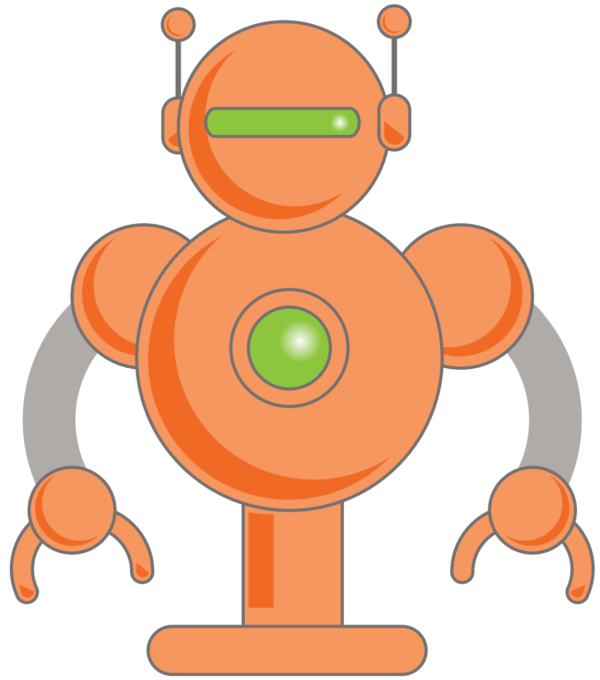
WHITE PAPER

# From Machine Intelligence to Deep Learning

abaco SYSTEMS

# From Machine Intelligence to Deep Learning



## Introduction

The massively parallel architecture of GPUs has taken them far from powerful graphics processors for ultrarealistic gaming experiences and into realms that were previously the domain of science fiction. This paper describes the continuing quest, begun by Alan Turing almost 70 years ago, to create computing systems that redefine human-independent autonomy through their ability to learn – making possible applications that will be truly transformative, not least for the defense industry.

The promise of intelligent machines has been around since the early 20th century. The idea that machines can serve humanity and take over mundane tasks such as house cleaning and automate manufacturing has been with us for almost a hundred years.

The idea was popularised back in the 1950s with robots such as Robbie the Robot (Forbidden Planet, 1956) and Gort (The Day the Earth Stood Still, 1951) either obediently serving mankind - or challenging our very existence (if you believe your science fiction).

Intelligent machines were supposed to be just around the corner - but machines that can truly mimic the human brain have been elusive. The problem in developing a synthetic 'brain' that can think for itself has been a far more complex one to crack – but now, we are starting to see massive gains in the field of deep learning.

### I propose to consider the question: "Can machines think?"

Alan Turing proposed a test in 1950 to assess the ability of a machine to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human. The test involves a human evaluator and a machine engaging in natural conversation over a period of five minutes. Originally, the machine was deemed to have passed the test if it could fool the human 70% of the time. The test did not measure correctness of the responses, but how close the responses are to those that would be given by a human. Today, the 'Turing Test' has been highly influential as well as widely criticized across the field of machine intelligence.

The Loebner Prize is an annual competition for practical Turing Tests which started in 1991 and restarted the conversation around AI (artificial intelligence) almost 40 years after Turing devised his test. It wasn't until 2014, during a competition in Brighton (UK), that a chatbot called Eugene Goostman managed to convince 33% of humans over a five-minute period - convincing organisers that the Turing Test had been passed for the first time. The event marked the 60th anniversary of the death of Alan Turing.

### The rise of vision computing

Small, cheap digital CCDs (charge coupled devices) as used in webcams and mobile phones became ubiquitous in the late 1990s, popularizing a new area of research in machine intelligence. This new discipline - referred to as computer vision - gave machines the ability to 'see' the world around them as we do. Computers and robots were quickly equipped with cameras to make them 'aware' of their environment and the world around them.

Computer vision systems and computer graphics (driven by the computer gaming industry) have a close - almost symbiotic - relationship. Cameras take information from the real world and then render and manipulate it in a host of different ways into the virtual world. Computer vision applications today are wide ranging from panoramic image stitching, light field distortion and object classification to 360° virtual reality.

Open source middleware such as OpenCV offers a wide variety of API calls for feature detection, ISP processing, object detection, image stabilization and for those wanting to get started with sophisticated vision processing algorithms. OpenCV is also optimized for real-time processing and can utilize GPUs to greatly increase the performance of its algorithms.

# From Machine Intelligence to Deep Learning

OpenVX is another low level programming framework (published by the open standard group Khronos, who also manage OpenGL/OpenCL). It is designed to enable software developers to efficiently access computer vision hardware acceleration with both functional and performance portability.

## It's not about robots any more

One company has emerged as the industry leader in computer graphics, artificial intelligence and deep learning. Unknown in the early 1990s, it developed GPU (graphics processing units) that outperformed the competition and won the contract to provide the GPU into Microsoft's Xbox gaming console, released in 2001 and later, to Sony's PS3 in 2005. The company is NVIDIA. It now has a turnover of $7 billion and has become known for its innovation and cutting edge technology - from embedded processors targeted at automotive to high end data center accelerator cards as well as traditional GPUs.

A step change in the GPU market occurred around 2005 when GPUs began to be used to perform compute tasks on the GPU - accelerating application code instead of rendering graphics. NVIDIA quickly latched on to this opportunity to re-purpose the GPU as a general purpose parallel processor for data parallelism. In 2005, NVIDIA released CUDA (Compute Unified Device Architecture) for general purpose computing on GPUs (GPGPU). Now, anyone could write massively parallel code and develop algorithms cheaply using off-the-shelf GPUs.

Using the enormous compute capability delivered by their massively parallel architecture, GPUs were soon put to task in cryptography, medical imaging, signal intelligence, radar processing, AI and deep learning.

## The AI revolution

Jen-Hsun Huang, CEO and Co-founder of NVIDIA, believes the AI revolution is just at the beginning of the 4th Major Industrial Revolution. Computers are performing tasks never thought possible and pushing the boundaries of what computers can do at an accelerated rate.

Deep learning (also known as deep structured learning, hierarchical learning or deep machine learning) is a branch of machine learning that has applications ranging from natural language processing and image recognition to drug discovery and toxicology, recommendation systems CRM, life sciences…

Applications using deep learning are built on deep neural networks (DNN), an artificial neural network with multiple hidden layers of units between the input and output layers. Artificial neurons loosely model an artificial brain, with each neuron connected to many others with an individual neural unit performing a summation function that combines the values of all its inputs together. The effect is that it is possible to train systems to be self-learning and that can perform specific tasks, given enough training data and time. These systems can often excel in areas where the solution or feature detection is difficult to express as a traditional computer algorithm.

Convolutional neural networks (CNN: see also the cuDNN implementation) model animal visual perception, and can be applied to visual recognition tasks. CNNs can take advantage of the 2D structure of the input data. In comparison with other deep architectures, convolutional neural networks have shown superior results in both image and speech applications.

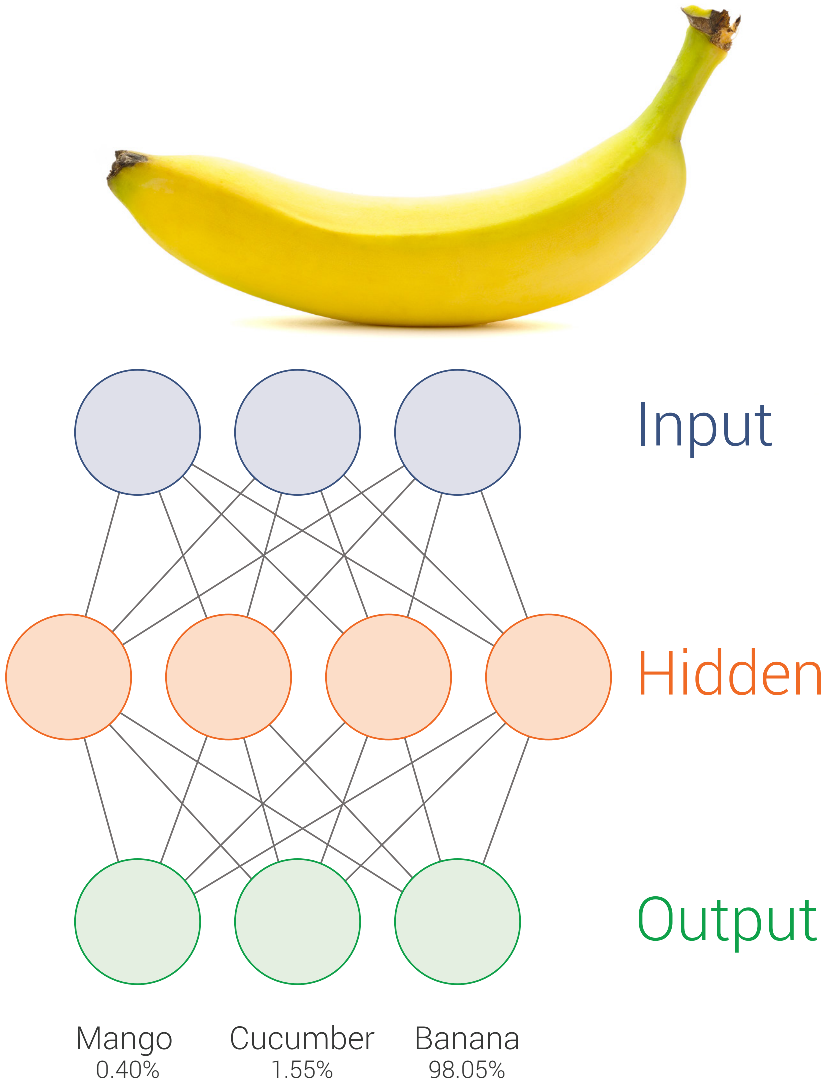
CNNs are easier to train than other regular, deep, feed-forward neural networks as they have significantly fewer parameters to estimate, making them a highly attractive architecture to use. They can also be trained with standard backpropagation for iterative training using large data sets.

## Inception

Google developed its own CNN (codenamed 'Inception') for the Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The challenge is for a machine to learn and classify images from a database of 1,000 object classes. The competition has been run annually since 2010 with GoogLeNet (based in the Inception model developed by Google) winning the classification category in 2014 with only a 6.66% error rate. Competent human image annotator error rate is around 5%.

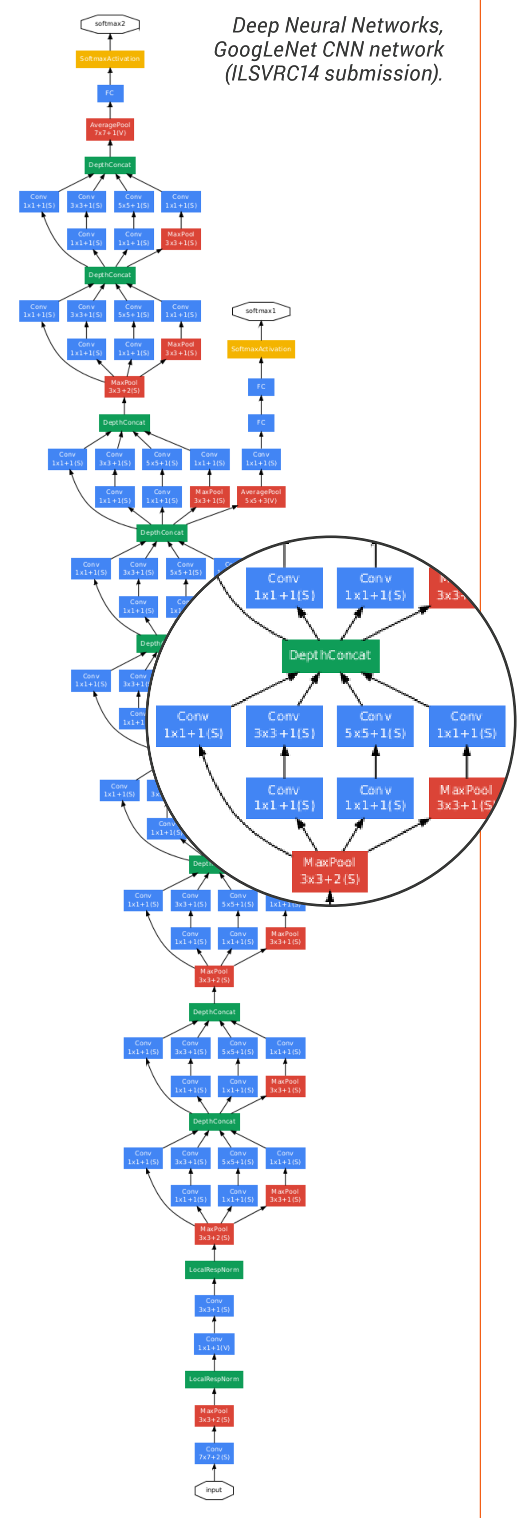
# From Machine Intelligence to Deep Learning

“The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) evaluates algorithms for object detection and image classification at large scale. One high level motivation is to allow researchers to compare progress in detection across a wider variety of objects.”



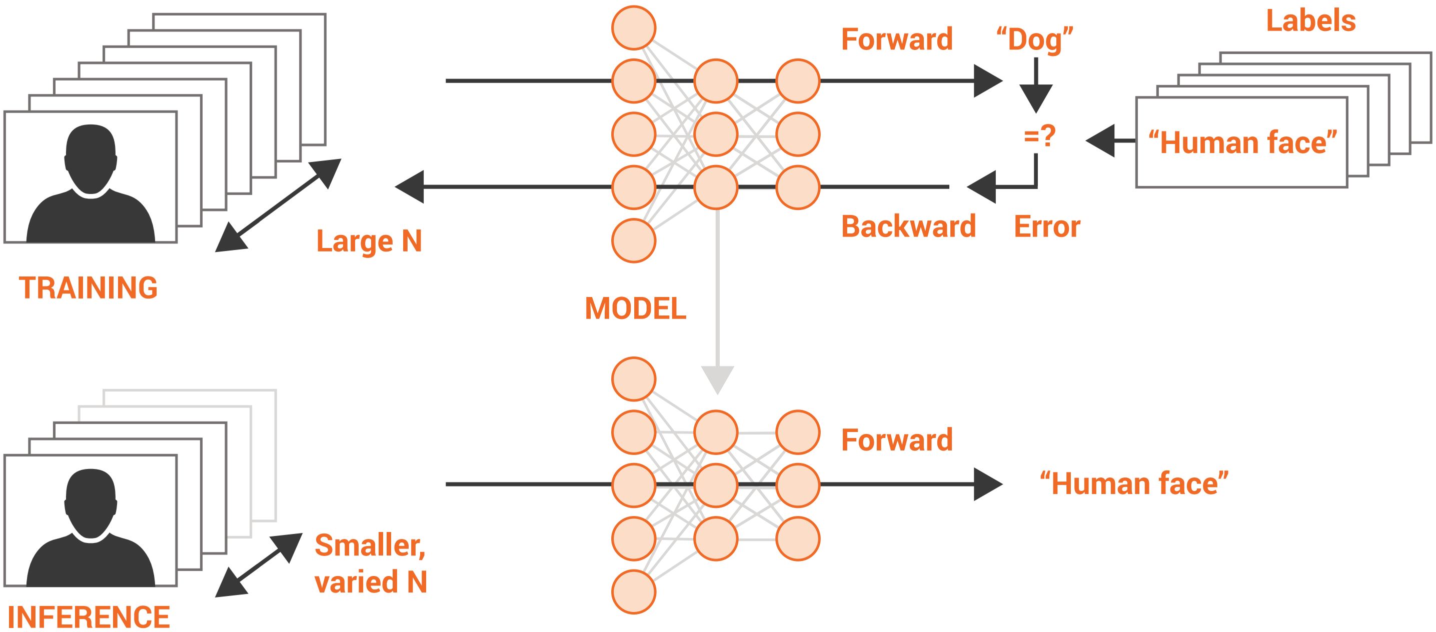

***Shallow Neural Network.***

CNNs are initially seeded with random weights, taking a long time to train using back propagation. GPU are used to accelerate the training of networks and can reduce the training time from weeks to days - even hours - depending on the size of the training data and the configuration of the network. To simplify the training phase, multiple GPUs may be used and networks can be re-trained - shortening the training cycle.




*Deep Neural Networks, GoogLeNet CNN network (ILSVRC14 submission).*

# From Machine Intelligence to Deep Learning




*Training and Inference.*

NVIDIA has released DIGITS to simplify the process of training. This web interface allows users to configure their datasets and monitor the training phase (in real time) to make the required optimizations on the fly. Multiple GPUs can be deployed to accelerate the training.

The process of developing a deep learning application has two separate phases. The first is the training phase, where the requirement is for vast amounts of training data to be catalogued for the training and validation of the network. Curation of the data can be very time consuming and collecting 1,000 – 10,000 images can be difficult - as well as optimizing the network to get the performance required.

Luckily, there are some shortcuts for those wanting to learn more about CNNs trained for large scale visual classification. Pre-trained models are available, so it is possible to jump straight to the second phase of deployment – inference. (Search Google for the 'Berkeley Caffe Model Zoo' for shared models).

Once training has been optimized, the network (model) can be deployed and exposed to visual imagery not previously seen and infer what that object may be with a degree of confidence. Inference is much like the first part of the training phase, where the network is run forward but without the feedback (backwards propagation) updating the weights within the network. For this reason, inference requires substantially less parallel processing and can be performed by state of the art embedded processors (utilizing the embedded GPUs) or FPGAs.

One such embedded GPU processor is the NVIDIA Tegra TX2 system-on-module which is designed to perform deep learning and AI as well as having desktop-performance graphics.

This module is used in the Abaco Systems GVC1000 SFF (small form factor) graphics and vision computer, providing a complete COTS solution with increased I/O performance with six ARM CPU cores running Linux, 10GigE, H.265 hardware compression and dual head display.

# From Machine Intelligence to Deep Learning

*Training - Abaco Systems Deep Learning demo was trained on 1,116 images of a banana.*


*Inference - live video running 1280x1024@30Htz.*

In the Abaco Systems inference demonstration seen at trade shows, the application can distinguish between over 1,000 individual object types. The GoogLeNet model is used to classify every single video frame in the video stream coming from a USB webcam, DEF-STAN 00-82 (GVA-compliant RTP) or GigE Vision camera source at 30 frames per second at 1080p resolution.

This code has been published on GitHub as a quick start demo and can be run on the TX2 development board, GVC1000 or one of Abaco's embedded 3U VPX GPGPU boards such as the GRA113. Other demos are also available for motion estimation and people detection (search 'Abaco Systems' on GitHub) as well as the standard demos included in the Jetpack installer.

To get developers started, NVIDIA provides the Jetpack installer to quickly download and install all the latest libraries, updates and SDK examples needed. The installer can flash the operating system to the board and download all the required software updates.

Jetpack includes the following software in one convenient installer:

- L4T (Linux For Tegra, currently based on Ubuntu 16.04 LTS)
- Visionworks
- OpenCV4Tegra
- cuDNN
- TensorRT
- CUDA
- Multimedia API
- Sample Code
- Profiling Tools

When the developer is ready to start training datasets, the DIGITS deep learning GPU Training System will run on most high-end GPU (must be CUDA Compute capability 3.0 or higher).

NVIDIA has specified a build standard for a state of the art training machine that can be provided pre-built DIGITS DevBox and purchased directly from NVIDIA (or it can be self-built using the how-to-build guide available to registered developers).

The DIGITS DevBox features a high end 6th Gen Intel Core i7 CPU and four TITAN X GPUs with 12GB of memory per GPU connected using 16 lanes of Gen3 PCIe - maximizing both bandwidth and performance.

*NVidia DIGITS™ DevBox.*

# From Machine Intelligence to Deep Learning

## Conclusion

### Need for speed

Deep learning and AI systems are literally driving the automotive industry forward towards the future of fully autonomous vehicles. In-car sensors are coupled with vision systems, utilizing a variety of traditional computer vision techniques, deep learning and AI to make vehicles aware of the world around them.

These same processors, libraries and techniques are also deployed in UAVs (unmanned aerial vehicles) from military drones to smaller commercial platforms performing search and rescue and other field operations for the emergency services.

From border security and ground vehicles to robotics and UAVs, the applications in which embedded deep learning can be deployed are practically endless. Systems such as the GVC1000 are small and cost-effective with high bandwidth Ethernet interfaces that can integrate into network clusters. These can be connected using high speed 10Gig-enabled rugged switches such as the RES3124 managed Ethernet switch, disseminating high bandwidth video and data.

### Summary

The rate of development and innovation in deep learning and AI is relentless and what seemed impossible a few years ago is commonplace today, thanks to innovations in processor technology and new techniques being deployed in machine learning. It's an exciting time to be working in this field and there are few industries not interested in how deep learning can be of benefit to them.

P.S. If the robots should rise up and come to destroy humanity, then remember these words: " Klaatu barada nikto"...