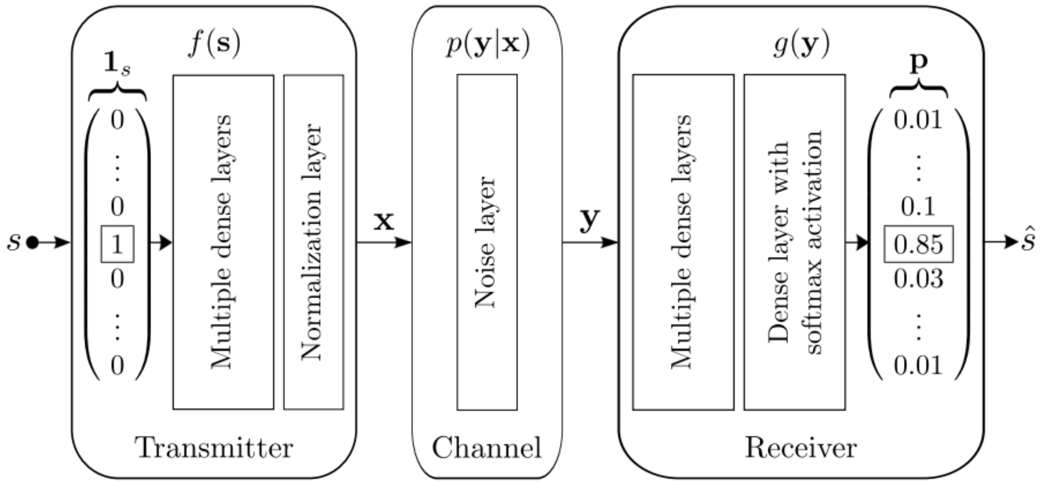


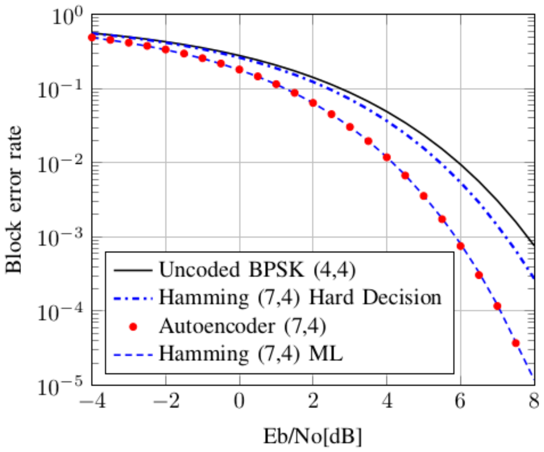
Error performance under AWGN/Fading for varying SNR & estimation block length

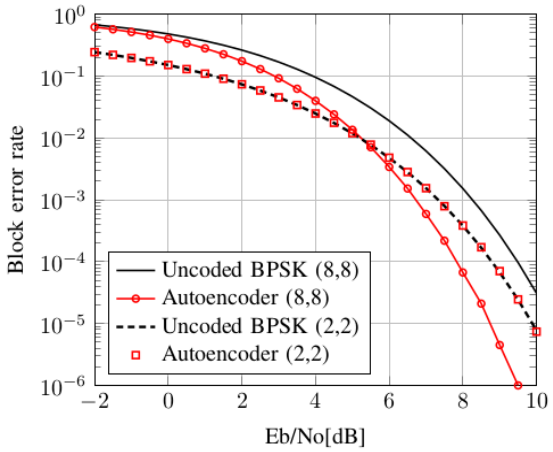


Resilient Communications System Learning

- For the past 70 years, communications systems have relied on closed-form analytic algorithm engineering
  - Simplified toy models for propagation
  - Disciplines of research dedicated to each task: modulation, coding, estimation, equalization
  - Lacking real data and full system optimization
- Need to re-think the basic problem
  - Reconstruction of transmitted information bits
  - Real data to represent propagation effects
  - Full-system optimization through end-to-end learning
  - Jointly optimize transmitter, receiver, synchronizer, over the air information representation.
  - This works surprisingly well!

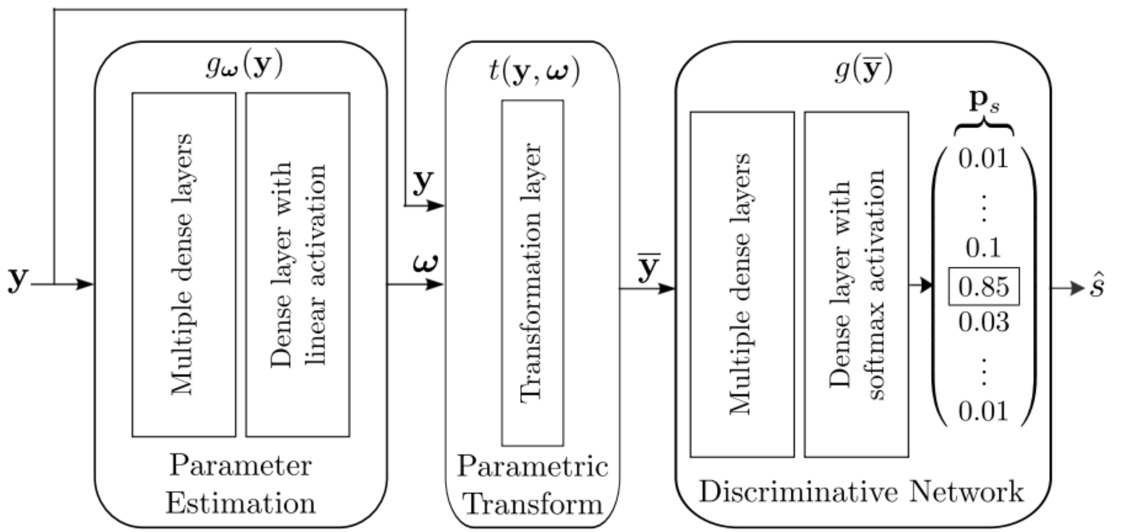


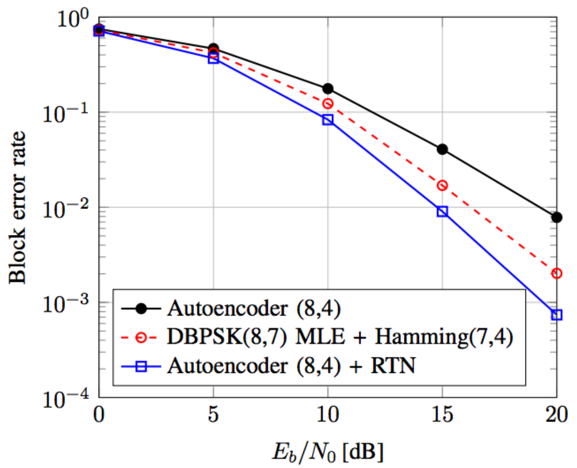


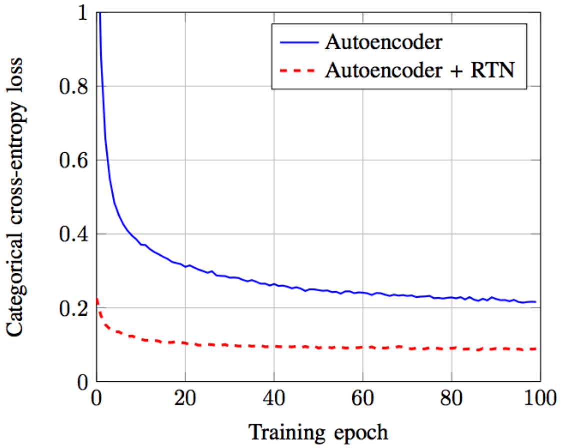


Resilient Communications System Learning

- Merging end-to-end learning with expert knowledge about the propagation domain
  - Decomposition of receiver
    - Learned estimation modules (Attention model)
    - Expert transformation modules to match physical world propagation models/effects
    - Learned demapping/representation modules
  - Joint learning of encoder/modulator, synchronizer, decoder/demodulator, and over the air representation
    - Lower complexity learning problem
      - Converges faster, less overfitting
      - Only imparts propagation medium expert knowledge (things we can't change)
      - Learn everything else end-to-end

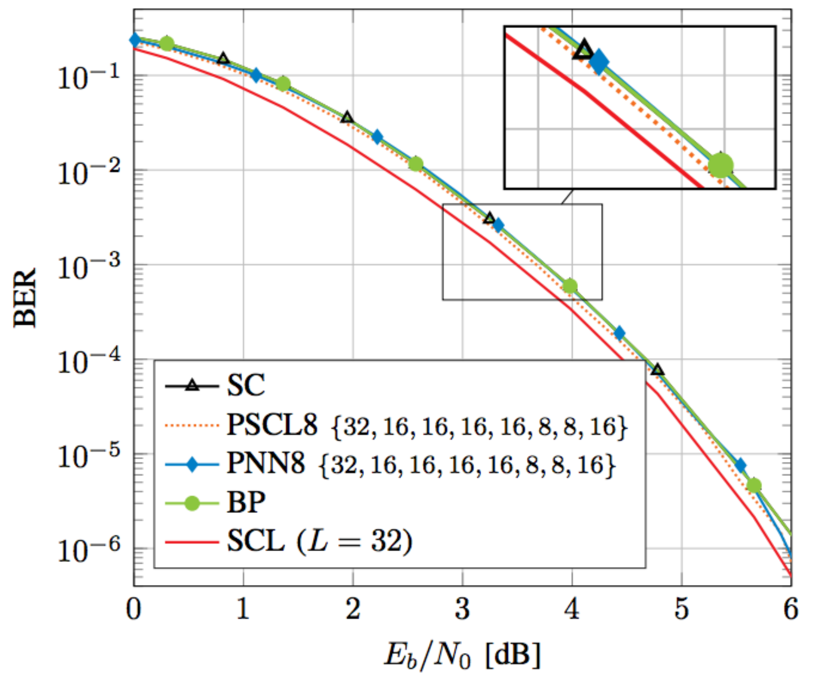


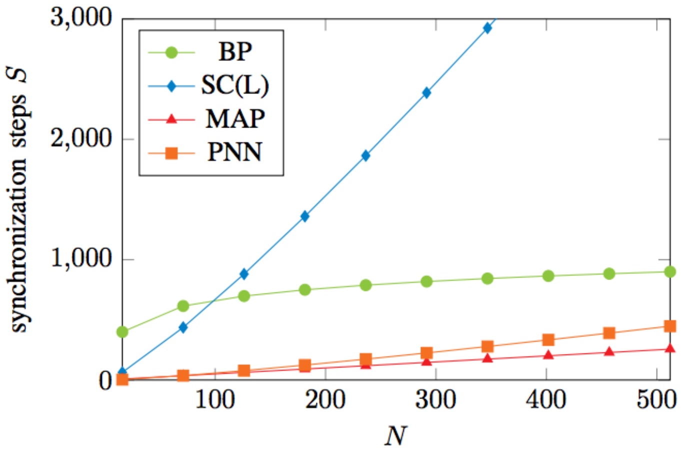




Resilient Communications System Learning

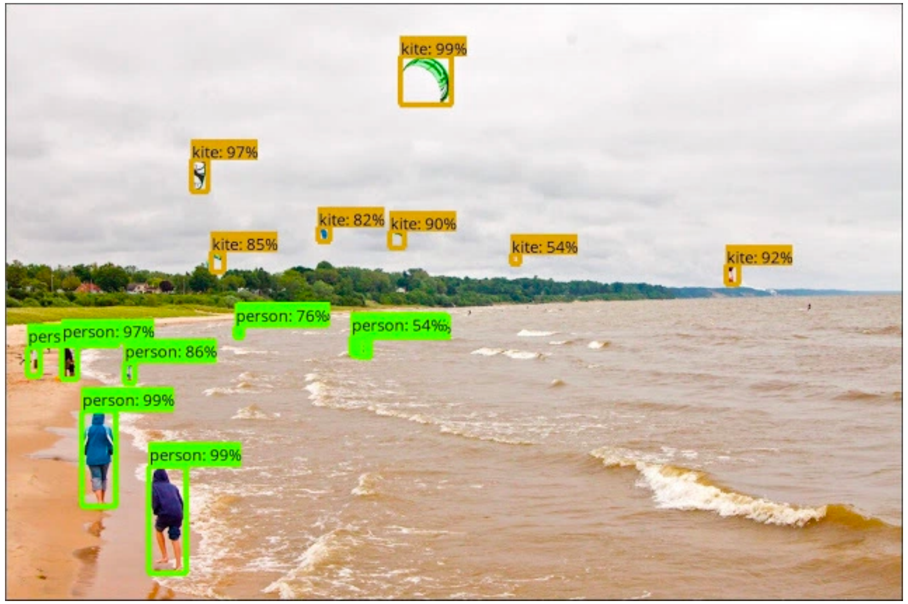
- Learning decoders for 'near optimal' error correction codes! (Partitioning to scale)
  - FEC decoding / detection currently the #1 power consuming operation in radio baseband devices
  - Work from Cammerer, Gruber, Hoydis, ten Brink (University of Stuttgart / Nokia-Bell Labs)
  - Shows near-optimal polar code decoding performance with partitioned neural networks at lower complexity than successive cancellation or Belief Propagation!
- Potentially a major advance in error correction
  - Learn approximate decoders on code-word sets
  - Low latency one-shot decoding at higher efficiency
- **"Scaling Deep Learning-based Decoding of Polar Codes via Partitioning"** https://arxiv.org/abs/1702.06901

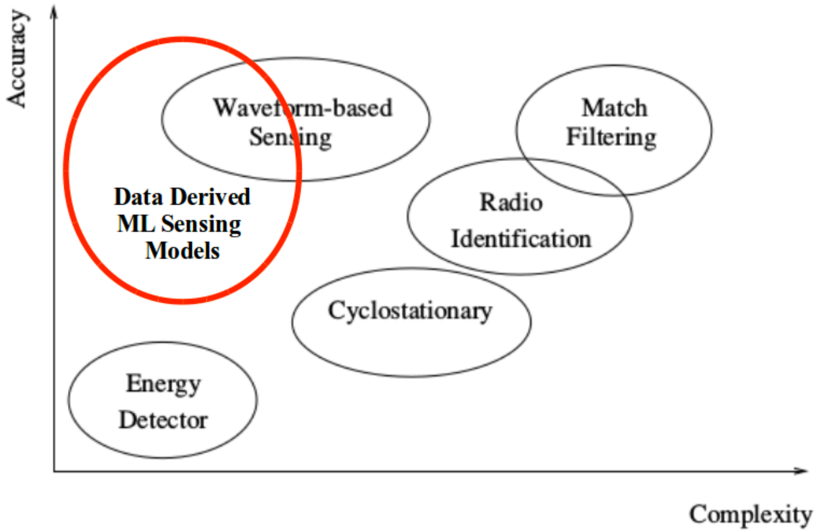


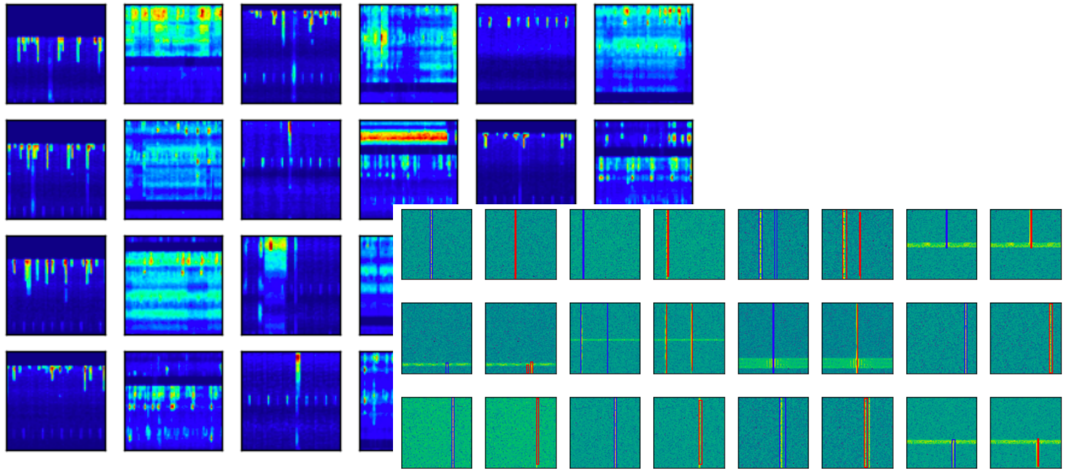


Signal Detection and Rapid Scene Understanding

- Most analogous problem to computer vision
  - Spectral situational awareness still a big problem
    - Highly specialized and brittle currently
    - Current solution lack context, simple energy detection methods are still pervasive!
  - Adopt learned detection methods for spectral event detection
    - Supervised bounding box regression
    - Rich context-aware detection and spectral-localization!
    - (Results to appear Asilomar SSC 2017)

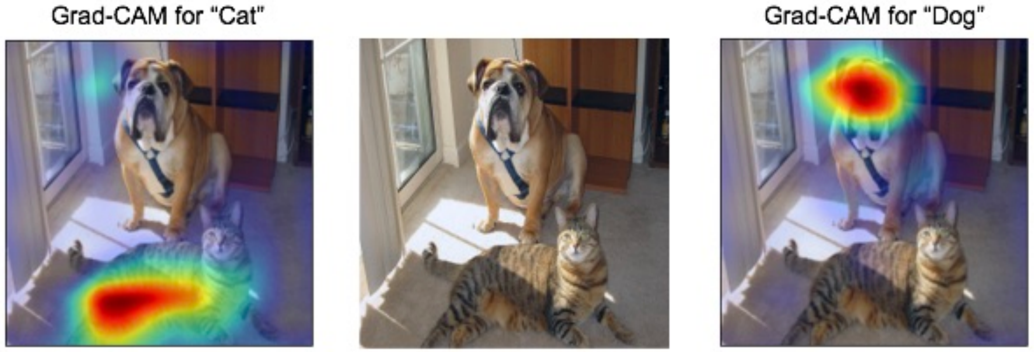


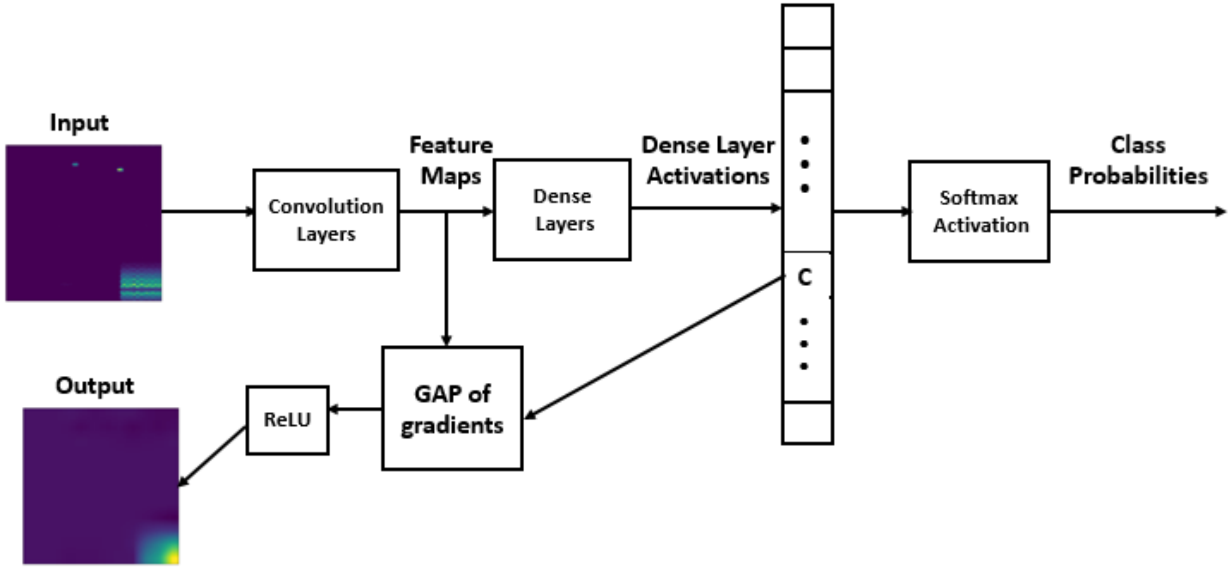


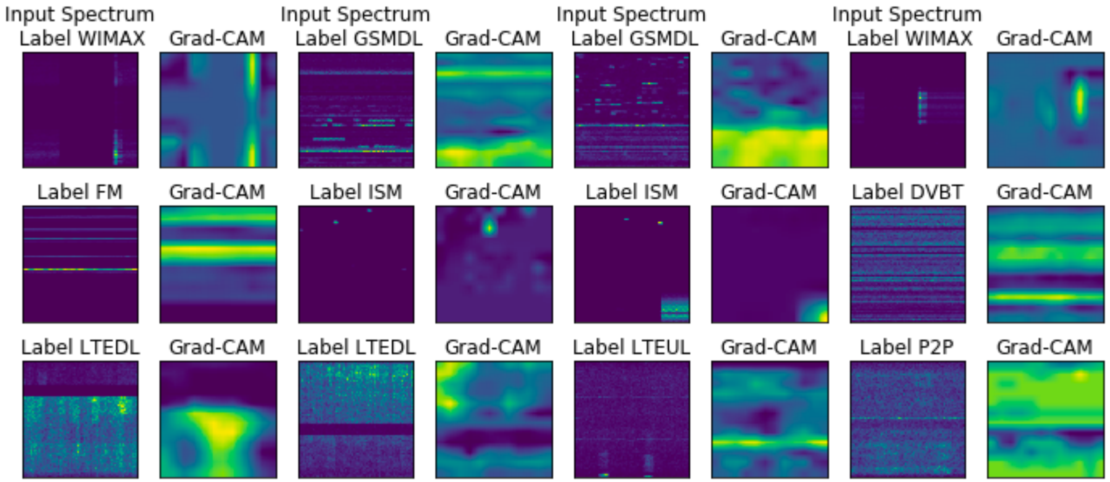


Signal Detection and Rapid Scene Understanding

- Semi-supervised spectral event detection and localization without labels
  - Localization of features used for classification
  - Does not require intensive training process, rapid learning and adaptation on new signals
  - Systems where detection generalized well to unseen signal types or operating modes!
  - (results to appear EUSIPCO 2017)

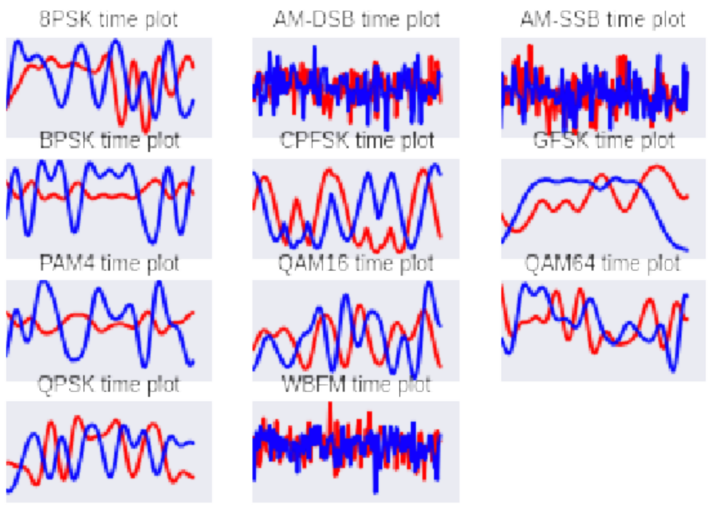


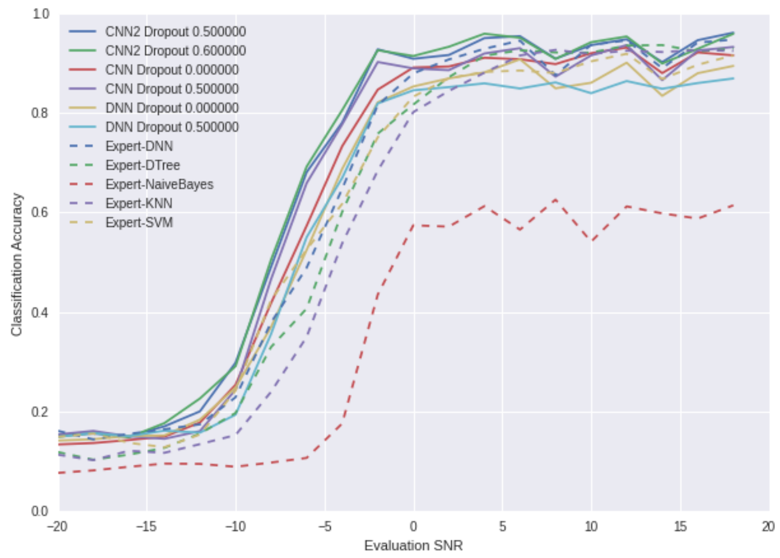


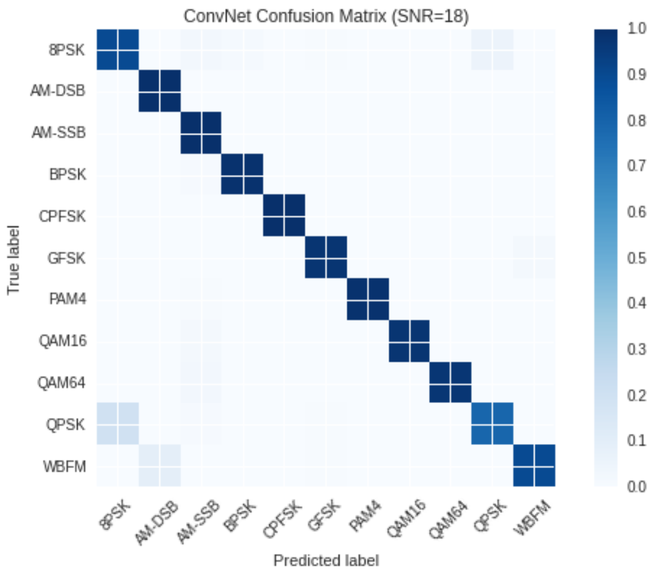


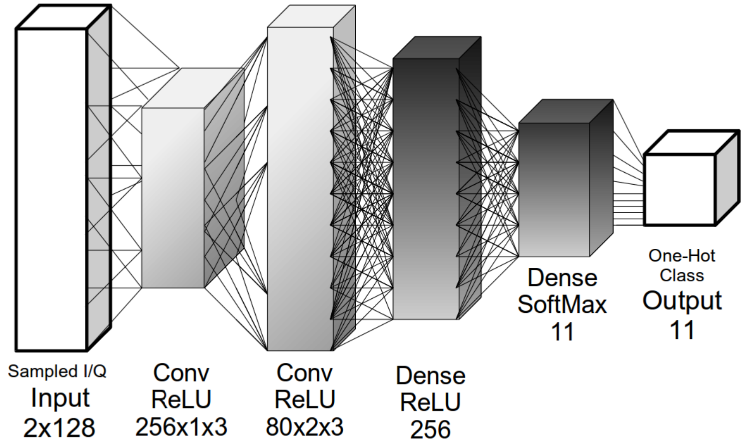
Signal Classification and Recognition

- Convolutional neural feature learning extremely well suited for radio signal identification & learning!
  - Simple CNN outperforms all traditional classifiers on modulation recognition task (SVM, D-tree, Naïve Bayes, KNN, etc). (2016 paper)
  - radioml.org dataset (radio dataset and resources for ML)
  - Numerous works since then (DARPA + DySpan workshop on the topic) in Q1 2017 (DARPA logistical failure, results never announced)







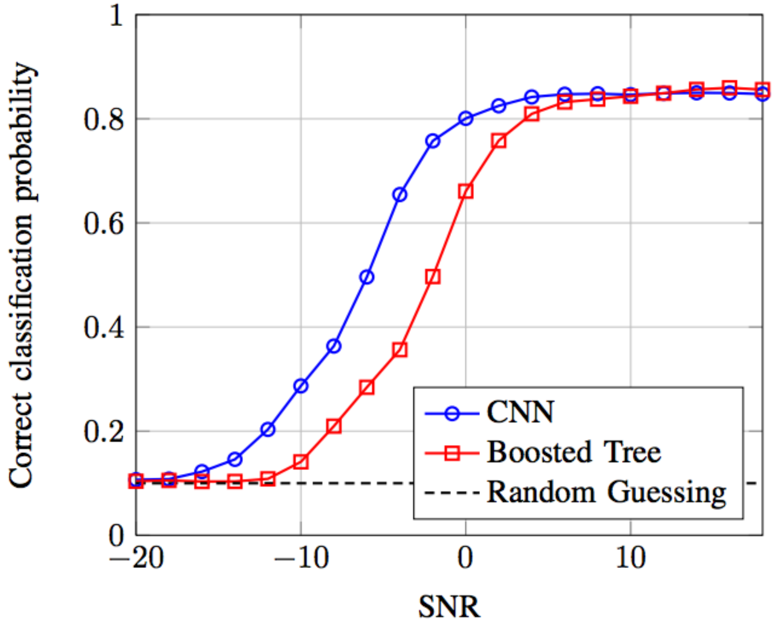


Signal Classification and Recognition

- Refinements in Q4 2016 / Q1 2017 Journal paper
  - Rigorous higher order moment classifier /w boosted gradient tree classifier (XGBoost is the gold standard for numerous Kaggle competitions)
  - Significant time spent tuning and optimizing this baseline for rigorous comparison
  - Tuned CNN outperforms feature based classifier by 3-4 dB accuracy across the board
  - Field has not had rigorous dataset and baseline methods release openly in the past, hard to perform head to head comparisons without this
- Pending release of 2017 dataset
  - 30 modulations and improvements in harsh channel models!

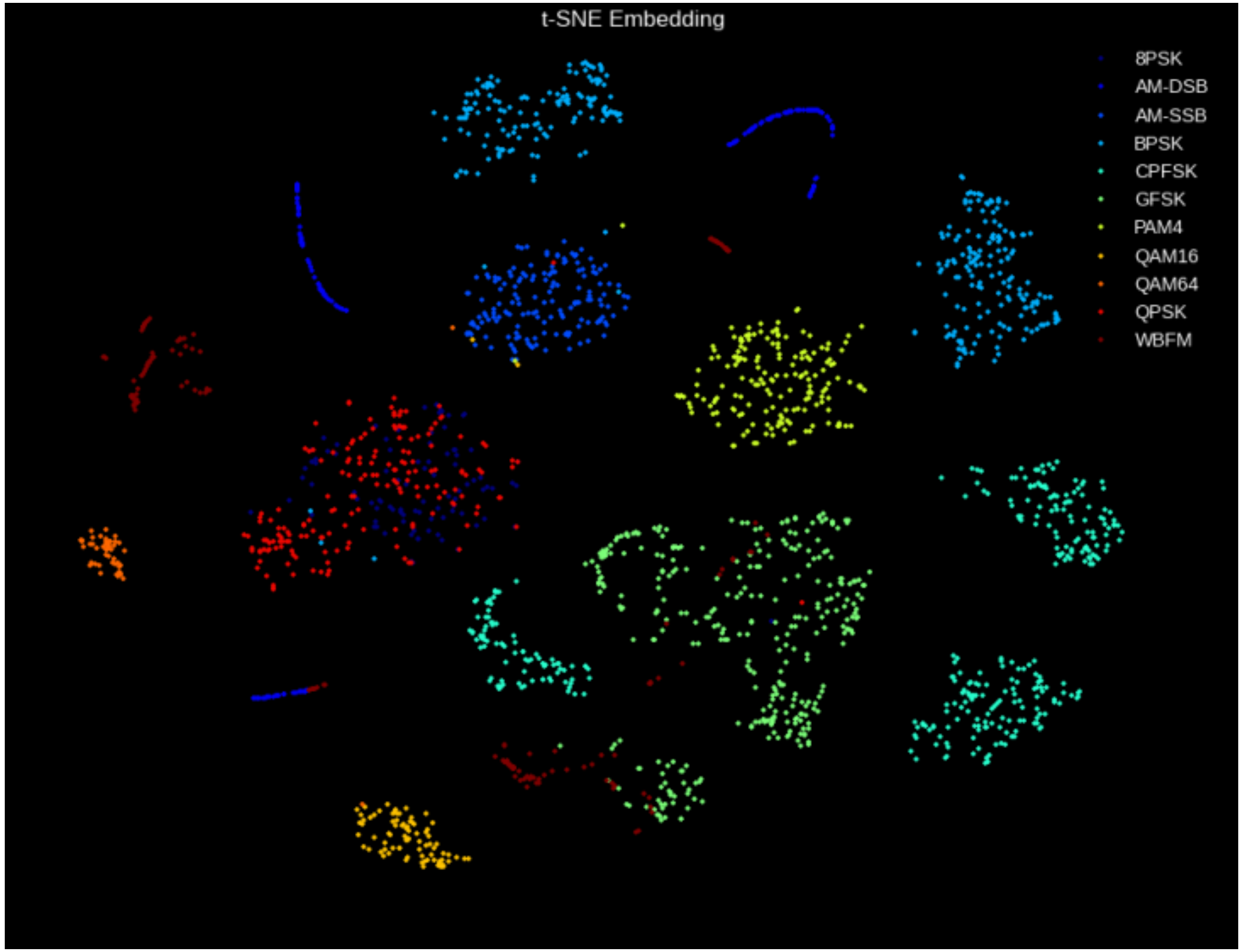
TABLE V: Layout of the CNN for modulation classification with $324,330$ trainable parameters.

Layer	Output dimensions
Input	2×128
Convolution (128 filters, size 2×8) + ReLU	128×121
Max Pooling (size 2, strides 2)	128×60
Convolution (64 filters, size 1×16) + ReLU	64×45
Max Pooling (size 2, strides 2)	64×22
Flatten	1408
Dense + ReLU	128
Dense + ReLU	64
Dense + ReLU	32
Dense + softmax	10



Semi-Supervised Signal Identification Learning

- Training data is the major challenge and limitation for this class of systems
  - Learning must occur on representative experiences and data examples
  - Need a way to accumulate new experience and identify new class labels
  - Visualization of new class identification through clustering and representation in a reduced dimensionality space using t-SNE.
  - Promising basis for EM recognition systems which learn and evolve classes over their lifetime
  - Rapidly recognize and reason with new unseen class types



Future Research Directions

Priorities and Impact

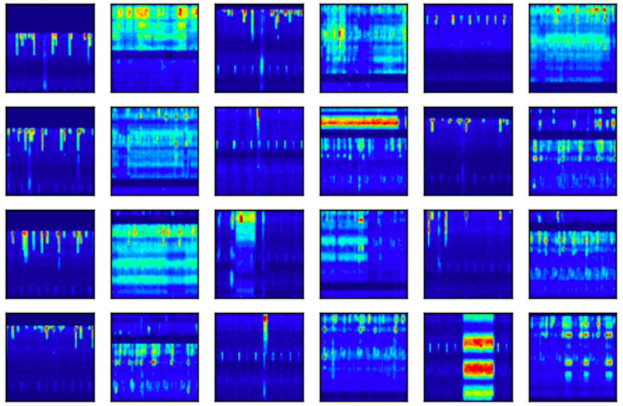
Ways Forward

LIBERAL-ARTS MAJORS MAY BE ANNOYING SOMETIMES, BUT THERE'S *NOTHING* MORE OBNOXIOUS THAN PROGRAMMERS ENCOUNTERING A NEW SUBJECT.

Approximate DL methods actually competitive with state of the art in many fields quite rapidly …

Ways Forward

- **Better Signal Processing Algorithms Than Possible Before**
  - Better models for the environment, context aware, data driven
  - End-to-end optimized and computationally efficient
  - Naturally concurrent and efficiently realizable
- **High Level Contextual Understanding of Radio Environment**
  - We have never been able to achieve this due to complexity of representation and diversity of expert signal processing approaches needed today
  - Exciting possibilities high level understanding and decision making
- **Highly Tuned Domain Adaptive Communications Systems**
  - Communications systems which are optimized for their exact operating environment
  - Better aggregate spectral efficiency, lower power consumption
- **Drawbacks & Difficulties**
  - Explainability is still a difficult problem
    - Effect is dominated by the variance in datasets not models
  - Hyper-parameter and architecture tuning is an unsolved problem (but we have some good approaches today)
- **Hostility in Degree of Departure from Status Quo**
  - Significant departure from signal processing status quo
  - Closed form analysis is nice when its possible, but often has to make simplifying assumptions which make it equally ‘approximate’!



Virginia Tech

Hume Center for National Security and Technology

DEEPSIG

http://deepsig.io

http://radioml.org