Unmatched Global Market Coverage

aiworld

Artificial Intelligence UPDATE

Machine Learning - Deep Learning - Software Research
Products - Technology - Applications - Industry News

AI World is a Global Publication of e2mos

Mar-Apr 2019

Bringing the next evolution of machine learning to the edge

The Automation (EDA) of AI 6 Experts a the Table

Adlink & Touch Cloud Deliver AI Solutions Powered by Intel® Vision Products

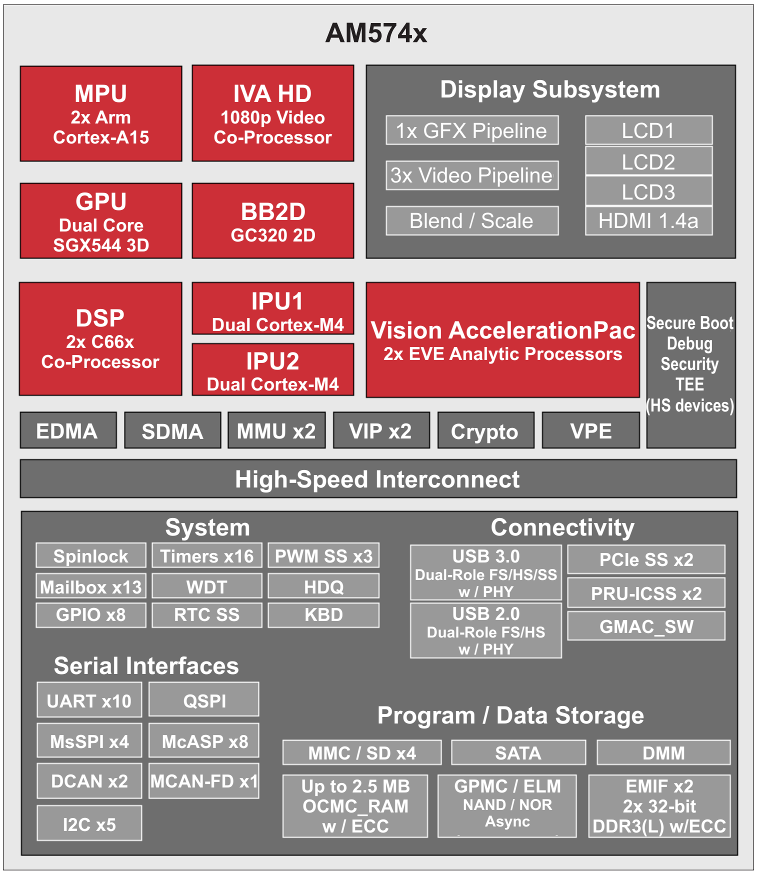
Products & Technology from Silicon to Application-ready Platforms

In this Edition:

• Bringing the next evolution of machine learning to the edge.
By Texas Instruments

• Sitara processor: dual arm Cortex-A15 & dual DSP, multimedia, ECC @ DDR, secure boot & deep learning.

AM574x Sitara Arm applications processors are built to meet the intense processing needs of modern embedded products.
From Texas Instruments



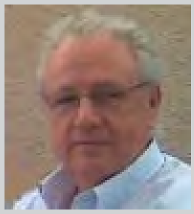
• The Automation of AI
Experts at the Table: Will the separation of hardware and software for AI cause problems and how will hardware platforms for AI influence algorithm development?

• ADLINK and Touch Cloud Deliver AI Solutions Powered by Intel® Vision Products

SOLUTION BRIEF: Intel® Vision Products AI Solutions for Manufacturing and Smart Cities

Daniel Dierickx
CEO & co-Founder
at e2mos
Acting Chief Editor

*Over 3 Decades
Semiconductors & Computer
Systems Market Expertise*



Dear Reader,

Here is your free copy of **AI World** one of our six e-magazines

Your benefits:
• Selected subjects
• Easy reading
• Many direct links for more
• Efficient and time saving
• FREE Worldwide

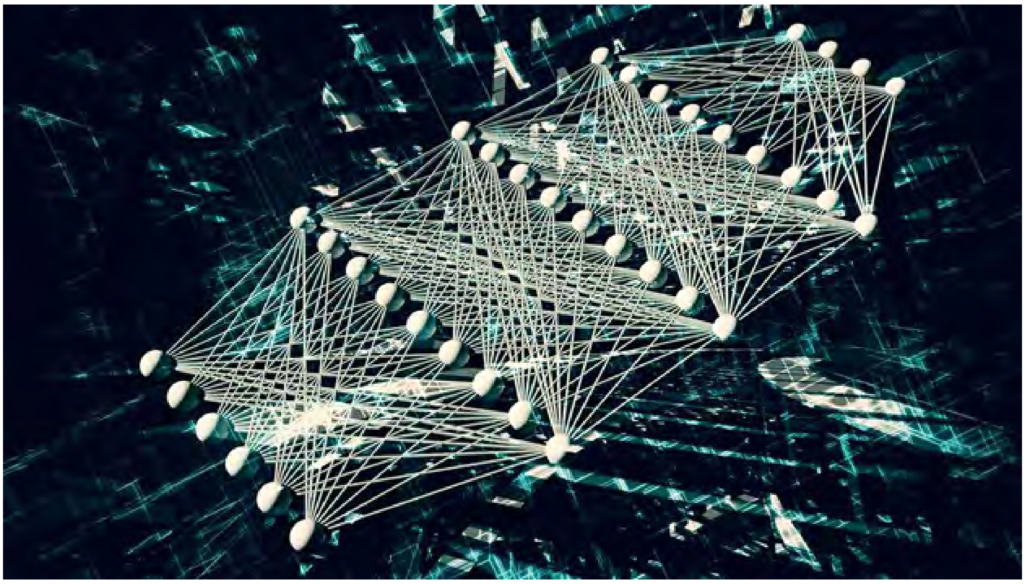
FREE Subscription
Click on the logos below

Semi Update World
aiworld
IoT World
Embedded Systems World
Telecom COTS World
Broadband Broadcast IoT AI Convergence
ATCA World

Editor/Publisher: e2mos
WEB: www.e2mos.com
Contact: mgt@e2mos.com

Bringing the next evolution of machine learning to the edge

From Texas Instruments



Even if you're not very familiar with deep learning, you've probably heard about it and how it can, among other things, help automate the driving experience, increase manufacturing efficiency and change the consumer shopping experience. Deep learning is the latest evolution of artificial intelligence (AI) and machine learning. Both machine learning and deep learning are subsets of AI. While traditional machine learning algorithms need to be specifically programmed by people with domain-level expertise, deep learning algorithms utilize neural networks that can be trained by feeding them data. The same network can be used to solve very different problems by training it with different data, removing the requirement of having specific expertise in the problem that needs to be solved. Because of this, deep learning is considered a foundational technology that has the potential to make an impact in many industries.

For those new to deep learning, think of it as a technology that can classify things. Many real-world problems can be broken down into a classification problem, for example:

Autonomous and semi-autonomous vehicles need to be able to classify roads, other vehicles, people and road signs as shown in Figure 1.

Smart factories need to be able to classify a product as defective

Automated retail outlets need to classify if a consumer is removing a product from a store for purchase or just looking at it

Figure 1: Classification example of an automotive use case. Every object in the scene is determined to be one of the five different classes.

... to next page

TEXAS INSTRUMENTS

... from previous page

Bringing the next evolution of machine learning to the edge

There are two main parts to deep learning: training and inference, as shown in Figure 2. During training, a model of a neural network that has been created to solve a classification problem is fed huge amounts of labeled data, from which it learns how to classify things. In the example scenario shown in figure 1 of an autonomous vehicle, the model would be fed images of pedestrians, road signs, cars and roads with each of these objects correctly labeled. Real-time performance or power is not an issue during training, so it usually occurs on a desktop or cloud platform.

During inference, the trained network that has been deployed into its end application makes a classification decision, such as determining whether or not a part on an assembly line is defective.

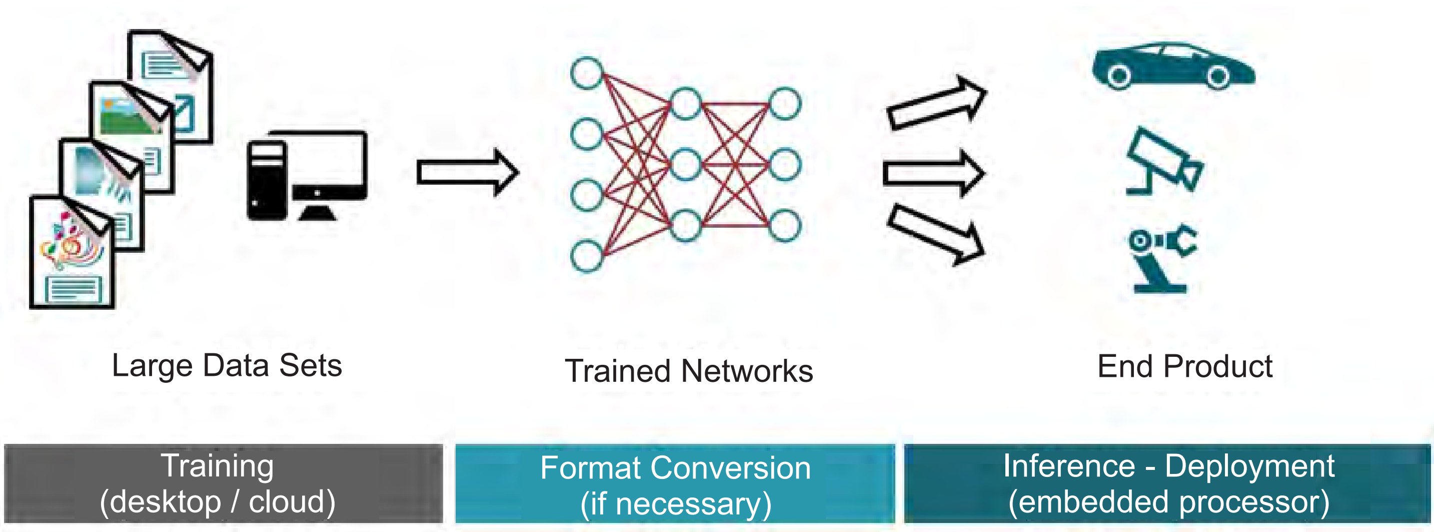


Figure 2: Typical development flow of a deep learning network.

For a combination of reasons – including reliability, low latency, privacy, power and cost – it is beneficial for inference to run close to where sensors gather data; that is, at the “edge” of the network. Advanced sensors that can handle some amount of AI are capable of running a simple inference in the sensor itself. For a more complex inference, a separate edge processor is needed.

TI Sitara™ processors are great devices for edge processing. With a scalable portfolio integrating Arm® Cortex®-A cores with flexible peripherals, Sitara processors provide industrial-grade solutions with an extensive mix of processing, integration and power efficiency. With TI Deep Learning (TIDL) software, Sitara AM57x processors leverage hardware acceleration to improve the performance of machine learning and deep learning inference at the edge.

... to next page

Bringing the next evolution of machine learning to the edge

TIDL software is a set of open-source Linux software packages included on the processor software development kit (SDK) and tools that enable the hardware-accelerated offloading of deep learning inference in an AM57x device. It's possible to accelerate inference across embedded vision engine (EVE) subsystems, C66x digital signal processor cores or a combination of both, using a set of application programming interfaces built with Open Computing Language (OpenCL).

For those who already have an existing neural network framework, TIDL software bridges the gap between the popular Caffe and TensorFlow frameworks with accelerated hardware on the AM57x through an import tool, as shown in Figure 3. To maximize efficiency on the hardware, or for those who don't have an existing neural network framework, Caffe Jacinto, a TI-developed framework forked from NVIDIA/Caffe and designed for the embedded space, can perform the training part of deep learning.

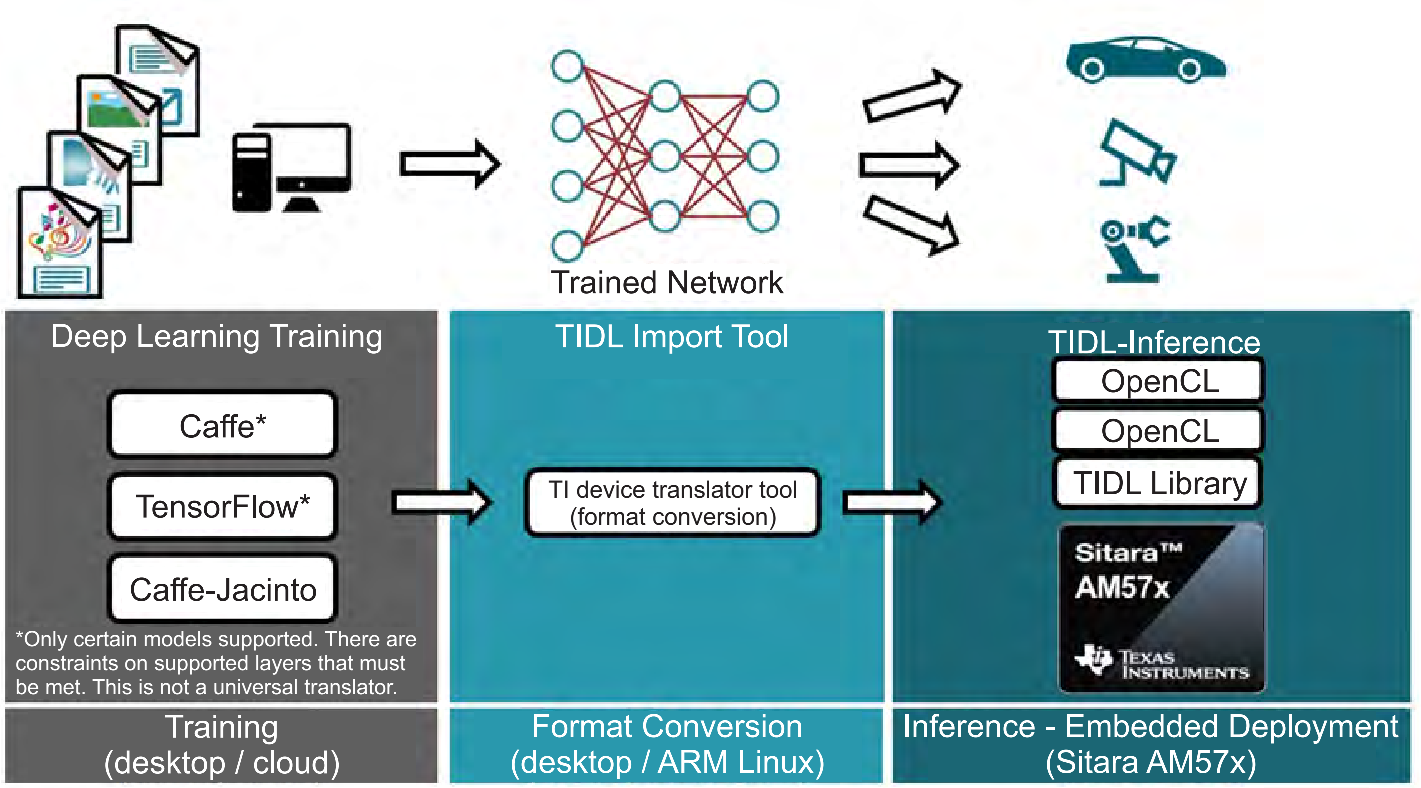


Figure 3: Texas Instruments Deep Learning (TIDL) software development flow.

Note that while TI has developed a framework that can be used to train a network model for an embedded application, the training itself still takes place on a desktop or cloud platform. It is the inference that runs on the TI embedded processor.

With free software as part of TI's Processor SDK, training videos and a deep learning TI reference design, whether you're new to deep learning or a seasoned expert, TI provides the building blocks to get you started quickly and easily.

Additional resources:

- Visit the TI Sitara Machine Learning page.
- Read more about machine learning:
    - Bringing machine learning to embedded systems
    - AI in Automotive: Practical deep learning

- Download the Deep Learning Inference for Embedded Applications Reference Design
- Watch these TI training videos: TIDL Overview for Sitara Processors

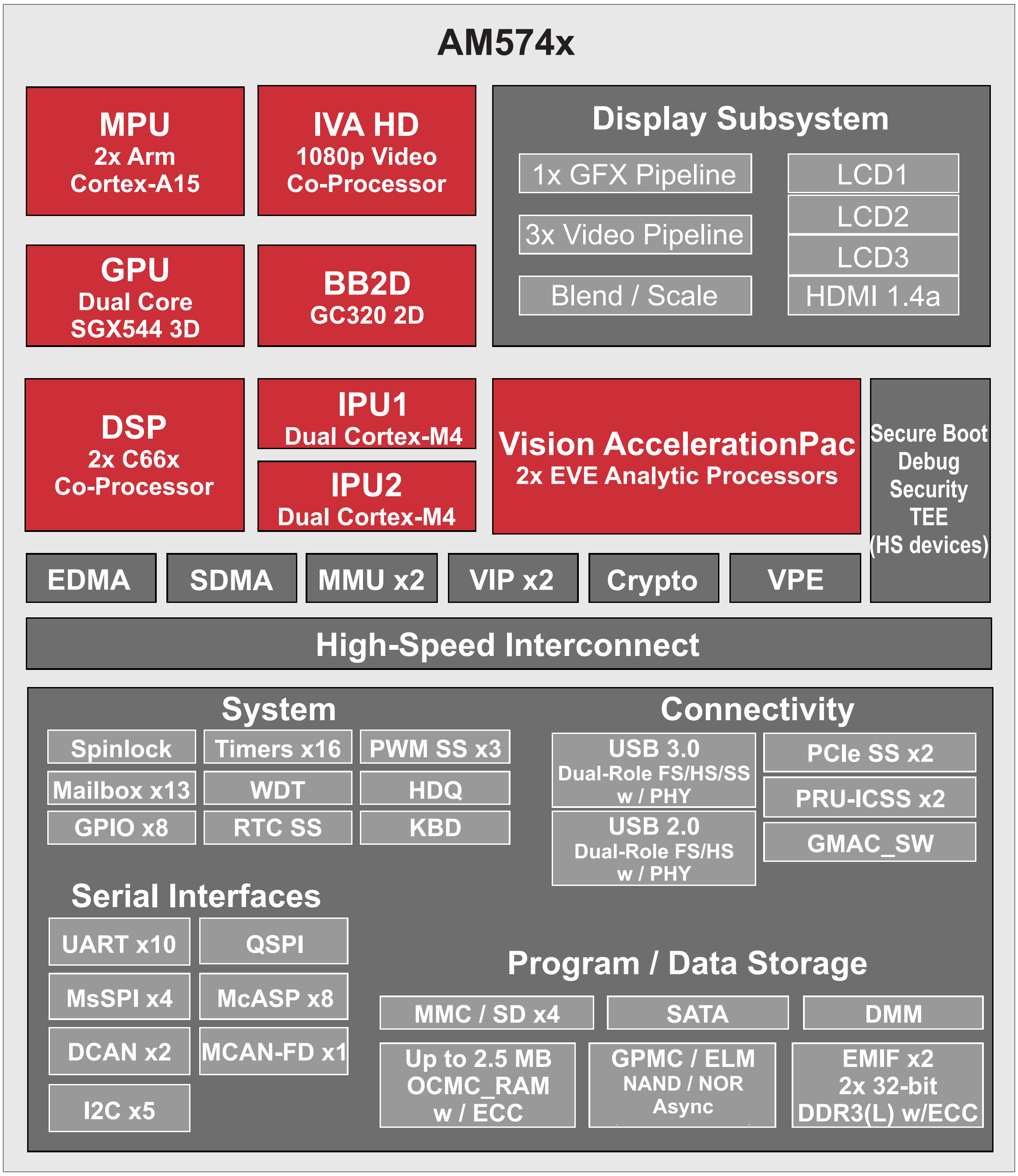
Sitara processor: dual arm Cortex-A15 & dual DSP, multimedia, ECC @ DDR, secure boot & deep learning

AM574x Sitara Arm applications processors are built to meet the intense processing needs of modern embedded products.

AM574x devices bring high processing performance through the maximum flexibility of a fully integrated mixed processor solution. The devices also combine programmable video processing with a highly integrated peripheral set. Cryptographic acceleration is available in every AM574x device.

Programmability is provided by dual-core Arm® Cortex®-A15 RISC CPUs with Arm® Neon™ extension, and two TI C66x VLIW floating-point DSP cores. The Arm allows developers to keep control functions separate from other algorithms programmed on the DSPs and coprocessors, thus reducing the complexity of the system software.

Additionally, TI provides a complete set of development tools. ACCESS ALL INFO, TOOLS & DEMO BOARDS



IVA HD	Image and Video Accelerator
BB2D	BitBLT 2D Hardware Acceleration Engine
IPU	Image Processing Unit

EVE	Embedded Vision Engine
VPE	Video Processing Engine
VPU	Video Processing Unit

The Automation of AI

Experts at the Table, part 1: Will the separation of hardware and software for AI cause problems and how will hardware platforms for AI influence algorithm development?

March 28th, 2019 - By: Brian Bailey

Semiconductor Engineering sat down to discuss the role that EDA has in automating artificial intelligence and machine learning with Doug Letcher, president and CEO of Metrics; Daniel Hansson, CEO of Verifyter; Harry Foster, chief scientist verification for Mentor, a Siemens Business; Larry Melling, product management director for Cadence; Manish Pandey, Synopsys fellow; and Raik Brinkmann, CEO of OneSpin Solutions. What follows are excerpts of that conversation.

SE: The EDA industry has struggled with the separation of hardware and software. While it is now trying to bring certain aspects of software back into the fold, the industry still struggles with system-level problems that are combinations of hardware and software. With machine learning and AI, the industry appears to be making the same choice to allow hardware and software to separate, which is potentially the same mistake.

Foster: I don't think the separation of hardware and software was a mistake, and I don't think that is the biggest issue right now. We are moving from Turing architectures to statistical architectures. That is the real challenge. It is not hardware/software separation. Consider high-end servers. There is a clear separation with very good concurrent processors to do the development, and it is done this way to facilitate verification and development. For example, there are APIs between the hardware and software worlds so development can progress independently and concurrently. I do see the point in terms of optimization, but I don't think that is the biggest issue.

Pandey: The separation between hardware and software has always existed. Much as we like to do hardware/software codesign, in reality, if you want to build a high-end transaction system, such as a database system, it does not make sense to start verifying whole end-to-end stacks. We do not have the capability to do that. We have a separation of abstraction; we verify the ISA, maybe verify the compilers, but verifying that a transaction goes through the database—that is too low a level of abstraction to be verifying that at the ISA-level. Taking the same example when you look at machine-learning applications, and you look at the quintessential application—self-driving cars—you have the sensors, you do sensor fusion, you try and recognize objects using CNNs. That level of verification—how do you even capture that? The best that we can do today, and that is a pretty reasonable separation, is that [when] you verify the neural-network processor, you make sure the individual operations work. The behavior of the system as an aggregate is a very different thing—we can't even characterize it well. That is an on-going topic of research. How do you even explain what is happening, so if you can't explain it, there is no way you can write an assertion to verify it. It would be nice to have a technique, but it does not exist today. You cannot verify an AI system end-to-end.

Foster: You stated that we have problems even describing it—a deterministic system means that you put A in and you expect Y. That is not true in an artificial neural network. You put A in, you have a probability of getting Y. Is that a bug in the design or is it outside of the probability?

... to next page

The Automation of AI

... from previous page

Experts at the Table, part 1

Pandey: Actually, these systems are deterministic. Given an example or an image, if the system is operating correctly, it will always provide the same answer, but you are right in that the answer is based on a correctness probability of it being a cat or a pedestrian crossing the road, and it might be 99% probability—but the hardware is deterministic. It is just that the results may not be right.

Foster: When you get down to a simple neural network, it is just an array of MACs. Those are easy to verify, and we have been doing that for a long time.

Hansson: The key with AI is that it is statistics when we talk about the system-level results. You need a deterministic framework to verify the outcome because sometimes it will do some something smart and at other times it will be stupid. You need the deterministic framework to know if you should make a U-turn while driving at high speed. You can verify the framework, but you cannot verify the outcome of every result and expect it to be within certain ranges.

Pandey: These systems are deterministic, but it is that real-world situations that are different all the time. If your car is deciding if it should do a lane change, given the exact set of inputs, the hardware and software will do exactly the same thing, except there are so many variations.

Hansson: You make a deterministic piece of hardware, but within the scenario there may be input stimulus that has never been seen and as a result you may get random output.

Letcher: You are verifying a general-purpose processor. The training that is done on that is a separate thing. Until it has been trained you cannot verify what it will do in an application space. Just like any processor, the state space to verify all applications is huge. So the question is, 'Do you verify it for a specific application after training or do you verify it as a processor that can be trained to do anything?'

Brinkmann: There are two drivers for the two cycles. One is the software side that has the algorithm as a driver. Algorithmic innovation is progressing faster than Moore's Law by a large margin on the machine learning side. So advancements in algorithms are more relevant at the moment than the silicon. The other side is the economics. The hardware platform is something that you do not want to change for each application individually. You need to reuse that, or you will not get enough leverage from the software. If you have a platform that is versatile enough, such that you can map different algorithms to it, then that is what people are trying to build today with programmable logic, processors, accelerators, etc. You want multiple versions of algorithms mapped to that, and leverage the innovation in software. When you look at verification and the separation—you have to verify the platform and that has multiple aspects. Is the fabric of the FPGA properly built? Is the processor doing its job? You may not have the application in mind when you do that. You verify the characteristics of the platform so that you can provide some guarantees, perhaps just statistical in the future if you think about analog blocks being used for machine learning. But still, they are guarantees about the platform. Second, you need to verify the algorithm and the application. Now you verify if you get a cat or a pedestrian. Third, you verify if the mapping to the platform works. If it goes onto an FPGA, is the generated RTL mapped properly to the FPGA? These are different lifecycles that are interlocked. The verification on the algorithmic side will be repeated for multiple versions of the algorithm based on new data from the field, and you have to repeat the mapping process and the verification while the platform remains unchanged. That is why the platform has to be robust. So the separation is actually getting stronger.

Melling: One of the great comparisons for AI is what has happened in GPUs. If you look at it, the GPU had the same problem. It was not an absolute answer. It was up to the visual judgment of what was being represented on the screen as to whether it was smooth enough, the right shading, etc. AI has the same problem. It is the microcode that is to be put on the hardware that will produce the system that we can ask, 'Is it giving me the right experience? Am I getting the expected answers?' With a GPU you had a golden eye—the people who were the aesthetic judges. We do not have that with AI because the algorithms are so diverse. We may have to borrow some verification technology and move it into the software world, where they are doing algorithms and try to work out how to randomly expand their test sets or their training sets in order to get higher-quality algorithms. At the end it will be statistics that judge the outcome and if we have achieved the desired results.

Hansson: It is not that different from random testing really. Random tests are a subset of an enormous space. Machine learning also can be validated as a subset of an enormous space.

Melling: We have all seen the example of someone putting two pieces of tape on a stop sign and it is no longer recognized.

Brinkmann: There are many new verification challenges for machine learning, such as characterizing the data set, looking for outliers, looking for bias. Being statistical is one thing, but understanding the space well enough to make some good predictions about how well it will perform in uncertain situations is a big challenge, and that is where we need a new type of verification.

... to next page

... from previous page

Experts at the Table, part 1

Foster: And we do not have good metrics in that area. The whole notion of coverage is being debated today both from an application and algorithmic point of view.

SE: In the hardware/software world, software quickly became king. Hardware could not change in any way that breaks software. As a result, we have been stuck with existing ISAs purely because of existing software. With AI, we started by using existing ISAs, but they did not have enough compute power, so we migrated to GPUs. The algorithms that are being developed for AI are now being heavily influenced by GPUs. However, we also realize that GPUs do not have enough performance and consume too much power. Algorithms appear to be saying we do not have the right platforms for either inferencing or training. This looks like an unstable situation. Who will make the breakthrough—hardware or software?

Brinkmann: Would you disagree if I said the new ISA is actually Caffe and TensorFlow? The abstraction has been virtualized. You are not looking at processor abstraction anymore. When talking about machine learning, you are talking mapping a Caffe or Tensorflow algorithm to the hardware that is available. It doesn't care about the ISA or if it is a GPU or FPGA. You have abstracted away from that. But you have to agree on something. Otherwise, you will not manage to use the same software on different platforms. The interface becomes virtual.

Letcher: It started out with things like CUDA and OpenCL—lower level standards, but they have been replaced largely by TensorFlow or an equivalent higher-level interface.

Pandey: The language of today—Tensor—has risen quickly. Some talk about a different language being required for AI, but the fundamental unit that we deal with are these tensors. Multi-dimensional matrices of numbers. We had the question of software being king, and now we are in an unstable situation that is driven by economics. Ten or fifteen years ago, you had processor performance increasing from technology improvements and micro-architectural advances. When you can jump 50% a year without needing to do anything in the software, then conventional applications will not change—that is just a matter of economics. But the situation today is different. First, the core performance is not increasing fast enough for machine-learning applications. You have to think about new micro-architectural techniques. Nvidia's current processor is doing whole matrix multiplication. These tensor units in the GPUs—everyone is coming up with different types of neural-network processors—these are all about doing large tensor operations efficiently. We are at a point where, when you think about applications, think of what the hardware architecture is going to be. I suspect that with so many startups working in this field, that there is no place in the market over the long run for 30 or 40 companies. The numbers will whittle down. Probably, as with the markets in general, we will settle on standard forms, languages, micro-architectures. But we are a long way from that today. That is what makes the area exciting right now.

Foster: I suspect it will grow before it shrinks. At DAC this year there have been 92 papers submitted on the subject of machine-learning architectures. This is up 64% from last year. A tremendous amount of research is going on in this area, and we will see a lot more startups until it all collapses.

Hansson: AI is extremely driven by software. This means that we have to improve the hardware, but it is extremely driven by software—even culturally. If you look at AI/ML, everything going on is open source. It is downloadable, and you can play around with it. There are no expensive EDA companies involved in AI algorithms. It is very much an open environment. The hardware world tries to provide hardware that will fit into that, but that is where the drivers are.

Melling: We are trying to provide fast-enough, low-power enough deployment platforms. We also have to distinguish the two sides of it—training and inferencing. Training, which is very hefty, compute-intensive, and inference and deployment, which are much more about low-power. It still has to have the performance to get the necessary response times. Those two types of systems are going to be different. The tradeoffs are different. People are looking at new memory architectures and how to implement in a single-bit algorithm. These are all things that are being looked at so that algorithms can be deployed in a low-power and cost-effective manner.

Pandey: Standards are slowly emerging. You have the problem of creating a model, then using it somewhere else, such as ONNX. (ONNX is an open format to represent deep-learning models. With ONNX, AI developers can more easily move models between state-of-the-art tools and choose the combination that is best for them.) Not everyone supports it yet, but such standards for model interchange are emerging. It is not clear what the final standard will be, but if we talk about data models being the currency of machine learning, then this is an area that needs to be standardized. It is very much a software-driven culture, but there are two aspects to it: during development of the model, and the deployment for inferencing, which gets a lot closer to the hardware because there is almost an equivalent of microcode.

... to next page

The Automation of AI

... from previous page

Experts at the Table, part 1

Hansson: And maybe that will get standardized in the future. There isn't room for 40 different chips.

Pandey: The standard for the models will help with this. It shows how to represent your values and the data. From my observations, the frameworks are open source. What is being closely guarded is the dataset. That is a very important component, and when you think about a verification framework, it is the owner of the dataset who has the power. So even when we know the API, we do not know the dataset or potentially the application.

Melling: That is where EDA comes in—the automation and expansion of the datasets, the improvement of the datasets. Some of the lessons learned from verification, and how to apply those to that problem, are going to be necessary to accelerate the quality of the algorithms.

Brinkmann: But the question remains—who is going to set the standards or expectations on this? So far, people in regulatory committees are still far away from perceiving it as a problem for verification. The dataset—that is where verification needs to go and that is where the real problems are.

Hansson: In academia, in some areas, they upload on Github datasets so they can do research on those and measure their algorithms, but that is not used for verification.

Foster: There is a lot of concern about bias in the datasets, and there is also aging of the dataset where we constantly have to drop data because the world evolves. This also has to be factored into the verification of the dataset.

Brian Bailey (all posts)
Brian Bailey is Technology Editor/EDA for Semiconductor Engineering.

MORE FROM BRIAN:

Defining Verification

By Brian Bailey - 25 Apr, 2019

There was a time when the notion of rigorous verification was seen as being unnecessary and even wasteful. I can remember early in my career working on flight control computers. We did no functional verification and created no models. We prototyped it and ran some engineering tests through it, primarily to structurally verify the system. We did not test the functionality of the system – that ... » read more

When Verification Leads

By Brian Bailey - 25 Apr, 2019

Semiconductor Engineering sat down to discuss the implications of having an executable specification that drives verification with Hagai Arbel, CEO for VTool; Adnan Hamid, CEO for Breker Verification; Mark Olen, product marketing manager for Mentor, a Siemens Business; Jim Hogan, managing partner of Vista Ventures; Sharon Rosenberg, senior solutions architect for Cadence Design Systems; and Tom... » read more

The Case For Embedded FPGAs Strengthens And Widens

By Brian Bailey - 25 Apr, 2019

The embedded FPGA, an IP core integrated into an ASIC or SoC, is winning converts. System architects are starting to see the benefits of eFPGAs, which offer the flexibility of programmable logic without the cost of FPGAs. Programmable logic is especially appealing for accelerating machine learning applications that need frequent updates. An eFPGA can provide some architects the cover they ne... » read more

SOLUTION BRIEF

Intel® Vision Products
AI Solutions for Manufacturing and Smart Cities

intel®

ADLINK and Touch Cloud Deliver AI Solutions Powered by Intel® Vision Products

AI catalyzes operational improvements for manufacturing and smart cities

"ADLINK is committed to bringing AI to the edge. We are taking a heterogeneous computing approach to AI deployments and building the right solution for our customer based on their computing needs. Using Intel® Vision Accelerator Design products with Intel® Movidius™ VPUs and the OpenVINO™ toolkit, Adlink's AI platforms can speed up image processing, computer vision, and deep learning inferencing with power efficiency not yet fully recognized in the market. With these advantages, ADLINK can better target edge applications and infuse the power of AI in vertical markets including manufacturing, transportation, healthcare, retail, smart city, and more. Specifically, by leveraging the OpenVINO™ toolkit we saw more than an 11x increase in performance1 on the CPU vs. running without the OpenVINO™ toolkit. Even more compelling is the 19x performance increase2 we saw when we added our own EDL-mPCIe-MA2485*, a mini-PCIe* accelerator card based on the Intel® Vision Accelerator Design."

—Edgar Chen, general manager, Embedded Platforms and Modules, ADLINK Technology

Executive summary

Video data is informing the next era of IoT, but aggregating, filtering, indexing, and classifying this data in near-real time requires advanced vision capabilities and technologies. Together, ADLINK, Touch Cloud, and Intel provide a turnkey AI engine to assist in data analytics, detection, classification, and prediction for a wide range of use cases. Smart city and manufacturing operations are prime examples of both the complexity and opportunities enabled by the convergence of vision capabilities, IoT, and AI solutions.

Challenges

Manufacturing facilities and smart cities share many of the features that make deploying IoT challenging across vertical sectors—from the complexity of gathering relevant, coherent data from a plethora of equipment, systems, devices, workflows, and facilities to cost-effectively filtering data for analysis at the edge or in the cloud to utilizing the resulting insight to take action in near-real time. Whether trying to evaluate operations across multiple factories or coordinate responses across diverse agencies, manufacturers and city planners require a holistic, accurate, and continual view of operations.

Industrial automation is a key facet of global manufacturing, with enterprises facing competitive pressures to maximize output while lowering costs. Factory operations are time sensitive and low latency is critical, as is achieving the right balance of power consumption and performance. Investments in legacy infrastructure, proprietary equipment that is not designed for compatibility, and the merger of OT and IT add to the complexity of implementing a cost-effective and manageable solution that meets compliance and deterministic requirements.

Likewise, today's cities must align a vast network of services, providers, agencies, equipment, and infrastructure in order to improve quality of life and increase efficiency. Legacy city systems are often decades old and unsuitable for rapidly increasing urban populations. Cities require timely, accurate data to ensure essential capabilities run without disruption, including emergency responders, police and fire departments, transit services, and utilities.

Video data offers valuable insight, but access to streaming video is only the beginning. In order to be relevant, and timely and inform strategy, video data must be accessible in near-real time, indexed, classified, and searchable.

ADLINK Leading EDGE COMPUTING

Touch Cloud Inc.

... from previous page

ADLINK & Touch Cloud Deliver AI Solutions Powered by Intel® Vision Products

Because this data also brings significant bandwidth demands, it also must be filtered for cost-effective transmission and actionable intelligence at the edge.

While AI unlocks the insight to address these challenges, no single solution can satisfy all applications.

Solution

Intel, ADLINK, and Touch Cloud are working together to bring inference acceleration to a wide range of industrial and commercial technologies.

In the integrated AI solution, ADLINK provides the optimized hardware platform and connectivity; Touch Cloud, the software application and analytics; and Intel, the IoT gateway processor, Intel® FPGA, and Intel® Movidius™ Myriad™ X vision processing unit (VPU), as well as the OpenVINO™ toolkit for smart vision application development. The end result allows implementation of AI and leverages legacy infrastructure, while achieving the benefits of IoT.

To enable AI and seamless integration with an Intel® architecture-based gateway at the edge, ADLINK's MXE- 210*—with its compact size, wide operating temperature range, and EMC-certified protection—can be easily installed in space-confined environments for reliable 24/7 operation. WLAN/WWAN support makes it possible for MXE-210 to communicate with back-end servers.

Intel® Vision Accelerator Design with Intel® Movidius™ VPU

Intel Vision Accelerator Designs with Intel® Movidius™ VPUs provide power-efficient deep neural network inference for fast, accurate video analytics. These accelerators are capable of operating on customizable complex networks and network layers with high compute and ultra-low power consumption, resulting in industry-leading performance/watt/$.
These VPUs support ecosystem solutions for high-quality image processing, computer vision, and deep neural networks. They drive a demanding mix of vision-centric tasks in modern smart devices. AI solutions can scale simply by adding Intel Vision Accelerator Designs with Intel Movidius VPUs, while still retaining their core efficiency. The elegant balance of performance and efficiency enables deployment for well-defined deep learning and machine vision workloads. Highly parallel programmable compute is co-located on a common intelligent memory fabric with workload-specific hardware acceleration.
ADLINK is currently utilizing Intel Vision Products for multiple AI solutions, and is working closely with Intel to ensure that the next generation of its deep learning accelerators and inference platforms leverage the power of the new Intel Vision Accelerator Design with Intel Movidius VPU. This accelerator design is expected to bring significant increases in the performance and processing of video data.

... to next page

... from previous page

ADLINK & Touch Cloud Deliver AI Solutions Powered by Intel® Vision Products

Addressing critical use cases

Touch Cloud AI models running on ADLINK's edge computing platforms powered by Intel architecture bring near-real-time analysis, indexing, and search to streaming video to support a breadth of IoT use cases.

Manufacturing

AI-enabled machine vision is addressing scenarios that cannot be tackled with rule-based machine vision. Solutions from ADLINK and Touch Cloud, powered by Intel® Vision Products, are able to mimic the human eye and brain, for instance, detecting manufacturing defects in hard-to-discern materials such as solar panels, textiles, wood, or food at high levels of accuracy for better quality assurance.

Automatic meter reading (AMR)

Unconnected meters are often inspected manually, requiring time-consuming and costly data collection. With the ADLINK MXE-210* inference platform and cameras at the edge, meters for industrial gas, oil, and other fluids can be read automatically. By sending the data to the back-end servers over wireless or wired networks, MXE-210 enables near-real-time remote monitoring of automated processes and industrial utilities, including water, electricity, and gas consumption.

AMR is ideal for use in oil and gas, petrochemical, water treatment, pharmaceutical, and food and beverage industries.

Meters

Cameras capture meter images and send them to the MXE-210.

MXE-210 with Intel® Movidius™ Myriad™ X VPU acts as an inference platform to identify the reading.

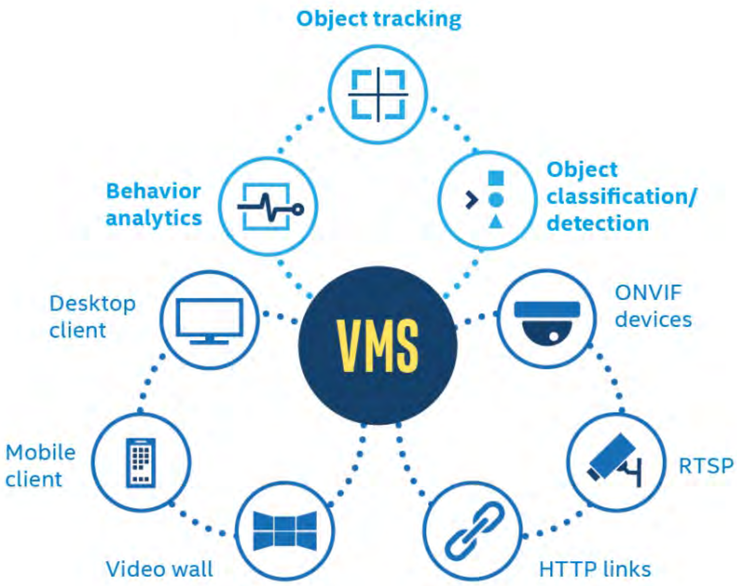
MXE-210 also serves as a gateway, sending the reading to back-end servers for data acquisition and near-real-time monitoring.

The meters can be monitored remotely.

Automated meter reading for manufacturing integrates AI for fast, accurate inspection

Image classification and segmentation
The AI solution leverages algorithms developed using the OpenVINO toolkit to classify images into predefined categories. In order to simplify and/or change the representation of an image into something that is more meaningful and easier to understand, image segmentation divides images into segments or regions belonging to the same class and category.

Automated optical inspection (AOI)
During the manufacturing process, defects that harm product yield and quality need to be detected and classified to protect production quality and reduce the cost of rework. AOI automates visual inspection of product manufacture (e.g., PCB, LCD, or transistor) via a camera that autonomously scans the device under test for quality defects, such as fillet size or shape or component skew. With ADLINK's high-performance expandable edge computing platforms, the solution improves near-real-time defect detection and identification of AOI machines and augments the defect classification capability with domain knowledge learned using AI.



Complex image classification and segmentation support fast indexing and searching of visual data streams

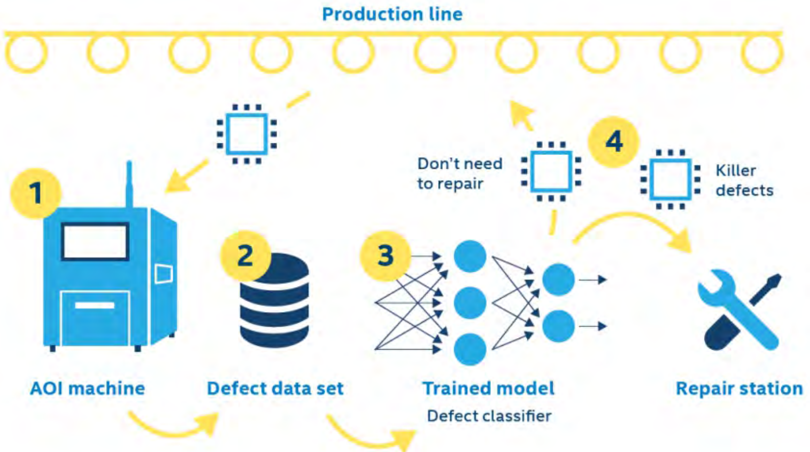
... to next page

... from previous page

ADLINK & Touch Cloud Deliver AI Solutions Powered by Intel® Vision Products

Defect classification

During the manufacturing process, defects can be introduced that impact product quality. It is necessary to classify the defects detected by AOI appropriately—the higher accuracy in classification, the less cost spent on rework. To perform fast inspection and classification of surface defects, AOI needs to execute near-real-time image processing via acceleration technologies, such as Intel® FPGAs.



Accurate detection of defects is critical to improving manufacturing efficiency and quality

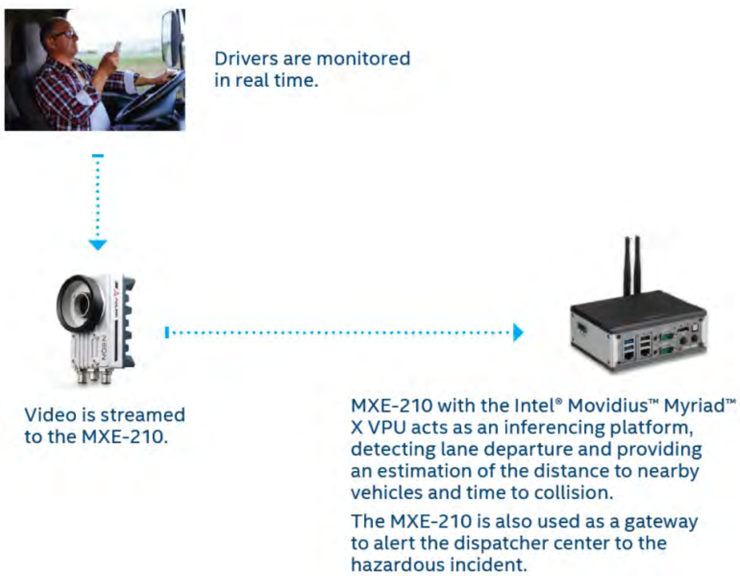
---

Smart cities

Traffic management solutions support vehicle and pedestrian detection for cities monitoring traffic flow, reducing congestion, and increasing citizen safety. With AI, fleet managers can easily be notified of hazardous driving. In these cases, the MXE-210 sends an alert to the dispatch center, which can implement driver training, reschedule the shift, or deploy any other necessary correction programs.

Dangerous driving behavior detection

Fleet managers can use dangerous driving behavior detection to identify hazardous driving behavior by bus and truck drivers due to distraction.



Dangerous driving detection uses AI to identify dangerous driving behavior due to distraction

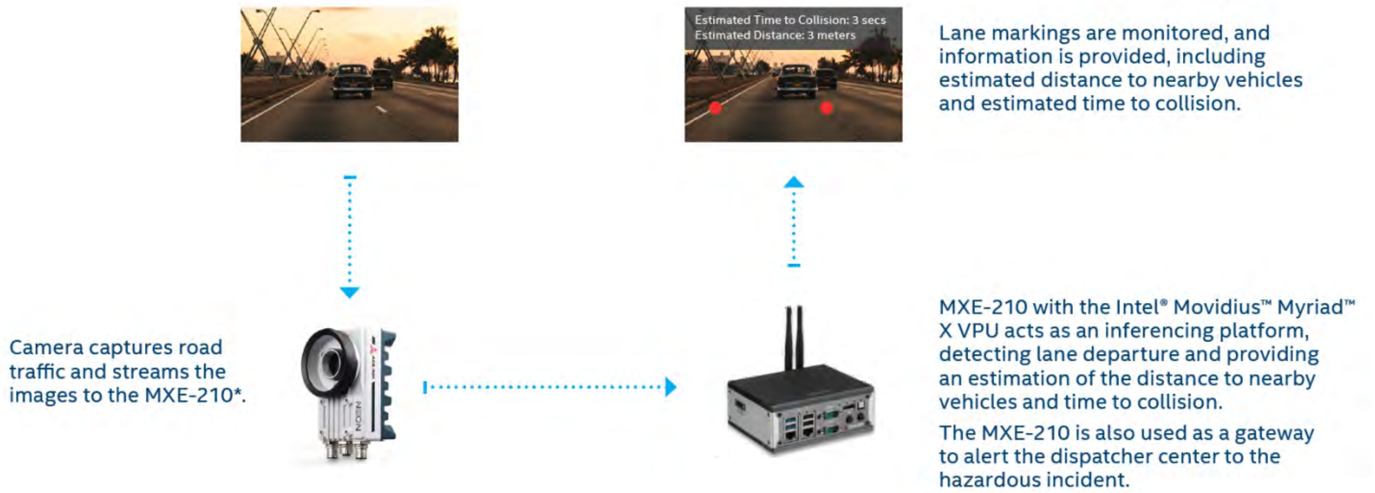
... to next page

... from previous page

ADLINK & Touch Cloud Deliver AI Solutions Powered by Intel® Vision Products

Lane departure detection

Lane departure detection allows fleet managers to monitor hazardous driving behavior due to distraction or fatigue. Lane departure detection solutions are ideal for use in fleet management for trucks, buses, and school buses.



Lane departure detection uses AI to monitor potential collisions and inform dispatchers

---

How it works

Edge compute helps customers turn massive volumes of machine- and device-generated data into actionable intelligence closer to the source of the data. The ability to filter raw edge data with minimal human intervention helps reduce the amount of data which must be stored and transmitted to downstream systems. It also helps reduce the impact of network latency and cost by taking the computing power from the server-side to the field and minimizing resources consumed.

Touch Cloud's AI software engine, developed using Caffe* and TensorFlow*, conducts video data analysis and processing via the ADLINK edge computing platform powered by the Intel Atom® processor. The result is near-real-time image and video streaming from ADLINK devices, coupled with detection, trigger event monitoring, and back-end system response.

Neural networks based on Intel FPGAs are designed and modified by Touch Cloud to meet specific manufacturing or smart city requirements with customization leveraging different AI and deep learning algorithms. Touch Cloud provides the visual recognition for detection, classification, and pose estimation tasks. Deep learning and AI model training improves performance and accuracy. Intel's OpenVINO toolkit enables Touch Cloud to convert trained AI models to run on the Intel® Core™ processors, Intel Atom® processors, and Intel Movidius Myriad X VPUs.

... to next page

ADLINK edge computing platforms enable seamless connection, aggregation, filtering, and data transmission to the cloud. Designed for ruggedized, embedded systems, edge computing platform features include:

- High performance per watt and per dollar
- Field protocol control interfaces
- Edge-to-cloud connectivity
- Function expansion for image acquisition, I/O control, and motion control
- Embedded components ensure long term availability

Intel® Vision Products integrate advanced software and hardware to capture complex, dynamic visual content from the edge to the cloud, with exceptional richness and accuracy. By delivering data processing flexibility at the edge—both in cameras and on-premise servers—as well as scalability in the cloud, these solutions are driving next-generation artificial intelligence and analytics, and enabling powerful deep learning inferencing capabilities across various industries.

... from previous page

ADLINK & Touch Cloud Deliver AI Solutions Powered by Intel® Vision Products

Open Visual Inference & Neural Network Optimization (OpenVINO) toolkit

This toolkit from Intel enables developers to easily integrate deep learning inference into their applications using industry-standard AI frameworks and standard or custom layers.3 These can then be deployed across the continuum of Intel-based product lines—from camera to cloud— irrespective of the target platform on which they will be run.2 With OpenVINO, developers can write code once and make it future-proof for fast, seamless deployment across current and future Intel® hardware—eliminating application redevelopment.4 Based on convolutional neural networks (CNN), the toolkit extends workloads across Intel® hardware and maximizes performance.

Intel FPGAs

The Intel FPGAs provide the flexibility to program and reprogram the rate of acceleration as needs evolve, all with less power than can be obtained with fixed-function GPUs. They deliver hardened floating point digital signal processing to speed complex machine learning algorithms using parallel processing.

Intel® Movidius™ VPUs

Intel Movidius VPUs are designed to bring vision technology out of the data center and into devices at the edge. Intel Movidius VPUs have a dedicated architecture that offers power efficiency for high-quality image processing, computer vision, and deep neural networks, making them suitable for the demanding mix of vision-centric tasks in modern smart devices. These VPUs are ideal for conditions constrained by size, power, and cost, offering the optimal balance of power efficiency and high computing performance.

ADLINK

ADLINK offers emerging edge computing platforms that converge data and video processing to support AI applications. Its solutions integrate Intel® processors and Intel Movidius Myriad X VPU to deliver hardware-accelerated deep learning processing with high performance per watt and per dollar, end-to-end connectivity to break down information silos for data-to-decision application enablement, and industrial environmental compliance to meet the extended life-cycle requirements of industrial and urban applications.

The company's robust, reliable deep learning accelerators and platforms ensure system compatibility and solution scalability to enable seamless transition to AI integration for operational improvements, performance boost, and efficiency gains across industries. ADLINK builds intelligent middleware for remote monitoring into all of its boards and modules and offers certified solutions for highly regulated markets.

Touch Cloud

Touch Cloud creates AI software for industry 4.0 and surveillance using deep learning, providing massive and high-dimensional numerical data analysis for automated plants and cities. The company's AI processes leverage continual streaming data at the time and place when it is acquired. Edge capabilities eliminate the roundtrip to the back-end server, reducing latency and accelerating response. The AI solution provides near-real-time insights, so that operators can make decisions faster, more efficiently, and with more accurate information. The solution enhances security, since data is not transported across the network or stored in data centers, and helps industrial facilities and city agencies to optimize costs.

The Touch Cloud solution supports modern scenarios requiring real-time observation and response times in milliseconds (such as defect detection) that cannot be achieved with a traditional server-based AI setup. Touch Cloud runs AI on ADLINK edge devices, with high levels of performance and accuracy, to help facilitate abnormality prediction and root cause analytics.

Conclusion

Video is a critical source of AI data—providing rich, near-real-time data that is both ongoing and contextual. For today's organizations, gathering streaming video data is insufficient—they need to be able to quickly analyze, index, and search.

ADLINK, Touch Cloud, and Intel are helping a broad spectrum of vertical sectors, such as manufacturers and smart cities, attain connected, fault-free performance, actionable intelligence, and transformative insight.

"The Touch Cloud AI engine benefits IIoT, bringing cost savings, operational efficiency, and more reliable, high-quality defect classification to critical manufacturing processes."
—Simon Lee, CEO, Touch Cloud

Learn More

See the latest Intel® Vision Products at intel.com/visionproducts.
Download the OpenVINO toolkit for designing computer vision solutions at software.intel.com/en-us/openvino-toolkit.
For more information about Intel IoT Technology and the Intel® IoT Solutions Alliance, please visit intel.com/iot.
Discover ADLINK IIoT, smart vision, and AI solutions at adlinktech.com or contact us at info@adlinktech.com.
Explore Touch Cloud AI solutions and applications at touchcloud.com.tw or contact us at info@touchcloud.com.tw.