

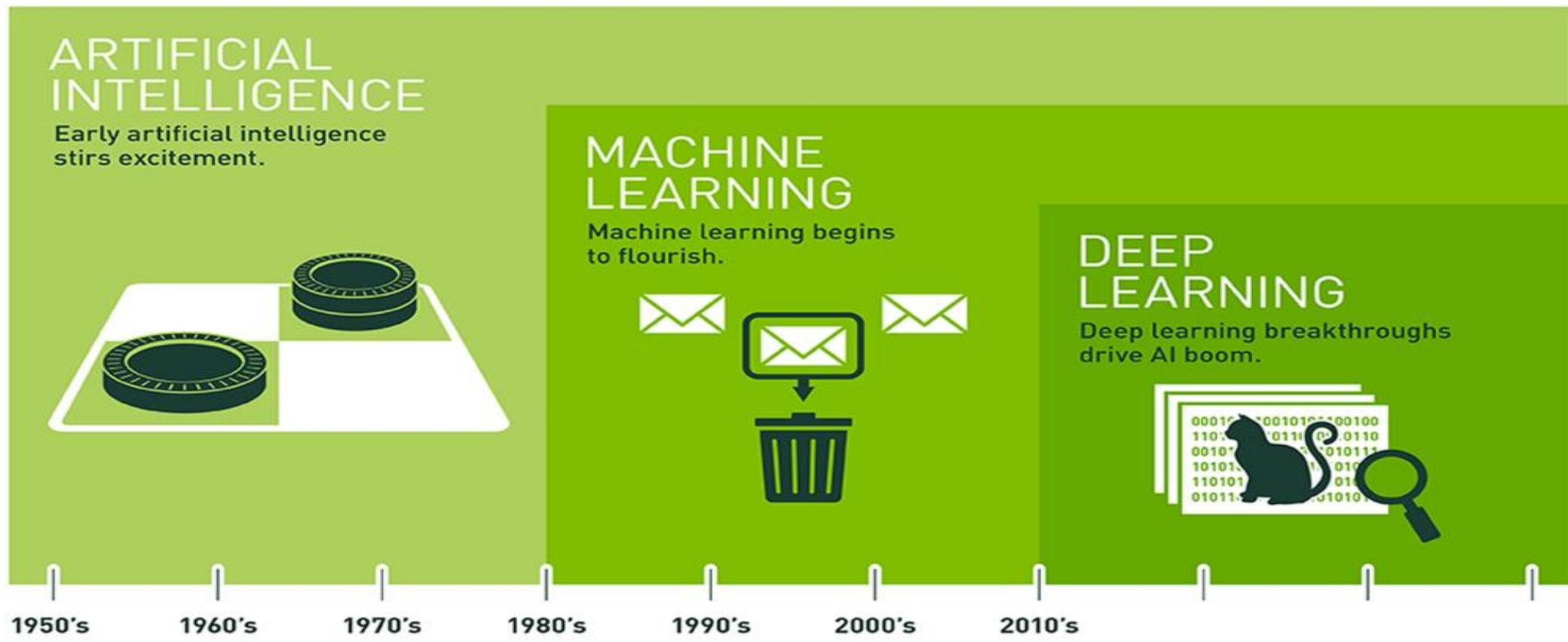


DEEP
LEARNING
INSTITUTE

DEEP LEARNING DEMYSTIFIED

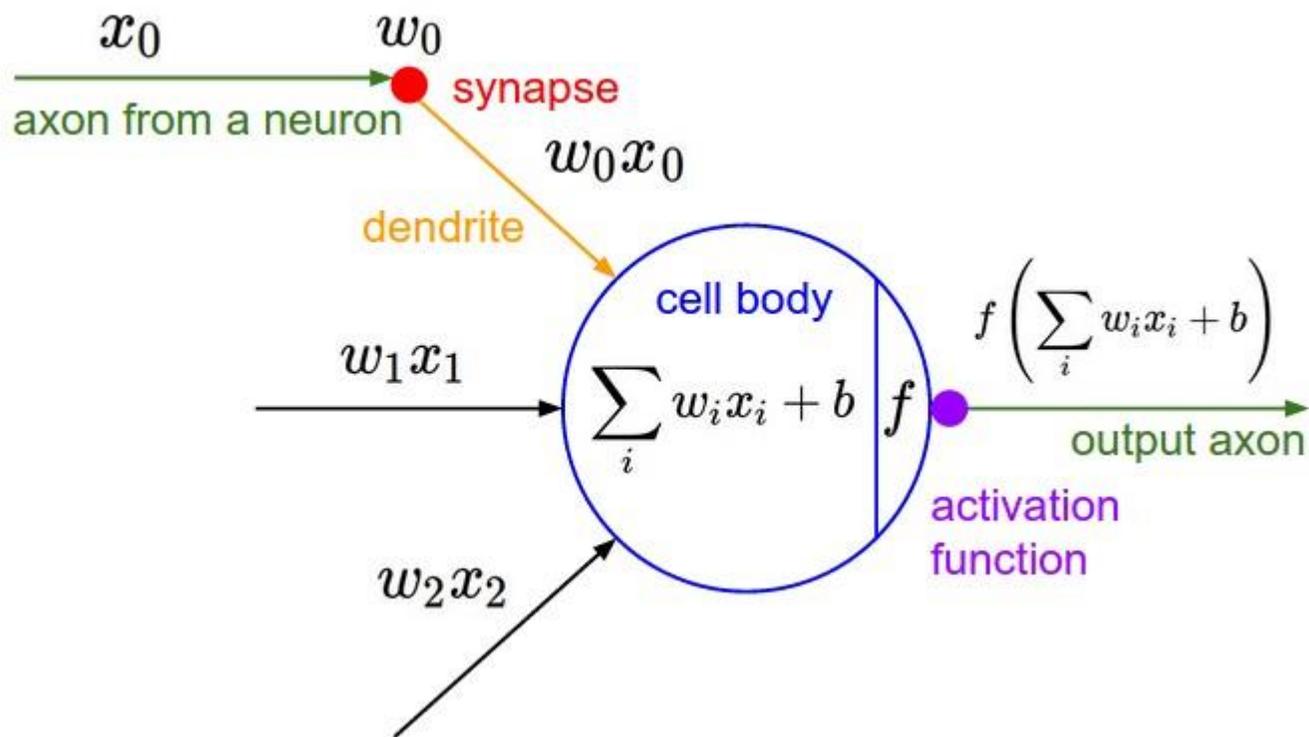
Frédéric Parienté
NVIDIA Corporation

DEFINITIONS



PERCEPTRON

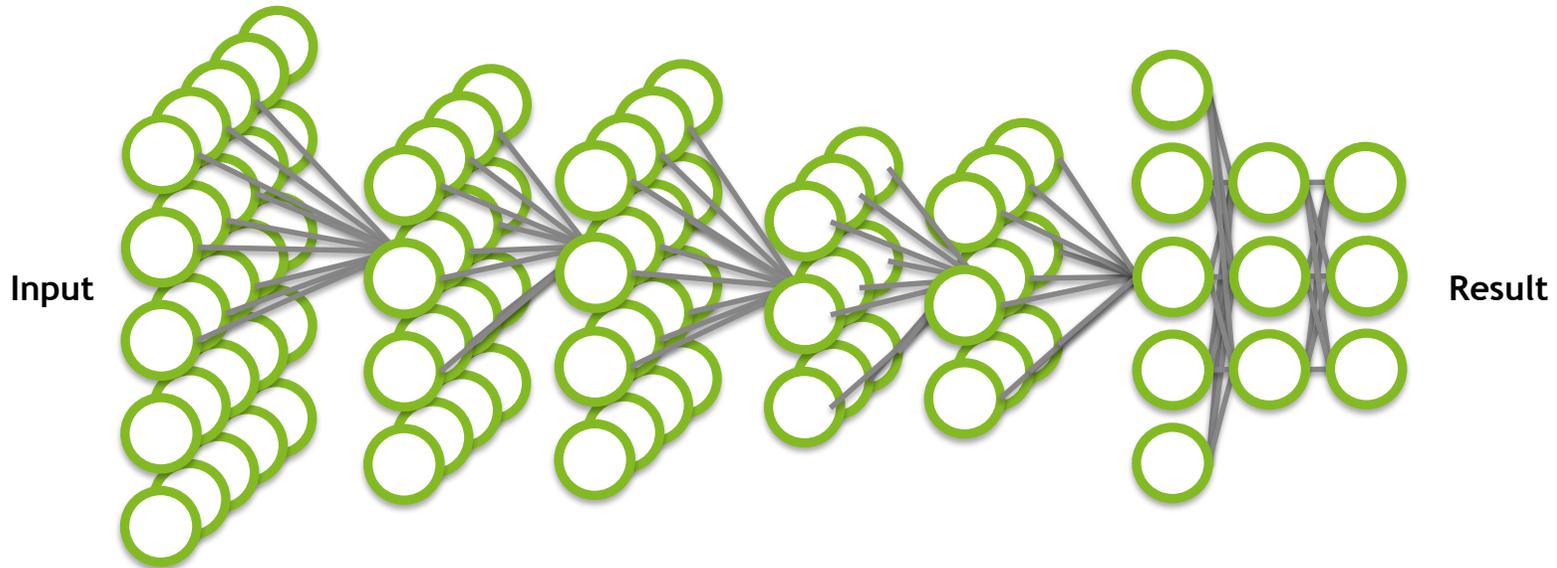
Rosenblatt, 1958



Linear Classifier

- Input
- Weight
- Linear combination
- Bias
- Activation function σ
- Output

DEEP NEURAL NETWORK (DNN)



HOW DL CAN BE APPLIED

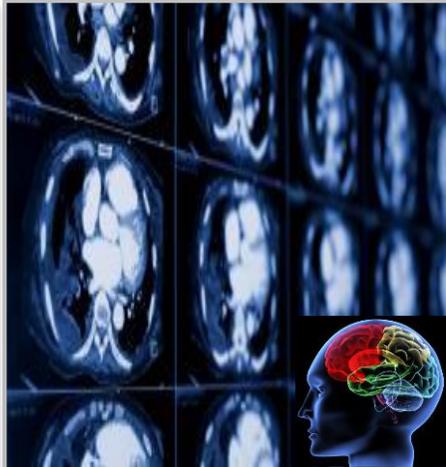
INPUTS		BUSINESS QUESTION	AI/DL TASK	EXAMPLES IN HEALTHCARE	EXAMPLES IN RETAIL	EXAMPLES IN FINANCE
 Text Data	 Images	Is “it” <u>present</u> or not?	Detection	Cancer Detection	Targeted ads	Cybersecurity
		What <u>type</u> of thing is “it”?	Classification	Image Classification	Basket Analysis	Credit Scoring
 Video	 Audio	To what <u>extent</u> is “it” present?	Segmentation	Tumor Size/Shape Analysis	Build 360° Customer View	Credit Risk Analysis
		What is the likely outcome?	Prediction	Survivability Prediction	Sentiment & behavior recognition	Fraud Detection
		What will satisfy the objective?	Recommendations	Therapy Recommendation	Recommendation Engine	Algorithmic Trading
		What is the speaker saying?	Natural Language Processing	Expert diagnosis	Virtual personal assistants	Robo Advisors

DEEP LEARNING IS SWEEPING ACROSS INDUSTRIES

Internet Services



Medicine



Media & Entertainment



Security & Defense



Autonomous Machines



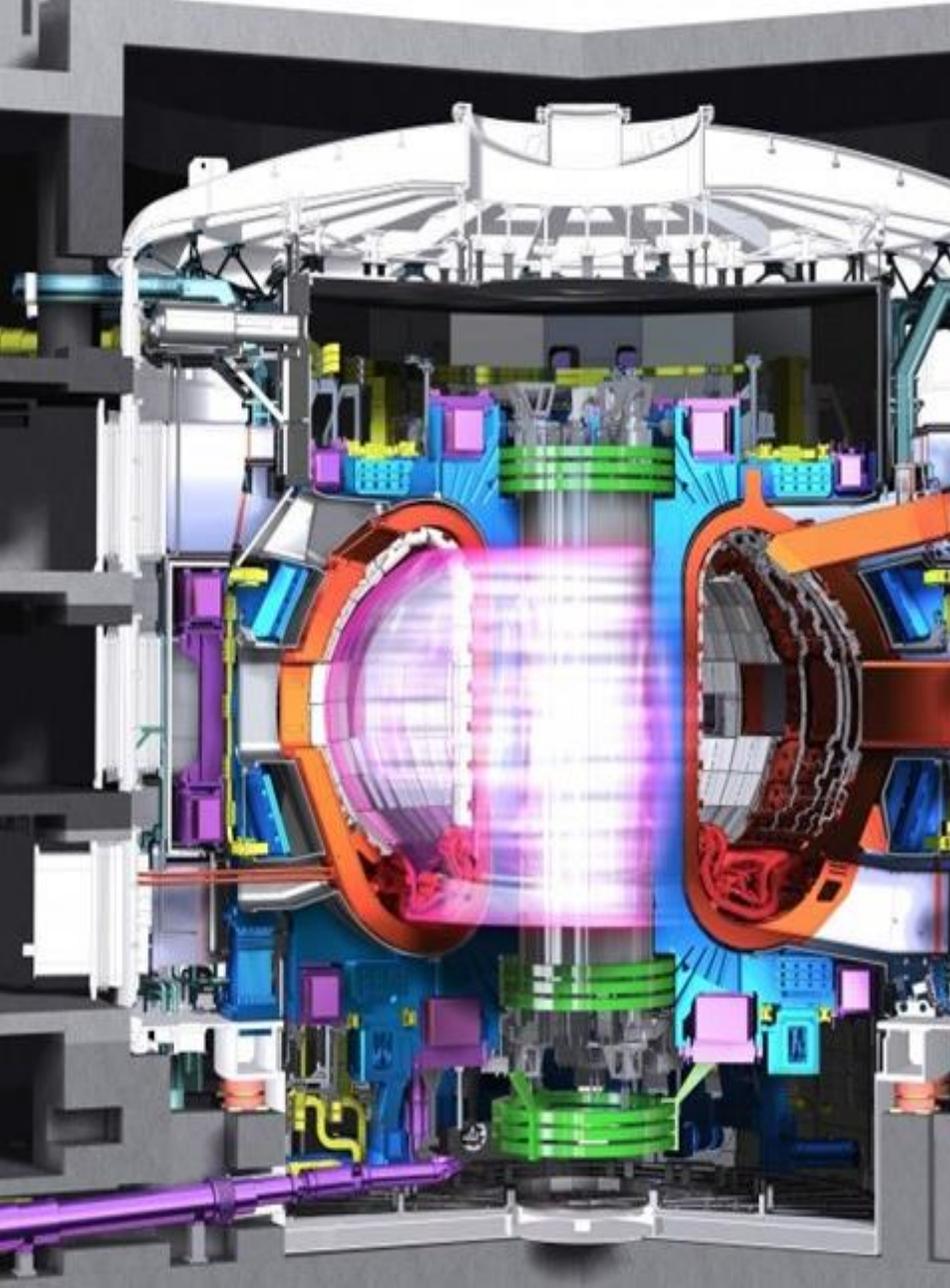
- Image/Video classification
- Speech recognition
- Natural language processing

- Cancer cell detection
- Diabetic grading
- Drug discovery

- Video captioning
- Content based search
- Real time translation

- Face recognition
- Video surveillance
- Cyber security

- Pedestrian detection
- Lane tracking
- Recognize traffic signs



Predicting Disruptions in Fusion Reactor using DL

Background

Grand challenge of fusion energy offers mankind changing opportunity to provide clean, safe energy for millions of years. ITER is a \$25B international investment in a fusion reactor.

Challenge

Fusion is highly sensitive, any disruption to conditions can cause reaction to stop suddenly. Challenge is to predict when a disruption will occur to prevent damage to ITER and to steer the reaction to continue to produce power. Traditional simulation and ML approaches were 65% to 85% accurate with 5% false alarm rate

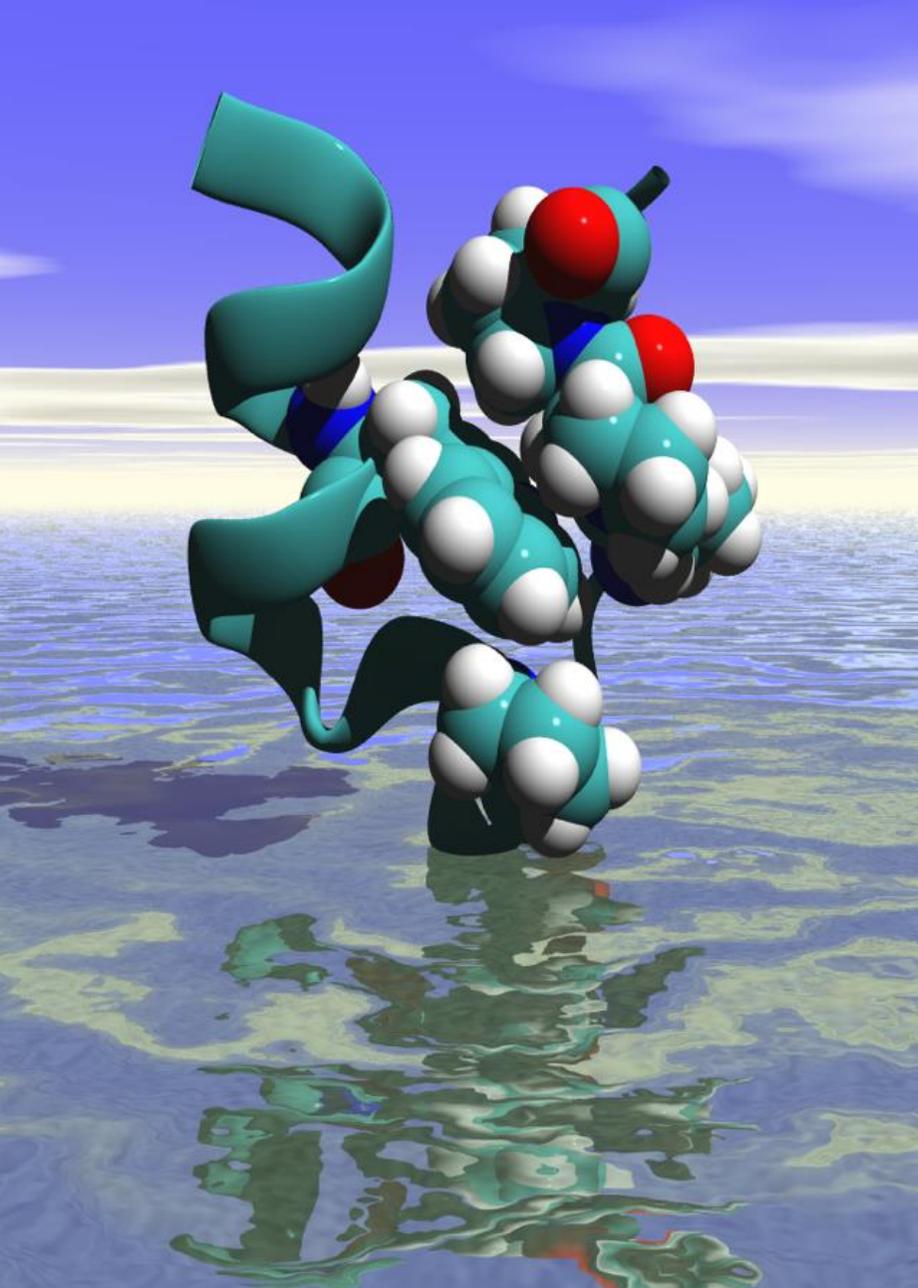
Solution

DL network called FRNN using Theano exceeds today's best accuracy results. It scales to 200 Tesla K20s, and with more GPUs, can deliver higher accuracy. Current level of accuracy is **95% prediction with 5% false alarm rate**.

Impact

Vision is to operate ITER with FRNN, operating and steering experiments in real-time to minimize damage and down-time.

http://www.ncic.ac.cn/codesign/codesign_ppt/W_Tang_HPC_China_2016_ML_Disruption%20Talk.pdf



AI Quantum Breakthrough

Background

Developing a new drug costs \$2.5B and takes 10-15 years. Quantum chemistry (QC) simulations are important to accurately screen millions of potential drugs to a few most promising drug candidates.

Challenge

QC simulation is computationally expensive so researchers use approximations, compromising on accuracy. To screen 10M drug candidates, it takes 5 years to compute on CPUs.

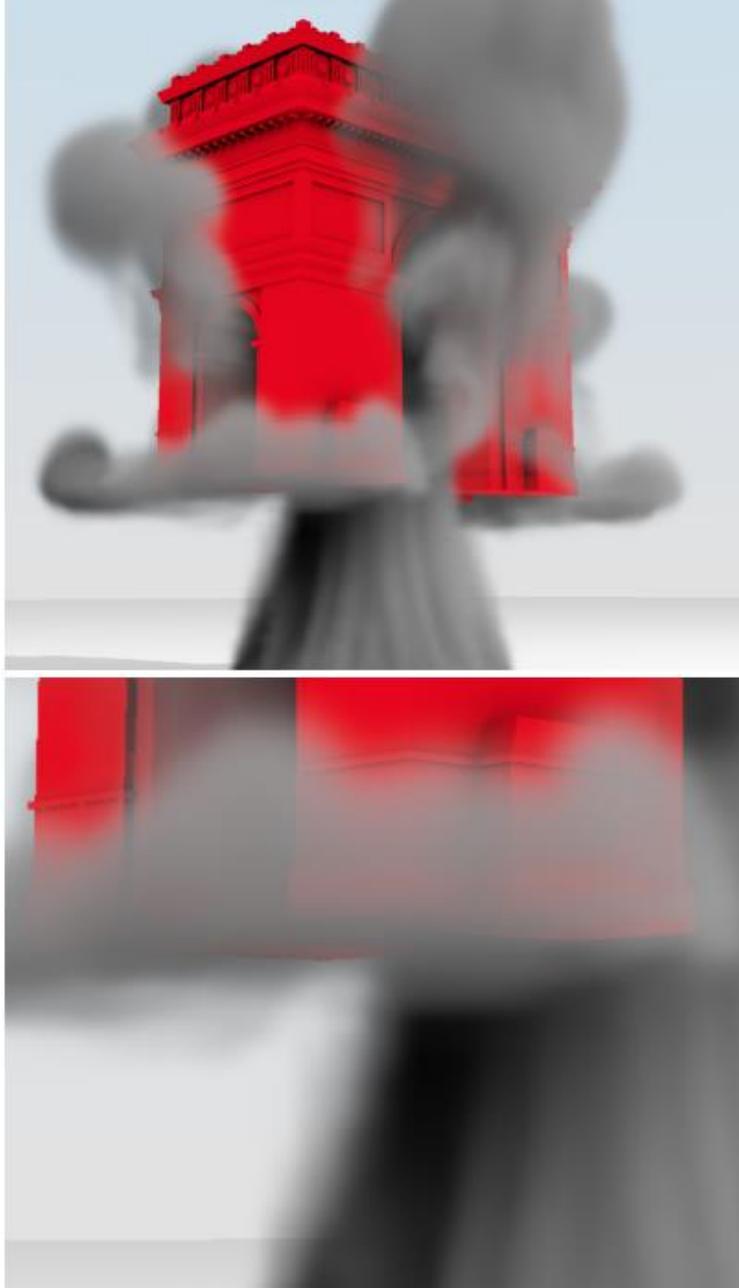
Solution

Researchers at the University of Florida and the University of North Carolina leveraged GPU deep learning to develop ANAKIN-ME, to reproduce molecular energy surfaces with super speed (microseconds versus several minutes), extremely high (DFT) accuracy, and at 1-10/millionths of the cost of current computational methods.

Impact

Faster, more accurate screening at far lower cost

<https://www.ncbi.nlm.nih.gov/pubmed/28507695>



Real-Time Simulation of Fluid Flow

Background

Eulerian methods are able to produce accurate results simulating fluids like water with high compute costs. The most demanding portion of this method is the “pressure projection” step, which satisfies the incompressibility constraint. It involves solving the discrete Poisson equation and leads to a well-known sparse, symmetric and positive-definite linear system.

Challenge

Real-time simulation of fluid and smoke is a long standing problem in computer graphics, where state-of-the-art approaches require large compute resources, making real-time applications often impractical.

Solution

Researchers at Google and the New York University (NYU) used CNN to solve the linear system. “Obtained results are realistic and show good generalization properties.”

Impact

Produces very stable divergence-free velocity fields while being **orders of magnitude faster**

<https://arxiv.org/pdf/1607.03597.pdf>



AI-HPC CONVERGENCE: A TAXONOMY

Transformation

HPC + AI couple simulation with live data in real time detection/control system

Experimental/simulated data is used to train a NN that is used to for detection/control of an experiment or clinical delivery system in real time. The NN is improved continuously as new simulated / live data is acquired

Augmentation

HPC + AI combined to improve simulation time to science > orders of magnitude

Experimental/simulated data is used to train a NN that is used to replace all or significant runtime portions of a conventional simulation. The NN is improved continuously as new simulated / live data is acquired

Modulation

HPC + AI combined to reduce the number of runs needed for a parameter sweep

Experimental/simulated data used to train a NN which steers simulation/experiment b/w runs

The steering NN can be trained continuously as new simulated / live data is acquired

Potential for Breakthroughs in Scientific Insight

THE EXPANDING UNIVERSE OF MODERN AI

"THE BIG BANG"

Big Data
GPU
Algorithms

RESEARCH



CORE TECHNOLOGY / FRAMEWORKS



AI-as-a-PLATFORM



START-UPS



1,000+ AI START-UPS
\$5B IN FUNDING

Source: Venture Scanner

INDUSTRY LEADERS



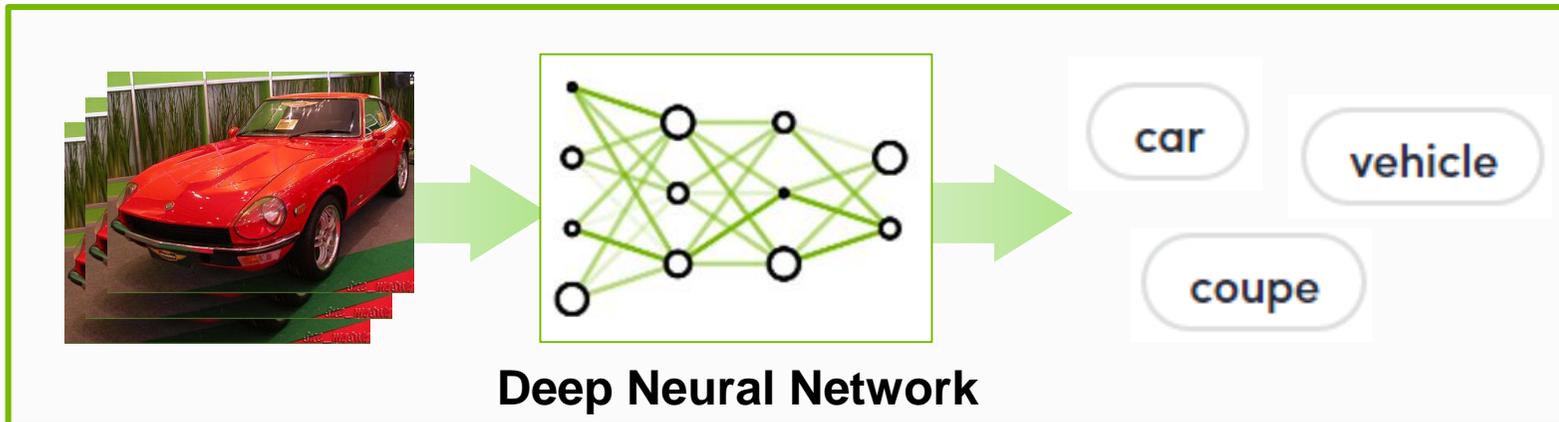
A NEW COMPUTING MODEL

Algorithms that Learn from Examples



Traditional Approach

- Requires domain experts
- Time consuming
- Error prone
- Limited accuracy
- Not scalable to new problems

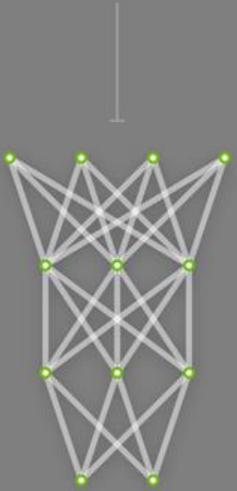


Deep Learning Approach

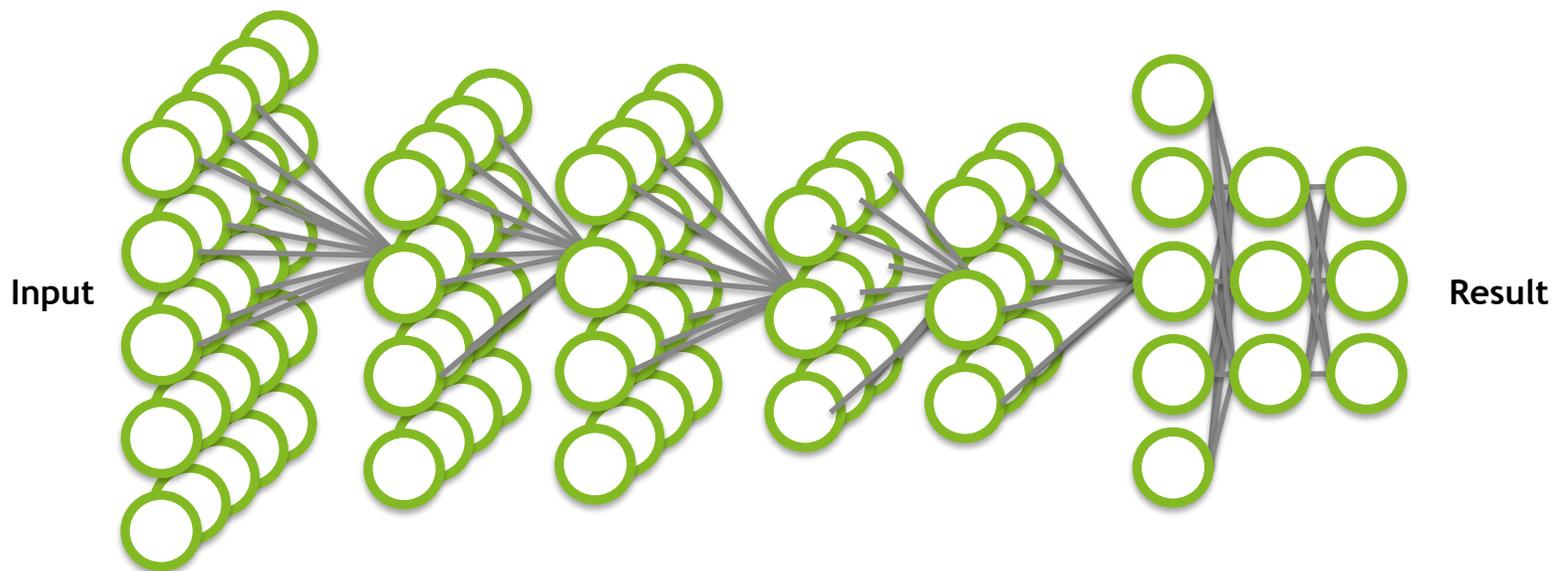
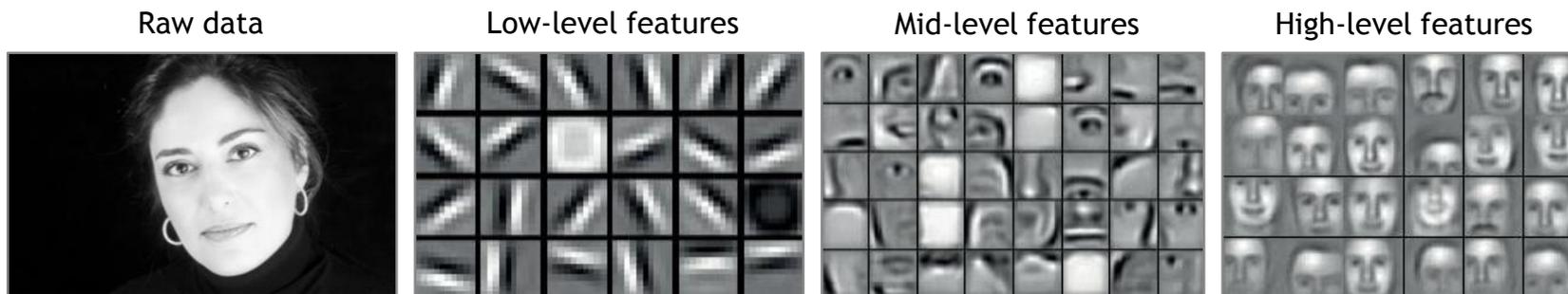
- ✓ Learn from data
- ✓ Easily to extend
- ✓ Speedup with GPUs

DEEP LEARNING

Untrained
Neural Network
Model



DEEP NEURAL NETWORK (DNN)



Application components:

Task objective
· e.g. Identify face

Training data
· 10-100M images

Network architecture
· ~10s-100s of layers
· 1B parameters

Learning algorithm
· ~30 Exaflops
· 1-30 GPU days

CHALLENGES

Deep Learning Needs	Why
Data Scientists	Demand far exceeds supply
Latest Algorithms	Rapidly evolving
Fast Training	Impossible -> Practical
Deployment Platform	Must be available everywhere

NVIDIA DEEP LEARNING INSTITUTE

Hands-on Training for Data Scientists and Software Engineers



Training organizations and individuals to solve challenging problems using Deep Learning

On-site workshops and online courses presented by certified instructors

Covering complete workflows for proven application use cases

Self-driving cars, recommendation engines, medical image classification, intelligent video analytics and more

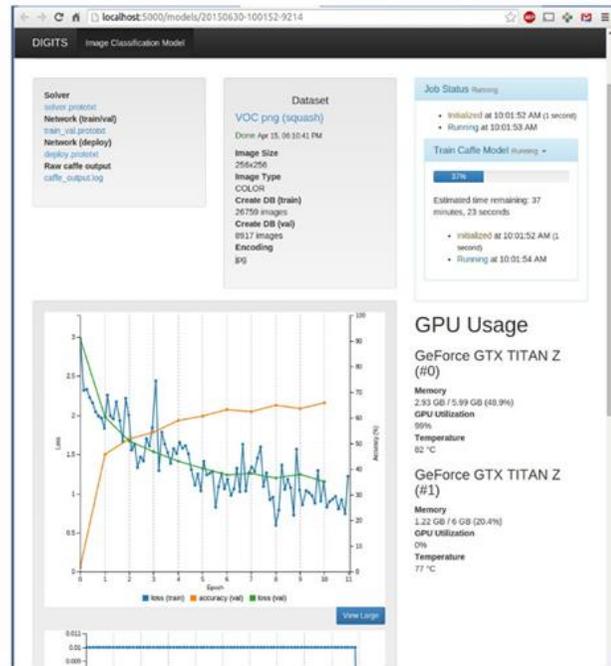
www.nvidia.com/dli

DEEP LEARNING SOFTWARE

NVIDIA DIGITS™

Interactively manage data and train deep learning models for image classification without the need to write code.

[Learn more](#)



Deep Learning Frameworks

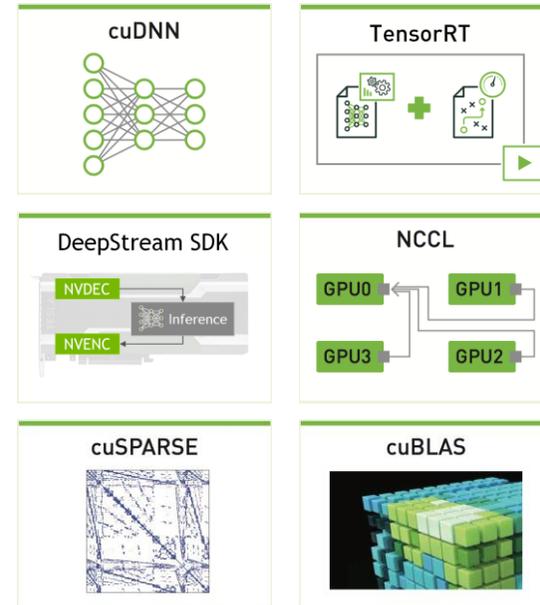
Design and train deep learning models using a high-level interface. Choose a deep learning framework that best suits your needs based on your choice of programming language, platform, and target application.

[Learn more](#)



NVIDIA Deep Learning SDK

This SDK delivers high-performance multi-GPU acceleration and industry-vetted deep learning algorithms, and is designed for easy drop-in acceleration for deep learning frameworks.



END-TO-END PRODUCT FAMILY

TRAINING

INFERENCE

FULLY INTEGRATED DL SUPERCOMPUTER



DGX-1

DESKTOP

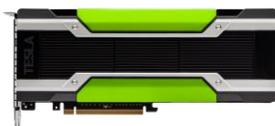


Titan X Pascal

DATACENTER



Tesla P100



Tesla P40

DATACENTER

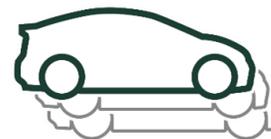


Tesla P4



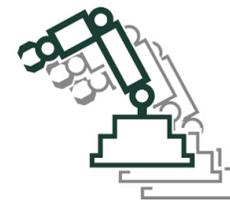
Tesla P40

AUTOMOTIVE



Drive PX2

EMBEDDED



Jetson TX1

CHALLENGES

Deep Learning Needs	NVIDIA Delivers
Data Scientists	DIGITS, DLI Training
Latest Algorithms	DL SDK, GPU-Accelerated Frameworks
Fast Training	DGX, P100, P40, TITAN X
Deployment Platform	TensorRT, P40, P4, Jetson, Drive PX

KEY TAKEAWAYS

AI is everywhere

DL is the key technology for Modern AI

NVIDIA provides the products and solutions powering the revolution

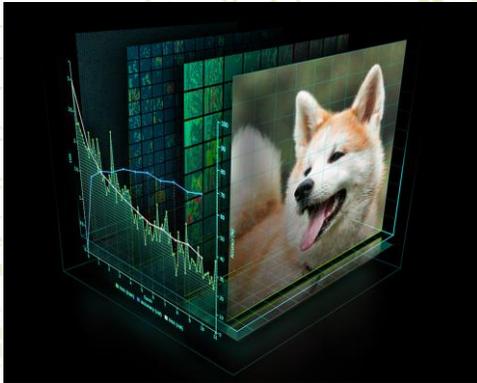
GPU TECHNOLOGY CONFERENCE

Oct 10-12, 2017 | ICC Munich

www.gputechconf.eu

Use promo code **FredericParienteGTCEU17**
for a discount on the conference fee!

Europe's Brightest Minds & Best Ideas



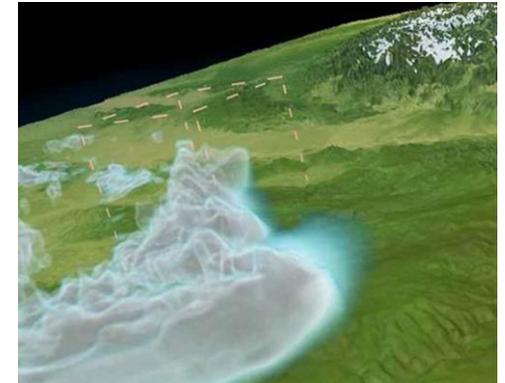
DEEP LEARNING &
ARTIFICIAL INTELLIGENCE



SELF-DRIVING CARS



VIRTUAL REALITY &
AUGMENTED REALITY



SUPERCOMPUTING & HPC

GTC Europe is a three-day conference designed to expose the innovative ways developers, businesses and academics are using artificial intelligence and parallel computing to transform our world.

3 Days | 2000+ Attendees | 50+ Exhibitors | 3 Keynotes | 150+ Sessions | Workshops | 1-to-1 Meetings



DEEP
LEARNING
INSTITUTE

www.nvidia.com/dli