SCALING DEEP MACHINE LEARNING SYSTEMS USING ANALOG COMPUTATIONS

Itamar Arel, Jeremy Holleman
Department of Electrical Engineering & Computer Science
The University of Tennessee
The Challenge

- We have great video processing SW
  - Real-time object recognition
  - Situation awareness

- **Problems:** consumes too much power!
  - Limits flight time
  - Dictates size/weight

- **Challenge:** can we reduce power by 100x-1000x while retraining performance?
Existing Technological Landscape

- Most DSP solutions use digital CMOS technology
- **Storage** - registers
  - e.g. 20-bit register requires ~200 transistors
- **Computation** - digital logic (gates, etc.)
  - Simple gates require >5 transistors + misc
- **Power** $\sim$ frequency
Solution: Analog Computations

- What if we could replace a register with a couple of transistors?
  - Substitute digital values with analog charge

- What if we can obtain functionality using analog circuits?
  - Leverage characteristics of devices instead of synthesized logic
DARPA UPSIDE

- 4 year/$4.8M project (begins May 2013)
- Team: UTK (lead), ORNL, U. Michigan

Highlights:
- Power gain of 2-3 orders of magnitude
- Two phases: existing analog CMOS + “Emerge devices”
- Deliverables: fabricated chips that demonstrate performance/power
DARPA UPSIDE Program Requirements

- “...real-time (video-based) target recognition and tracking over large numbers of objects ...”

- “to create a high-level non-Boolean computational ... unique operational properties of new, power efficient, non-CMOS, nanoscale devices.”

- “employ biologically-inspired hierarchical architectures for information representation ...”
Dealing with High-dimensionality

- Images contain millions of pixels
- No classifier can handle such dimensionality
  - Curse of dimensionality
- Standard techniques: three-phase process

ROI detection  \[10^6\]  Feature Extraction  \[10^4\]  Classification  \[10^2\]
Limitations of this approach:

- Domain specific
- Requires extensive hand-crafting of features

Dealing with High-dimensionality
Deep Machine Learning*

- Biologically-inspired computational intelligence approach
- **Goal:** learning to represent complex observations
- **Hypothesis:** brain represents the world by exploiting a hierarchy of abstraction

DeSTIN Deep Learning Framework
Benchmarking DeSTIN

- **Deep SpatioTemporal Inference Network (DeSTIN)**
- **State-of-the-art results** on many benchmarks

- **MNIST** - 98.8%
- **TIMIT Phoneme Recognition** - 80.1%
- **CIFAR-10** - 79%
DeSTIN Under the Hood

- Each node in DeSTIN captures spatiotemporal regularities
  - Modality independent (not just images)

- A node maintains a belief — represent similarity of current input to learned templates
DeSTIN: Belief Construct

- Combination of online clustering with feedback-based Bayesian inference
Low-power Analog Computations

- Computation from transistor physics → extremely low power dissipation
  - Example: Bump Circuit: 200 comparisons/sec @ 200 fA
  - 1 Inverter @ 200 Hz → 4x current!
- Floating gates: analog storage
  - Update via tunneling, injection
- Many applications can tolerate inaccuracies
  - Natural feedback in ML algorithms
The Challenge

- Many imperfections in analog computations
  - Inaccurate computations
  - Mistmatches across many signals
  - Additive and multiplicative noise
- Research focused on resolving the above
- Involves ML and electronics students closely collaborating
Multi-target tracking from monocular camera

- **Goal:** Jointly track multiple targets and estimate the camera pose parameters.
- **Novelty:** Assumes: (1) the camera is moving; (2) the camera is un-calibrated and monocular; (2) targets can move in arbitrary directions in a cluttered environment; 3) Efficiency is a critical requirement; 4) The algorithm can use multi-modal observations (images + 3D range data) if available.

- **Technical approach:**
  - Inputs: target detections (confidence map) frame by frame
  - The unknowns (target trajectories in 3D, associations and camera parameters) are estimated using RJ-MCMC
  - Track associations may be guided by high level semantics (activity types, interaction models, etc...)

- **Results:**
  - State-of-the-art tracking and camera localization results
  - Tested on indoor and outdoor video sequences
  - Extensive qualitative and quantitative analysis on several public and in-house datasets

Our approach allows to localize the camera in the 3D world coordinates

Our approach enables accurate tracking of multiple individuals from a video sequence
Thank you