Integrating Deep Learning and Model-Based Reasoning for Robust Sensor Data Fusion

Chee-Yee Chong
Independent Researcher
Los Altos, California
USA

cychong@ieee.org

Presented at NATO SET-262
Budapest, Hungary
5-6 November, 2018

Motivation

- Humans are good at sensor and data fusion
  - Low level perception from sight and sound
  - High level reasoning to understand situation, cause and effect
- Artificial intelligence (AI) attempts to replicate human sensor data fusion capabilities
- Traditional AI approaches rely on domain knowledge
  - Represent and reason with models
  - Models provide explainable results
- Deep learning (deep neural networks) relies only on data
  - Neural networks learned from large amounts of data
  - Excellent performance in some applications but results are hard to explain
- Robust sensor data fusion will require both deep learning and model-based reasoning

Face Grouping by Google Photos - Adults

2013

2010

2009

2009

2008

2007

Adult faces are correctly grouped into a “track” over time

Face Grouping by Google Photos - Babies

March 2018

December 2014

- Faces of two babies are incorrectly grouped
- Incorrect grouping can be detected easily from
  - Metadata (dates when photos are taken)
  - Knowledge/model that babies grow fast and appearance change quickly

Outline

- Model based AI for sensor and data fusion
- Deep learning as function-based AI
- Approach to integrate deep learning and model-based reasoning for fusion

Advances in Computing/Communication and Sensor/Data Fusion - Perspectives from A Silicon Valley Researcher

1980

Vax 11/750 Minicomputer
3 Mhz, 120K FLOPS, 2 MB RAM
1200 baud dialup modem

- Government/non-profit
  - SRII ← Garvey (AI)
  - NASA Ames
- Large aerospace/defense
  - Lockheed ← Reid (MHT)
  - TRW Ford Aerospace
  - Loral GTE
- Small R&D companies
  - Systems Control ← Bar-Shalom
  - Advanced Information & Decision Systems (AI&DS) ← Chong

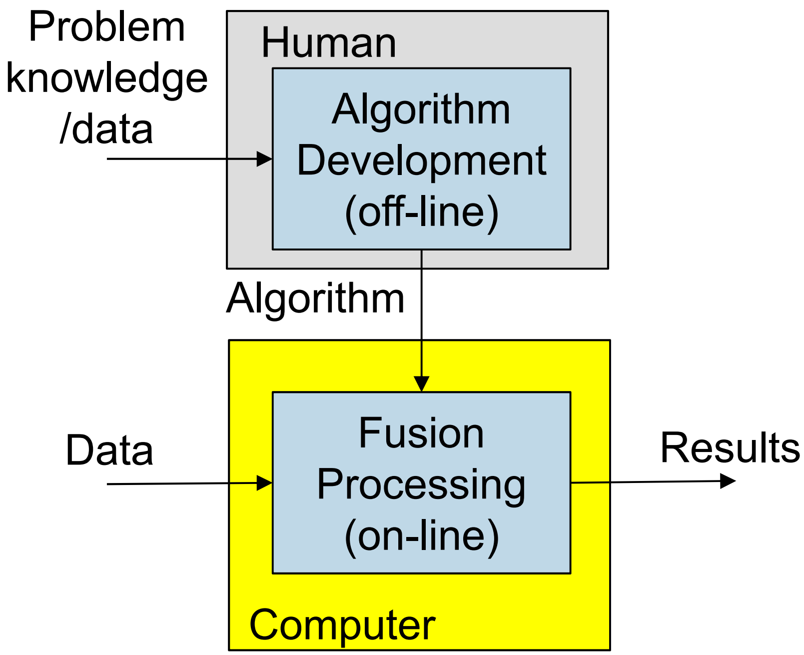
2018

IPhone XS
2.49 Ghz, 409 GFLOPS, 4 GB RAM
1 Gb network

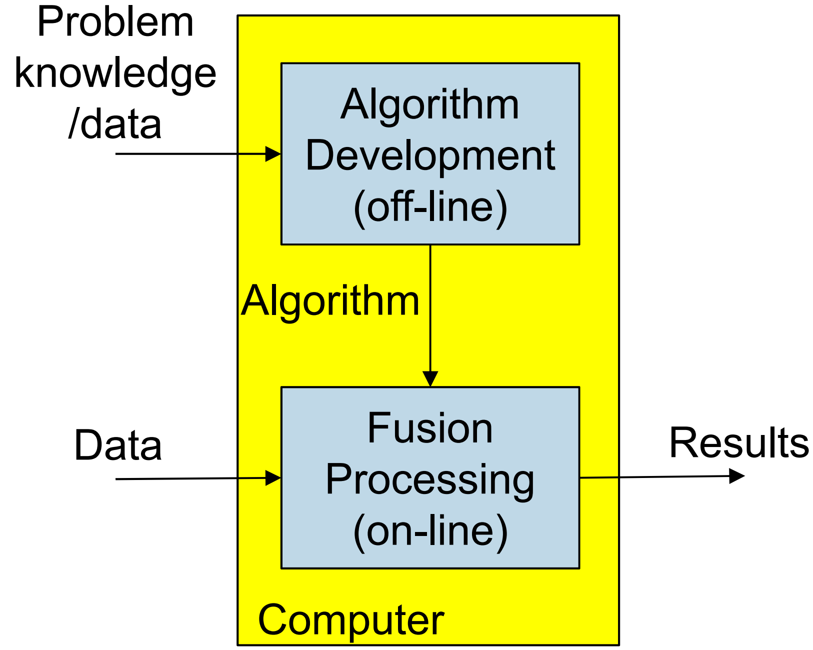
- Government/non-profit
  - SRII
  - NASA Ames
- Large aerospace/defense
  - Lockheed Martin
  - Loral
- Computing
  - Apple
  - Intel (Mobileye)
  - Nvidia
- Automotive
  - Tesla
  - BMW
  - Mercedes
  - Audi
- Internet/software
  - Google
  - Facebook
  - Microsoft
  - Amazon
  - Netflix

- Revolutionary advances in computing and communication
- More commercial applications in sensor and data fusion, easy and difficult
- Government no longer drives technology development

AI Exploits Computing Power to Develop Algorithms



Traditional approach relies on humans to develop algorithm



AI automates algorithm development

AI Approaches for Sensor Data Fusion Are Evolving

- Expert systems (mimic human experts)
  - Speech understanding
  - Signal understanding
- Probabilistic reasoning (with models)
  - Situation understanding
  - Object recognition
- Neural networks / deep learning
  - Speech understanding
  - Object recognition
  - Video tracking

First wave of AI - **Model-based** with handcrafted knowledge

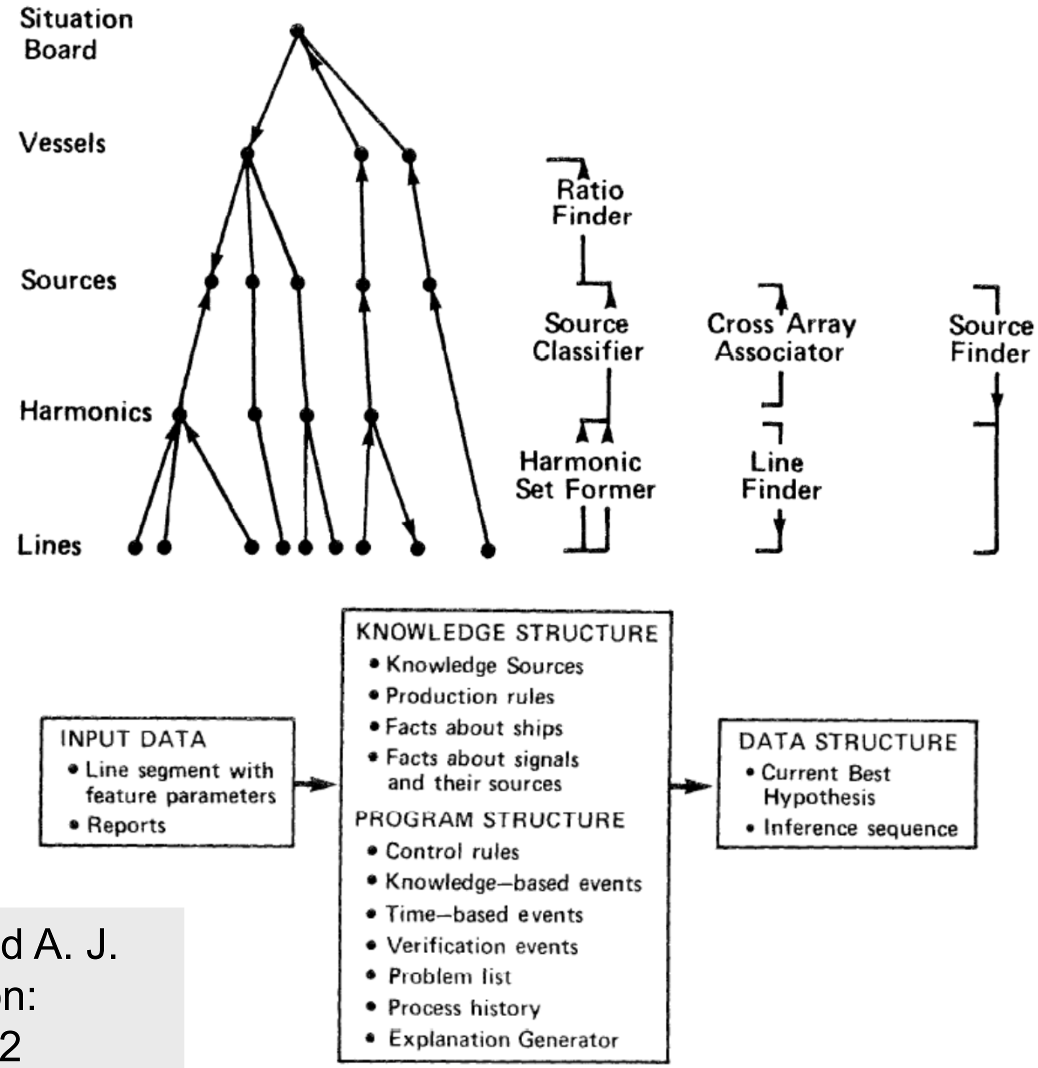
Second wave of AI - Statistical learning /**function-based** from data

Third wave of AI - Contextual Adaptation

J. Launchbury, "A DARPA perspective on artificial intelligence," 2017
http://www.darpa.mil/attachments/AIFull.pdf

Expert Systems for Sensor Fusion – HASP/SIAP (1970's)

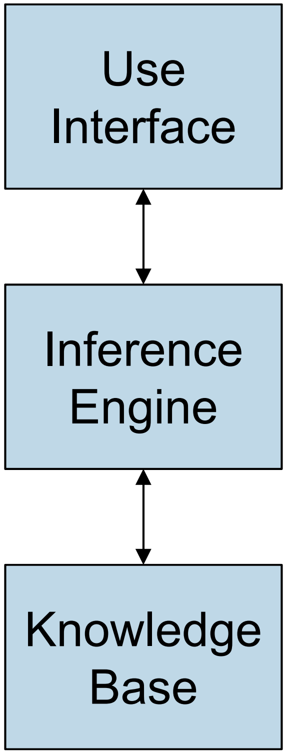
- Signal understanding system
  - Inputs: acoustic signals from hydrophones
  - Outputs: detection, location and type of vessel
- Rule-based system
  - Hierarchy of rules for signal to symbol transformation
  - Inference by knowledge sources responsible for different levels of processing



H. P. Nii, E. A. Feigenbaum, J. J. Anton, and A. J. Rockmore, "Signal-to-symbol transformation: HASP/SIAP case study," *AI Magazine*, 1982

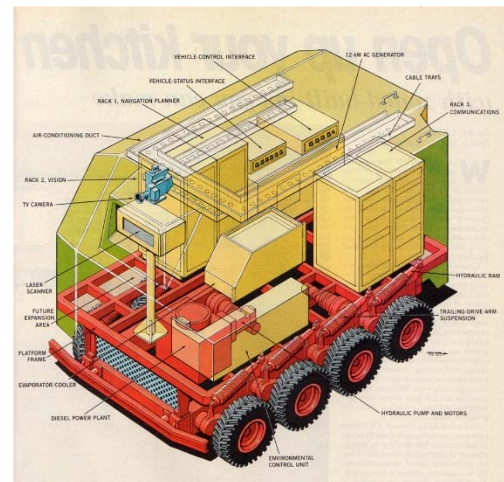
Issues with Expert Systems for Sensor Data Fusion

- Knowledge acquisition
  - Finding experts that can articulate their reasoning; experts with good intuition are not suitable
  - Extracting knowledge from experts
- Knowledge representation
  - Consistency and completeness of rules
  - Representation of uncertainty
- Inference engine
  - Control of inference
  - Reasoning with uncertainty
  - Processing speed
- Expert systems (*Scruffies*) have more success in other commercial applications than in sensor data fusion



AI in 1980’s

- 1981 – Japan started “Fifth Generation Computer” project to build intelligent computers
- United States responded with “Strategic Computing Initiative” with AI as main objective, including Autonomous Land Vehicle (ALV)
- ALV followed road in 1985 demo but vision system was very brittle and sensitive to
    - Light and shadow – detect road edge at noon, but not with shadow at dusk
    - Environmental change (like mud left along road by another vehicle)
- Booming AI industry (software, hardware) became a bust with AI winter in late 1980’s



J. E. Franklin, C. L. Carmody, K. Kellter, T. S. Levitt, B. L. Buteau, ”Expert system technology for the military: selected samples,” *IEEE Proceedings*, Oct. 1988.

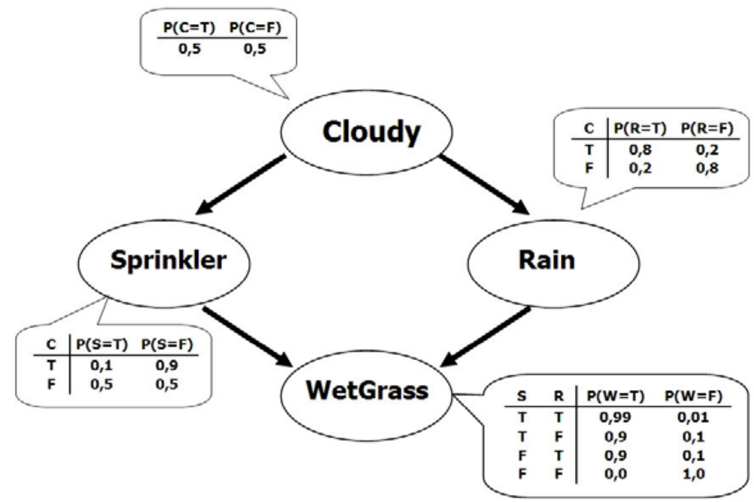
Uncertainty Reasoning in AI

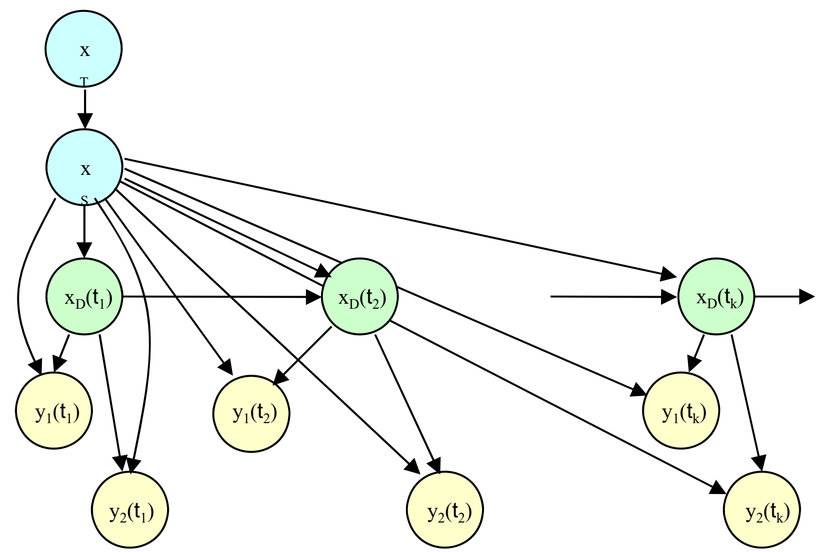
- Sensor data fusion has to deal with uncertainty in evidence (input) and knowledge
  - Drawback of neural network approach, which was also popular in 1980s
  - Recognized very early by expert system developers
- Competing uncertainty reasoning approaches include
  - Rule-based methods
  - Probabilistic reasoning
  - Evidence theory
  - Fuzzy sets
- Probabilistic graphical models became dominant uncertainty reasoning approach in AI
  - Graphical model
  - Distributed inference

J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

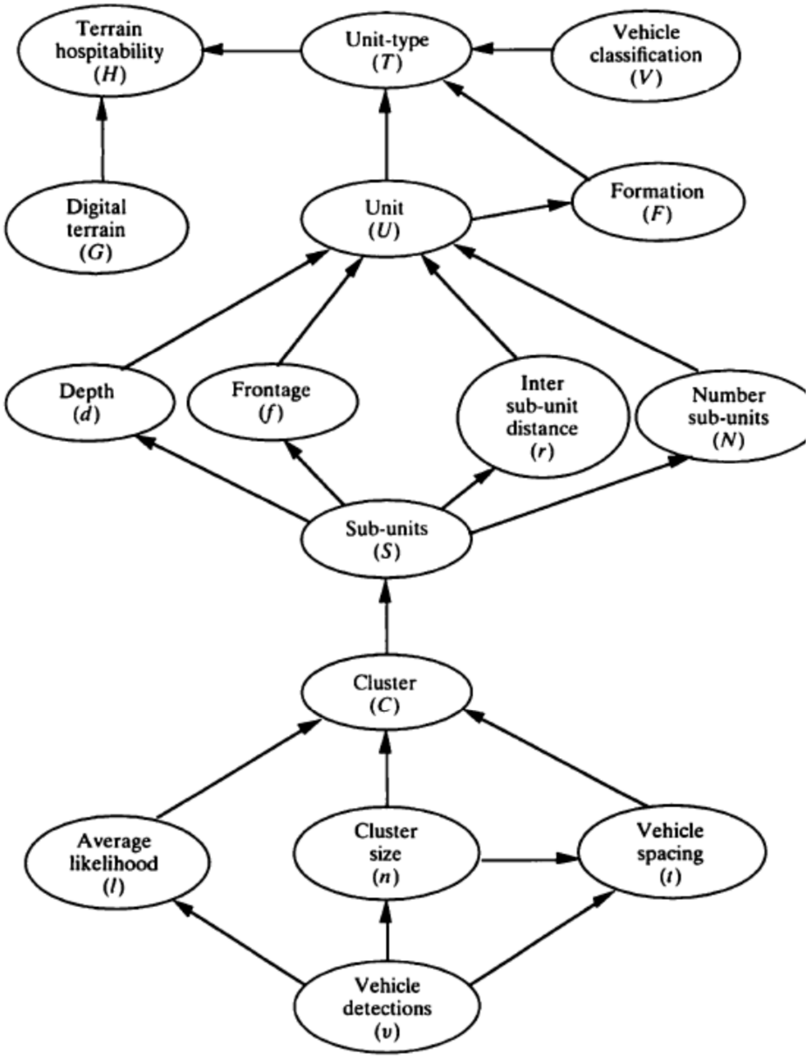
Probabilistic Reasoning/Graphical Models

- Probability model expressed graphically as networks
  - Nodes are random variables
  - Weights on edges represent conditional probabilities
- Inference computes conditional probabilities given evidence
  - Node elimination
  - Junction tree
  - Markov chain Monte Carlo
- Very natural for researchers with background in estimation theory
- Considered AI because of separation into knowledge and automatic inference





Military Unit Detection from Synthetic Aperture Radar (SAR) Imagery

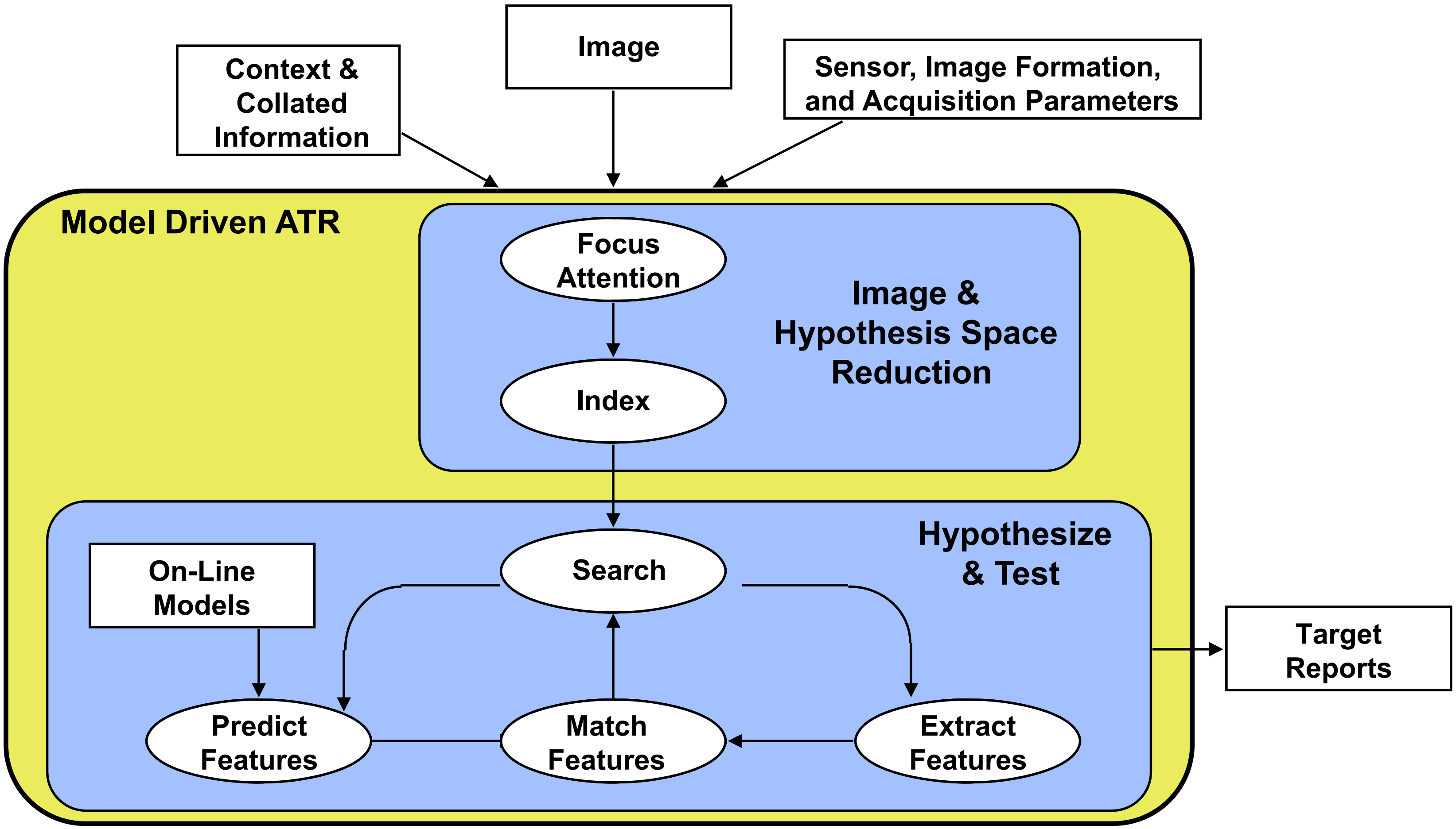


Joint probability of all IES/BTI random variables =
$p(T, U, H, G, F, V, d, f, r, N, S, C, l, n, t, v)$ = product of the terms:

	Probability semantics	Processing component
$p(T \mid U, F, V)$	/force type	Inference
$p(V)$	/vehicle classification	Vehicle classification
$p(H \mid T, G)$	/terrain hospitability	Terrain hospitability
$p(G)$	/digital terrain data	Terrain database constructor
$p(F \mid U)$	/formation	Formation matching
$p(U \mid d, f, r, N)$	/unit existence, location	Inference
$p(d \mid S)$	/formation depth	Formation matching
$p(f \mid S)$	/formation frontage	Formation matching
$p(r \mid S)$	/sub-unit spacing	Formation matching
$p(N \mid S)$	/number sub-units	Formation matching
$p(S \mid C)$	/sub-units' existence, location	Inference
$p(C \mid l, t, n)$	/cluster existence	Detection clustering
$p(l \mid v)$	/mean likelihood	Detection clustering
$p(t \mid v, n)$	/detection spacing	Detection clustering
$p(n \mid v)$	/number detections	Detection clustering
$p(v)$	/vehicle existence	Vehicle detection

T. S. Levitt, C. L. Winter, C. J.Turner, R. A. Chestek, G. J. Ettinger, and S.M.Sayre, "Bayesian inference-based fusion of radar imagery, military forces and tactical terrain models in the image exploitation system/balanced technology initiative," *Int. J. Human Computer Studies*, 1995

Model-Based Object Recognition



J. Diemunsch and J. Wissinger, “Moving and stationary target acquisition and recognition (MSTAR) model-based automatic target recognition: search technology for a robust ATR", *SPIE* 1998.

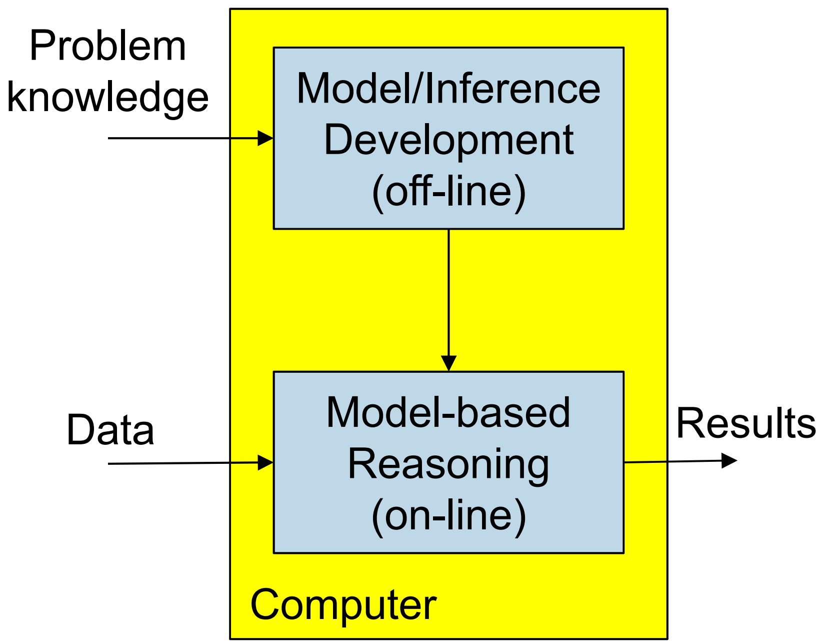
AI for Sensor Data Fusion from 1990’s to 2010

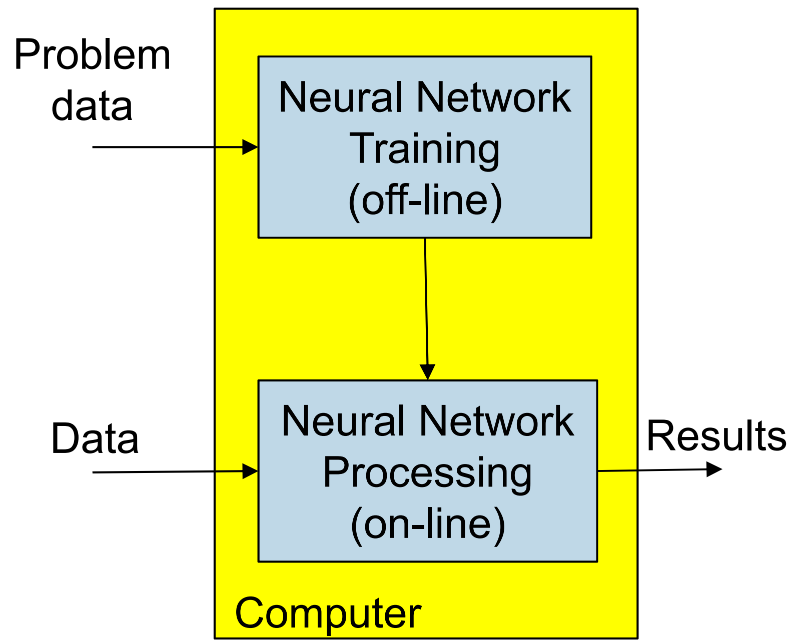
- Model-based reasoning by probabilistic graphical models becomes popular
  - Rigorous representation of uncertainty and inference from evidence
  - Model-based approach is explainable
  - Models can be learned from data
  - Can be extended to handle logical relationships
- However, successful implementation requires good models
  - Manageable for problems such as force structure analysis, intent assessment
  - Difficult for object recognition and speech recognition
- Meanwhile, neural networks were used for many low level functions where modeling is difficult and training is easy
- Then computers become more powerful and massive amounts of data are available

AI/Machine Learning is Popular and Deep Learning is Main Approach

- AI has recent spectacular successes defeating humans
  - Deep Blue beat world champion in chess (1997)
  - Watson won US Jeopardy quiz show (2011)
  - AlphaGo beat world's top player (Ke Jie) in go (2017)
- Deep learning (deep neural networks) has been successfully used in
  - Speech recognition
  - Computer vision
  - Cyber security (where modeling is very difficult and data is plentiful)
- Entry cost to using deep learning is very low
  - Does not require deep domain knowledge
  - Open source software, e.g., Caffe, TensorFlow, Theano
  - Public domain data, e.g., ImageNet, Open Images Dataset
  - Inexpensive powerful hardware, e.g., Nvidia, Intel, Google

Shift from Model-Based to Learning (Function)-Based AI

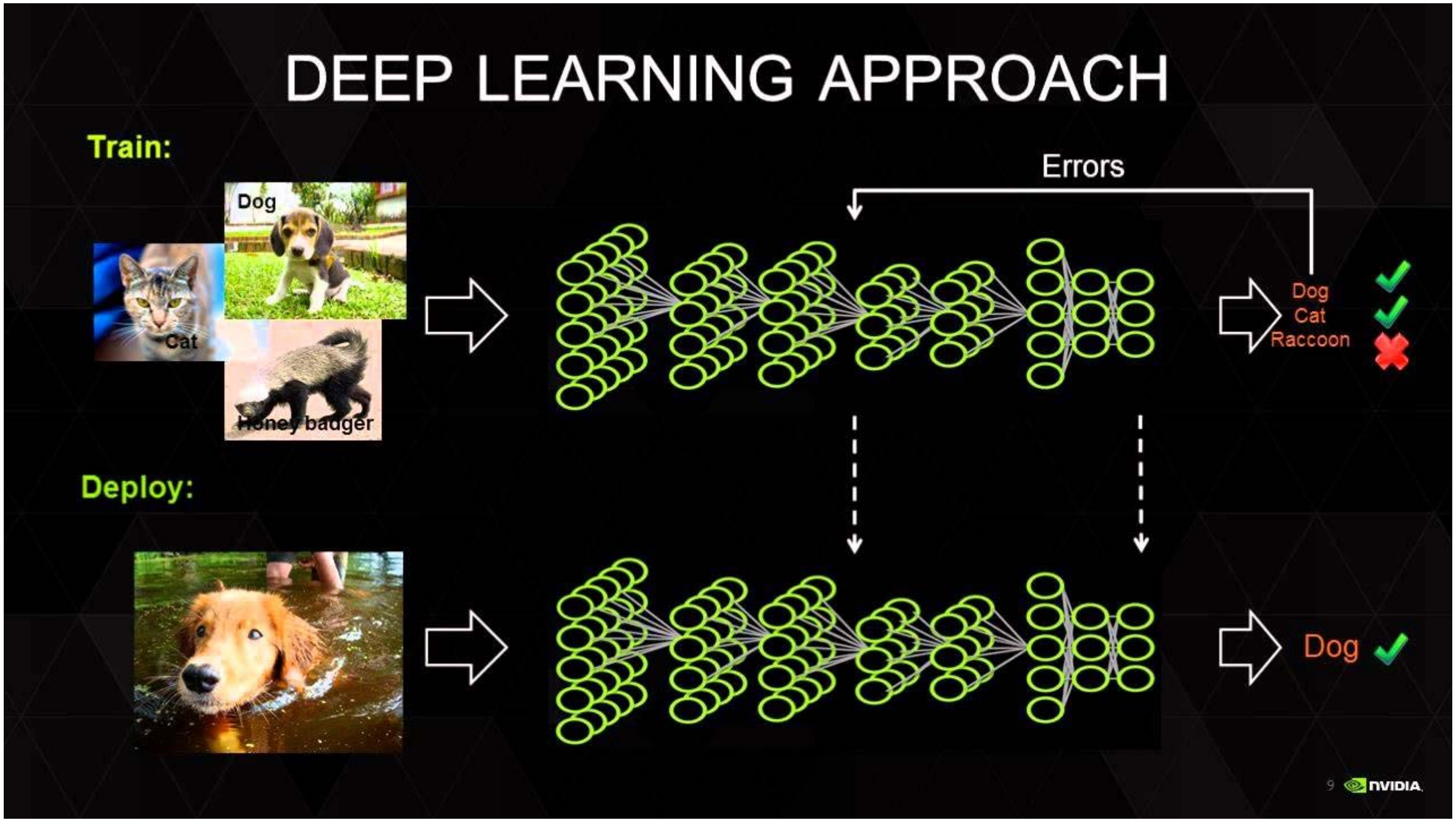




- Explicit problem knowledge
- Represent and reason
- Transparent fusion processing

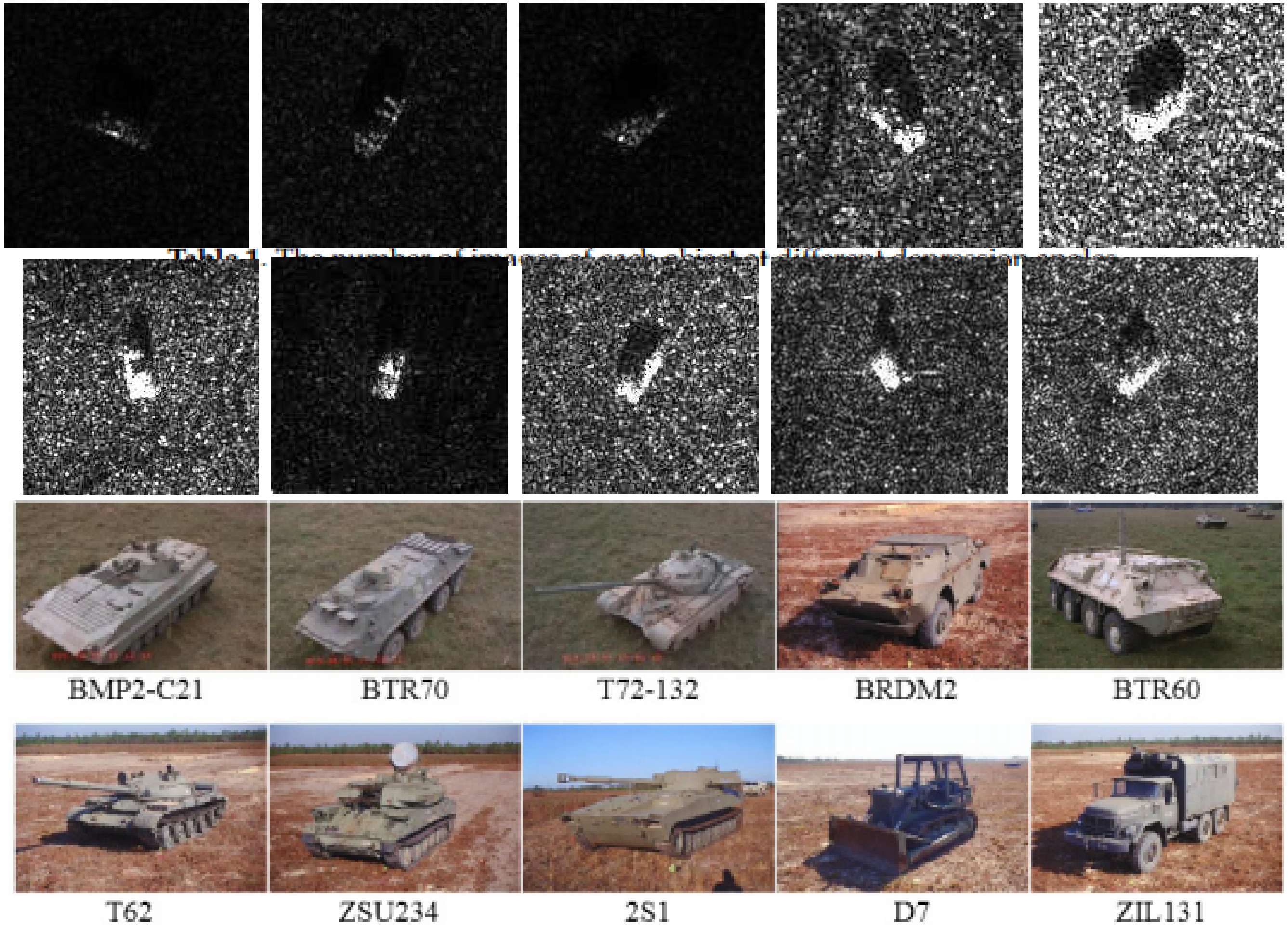
- Problem knowledge captured by data
- Train neural network from data
- Black box fusion processing

Deep Learning (Deep Neural Network)



- Deep neural network uses multiple hidden layers between input and output layers to model complex nonlinear relationships
- Input layers can be images or audio signals instead of features

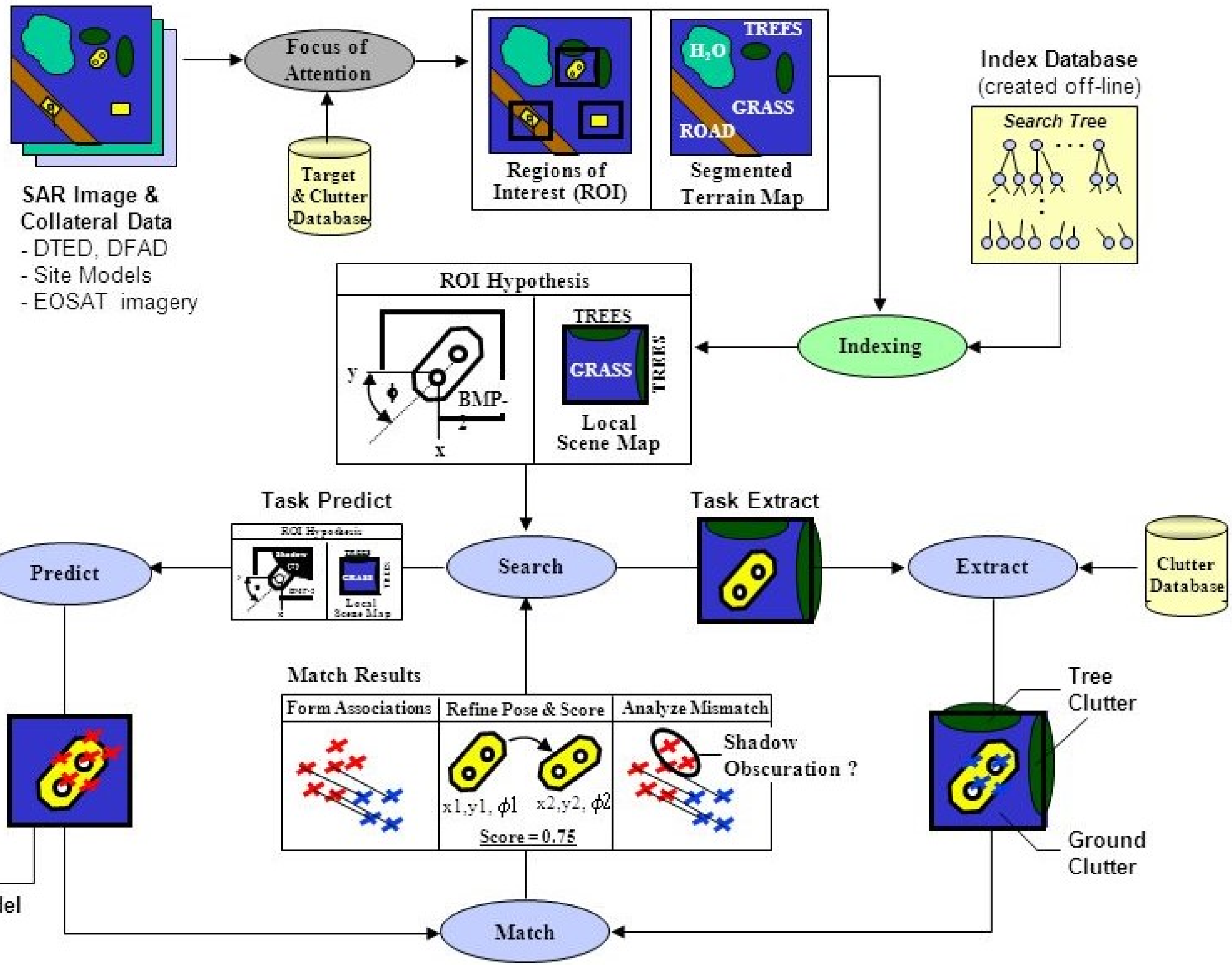
MSTAR Target Types and SAR Images



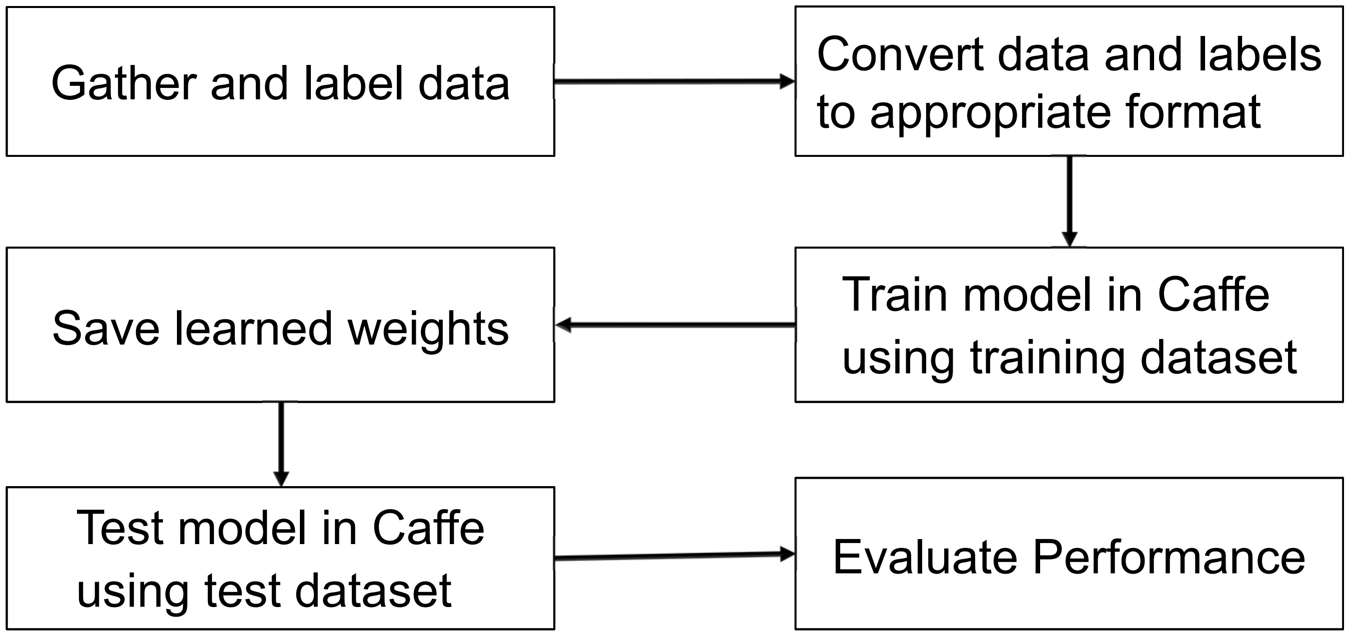
U. K. Majumder, “Deep Learning in AI and Information Fusion” *SPIE Panel Discussion*, 2018, Orlando, FL

Moving and Stationary Target Acquisition and Recognition (MSTAR) – Model-Based Approach

- Program manager said “neural network is not allowed”



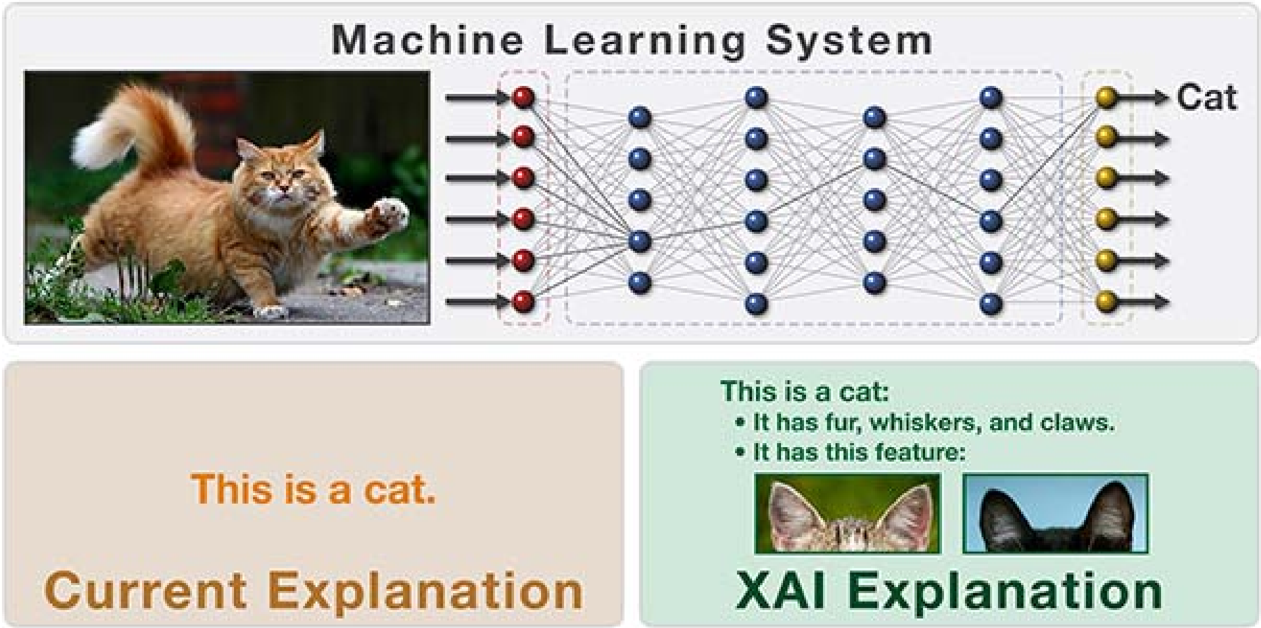
Deep Learning for MSTAR Using Caffe



- Deep learning MSTAR is much easier to develop than model-based MSTAR
- Performance is also very good (99% accuracy) for test data
- However, it is difficult to modify the network for other operating conditions with camouflage and deception

Challenges of Using Deep Learning for Sensor Data Fusion

- Performance is only as good as data
  - Large amounts of data are needed
  - Training data for rare events are sparse
  - Easily fooled by opposition data



DARPA Explaniable AI (XAI) Problem

- Lack rigorous representation and propagation of uncertainty
- Neural networks do not provide explanation for results
- Thus deep learning is popular in commercial application but seldom used in mission-critical systems

Why Explainable Algorithms are Important

- Assumptions provide operating conditions for algorithm
- Reasoning/processing approach allows generalization to other operating conditions
- Explainable algorithm model supports prediction of performance, e.g., error bounds from Kalman filter equations
- Fusion by other nodes requires models, e.g., 2D video tracks with other 2D video tracks
- Critical decisions have to be explained, e.g., right to explanation in EU General Data Protection Regulation (GDPR)
- Good engineering requires design review, not possible without explainable components

Robust Sensor Data Fusion

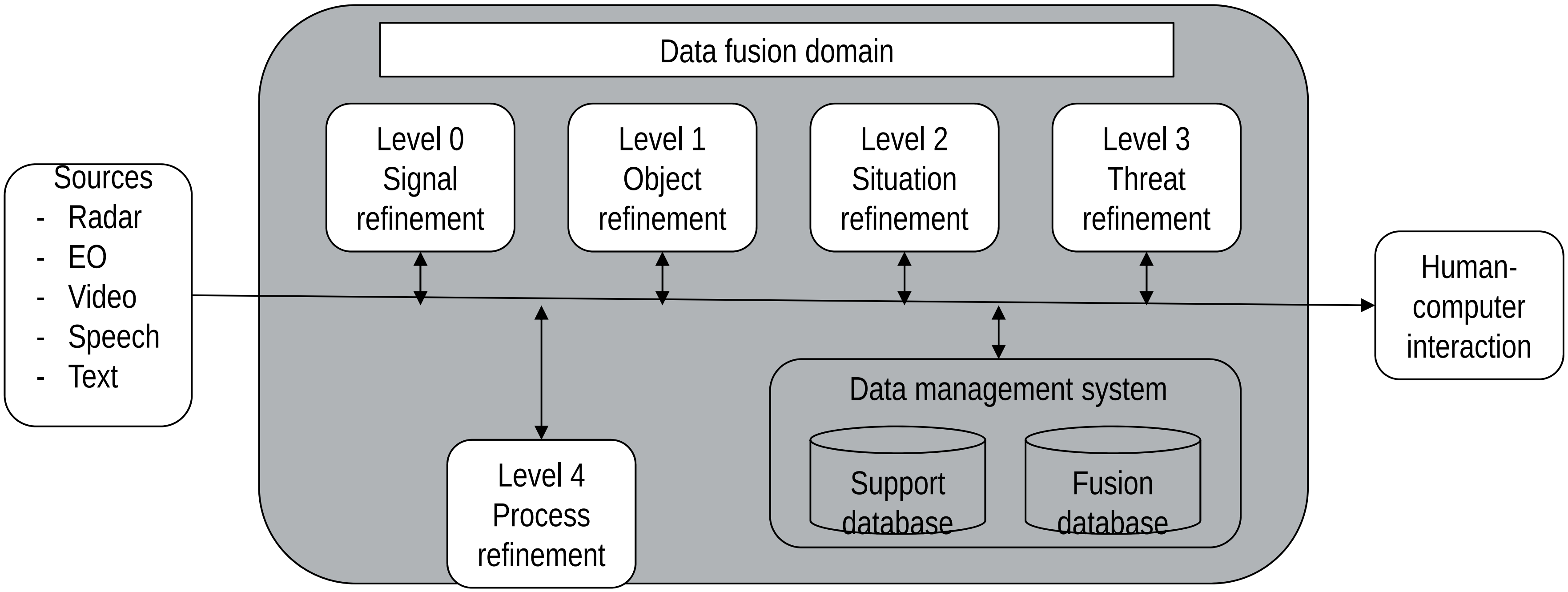
- “Robust” is hard to define precisely, but is opposite to “brittle”
- A trusted system has to be robust
- Properties of robust sensor data fusion
  - Well-defined operating conditions
  - Predictable performance
  - Insensitivity to small changes
  - Graceful performance degradation within operating condition
  - Adaptability to context
- Robust sensor data fusion requires components that are either
  - Explainable with models
  - Tested and statistically characterized with large amounts of data

Proposed Approach for Integrating Deep Learning and Model-Based Reasoning

- Map processing approach to required ability
  - Deep learning for low-level functions such as perception (animal-like abilities*)
  - Model-based reasoning for high level situation assessment (human-level intelligence*)
- Design rule
  - Model rich, data rich – model-based reasoning
  - Model rich, data poor – model-based reasoning
  - Model poor, data rich – deep learning
  - Model poor, data poor – get better sensors

*A. Darwiche, “Human-Level Intelligence or Animal-Like Abilities?” *CACM*, Oct. 2018

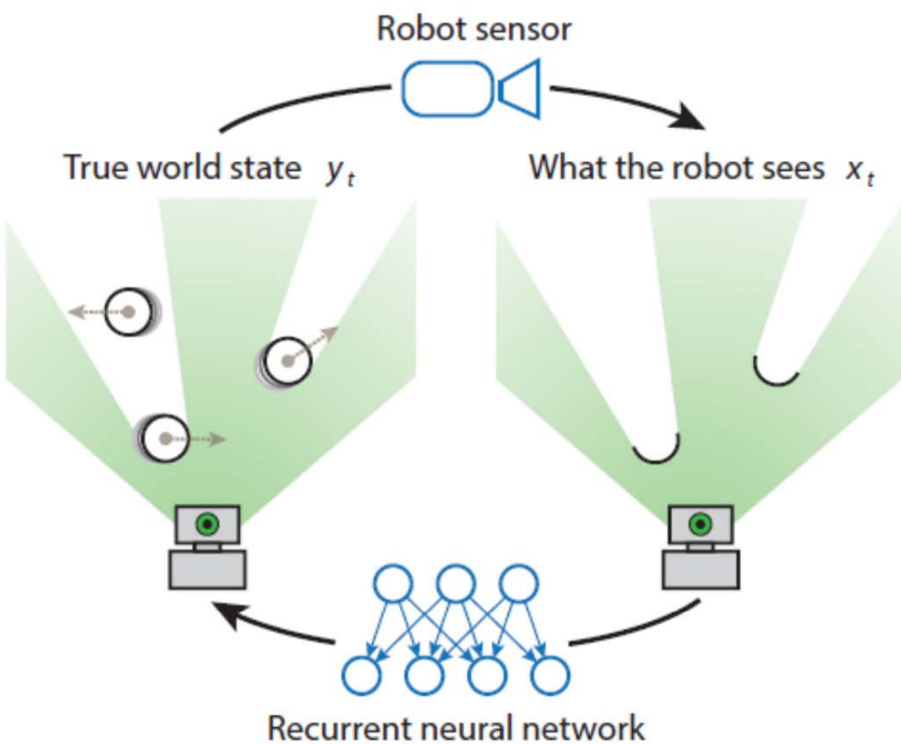
Allocating JDL/DFG* Fusion Levels to Learning and Models

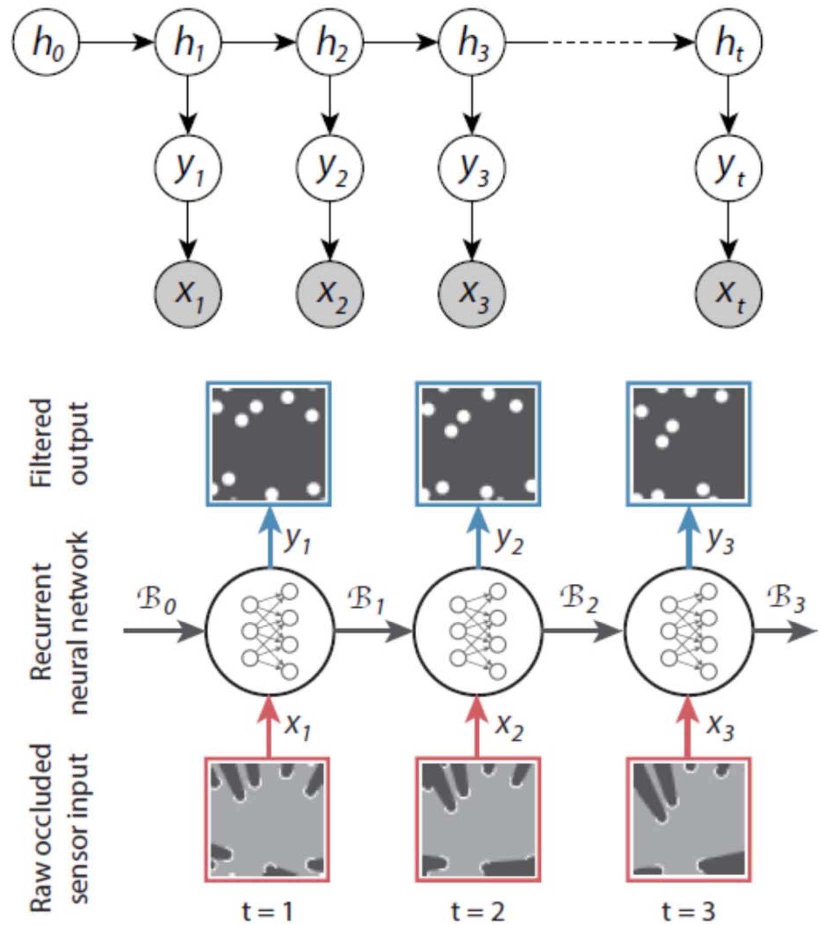


- Level 0: DL is suitable for speech recognition, feature extraction
- Level 1: DL can recognize objects/person and skip level 0; tracking can use both DL and model-based reasoning
- Level 2 and above: model-based reasoning (users want explanations)

*JDL/DFG – Joint Director of Laboratories/Data Fusion Group

Direct Learning of Sensor Input to Object Tracks

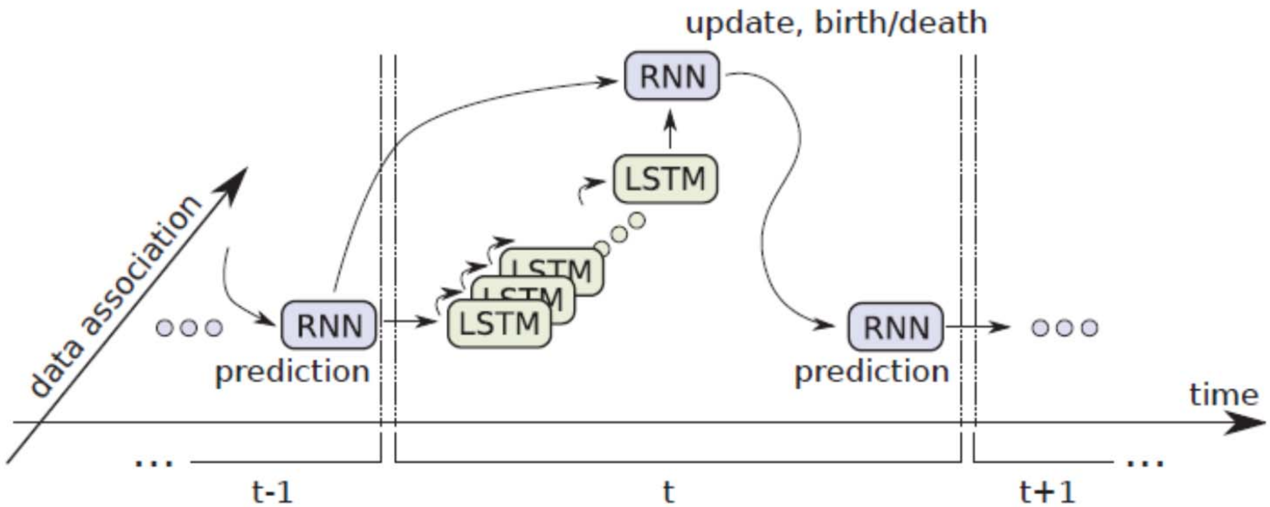




- Related to PHD
- Training and performance issues

P. Ondruska and I. Posner, “Deep tracking: Seeing beyond seeing using recurrent neural networks,” *AAAI Conference*, 2016

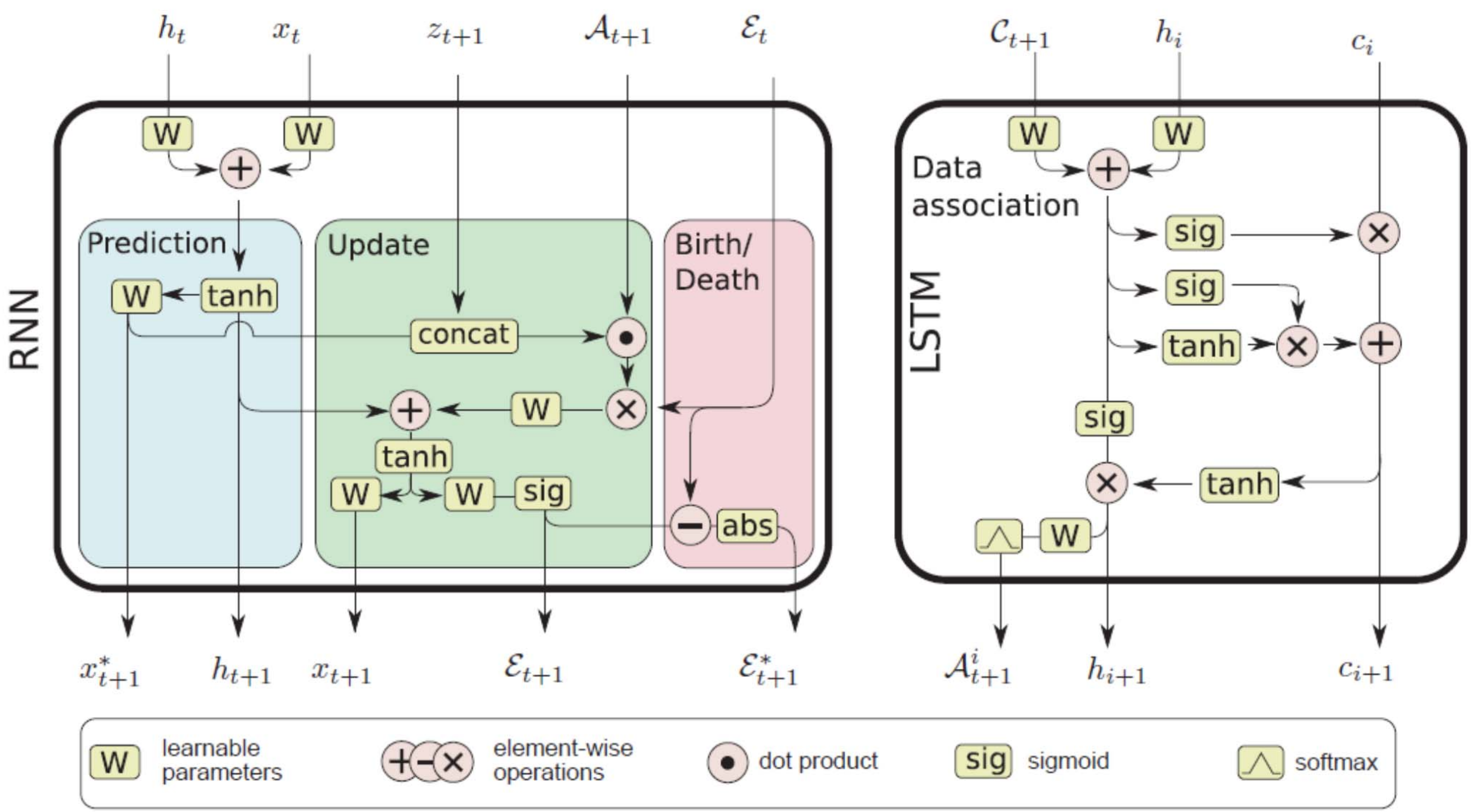
Multiple RNNs for Online Multiple Object Tracking



- RNN for prediction and update, target birth and death
  - Completely model free
  - Handles arbitrary dynamics, distributions
- LSTM for data association
  - Handles combinatorial problem
  - Controls memory in RNN

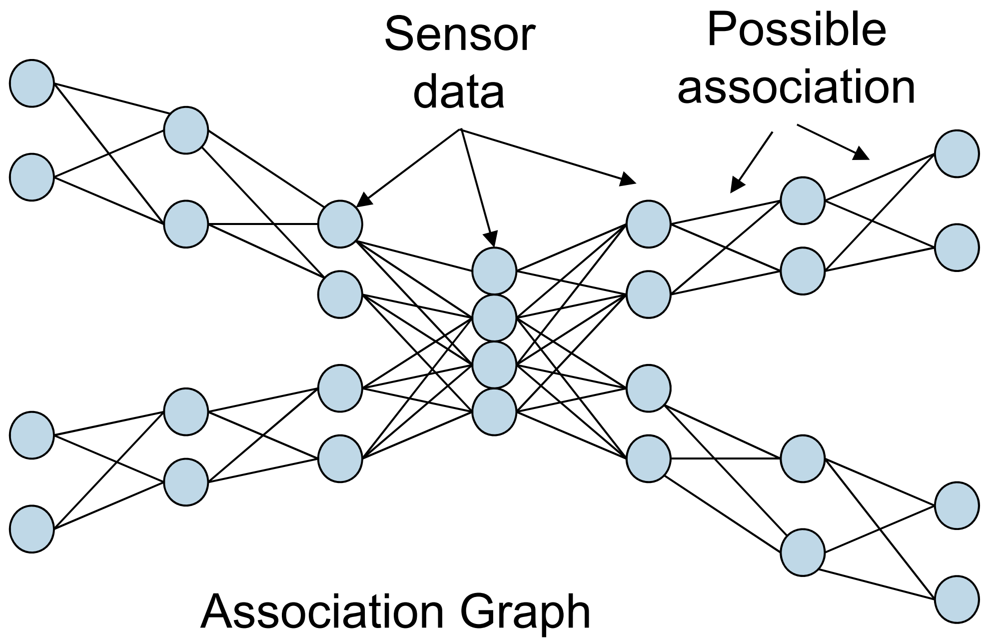
A. Milan et al, "Online multi-target tracking using recurrent neural networks," *AAAI Conf. 2017*

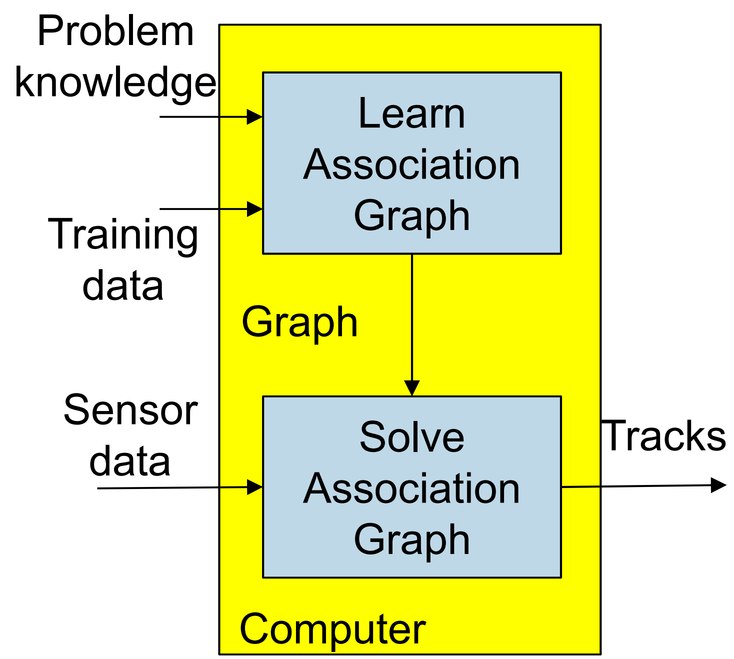
RNN and LSTM for Multiple Object Tracking



- Model-free approach is hard to explain and adapt to different situations

Feature-Aided Tracking using Deep Learning and Models





- Possible associations are found by deep learning
- Association costs are computed using models (object dynamics)
- Target tracks are computed by solving optimization problem on association graph

C. Y. Chong, "Graph approaches for data association," *Fusion 2012*.

Conclusions

- Deep learning performs well in some essential fusion functions (e.g., speech recognition, image recognition) but not suitable for other functions
- Model-based reasoning is needed for high-level fusion but has limited success with unstructured high-dimensional data, e.g., image, speech
- Robust fusion systems should use
  - Models when they are available and can be represented for reasoning
  - Deep learning when models are weak and data are available
- Successful use of deep learning in sensor data fusion systems requires
  - Understanding capabilities and limits of function-based approaches
  - Characterizing deep learning algorithms (assumptions, operating conditions, and performance) for use by other fusion functions