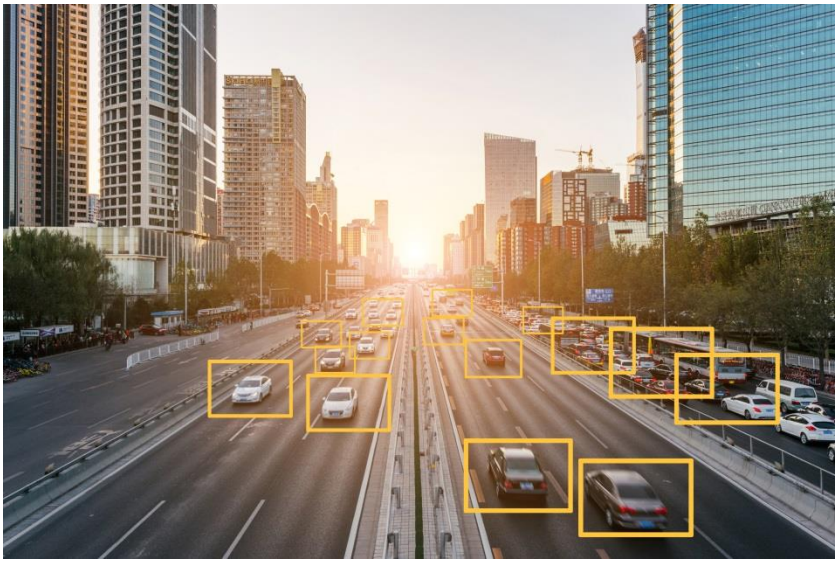
Driving Intelligence at the Edge

Axiomtek's Edge AI Solutions



From IoT to AIoT – Smart IoT Powered by AI

IoT data overload

While IoT allows enterprises to turn device data into actionable insights to optimize business processes and prevent problems, the ability to handle data in a timely, effective manner will determine whether an enterprise can fully enjoy the benefits of IoT. With numerous flows of data streamed from connected sensors and devices that are increasing by billion per year, it is only a matter of time that the enterprise clouds, growing slowly at an annual rate of thousands, will eventually be overwhelmed by enormous volumes of datasets which are beyond their capacity to digest.

Also, in applications like autonomous machine operation, security surveillance, and manufacturing process monitoring, local devices need to act instantly in response to time-critical events. Waiting for feedback from the cloud can result in a response delay and make devices less likely to accomplish the tasks in real time. To resolve the issues of data overload and response lag, a growing number of companies are now seeking to incorporate edge computing and AI solutions into their IoT systems.

Edge computing: processing data where it is needed

Edge computing is a distributed computing technology which brings computation to the edge of an IoT network, where local devices are able to process time-sensitive data as close to its source as possible, rather than having to send the data to a centralized control server for analysis. The primary benefit of bringing data processing back to the edge is that it allows sensor data to be processed right on the spot where it is generated, which eliminates latency and enables local devices and applications to respond instantaneously. Meanwhile, by filtering raw data near the source, edge computing can significantly reduce the amount of data to be sent to the enterprise cloud, alleviating both bandwidth usage and analytical burden.

AIoT: when IoT meets artificial intelligence

Although some IoT systems are built for simple event control where a sensor signal triggers a corresponding reaction, such as switching on/off light based on ambient lighting changes, many events are far more complex, requiring applications to interpret the event using analytical techniques in order to initiate proper actions. To make this work, a new IoT structure known as the Artificial Intelligence of Things (AIoT) comes into play. It applies intelligence to the edge and gives devices the ability to understand the data, observe the environment around them, and decide what to do best – all can be done with minimum human intervention. With the power of AI, AIoT devices are not just messengers feeding information to the control center, but have evolved into intelligent machines capable of performing self-driven analytics and acting independently.

Deep Learning – Machines Learn Like Humans

Deep learning is an advanced branch of artificial intelligence algorithms increasingly deployed at the edge for analyzing visual imagery. A key technology behind computer vision, self-driving vehicles, robots, and many other vision-enabled autonomous machines, deep learning teaches computers to learn complex patterns from image data in order to detect and identify objects in photos and videos – in a similar way that the human brain does.

Training a deep learning model

Inspired by how the neurons in the human brain process information, deep learning uses a deep artificial neural network – an input-output computational model containing many linked hidden layers of processing units – to train a computer to understand data and learn on its own. In the training for image recognition, a deep neural network is fed with a large number of labelled images. The neural network breaks down image input into various components through its hidden layers, where each layer is responsible for extracting certain features and characteristics like edges, colors or shapes (all represented by specific pixel values), then passing these features as output into the following layer. The neural network learns multiple layers of features through a trial-and-error approach, assigning the weight or parameter to each feature to determine how significant the feature is as it pieces together meaningful patterns. This process can be repeated many times, until the network generates a combination of features to figure out what the image represents and produces desirable results with the best accuracy possible through its output layer.

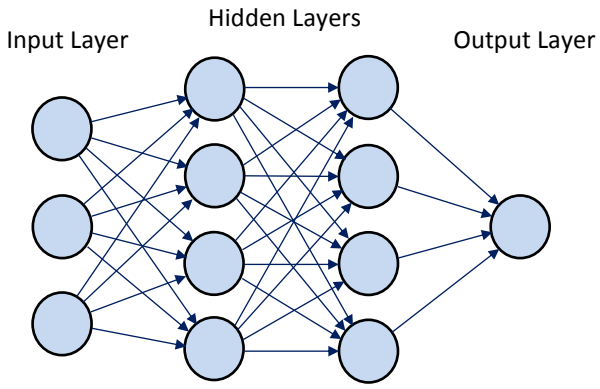


Figure 1: Representation of a classic deep neural network structure. Each hidden processing layer takes an input, performs feature extractions and then passes the output to the next layer. The term "deep learning" refers to the way the data flows through a neural network's hidden layers, where output is not visible. (Reference source: *"CS231n Convolutional Neural Networks for Visual Recognition"*.)

Running inference – the more the data, the higher the accuracy

Once a neural network model has been trained, it will have the ability to infer – to make a prediction or conclusion about real-world data by discovering the key features or patterns it has learned to determine what's in the data. For example, a model trained as an image classifier is able to identify and label unknown objects within an image: when presented with something it's never seen in the past, it will output the correct result of which category the object falls into.

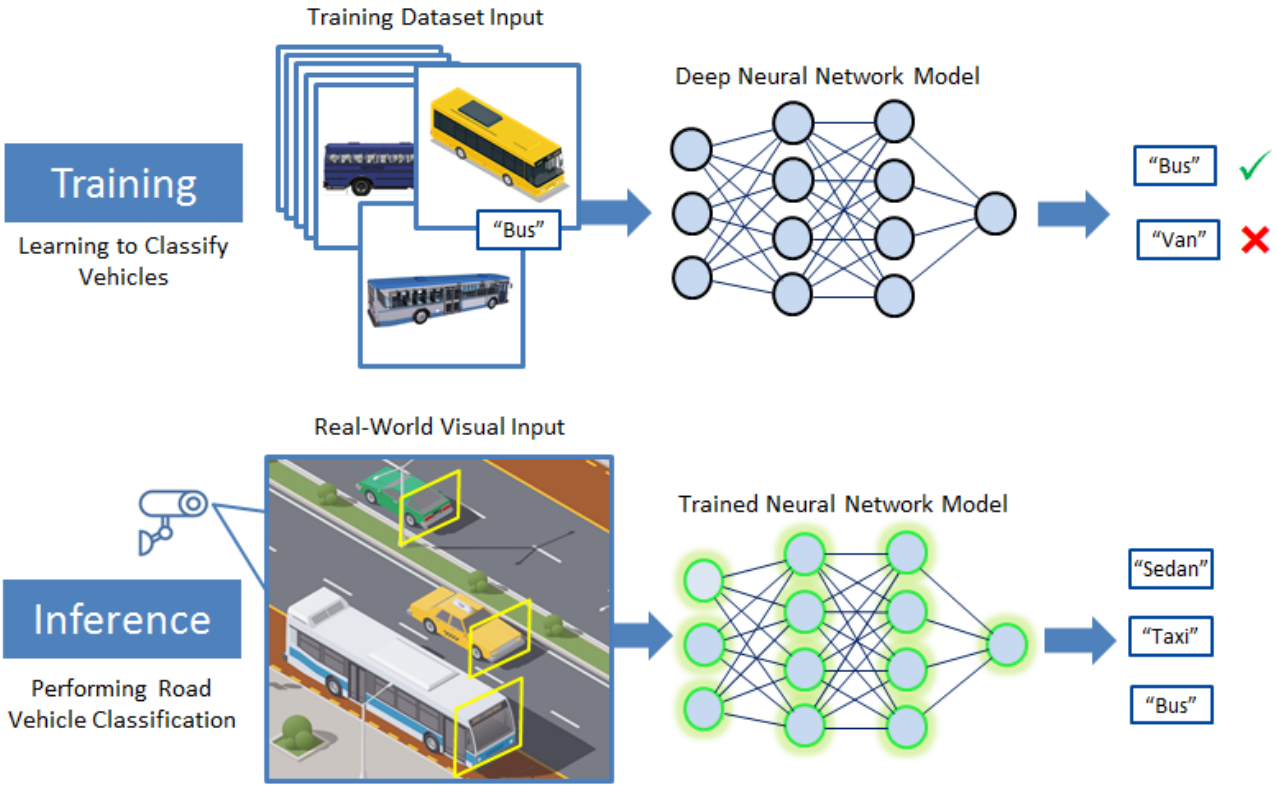


Figure 2: An example of deep neural network training vs. inference

Deep learning allows a computer to recognize intricate patterns much faster and with greater accuracy, in many cases surpassing human-level performance. It is also a highly data-driven technology because a deep neural network must take in tremendous amounts of training data in order to increase inference accuracy. This makes IoT a perfect environment for deep learning, where interconnected machines and sensors constantly feed tons of data from which deep learning models can learn and improve their performance. Deploying AI at the edge of an IoT network also gives deep learning models the ability to observe their surroundings more closely than they ever could before, allowing them to deliver better inference results.

Video analysis with deep learning

With the abundance of visual data available from video-capturing devices like surveillance and car cameras and even smart phones, different types of neural network models are now widely applied to video content analysis to help people obtain rich information from video footage and make better decisions based on real-time video streams. A computer with a trained neural network model can be used to develop video analytics techniques such as object detection, target tracking, and motion recognition. The computer is able to detect objects of interest in a video scene, locate and follow a moving object with specific features, or even estimate the direction in which the object is headed. The targeted items are often marked by a colored bounding box around them for easy identification, with descriptive tags showing object name/classification or certainty percentage.

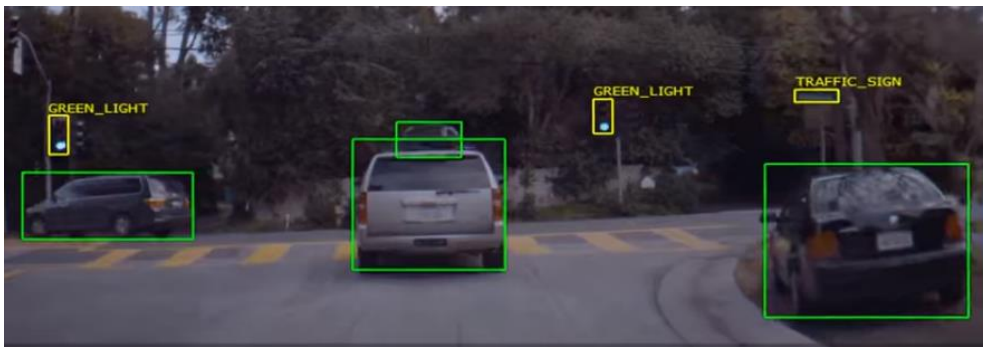


Figure 3: Video output of an object detector model that tracks vehicles and detects traffic lights, stop signs, and other road objects to help NVIDIA's self-driving car perceive its surroundings. (Source: *"NVIDIA Self-Driving Car Demo at CES 2017." NVIDIA.*)

NVIDIA® Edge AI Solutions

GPU – the heart of deep learning AI

Running deep learning algorithms requires intensive computational power to perform large amounts of simultaneous neural network calculations. A GPU (graphics processing unit) can run neural network models much faster because it has a parallel architecture with multiple cores designed to handle heavy multitasking computations. This makes the GPU an ideal AI accelerator capable of speeding up deep learning training and inferencing processes, especially for tasks involving analysis of visual inputs like photos and videos.

The NVIDIA Jetson™ platform

NVIDIA's Jetson™ embedded AI platform combines GPU-driven computation hardware and rich developer SDKs to deliver the performance and efficiency needed for accelerating AI deployment at the edge, where developers can quickly build and get deep learning models running for their AI projects.

NVIDIA® Jetson™ System-on-Modules (SoMs) come in three form factor series – NVIDIA® Jetson Nano™, NVIDIA® Jetson™ TX2, and NVIDIA® Jetson AGX Xavier™ – with specifications targeting particular AI performance needs. Each system is packed with a multi-core CUDA GPU with accelerated parallel processing, as well as a complete set of hardware components including CPU, DRAM and flash storage to save development time and cost. (*Source: NVIDIA Autonomous Machines*)

NVIDIA JetPack SDK is a unified software package supporting all Jetson series modules. It provides comprehensive software resources needed for deploying AI models to Jetson modules, including:

- Linux OS images
- Board support package (BSP)
- CUDA Developer Toolkit
- TensorRT deep learning inference accelerator
- CUDA libraries: CUDA® Deep Neural Network (cuDNN) library and computer vision libraries (VisionWorks, OpenCV)
- Multimedia APIs and samples

Through NVIDIA JetPack SDK, software architects and engineers can make full use of the GPU power to design and train their own neural network models on popular deep learning frameworks, meanwhile optimizing GPU-accelerated inference performance when running deep learning tasks across all Jetson platforms.
(*Source: NVIDIA JetPack*)

Axiomtek Brings the Power of AI to the Edge

As a member of the NVIDIA® Jetson ecosystem, Axiomtek has been in close collaboration with NVIDIA in driving AI innovation at the edge. Combining its strong edge computing expertise with NVIDIA's AI and deep learning technologies, Axiomtek delivers a full portfolio of GPU-optimized hardware platforms and data communication solutions for deploying AI to the edge, helping customers speed up implementation of deep learning capabilities into their IoT devices and turn Artificial Intelligence of Things (AIoT) into reality.

Axiomtek has rolled out a series of edge AI embedded systems specifically built around NVIDIA® Jetson Supercomputer-on-Modules, with integrated NVIDIA GPUs to deliver exceptional computing performance for processing AI data at the edge. These edge AI systems come bundled with NVIDIA JetPack SDK to provide convenient access to the most advanced tools, making it quick and easy to set up development environments for running AI and deep learning algorithms. Depending on what they wish to accomplish, developers and system integrators can utilize pre-trained inference models to build a variety of computer vision programs with their datasets, encompassing the tasks of image classification, object detection, facial recognition, surveillance video analytics, license plate recognition, vehicle tracking, industrial machine vision inspection, and so on.

AIE500-901-FL – Advanced AI Computer for Smart Manufacturing

Fanless Edge AI System with NVIDIA® JETSON™ TX2 Series SoM

The AIE500-901-FL is the latest addition to Axiomtek's edge AI platform, a new generation edge AI computer built specifically for edge computing and deep learning operations. This embedded system employs an NVIDIA Jetson™ TX2 module, with a 64-bit ARM® A57 processor, a NVIDIA® Pascal™ GPU with 256 CUDA cores, and 8GB of 128-bit LPDDR4 memory, delivering remarkable parallel compute power required for performing image processing, deep learning inference, video analysis, object classification, and other computer vision tasks for smart manufacturing. The AIE500-901-FL offers complete AI support via the bundled NVIDIA JetPack 4.2.1 SDK, which helps developers jumpstart their development environment and minimizes the workload for deploying AI-related applications on the NVIDIA Jetson platform. To accommodate massive AI datasets, besides the onboard 32GB eMMC that comes with the Jetson TX2 module, the edge AI computer provides one Micro SD slot and one M.2 Key M 2280 SSD slot, which supports optional 64GB or above PCIe and SATA SSDs.

The AIE500-901-FL features diverse I/O and expansion interfaces, including one HDMI 2.0 port, two USB 3.1 Gen1 ports, two GbE LAN, two COM ports or two CAN ports, and one OTG Micro USB port. It also has one full-size PCI Express Mini Card slot accompanied by a SIM slot for 3G/4G, GPS, Wi-Fi and Bluetooth connections. Other features include reset and power buttons, a recovery switch, and four SMA-type antenna openings. Meanwhile, to withstand the rigors of day-to-day operation, the AIE500-901-FL provides extended operating temperature range of -30°C to +60°C and vibration resistance of up to 3Grms with its strong construction, plus a 12 or 24 VDC power input to meet industrial and automation requirements.

AIE500-901-FL

Feature Highlights

- NVIDIA® Jetson™ TX2 with Pascal™, 256 CUDA cores GPU
- High-performance AI computing with GPU accelerated processing
- Ideal for intelligent edge applications, deep learning and computer vision
- Supports M2. PCIe & SATA SSD and MicroSD slot
- Supports 2 USB 3.1 Gen 1 and 2 COM or 2 CAN
- -30°C to +60°C wide operating temperature range and 3Grms vibration resistance
- JetPack 4.2.1 support

eBOX560-900-FL – Factory Automation

NVIDIA® Jetson™ TX2 Based Fanless Edge AI System

Axiomtek's eBOX560-900-FL fanless edge AI system is powered by an NVIDIA Jetson™ TX2 module, which features a 64-bit ARM A57 processor, an NVIDIA Pascal™ Architecture GPU with 256 CUDA cores, and 8GB of LPDDR4 memory to deliver powerful computing and graphics processing performance. This embedded system packs rich I/O and flexible expansion interfaces in a palm-size, IP40-rated box. It comes equipped with two Gigabit Ethernet ports, one HDMI 2.0 port and one onboard 32GB eMMC, plus one M.2 SSD PCIe 2.0 x4 socket which supports a high performance NVM Express interface for maximum storage needs. The eBOX560-900-FL also includes a PCI Express Mini Card slot for 3G/4G/LTE connections, as well as an 802.11ac Wi-Fi module with Bluetooth connectivity onboard.

The eBOX560-900-FL supports NVIDIA JetPack 4.2.1 SDK based on Linux Ubuntu 18.04, which bundles all Jetson platform software for building AI models. This AI-enabled system makes a perfect platform for realizing smart manufacturing solutions that rely on deep learning and computer vision techniques, covering everything from facial recognition and behavioral analytics to machine vision inspection and quality control.

eBOX560-900-FL

Feature Highlights

- NVIDIA® Jetson™ TX2 with Pascal™, 256 CUDA cores GPU
- High-performance AI computing with GPU accelerated processing
- Supports M2. NVMe SSD
- 1 PCI Express Mini Card expansion slot
- JetPack 4.2.1 support; ideal for intelligent edge applications
- Ultra-compact enclosure design

eBOX800-900-FL – Smart Traffic Control

Rugged IP67-rated Fanless Edge AI System with NVIDIA® Jetson™ TX2

Axiomtek's eBOX800-900-FL is a ruggedized edge AI embedded system built to survive the harshest outdoor surroundings for edge computing and intelligent traffic applications. This tough edge computer is enclosed in an IP67-rated extruded aluminum and heavy-duty steel case for dust protection and water resistance. Other ruggedized features include wide temperature range support of -30°C to 60°C (-22°F to +140°F), 10 kV surge protection, 3Grms vibration endurance, as well as M12 type I/O connectors and four N-jack waterproof antenna openings for enhanced operational stability in rugged environments.

The eBOX800-900-FL is powered by the NVIDIA Jetson™ TX2 module, which has a powerful 64-bit ARM A57 processor and a 256-core NVIDIA® Pascal GPU with NVIDIA's JetPack toolkit support for running AI and deep learning models. It also has 8GB of LPDDR4 memory and 32GB eMMC onboard. The edge computer features one M.2 SSD PCIe 2.0 x4 socket which supports a high-performance NVM Express interface for extensive storage needs, and has a PoE port to support the applications involving the use of IP cameras or any other PoE device for traffic flow monitoring, license plate identification, vehicle recognition, and machine vision. This advanced embedded system runs on Linux Ubuntu 18.04.

eBOX800-900-FL

Feature Highlights

- IP67-rating for outdoor use
- NVIDIA® Jetson™ TX2 with Pascal™, 256 CUDA cores GPU
- High-performance AI computing with GPU accelerated processing
- -30°C to +60°C wide operating temperature range
- 100 to 240 VAC wide range power input with 10kV surge protection
- 4 N-jack antenna openings with water proof design
- M12 lockable I/Os
- 1 PCI Express Mini Card expansion slot
- 1 IEEE 802.3at GbE LAN port (30W)
- Wall/VESA mounting

*For detailed specifications on Axiomtek's edge AI systems, visit www.axiomtek.com and go to: Products > IoT Solutions > Automation > Edge AI System.

How Axiomtek's Edge AI Systems Enable Deeping Learning in Various Markets

Use case 1: smart farming with eBOX560-900-FL

The customer has been developing a smart camera system that enables 24/7 cattle behavior monitoring and video data analysis to improve farming operations. Using Axiomtek's eBOX560-900-FL to combine both computer vision and artificial intelligence capabilities, the camera system is able to identify animals with specific eating or drinking activities from video footage, discover health and feeding patterns, and assess how environmental changes/farming practices impact livestock. The camera system delivers daily event notifications to farmers through their phones, while also providing remote access to detailed analytics about their herd and farm operations, which is helpful for farmers to turn visual information into actionable insights and make data-driven decisions to maximize productivity and profitability.

Use case 2: video-based traffic management with eBOX800-900-FL

The customer was initiating a traffic management program involving the implementation of an on-site video IoT solution with edge AI processing power to analyze live video feeds from street surveillance cameras. It aimed to provide timely traffic control via real-time video content analysis, meanwhile relieving the burden of transferring large video datasets back to the cloud for analysis.

Axiomtek's eBOX800-900-FL edge AI computing system, with its PoE camera connectivity, deep learning capability, as well as ruggedized design for harsh outdoor use, was deployed into the customer's platform to facilitate traffic management by performing the tasks of computer vision-enabled video analytics:

- Traffic flow measurement: vehicle counting, vehicle speed detection, etc.
- Vehicle Tracking: vehicle type classification, driving line identification, and moving direction predictions (going straight; turning right or left).

The traffic and vehicle tracking analysis based on live video streams would help transportation authorities detect incidents and give them a better understanding of the actual traffic volume on the road, allowing them to precisely estimate potential traffic jam areas/periods and take prompt actions to eliminate congestion or help drivers avoid it. The results of real-time traffic analysis can also be integrated with other intelligent traffic systems such as traffic light control to direct vehicles to alternate routes with less traffic.

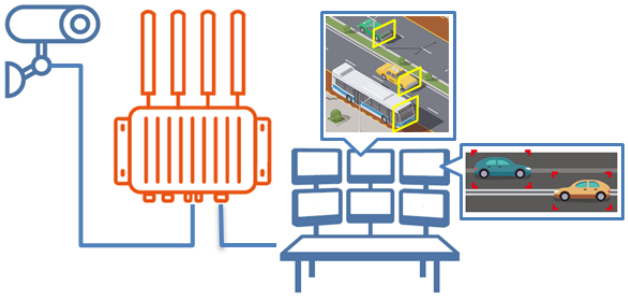


Figure 4: A conceptual illustration of the AI traffic management platform

About Axiomtek Co., Ltd.

As one of the world's leading designers and manufacturers of PC-based industrial computer products, Axiomtek specializes in data acquisitions and control systems of rich diversity and modularization. With the upmost enthusiasm in serving their customers, Axiomtek has mirrored PC evolutions in various industries by shifting its focus toward the design and manufacture of PC-based industrial automation solutions, standing as a trustworthy long-term provider of industrial computers.

Established in 1990, Axiomtek has partnered with more than 60 distributors globally, offering more than 400 products through product lines of Industrial PCs (IPCs), Single Board Computers (SBCs), System on Modules (SoMs), Fanless and Rugged Embedded Systems (eBOX and rBOX), Intelligent Transportation Systems (tBOX and UST), Industrial IoT Gateway, Touch Panel Computers (TPCs), Medical Panel Computers (MPCs), Digital Signage Solutions and Network Appliances (NAs).

Axiomtek is a Member of the Intel IoT® Solutions Alliance. A global ecosystem of more than 800 industry leaders, the Alliance offers its Members unique access to Intel technology, expertise, and go-to-market support—accelerating deployment of best-in-class solutions.