DIGITAL EBOOK

POWER UP YOUR VOICE AND VISION AI DEVELOPMENT

Artificial Intelligence (AI) is transforming every aspect of life, revolutionizing the business environment and making our lives easier with more products enabled by voice / speech recognition and vision AI. The global AI voice generator market size is expected to have a CAGR growth of more than 15%*, while the computer vision AI market should see a growth rate of 21.5%**. As this market expands, more AI-embedded products are required at a faster speed, however, with the constant technology enhancements, it takes time to learn new AI algorithms and deploy that innovation into end products.

Aligning with these market needs, Renesas equips designers with the right tools for Voice User Interface (VUI) and Vision AI solutions based on Arm-based RA and RISC-V microcontrollers and RZ microprocessors, along with easy-to-use software to simplify development. Renesas also collaborates with a wide range of ecosystem partners to provide cost-efficient turnkey solutions for immediate deployment to easily implement AI functions for product development.

We hope to get your exploration started with this collection of articles from Renesas. AI and machine learning can help solve problems in new ways providing benefits for users and we look forward to being part of your innovation journey.

WHAT'S INSIDE

02 Give Voice to Smart Products

06 No Internet Connection for that Smart Device Design? Not a Problem with Easy-to-Deploy VUI Solutions

08 RISC-V-Based ASSP EASY for Voice HMI – The Journey Continues

10 Revolution of Endpoint AI in Embedded Vision Applications

14 Take the First Step in Vision AI Development

19 Embedded AI-Accelerator DRP-AI

24 AI Applications and AI SDK for RZ/V Series MPUs

25 RZ/A2M DRP Realizes Efficient Masked Face Detection

and more >>

Give Voice to Smart Products

> *It is not only about **what** you say. It is also about **how** you say it.*

This old-age adage quite aptly sums up the need for human beings to communicate effectively with each other. The necessity of humans to interconnect with one another through voice and sounds has presented a future where communication with machines has become inevitable.

A key enabler for the increasing adoption of voice communication has been accelerated with the expansion of the Internet of Things and artificial intelligence. Integration of AI at the endpoint combined with advances in voice analytics is changing the availability of products and the consumption of product experiences and giving rise to a new ecosystem of companies that are participants and enablers of these products. Intelligent endpoint solutions are making it possible to implement both online and offline systems, reducing reliance on always-on internet/ cloud connections. This in turn is creating new opportunities to solve many challenges related to real-time voice analytics across several consumer and industrial applications.

The advances in psycholinguistic data analytics and affective computing make allowance for inferring emotions, attitudes, and intent with data-driven modeling of voice. With the voice medium becoming a natural way for humans to interact, it will lead to improvements in measuring intent from voice recognition and voice analytics.

Challenges of Using VUIs

Voice user interfaces (VUIs) allow the user to interact with Endpoint systems through voice or speech commands. Despite mass deployments across a wide range of applications, VUIs have some limitations.

- **Poor sound quality -**
  Inconsistent sound quality with continued background noise can make voice recognition a challenge. Voice controllers in IoT can only operate flawlessly if the sound is crystal clear which is a formidable task in a noisy environment. A voice-enabled assistant can only be truly effective if it is able to support different languages and accents, as well as isolate the human voice from the background noise.
- **Power consumption -**
  Voice Command systems are restrictive as they require the activation of at least one microphone as well as the processor that recognizes the wake word.
- **Real-time processing -**
  Slow or congested networks can result in command latencies that can impact the user experience. This issue may be addressed by implementing distributed intelligence at the endpoint with the ability to process the voice command in real time without any reliance on the centralized cloud system.
- **Accuracy and noise immunity -**
  Voice recognition accuracy and background noise immunity are always major concerns when designing any VUI system. Voice recognition presents a number of challenges as there can be multiple sound sources including interior and exterior noise and echoes from surfaces in the room etc. Isolating the source of a command, canceling echoes and reducing background noise requires some sophisticated technology depending on multiple microphones, beamforming and echo cancellation, along with noise suppression.

Give Voice to Smart Products CONTINUED

Renesas Electronics is addressing these challenges by using state-of-the-art microcontrollers and partner-enabled intelligent voice processing algorithms which makes it easier for product manufacturers to integrate highly efficient voice commands. Renesas Electronics is providing general purpose MCUs enabling cost-optimized VUI integration without compromising on performance and power consumption.

Requirements for Robust Voice Recognition

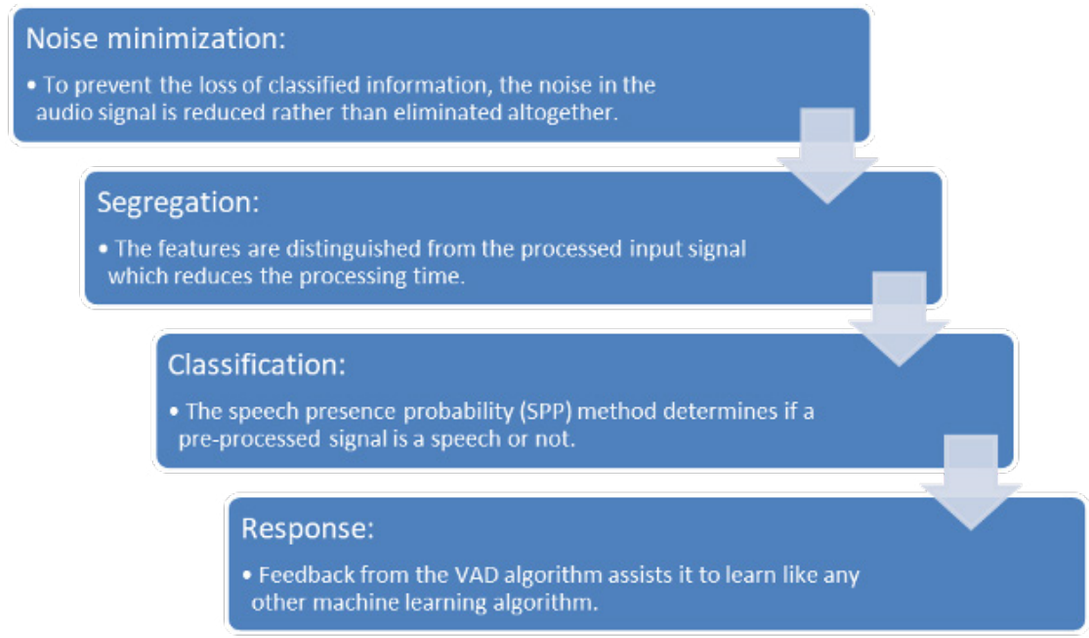
To make the experience compelling for the user, devices need to be equipped with several components to ensure robust voice recognition.

Command Recognition

One of the most significant features of a voice-enabled device is its ability to identify speech commands from an audio input. The speech command recognition system on the device is activated by the wake word which then takes the input, interprets it, and transcribes it to text. This text ultimately serves the purpose of the input or command to perform the specific task.

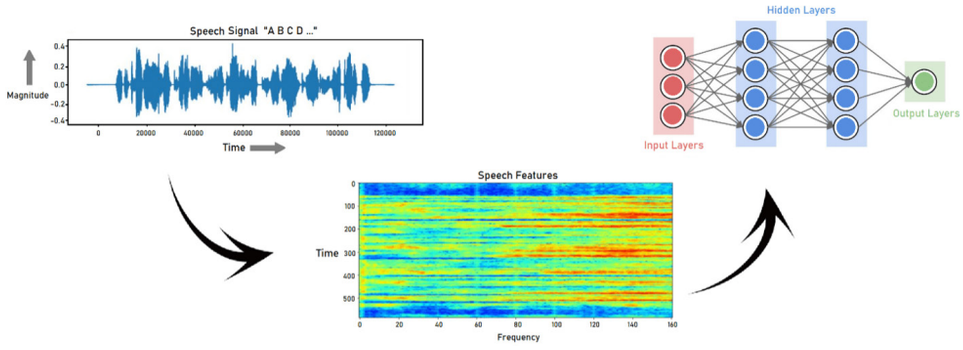
Voice Activity Detection

Voice activity detection (VAD) is the process that distinguishes human speech from the audio signal and background noise. VAD is further utilized to improve the optimization of overall system power consumption otherwise the system needs to be active all the time resulting in unnecessary power consumption. The VAD algorithm can be subdivided into four stages:

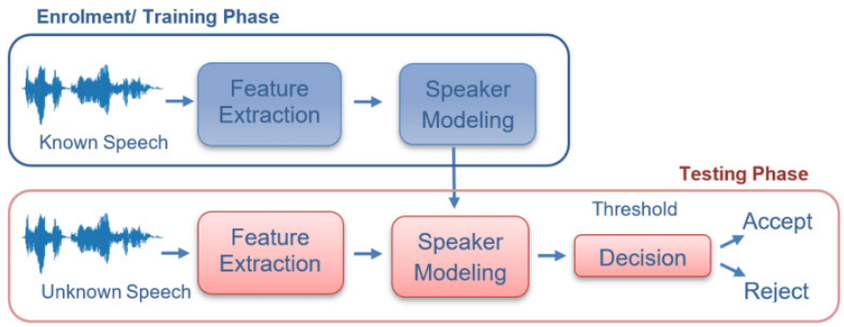


The Renesas RA voice command solution built on the RA MCU family and partner-enabled voice recognition MW boasts a robust noise reduction technique that helps in ensuring high accuracy in VAD. In addition, Renesas can help to address some of the key voice command features outlined below:

Give Voice to Smart Products CONTINUED



Keyword spotting process



Block diagram of a speaker identification system

Keyword Spotting

Keyword spotting systems (KWS) are one of the key features of any voice-enabled device. The KWS relies on speech recognition to identify the keywords and phrases. These words trigger and initiate the recognition process at the endpoint, allowing the audio to correspond to the rest of the query.

To contribute to a better hands-free user experience, the KWS is required to provide highly accurate real-time responses. This places an immense constraint on the KWS power budget. Therefore, Renesas provides partner-enabled high-performance optimized ML models capable of running on our advanced 32-bit RA microcontrollers. They come with pre-trained DNN models which help in achieving high accuracy when performing keyword spotting.

Speaker Identification

Speaker identification, as the name suggests, is the process of identifying which registered speaker has the given voice input. Speaker recognition can be classified as text dependent, text independent, and text prompted. To train the DNN for speaker identification, individual idiosyncrasies such as dialect, pronunciation, prosody (rhythmic patterns of speech), and phone usage are obtained.

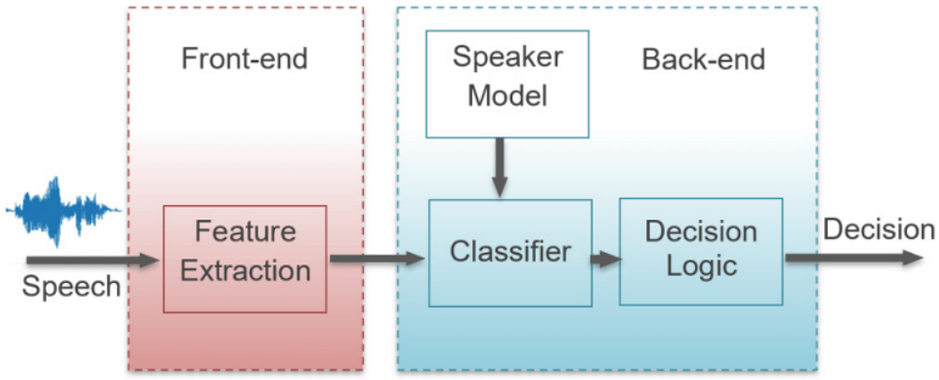
Voice/Sound Anti-Spoofing

Spoofing is a type of scam where the intruder attempts to gain unauthorized access to a system by pretending to be the target speaker. This can be countered by including anti-spoofing software to ensure the security of the system. The spoofing attacks are usually against Automatic Speaker Verification (ASV) systems. The spoofed speech samples can be generated using speech synthesis, voice conversion, or by just replaying recorded speech. These attacks can be classified as direct or indirect depending on how they interact with the ASV system.

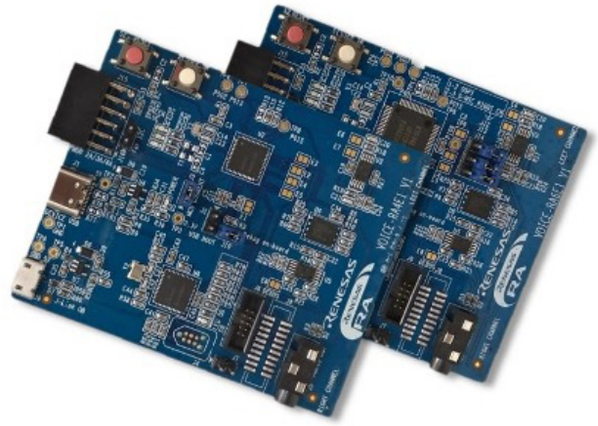
- **Direct Attack –** This can occur through the sensor at the microphone and transmission level and is also known as Physical Access.
- **Indirect Attack –** This is an intrusion into the feature extraction, models, and the decision-making process of the ASV system software and is also known as Logical Access attack.

Give Voice to Smart Products CONTINUED

Block representation of an automatic speaker verification



Renesas Voice User Interface solution boards



Multi-Language/Accent Recognition and Understanding

Accent recognition in English-speaking countries is a much smoother process due to the availability of training data, hence accurate predictions. The downside for organizations operating in countries where English is not the first language is less precision with speech recognition due to the availability of a limited amount of data. An inadequate amount of training data makes building conversational models of high accuracy challenging.

One of the Renesas VUI partner-enabled solutions supports more than 44 languages making it a highly adaptable speech recognition solution that can be used by any organization worldwide.

Renesas invites you to take advantage of this opportunity as it implements a comprehensive solution platform that simplifies voice integration adoption based on our advanced microcontrollers.

Our RA MCU-based and partner-enabled voice command solutions offer not only low BOM costs, integrated security, and low latency, but also provide the capability to run at an endpoint. We provide value-added features like local voice triggers, command recognition, robust noise reduction, voice activity detection, and multi-language support just to name a few. This is complemented by a fully featured voice solution development suite and endpoint capable MW for its partners that provide a DNN-based pre-trained voice model where the user can quickly create voice commands with a simple text input. In short, it can be used as it is or tailored to meet individual needs!

No Internet Connection for that Smart Device Design? Not a Problem with Easy-to-Deploy VUI Solutions

How many consumers crave a cloud-connected microwave? An Internet-enabled toaster? A toothbrush controlled with a mobile app? I think we all know the answer is – not many.

But rapid consumer adoption of connected smart speakers and voice assistants suggests something else; many people prefer to control products with speech rather than counter-intuitive button pushes. And many consumers also seem to appreciate devices and home appliances that are smart without being connected. After all, who wants to wait for an Internet connection to come back online just so you can pop some microwave popcorn?

The uptick in devices with edge or endpoint voice user interfaces (VUI) is growing along with the increasing sophistication of intelligence at the edge and the endpoint. You can read more on this in a recent blog by Renesas executive vice president Sailesh Chittipeddi. In other words, if your product needs a few good command phrases, all the processing can be done on the device.

But designing home-automation products with an optimal VUI can be tricky for engineers who don't already have deep expertise in voice systems. And for manufacturers without in-house design experience, outsourcing the design and development can be cost-prohibitive. That's why over the past few years, Renesas has teamed with industry-leading companies like Cyberon, DSP Concepts and Sensory to offer easy-to-implement software solutions that run on our cost-effective Renesas Advanced (RA) family of 32-bit MCUs. If your company wants, say, washing machines and refrigerators that understand simple voice commands, the Renesas VUI solution is your one-stop design, service and support shop.

Let Us Show You How

The Renesas VUI solution, a third-generation reference platform built on our scalable Arm-based RA2, RA4, and RA6 series of MCUs, is especially targeted to applications in home automation, white goods, and small appliances. Voice-enabled examples include ovens, refrigerator alarms, coffee makers, dishwashers, ceiling fans, and more. The VUI enables these products to respond in a standardized way to a user's voice command. This standardization is a unique approach to help developers avoid writing new code for each product configuration.

A FULL SUITE OF RA MCU EVALUATION KITS

Renesas RA microcontroller kits enable users to effortlessly evaluate the features of different RA MCU Groups and easily develop sophisticated IoT and embedded systems applications.

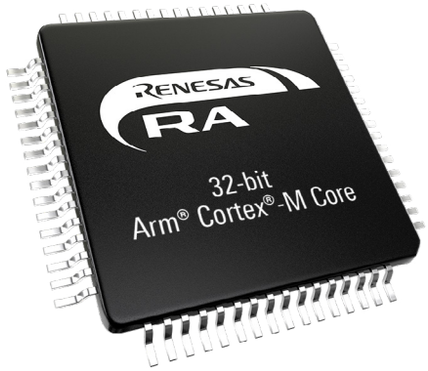
The kits are based on a novel architecture that provides an unparalleled combination of standardization and flexibility. The kit design helps users shorten the learning curve and accelerate development, providing more time for differentiated innovation or taking products to market faster. Users can utilize rich onboard features along with their choice of popular ecosystem add-ons to bring their big ideas to life.

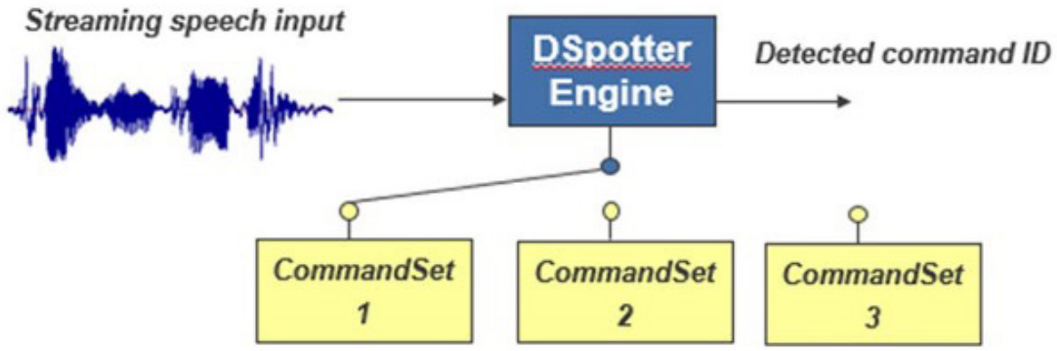
GET AN RA KIT AND START YOUR NEXT DESIGN TODAY

No Internet Connection for that Smart Device Design? Not a Problem with Easy-to-Deploy VUI Solutions CONTINUED

To see how simple it can be to incorporate an advanced voice user interface reference design, check out DSP Concepts' RA6M5 advanced VUI implementation webinar. In this webinar, the moderator demonstrates how microphone and MCU hardware can be combined with far-field VUI software and Sensory's wake word and command sets to capture and process voice – offline – even in noisy environments. For a preview, take a look at the reference design developed by DSP Concepts.

Renesas and Cyberon present an equally useful on-demand webinar about implementing the powerful Cyberon DSpotter recognition engine on RA MCUs to enable VUIs on devices – such as smart-home light switches – that are typically compute-, memory- or power-constrained. Cyberon's local voice trigger and command recognition solution features robust noise reduction, low-power consumption and highly accurate performance. See relevant product briefs and the Cyberon solution page for more details.



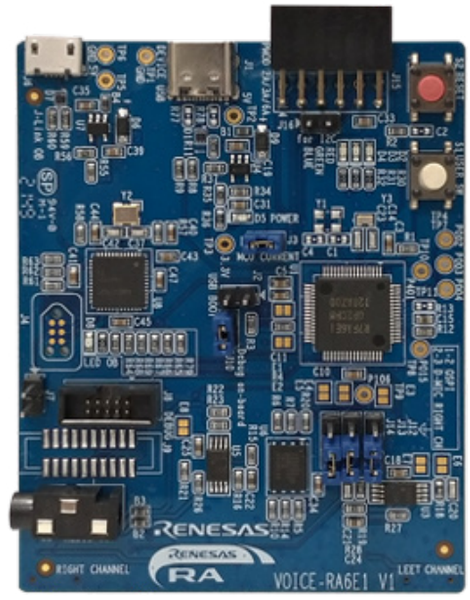


On-Demand Virtual Lab Instruction

After viewing the webinars, perhaps you'll be ready to create a voice-recognition project with DSpotter using your own commands. Visit the Voice Recognition page to request a Voice kit and then watch the virtual hands-on lab for step-by-step instruction on how to add a VUI in minutes.

In the lab, you'll experience the Cyberon DSpotter software library and create a project from scratch with Cyberon tools. You'll use one of our newest Voice kits and enable the interface for a digital microphone.

Whether you're just starting out with voice or want to customize the latest set of commands for your product, the Renesas VUI solution gives you access to a variety of developer tools and training resources that will take your product designs to the next level. Visit the Voice Recognition page to learn more and get started on your design.



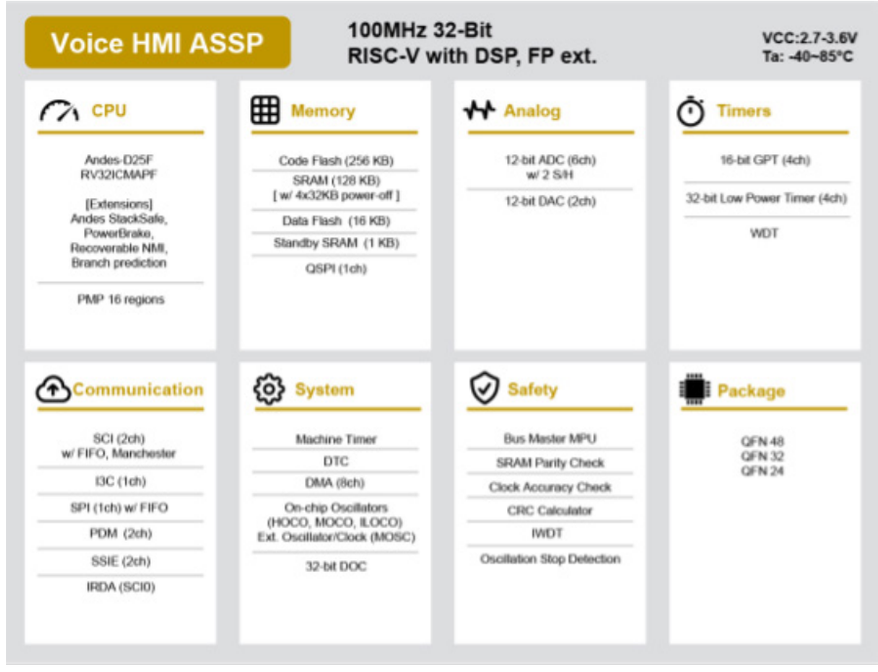
RA6E1 Voice Kit Board

RISC-V-Based ASSP EASY for Voice HMI – The Journey Continues

Human-machine interfaces based on voice commands provide a convenient way to interact with systems like appliances, displays, home accessories, and other gadgets and can replace or augment traditional means of control like buttons, knobs, or sliders. The recent worldwide pandemic has obviously increased the attention for such touch-free interaction methods but even without such urgency, for modern systems it is very desirable to implement such advanced controls to improve the end user experience.

Implementing such technologies, however, requires specialized expertise in many complementary fields of embedded systems development. For such applications, the environmental audio content is acquired in real time and the stream of samples gets processed to extract and filter relevant information to recognize specific patterns. For example, to recognize a pattern associated with a certain spoken phoneme, in a specific language. The further composition of those patterns and sequences can be associated with the likeliness of a specific keyword being recognized within the acquired audio stream. At such moments, the system designer might decide to have detected a keyword he was actively listening for and decide to perform further actions or give the user feedback. What those actions will be is totally dependent on the specific appliance. Implementing the application layer to input the audio samples and manage the results is something that an embedded software engineer is comfortable developing.

On the other hand, the method used to detect those keywords within the audio stream is hard to implement without specific expertise. One approach is to use neural network-based software models which get trained to recognize specific phonemes with a desired accuracy and the application can use those to perform keyword detection. Although much



RISC-V Voice ASSP block diagram

RISC-V-Based ASSP EASY for Voice HMI – The Journey Continues CONTINUED

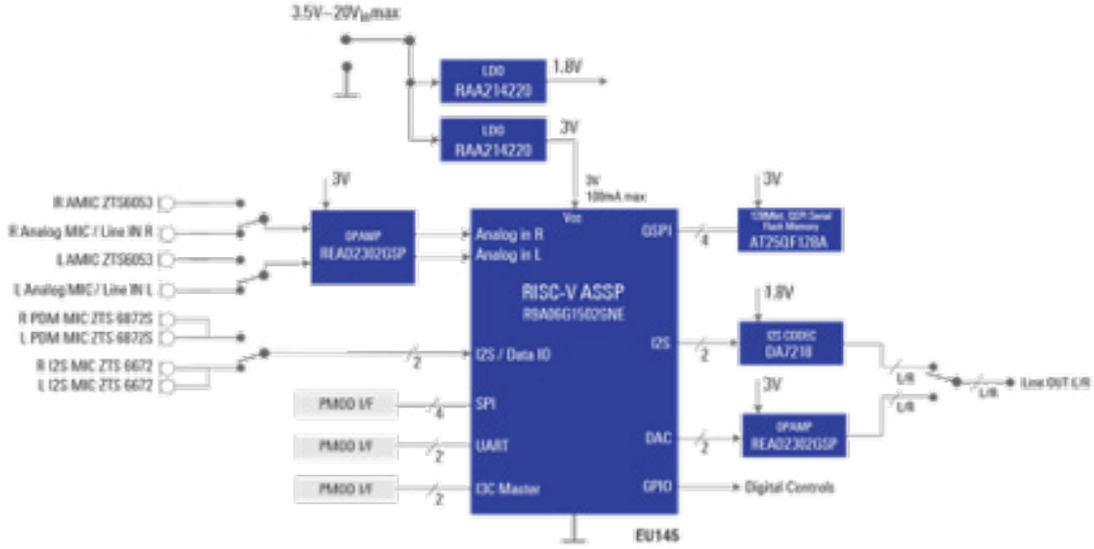
technical literature is available on the subject, it is not trivial to develop, train, and deploy such e-AI models to implement the keyword spotting functionality. It can require a significant investment to fine-tune and obtain the desired accuracy to prepare a prototype for mass production.

For most users, it would be ideal if this technology could be easily retrofitted into existing applications, to simply enjoy the benefits of this emerging HMI technology without prohibitive levels of development effort.

RISC-V as an emerging free and open ISA is creating a lot of momentum within the engineering community and Renesas is leading commercial adoption within the embedded MCU space. The release of the latest RISC-V ASSP for voice HMI provides an ideal platform for the above-mentioned applications.

The ASSP hardware integrates a CPU equipped with instructions suitable for high-performance digital signal processing, plenty of memory, and several peripheral interfaces ideal to interface with digital and analog microphones, provide audio feedback output, and implement further control possibilities.

The hurdle of developing the required embedded software is eliminated by cooperation with trusted Renesas partners like Orbstar and Cyberon which have deep and specific expertise in the field and proven capability to deploy such technology for a variety of applications in consumer and industrial spaces. Those partner companies were the first to use the Renesas ASSP hardware platform to offer their solutions, and others will follow.



RISC-V ASSP EASY Voice HMI Reference Design

From the end user perspective, the choice to make is about the supported language, defining the keywords to be recognized, and the desired resulting behavior. Our partner companies can customize their production-ready solution according to user needs.

Interfacing to the ASSP solution can be done with simple serial interfaces that make it easy to connect the solution to an existing system and naturally extend its functionality without the need to radically redesign, modify and re-test an existing application. At the system level, no connectivity is necessary to implement the functionality, making this solution an ideal add-on option for any type of appliance. An evaluation kit is supported by sophisticated but easy-to-use PC tools to test and monitor the performance to ensure it matches expectations. Additionally, many complementary analog and digital devices from Renesas are suggested in the form of a complete reference design to be used in combination with this solution.

Visit renesas.com/risc-v to find out more.

Revolution of Endpoint AI in Embedded Vision Applications

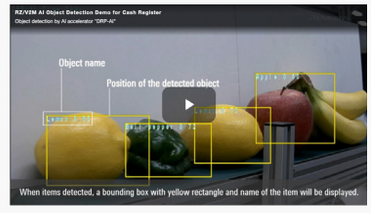
Endpoint AI is a new frontier in the space of artificial intelligence that takes the processing power of AI to the edge. It is a revolutionary way of managing information, accumulating relevant data, and making decisions locally on a device. Endpoint AI employs intelligent functionality at the edge of the network, in other words, it transforms the IoT devices that are used to compute data into smarter tools embedded with AI features. This in turn improves real-time decision-making capabilities and functionalities. The goal is to bring machine-learning based intelligent decision-making physically closer to the source of the data. In this context, Embedded vision shifts to the Endpoint. Embedded vision incorporates more than breaking down images or videos into pixels - it is the means to understand pixels, make sense of what is inside and support making a smart decision based on specific events that transpire. There have been massive endeavors at the research and industry levels to develop and improve AI technologies and algorithms.

What is Embedded Vision?

Embedded computer vision is a technology that imparts machines with the ability to see, i.e., the sense of sight, which enables them to explore the environment with the support of machine-learning and deep-learning algorithms. There are numerous applications across several industries whose functionality relies on computer vision, thus becoming an integral part of technological procedures. In precise terms, computer vision is one of the Artificial Intelligent (AI) fields that enables machines to extract meaningful information from digital multimedia sources to take actions or make recommendations based on the information that has been obtained. Computer vision is, to some extent, akin to the human sense of sight. However, the two differ on several grounds. Human sight has the exceptional ability to understand many and varied things from what it sees. On the other hand, computer vision recognizes only what it has been trained on and what it is designed to do exactly, and that too with an error rate. AI in embedded vision processes trains the machines to perform supposed functions with the least processing time and has an upper edge over human sight in analyzing hundreds of thousands of images in a lesser timeframe.

Embedded vision is one of the leading technologies with embedded AI utilized in smart endpoint applications in a wide range of consumer and industrial applications. There are a number of value-added use cases examples; such as counting/analyzing the quality of products on a factory line, keeping a tally of people in a crowd, identifying objects, and analyzing the contents of a specific area in the environment to name a few.

Watch the video on Renesas website to learn more about the **RZ/V2M AI Object Detection Demo for Cash Register**



Revolution of Endpoint AI in Embedded Vision Applications CONTINUED

While considering the processing of embedded vision applications at the endpoint, the performance of such an operation may face some challenges. The data flow from the vision sensing device to the cloud for the purposes of analyzing and processing could be very large and may exceed the network available bandwidth. For instance, a camera with 1920 x1080 and operating with 30FPS (Frame Per Second) may generate about 190MB/S of data. In addition to privacy concerns, this substantial amount of data contributes to latency during the round trip of data from the edge to the cloud, then back again to the endpoint. These limitations could negatively impact the employment of embedded vision technologies in real-time applications.

IoT Security is also a concern in the adoption and growth of embedded vision applications across any segment. In general, all IoT devices must be secured. A critical issue and concern in the use of smart vision devices is the possible misuse of sensitive images and videos. Unauthorized access to smart cameras, for example, is not only a breach of privacy, but it could pave a way for a more harmful outcome.

Vision AI at the Endpoint

- Endpoint AI can enable image processing to infer a complex insight from a huge number of captured images.
- AI uses machine-learning and deep-learning capabilities within smart imaging devices to check a huge amount of previously well-known use cases.
- For optimum performance, Embedded vision requires AI algorithms to run on the endpoint devices and not transmit data to the cloud. The data here is captured by the imaging recognition device, then processed and analyzed in the same device.

The limitation of power consumption at the endpoint is still available, where the microcontrollers or microprocessors need to have more efficiency to take on high volumes of multiply-accumulate (MAC) that are required for AI processing.

Deployment of AI Vision Applications

There are unlimited use cases for the deployment of AI in vision applications in the real world. Here are some of the examples where Renesas can provide comprehensive MCU and MPU-based solutions inc. all the necessary SW and tools to enable quick development.

RENESAS READY PARTNER NETWORK

This trusted network of ecosystem partners provides pre-developed software and hardware building blocks to shorten your design cycle. All solutions include easy to understand collateral and demonstration projects and work out-of-the-box with Renesas MCU and MPU products to solve real-world problems. They are revised to keep up with every major release of the Renesas software platforms and tools.

- Expansive third-party solutions portfolioCommercial-grade building blocks
- Designed for problem solving

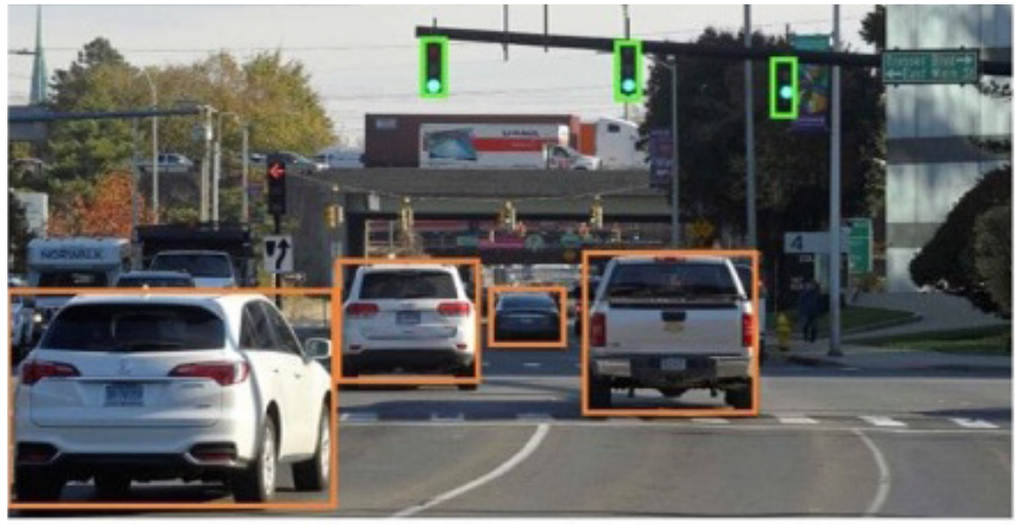
Scale your design with the cohesive, robust network of more than 200 trusted partners providing more than 300 solutions across Renesas MCU and MPU platforms.

EXPLORE THE NETWORK

Revolution of Endpoint AI in Embedded Vision Applications CONTINUED

Smart Access Control:
Security access control systems are becoming more valuable with the addition of voice and facial recognition features. Real-time recognition requires embedded systems with very high computational capabilities and on-chip hardware acceleration. To meet this challenge, Renesas provides a choice of MCU or MPU that offers very high computational power that also integrates many key features that are critical to high-performance facial and voice recognition systems such as built-in H.265 hardware decoding, 2D/3D graphic acceleration, and ECC on internal and external memory to eliminate soft-errors and allow for high-speed video processing.

Industrial Control:
Embedded vision has a huge impact as it utilizes many applications, including product safety, automation, sorting of products, and many more. AI techniques can perform multiple operations in the production process such as packaging and distribution, which can ensure quality and safety during production in all stages. Safety is needed in areas such as critical infrastructure, warehouses, production plants, and buildings that require a high level of human resources.



Transportation:
Computer vision presents a large scale of ways to improve transportation services. In self-driving cars as an example, computer vision is used to detect and classify objects on the road. It is also used to create 3D maps and estimate movement around. By using computer vision, self-driving cars gather information from the environment using cameras and sensors, which then interpret and analyze the data to make the most suitable response by using vision techniques such as pattern recognition, feature extraction, and object tracking.

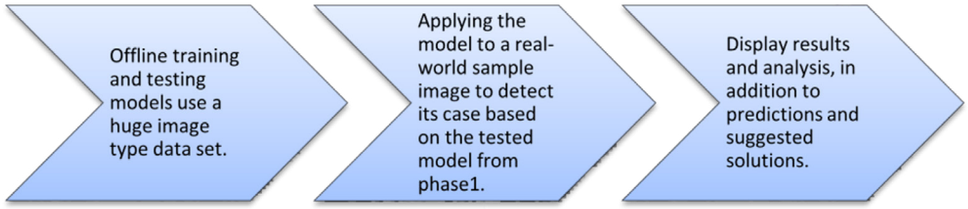
In general, Embedded vision can serve many purposes, and these functionalities can be used after customization and the needed training on different types of datasets from many areas. Functionalities include monitoring physical area, recognizing intrusion, detecting crowd density, and counting humans or objects, or animals. They also include identifying people or finding cars based on car numbers, detection of motion, and human behavior analysis in different cases.

Case study: Agricultural Plant Disease Detection

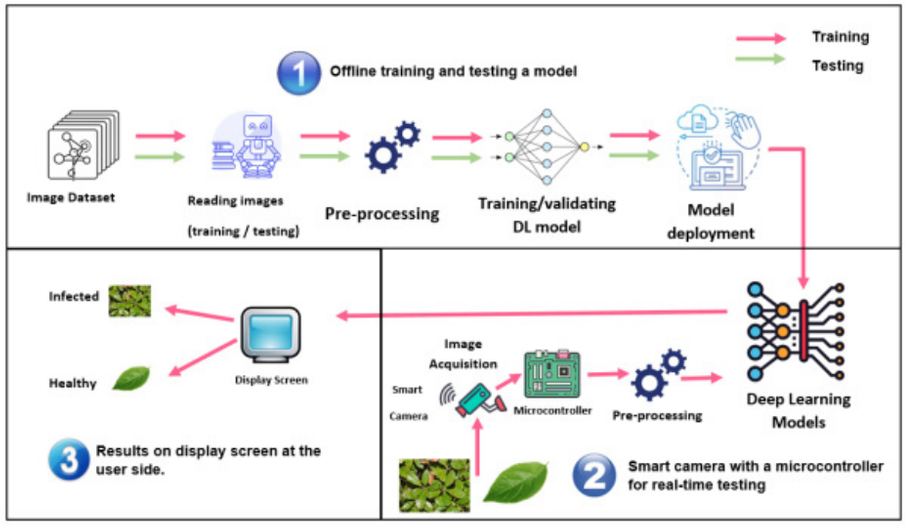
Vision AI and deep learning may be employed to detect various anomalies - plant disease detection is one example of this type of system. Deep learning algorithms -one of the AI techniques- are used widely for this purpose. According to research, computer vision gives better, accurate, fast, and low-cost results, as compared to the costly and slow labor-intensive results of previous methods.

Revolution of Endpoint AI in Embedded Vision Applications CONTINUED

The process that is used in this case study can be applied to any other detection. There are three main steps for using deep learning in computer/machine vision:



Step one is performed on normal computers in the lab, whereas step two is deployed on a microcontroller at the endpoint, which can be on the farm. Results in step three are displayed on the screen on the user side. The following diagram shows the process in general.



We are experiencing a revolution in high-performance smart vision applications across a number of segments. The trend is well supported by the growing computational power of microcontrollers and microprocessors at the endpoints, opening up great opportunities for exciting new vision applications. Renesas Vision AI solutions can help you to enhance overall system capability by delivering embedded AI technology with intelligent data processing at the endpoint. Our advanced image processing solutions at the edge are provided through a unique combination of low power, multimodal, multi-feature AI inference capabilities. Take the chance now and start developing your vision AI application with Renesas Electronics.

Take the First Step in Vision AI Development

Have you ever been asked, "Couldn't you use AI to improve the performance of our system?"

Or, "Can you add more value to our equipment by incorporating AI into it?"

Artificial intelligence, deep learning, neural networks...the AI applications are expanding exponentially and there is not a day that goes by without hearing these words. Also, there are many people out there who are interested but don't quite understand how to use AI in their work, and are not aware of the values it brings to their work.

This article will introduce AI evaluation using the software package of Renesas' embedded AI processor - RZ/V series.

What is AI?

Artificial Intelligence (AI) has been explained in various ways in many books and websites, such as AI can play chess like a pro or turn on the lights in a room through voice activation. But, I would like to focus on the narrowest definition of AI: artificial intelligence that uses neural network technology, specifically – Vision AI – that can perform various recognitions and judgments on images.

Neural network

As shown in *Figure 1*, the nerves in the brain of an organism receive inputs from multiple other neurons and transmit them to the next neuron. It is known that memory, recognition and judgment are based on a combination of two types of signals: how strongly the signal from the input side is used (weighting) and how the sum of each input after weighting is transmitted to the output side (activation function).

AI using neural networks imitates this mechanism and performs various recognitions and judgments by using an enormous amount of arithmetic processing, such as weighting multiple inputs, summing them up, and passing the results through an activation function to the next stage.

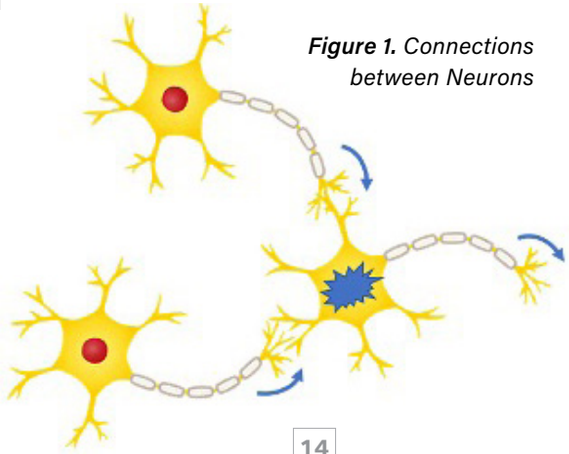


Figure 1. *Connections between Neurons*

For example, the 3x3 convolutional operation, which is often used in Vision AI to quantify the features of an input image. The image is first subdivided into images of 3 pixels x 3 pixels. Each value (pixel density) is weighted, and the sum of the values is made into an output value using an activation function and sent to the next stage. (*Figure 2*)

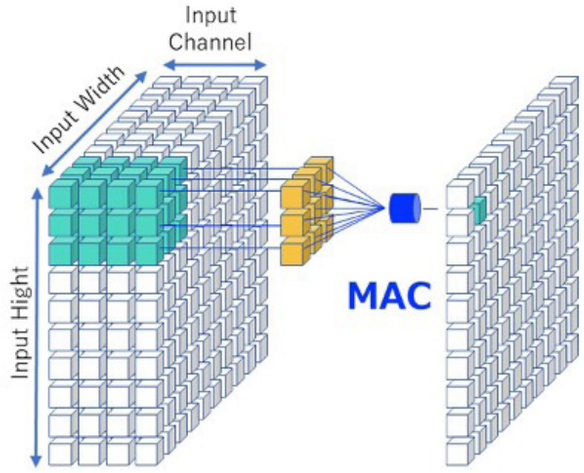


Figure 2. *Example of a 3x3 Convolutional Operation by a Neural Network*

Take the First Step in Vision AI Development CONTINUED

This process is first carried out over the entire surface of the input image, and then the next image, which was created by the output (so it is not exactly the data that can already be seen as an image, and it is called – feature value which means data that abstracts the features of the input image) is sent to the next neural network where the same process is repeated.

Even simple image recognition such as distinguishing numbers and letters requires several layers of this kind of computation, while general object recognition requires dozens of layers of this kind of computation, so by repeating numerous operations, we can achieve image recognition, such as determining the number 5 to be a 5 or identifying the location of a dog from an image of a dog.

Using conventional computer software without AI - people think about what they want to process (the algorithm) and build a program, - whereas AI that is using neural networks - uses a large amount of input and automatically prepares the internal data (weighting parameters) necessary for the processing. In other words, in AI the lead role is not the program, but the data. Thus, while conventional software development focuses on the programming, the most important and time-consuming part of AI development is the process of preparing the weighting parameters (learning) so that it can process the objects to be recognized and determined with the necessary accuracy and speed.

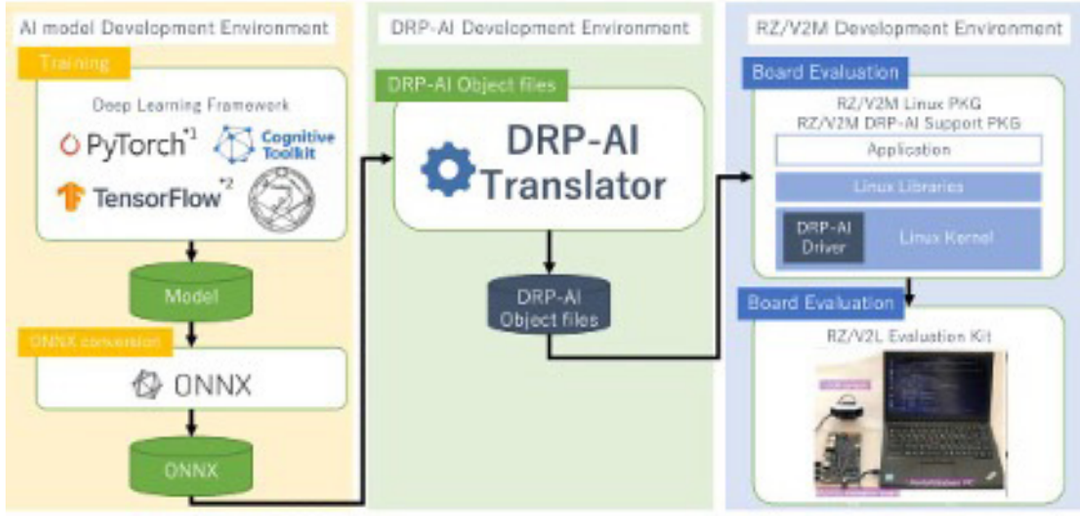
The software package for AI evaluation that we are introducing here uses pre-trained models provided by AI frameworks such as PyTorch, so you can evaluate the AI execution of image inference without the time-consuming learning process.

What is required?

There is no need to purchase an evaluation board from the start. The Vision AI implementation flow of the embedded AI processor RZ/V series, using only free open source software (hereafter referred to as OSS) and a software package is provided free of charge on Renesas Electronics website.

Figure 3 shows the overall tool flow. There are various industry standard frameworks for AI learning people, so we developed the RZ/V

Figure 3. *RZ/V2L AI Implementation Tool Flow Overall Structure*



*1 PyTorch, the PyTorch logo and any related marks are trademarks of Facebook, Inc.

*2 TensorFlow, the TensorFlow logo and any related marks are trademarks of Google Inc.

*3 DRP-AI Translator: ONNX conversion tool from Renesas Electronics

Take the First Step in Vision AI Development CONTINUED

series for designers to use existing AI frameworks they are familiar with for AI training and connect the trained neural network models to Renesas tools using a common format called ONNX - which is an AI development flow. ONNX is a widely adopted format, and most AI frameworks are able to output in ONNX format, either directly or with a conversion tool, but we will use PyTorch, an AI framework, as our example here.

In the AI world, Linux, not Windows, is the de facto standard and all OSs and software packages used here also require a PC with Linux (Ubuntu) installed.

There is a technology called WSL2 that runs Linux as a virtual OS on Windows, but the Renesas evaluation package is not guaranteed to work with WSL2, so please prepare a PC with Ubuntu installed.

First, a Linux PC and the DRP-AI Support Package

Getting the Linux PC ready

For the hardware, you can use your own PC with an x86 64-bit CPU and at least 6GB of RAM. As for OS, we will use Ubuntu version 20.04 which is one of the Linux distributions instead of Windows.

Download the ISO file of Ubuntu 2004 from ubuntu-20.04.6-desktop-amd64.iso.

You can use the Rufus tool for Windows to convert the ISO file into a bootable USB drive to install Ubuntu 20.04 on the PC.

(Please refer to https://rufus.ie/en/ for details on using Rufus.)

Getting the software ready

Next download the software package "RZ/V2L DRP-AI Support Package" from the Renesas website using a high-speed internet connection.

Open the implementation guide that has step-by-step guidance on how to evaluate the embedded trained neural network model, from conversion to actual operation on the evaluation board. Below are some points to be considered when using this guide to proceed with the evaluation.

Creating the ONNX file

From here, we will be using the command line in Linux.

(If you are new to Linux, you will need to learn the basic commands of the OS itself.)

The AI frameworks PyTorch and torchvision can be installed from the Linux command line using the command pip3. (Refer to the Implementation Guide, Chapter 2.2, Page 26)

Unzip the downloaded DRP-AI Support Package, you'll find the rzv2l_ai-implementation-guide_ver5.00.tar.gz compressed file.

Similarly, in the folder named pytorch_mobilenet, under rzv2l_ai-implementation-guide, you'll find the files

pytorch_mobilenet_en_rev5.00.pdf (hereafter referred to as MobileNet Guide) and pytorch_mobilenet_ver5.00.tar.gz which you should unzip as described in the Implementation Guide.

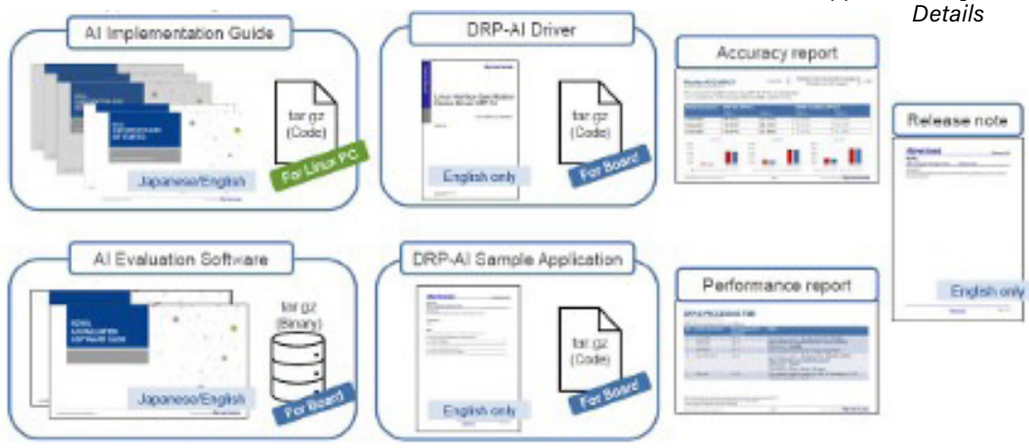


Figure 4. *DRP-AI Support Package Details*

Take the First Step in Vision AI Development CONTINUED

MobileNet is a lightweight and fast neural network for image recognition developed for mobile and embedded devices, which outputs the probability of a correct answer for an object in the image. For example, as shown in *Figure 5*, the probability that the object is a beagle is 93.53%, indicating that the determination is correct.

If you move on to Chapter 2 of pytorch_mobilenet_en_rev5.00.pdf, you will find a file called mobilenet_v2.onnx. This is a representation of the pre-trained MobileNet v2 neural network model with weighting parameters in ONNX format.

With have the trained ONNX files, let's move on to the next phase of creating DRP-AI Object files for DRP-AI and estimating the performance.

Converting to DRP-AI Object files in DRP-AI Translator

From here, we will use the ONNX conversion tool DRP-AI Translator by Renesas. First, download the DRP-AI Translator from the Renesas website.

Install the unzipped installer according to Chapter 3.1 of the Implementation Guide and follow the steps in the Implementation Guide and MobileNet Guide, to reach with a directory structure like the one in *Figure 6.*

Figure 5. *Image Recognition Image by MobileNet*

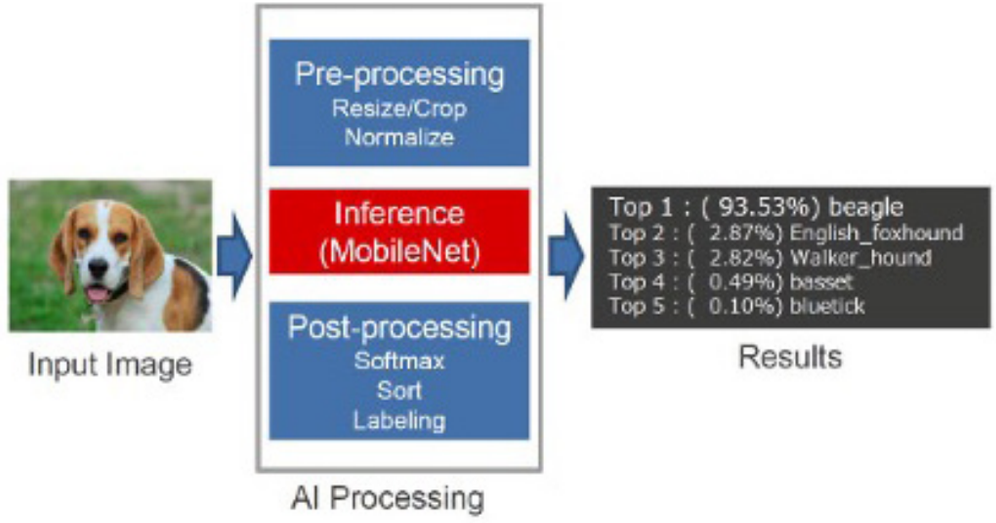
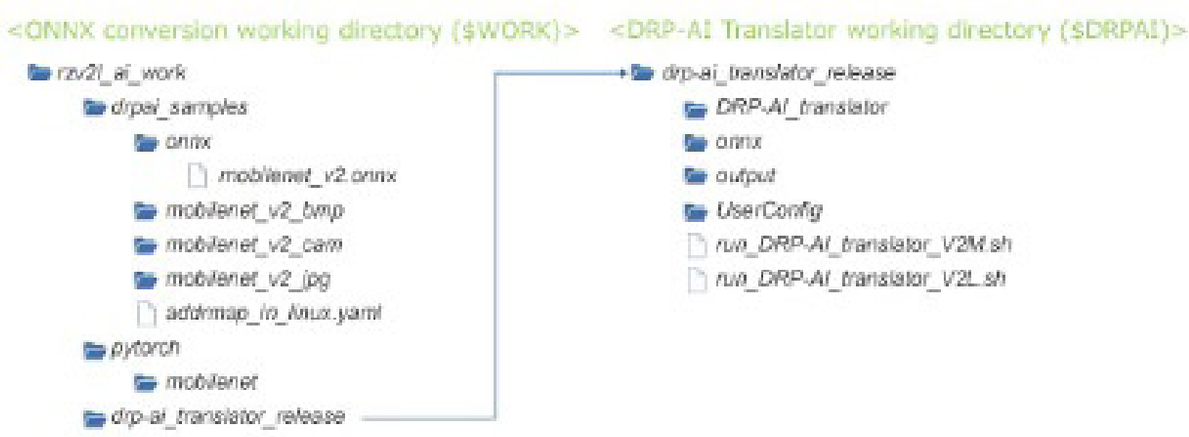


Figure 6. *Directory Structure of the ONNX Files and the DRP-AI Translator Directory Structure*



Take the First Step in Vision AI Development CONTINUED

Figure 7. *List of Files after Running DRP-AI Translator*

After this, prepare the file to be input to the DRP-AI Translator by copying/renaming and editing the file from the sample. Please refer to the MobileNet Guide, Chapters 3.3–3.5 for detailed instructions.

The ONNX translation itself can be completed with a single command. For RZ/V2L, from the shell, run

$ ./run_DRP-AI_translator_V2L.sh mobilenet_v2 -onnx ./onnx/mobilenet_v2.onnx

in the output/mobilenet_v2/ directory under the working directory, and this will generate the binary files needed to run MobileNet on a real chip. (*Figure 7*)

The DRP-AI Translator outputs the performance estimator into an Excel spreadsheet including the approximate processing time for each layer of the model.

With Renesas evaluation kits and complimentary software package, users can now start your first vision AI development at ease.

To continue reading, click the link below to find the rest of this article.

DEVELOPMENT ENVIRONMENT FOR RZ MPUS

RZ MPUs are supported by a full range of development kits, application resources and a global ecosystem of Arm® partners with solutions to support your designs.

This contributes to faster prototype development time and reduced development costs. Additional development tools, OS, middleware, boards, and other products are available from the ecosystem partner network.

GET AN RZ KIT AND START YOUR NEXT DESIGN TODAY

Embedded AI-Accelerator DRP-AI

Overview

With the remarkable advances in computing power in recent years, AI (Artificial Intelligence) is penetrating our lives, starting from cloud services. Reflecting this, according to Gertner's report1 , AI semiconductor market size is expected to grow from $12 billion in 2019 to $43 billion in 2024. A new trend is the implementation of AI into endpoints from cloud. This is because endpoints, such as IoT devices and robots, are required to be smarter and react in real time. The AI required for endpoints is inference processing based on deep learning that replaces human perception such as vision and hearing. To implement AI in endpoints, two major challenges need to be overcome: First, power consumption limitations, and second, flexibility. While the cloud can be equipped with sufficient power and cooling, endpoints are strictly required to limit power consumption which can cause shorter runtimes, generate heat, or increase costs. The idea of power consumption saving is to utilize dedicated hardware that is specialized for specific AI processing; however, the hardware will soon become obsolete since AI models are evolving day by day. Therefore, AI acceleration in endpoints is required to provide the flexibility to support newly developed AI models. Renesas has developed the DRP-AI (Dynamically Reconfigurable Processor for AI) as an AI accelerator with high-speed AI inference processing that achieves the low power and flexibility required by endpoints based on the reconfigurable processor technology it has cultivated over many years.

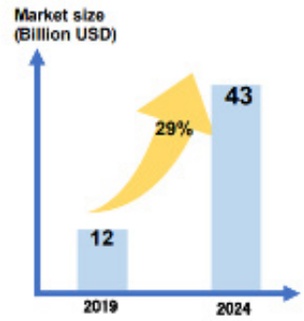


Figure 1: AI Semiconductors Market Dynamics 2019 2024 12 43 Market size (Billion USD)

DRP-AI Features

- AI accelerator dedicated AI inference
- High power efficiency by the collaboration HW (DRPAI) and SW (DRP translator)
- Supported AI model extension by continuous update of DRP-AI translator

The DRP-AI translator is a hardware dedicated to AI inference, but it achieves flexibility, high-speed processing, and power efficiency by utilizing Renesas' unique dynamic reconfigurable technology. The DRP-AI translator is provided to enable users to easily implement AI models optimized to maximize the performance of DRP-AI on this flexible hardware. Multiple executables output by the DRPAI translator can be placed in external memory. This makes it possible to dynamically switch between multiple AI models as a system. In addition, the DRP-AI translator can be continuously updated to support newly developed AI models without hardware changes.

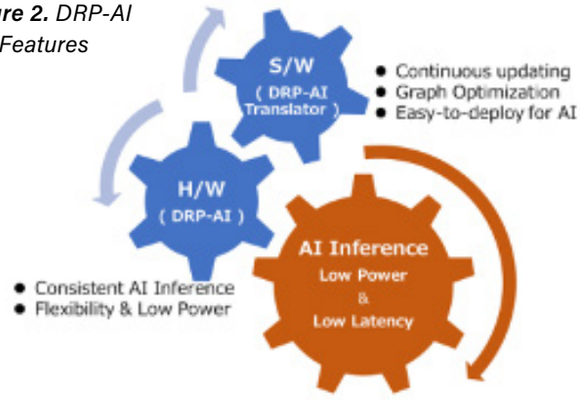
Embedded AI-Accelerator DRP-AI CONTINUED

DRP-AI Hardware Architecture

- DRP (Dynamically Reconfigurable Processor): Programable hardware
- AI-MAC (multiply-and-accumulate): Hardware dedicated for MAC computing
- DMAC (Direct Memory Access Controller)

DRP-AI is composed of AI-MAC and DRP (Dynamically Reconfigurable Processor), which can efficiently process operations in convolutional and all-combining layers by optimizing data flow with internal switches. The DRP can process complex processing such as image preprocessing and AI model pooling layers flexibly and quickly by dynamically changing the hardware configuration. The DRP-AI translator automatically allocates each process of the AI model to the AI-MAC and DRP, thus allowing the user to easily use DRP-AI without being aware of the hardware.

Figure 2. *DRP-AI Features*



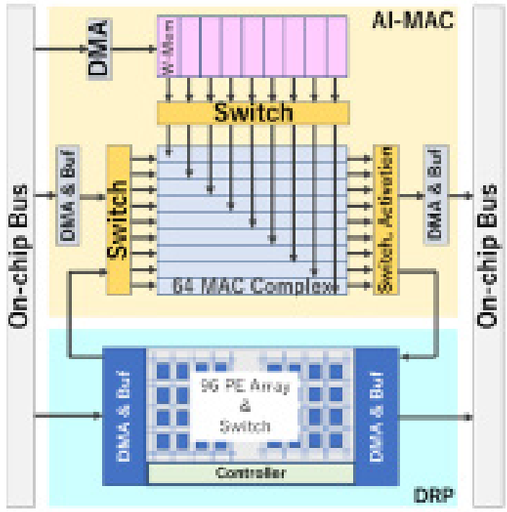


Figure 3. *DRP-AI H/W Architecture*

- Tool to generate DRP-AI optimized executables from trained ONNX model
- Optimizing the graph structure of AI models to minimize memory access and improve computing efficiency
- Extension of supported AI models by continuous updates

The DRP-AI translator is a tool that generates DRP-AI optimized executables from trained models based on the ONNX format, which is independent of various AI frameworks. The tool's internal tasks are:

1. Scheduling of each operation to process the AI model
2. Hiding the overhead such as memory access time that occurs during the transition of each operation in the schedule defined in 1.
3. Optimization of graph structure in the network (Layer fusion, DRP & AI-MAC processing allocation)

Using the DRP-AI translator, users can implement automatically optimized AI models from ONNX AI models into DRP-AI without knowledge of the DRP-AI hardware configuration. The user can then simply make calls through the supplied driver to run the high-performance AI model.

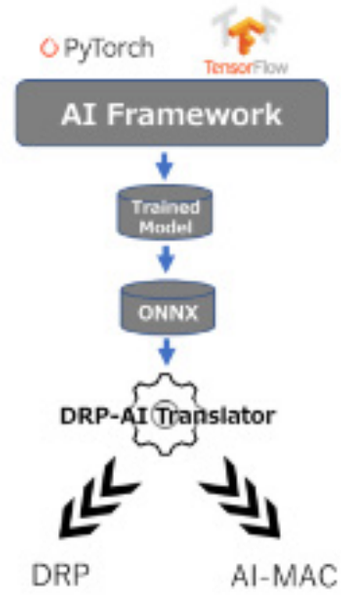


Figure 4. *AI model implementation flow by DRP-AI translator*

Embedded AI-Accelerator DRP-AI CONTINUED

Architecture for High Power Efficiency

- Reduction of external memory communication volume by data reuse technique
- Low power control using inputted zero data
- Scheduling of operation flow Reduction of external memory communication volume by data reuse technique

The power consumption of AI accelerators is not only due to the enormous matrix operations, but also the power components due to data transactions between the accelerator's internal components or external memory becoming large. In addition, as shown in *Figure 5*, the ratio of the amount of data related to weight/input/output are different depending on the image size or types of models etc., and the power bottleneck factors are diverse. Therefore, to comprehensively reduce the power consumption of AI model execution, it is necessary to reduce the amount of memory access for all data types. As an effective way to reduce the amount of external memory access, DRP-AI employs a technology that efficiently reuses data which input to AI-MAC one time. For example, in a convolutional operation using a 3x3 filter, one pixel of data is used for nine filter operations. im2col, which is widely used as a highly parallel operation method in GPUs, expands all the image data in the order of matrix operations as a pre-processing step for input to the GPU.

At this time, the data information of one pixel appears nine times, so the number of data increases by a factor of nine. This causes an increase in power consumption and communication bandwidth. On the other hand, AI-MAC can reuse the data by shifting the data taken into the register corresponding to the MAC arithmetic unit to the adjacent register. The specific flow is explained using *Figure 6*.

1. Data stored in the external memory (blue) is loaded into the AI-MAC buffer (*Figure 6-left*)
2. Transfer data (blue) from buffer to register (*Figure 6-center*)
3. The data in the register is used to perform operations in the corresponding MAC arithmetic unit (*Figure 6-right*)
4. Shift data (blue) to the next register down (data reuse)

By adopting this configuration, the number of data loads from external memory and internal buffer to the AI-MAC can be reduced by up to a factor of nine compared to the GPU. As a result, the power and communication bandwidth required for data movement is significantly reduced. In addition, AI-MAC can reuse data not only for input data, but also for output and weight information, reducing access to external memory by more than one order of magnitude.

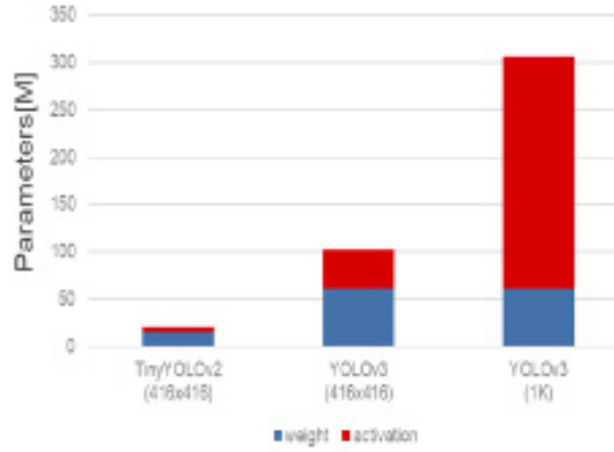
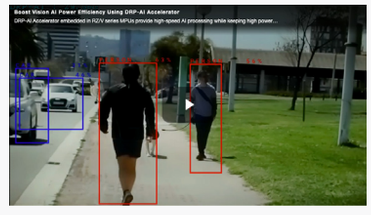


Figure 5. *Data Structure of AI Model*

Watch the video on Renesas website to learn more about **Boosting Vision AI Power Efficiency Using DRP-AI Accelerator**



Low power control using inputted zero data One of the characteristics in AI model computation is the high ratio of „zero" values in the weight data and input/output data of each layer (it is called sparsification). For example, as shown in *Figure 7*, in the image recognition model, more than 50% of the input and output data of all layers are zero values on average.

Figure 6. *Data Reuse Diagram*

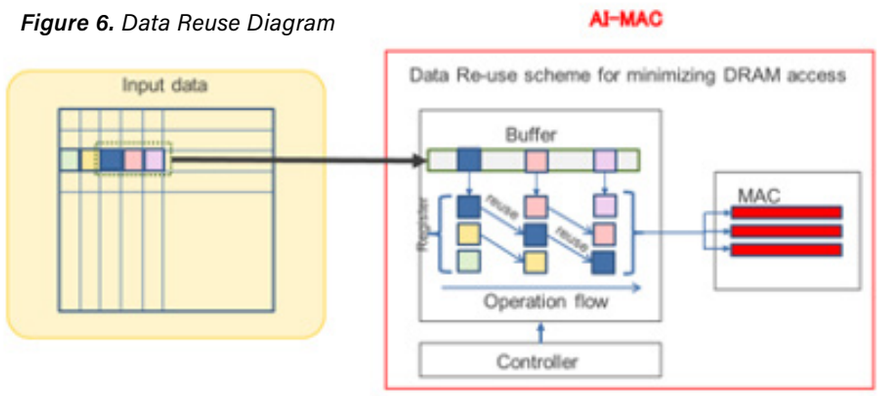
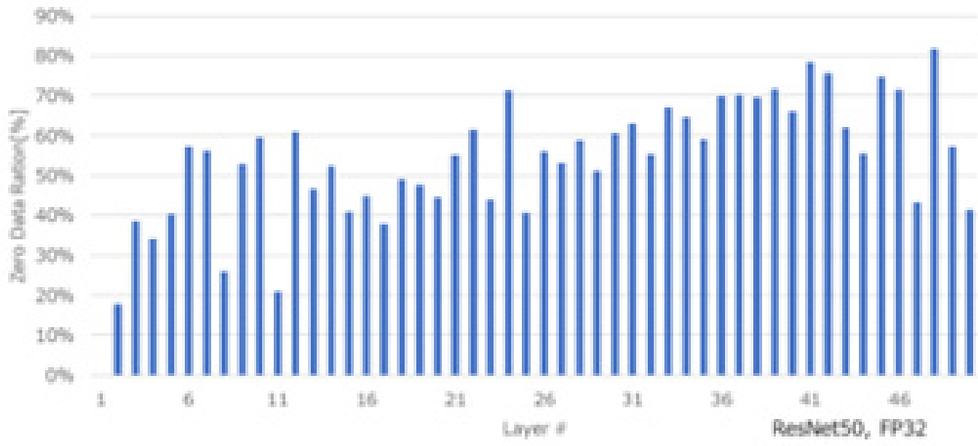


Figure 7: Zero Data Ratio of AI Models



This is because many AI models use an activation function (ReLU) that replaces all negative results of the sum-of products operation with zero. In DRP-AI, unnecessary computational power is reduced by introducing a switching technology that detects in advance when zero is entered in the input for each element of the operation and prevents unnecessary operations. AI-MAC

Scheduling of operation flow

In addition to the data reuse techniques mentioned above, optimization of the order and timing of operations such as external data access or MAC operation processing, etc. is essential for efficient AI execution. In other words, scheduling the operation flow can maximize the performance of DRP-AI. One example is described below. By scheduling the external memory access timing so that the weight information for the next operation is read ahead and stored in the buffer during AI-MAC operation, the external memory access latency can be hidden. Such cases also occur in the timing of internal memory access and any internal arithmetic processing, and scheduling can avoid unnecessary waiting time and power generation between each process. Since the DRP-AI translator automatically generates this optimized scheduling, the user can easily

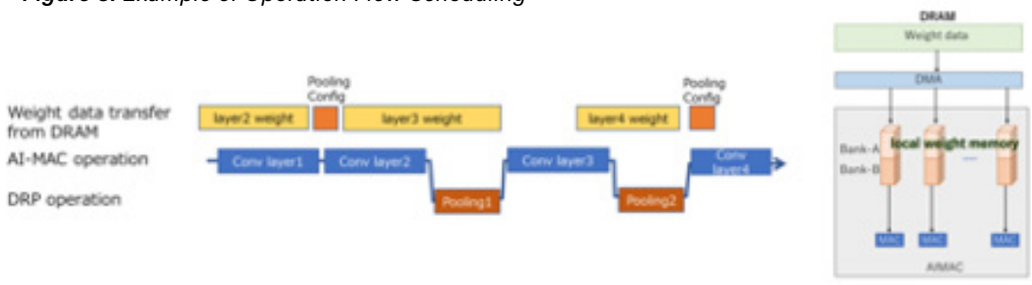
Embedded AI-Accelerator DRP-AI CONTINUED

handle the DRP-AI evaluation. We implemented TinyYolov22 in a DRP-AI test vehicle, the high power-efficient architecture which has been previously described and measured the surface temperature of the device using thermography. In the experiment, we ran the AI under the same conditions using a commercially available GPU as a comparison to express the performance of DRP-AI in a digestible way.

Results are shown in *Figure 8*. You can see that the surface temperature of our DRP-AI test vehicle is clearly lower than that of the commercial GPU. The surface temperature of the DRP-AI test vehicle was 40.9° without heat sink when running TinyYolov2 at 42 fps34. On the other hand, that of the commercial GPU was 79.0° despite having a heat sink when running AI at a lower framerate than DRP-AI. In this experiment, similar to a real use case, we hope that you have understood that the product with DRP-AI is a product that can withstand mass production of endpoint products under severe temperature constraints even when running practical level AI. 2 Tiny Yolov2: https://pjreddie.com/darknet/yolov2/ 3 The value is AI inference only 4 The performance varies from product to product due to the difference in device configuration

Renesas has developed the DRP-AI (Dynamically Reconfigurable Processor for AI) as an AI accelerator with high-speed AI inference processing that achieves the low power and flexibility required by endpoint devices. We will deploy MPU products equipped with this superior AI accelerator in a scalable manner to help endpoint products be equipped with AI that reacts intelligently and in real time.

Figure 8. *Example of Operation Flow Scheduling*



Visit www.renesas.com/rz to learn more about the RZ/V2M, Renesas' original AI-dedicated Accelerator (DRP-AI), 4K-compatible Image Signal Processor (ISP), Vision-AI ASSP for real-time human and object recognition; and the RZ/V2L general-purpose MPU with the same DRP-AI, 1.2GHz Dual-Core Arm® Cortex®-A55 CPU, 3D Graphics, and video codec engine.

*1 Graph created by Renesas based on Gartner Research, Source: Forecast Analysis: AI Neural Network Processing Semiconductor Revenue, worldwide, Alan Priestley, 20 Apr 2020, Revenue Basis.

AI Applications and AI SDK for RZ/V Series MPUs

The best solution for starting your AI applications. Renesas AI applications and AI SDK are quick and easy solutions for starting AI development. Various AI applications are provided free of charge.

View AI Applications > View AI SDK >

Getting Started

Renesas RZ/V AI software development kit (AI SDK) is the most comprehensive solution for building end-to-end accelerated AI applications. The AI SDK includes Yocto Linux with bootloader, Linux Kernel, Cross Compiler and a complete set of libraries for DRP-AI, graphics and codec.

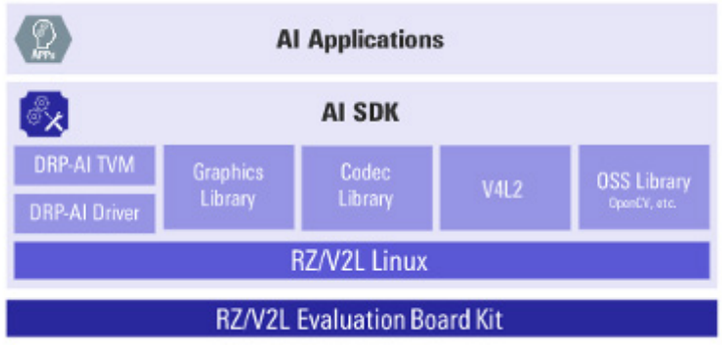
The Getting Started Guide is a complete overview to learn how to run AI applications. It explains the procedure for setting up the development environment, compiling AI applications, setting up the board and running AI applications.

Get your target board now and Get Started.

AI Applications

AI applications provide the source code, pre-built application binary and pre-trained AI model objects, which allow you to select the application from various use cases and run on the board immediately. Choose from among a variety of application categories:

- Agriculture
- Healthcare
- Industrial
- Smart Building
- Smart City
- Smart Home
- Retail

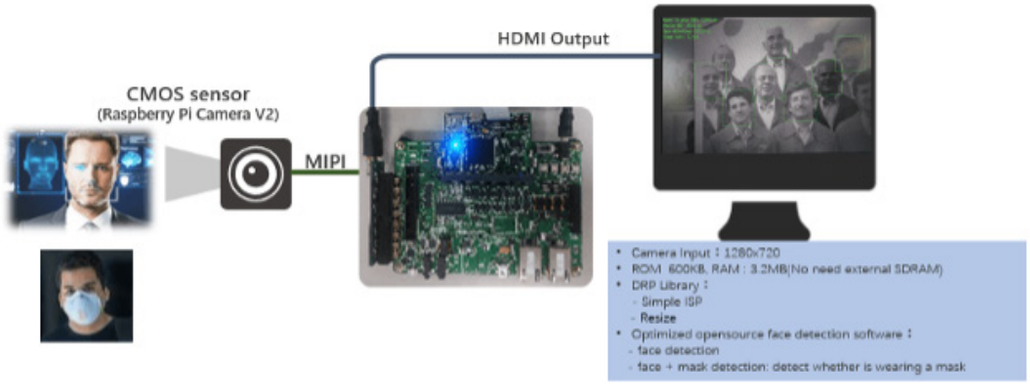


RZ/A2M DRP Realizes Efficient Masked Face Detection

Let's take a look at the combination of RZ/A2M's DRP and embedded AI to realize masked face detection.

In the following example, we connect the Sony IMX219 CMOS sensor through the MIPI interface, input a 1280x720 resolution image, using DRP in RZ/A2M to perform Simple ISP processing and image scaling processing on the input image, and run a light and efficient masked face detection model. It can achieve a detection speed of 30FPS in the face detection mode, and a detection speed of 20FPS in the mode of distinguishing whether to wear a mask. Let's take a look at how it is implemented.

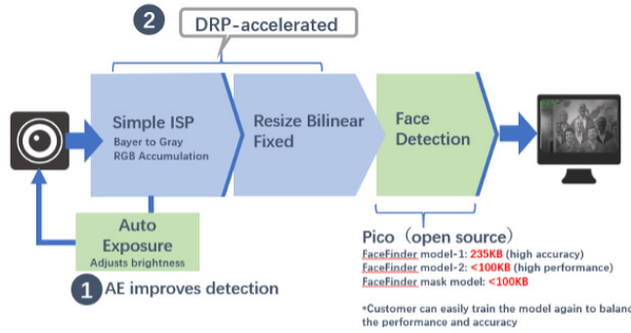
The following figure shows the data processing flow:



The processing of the portion shown in blue is realized by DRP hardware acceleration, and the Simple ISP library converts the Bayer format data of the CMOS sensor into gray-scale data and counts the average brightness of three preset areas in a frame of the image to adjust the automatic exposure parameters.

The second DRP library implements image scaling, compressing the 1280x720 resolution grayscale image into a 640x360 size image, which will greatly improve the face detection speed.

The green part in the picture is a light mask and face detection model run by the Cortex-A9 processor, which is used to calibrate whether the current frame has a human face and whether it has a mask.



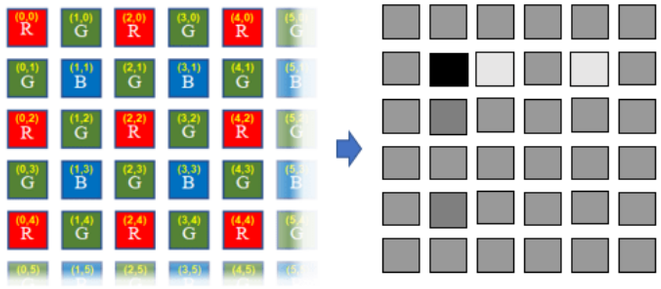
In this example, we will not rely on external RAM and only use RZ/A2M's 4MB on-chip high-speed RAM.

Step 1

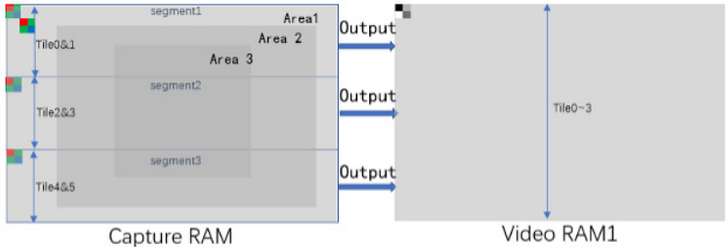
Since face detection only needs to use grayscale images, we need to convert the Bayer format image of the CMOS sensor into a grayscale image. At this time, we load a simple_isp_2_tiles DRP library. This DRP library needs to have the following characteristics:

- Occupies 2 tiles of DRP hardware resources
- Realize Bayer to grayscale
- Accumulate the brightness values of all pixels in 3 independent areas
- Support multi-tile parallel processing

RZ/A2M DRP Realizes Efficient Masked Face Detection CONTINUED



Since this library features multi-tile parallelized (segmented) processing, we can load it into 3 sets of DRP tiles. Among them, the **simple_isp_2_tiles** library of Tile 0 and 1 handles the top 1/3 of the image, the **simple_isp_2_tiles** library of Tiles 2 and 3 handles the middle 1/3 of the image, and the **simple_isp_2_tiles** library of Tile 4 and 5 handles the bottom 1/3 of the image. These three parts of images are processed in parallel at the same time, which increases the processing speed by 3 times.



Because the DRP library provides a very convenient API interface, the above functions can be realized by simple programming operations.

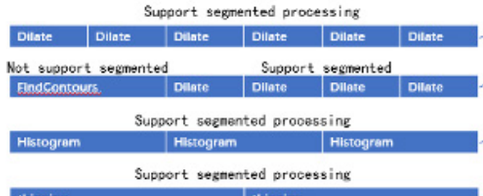
```
for (tile_no = 0; tile_no < 3; tile_no++)                    ➡ Process 3 segments in parallel
{
    R_MMU_VAtoPA((uint32_t)(in_adr + (width * (height / 3)*tile_no)), &(param_isp_fd[tile_no].src));
    R_MMU_VAtoPA((uint32_t)(out_adr+ (width * (height / 3)*tile_no)), &(param_isp_fd[tile_no].dst));   ➡ Convert Input/Outpu address
                                                                                                          to Phicial address
    /* Set Image size */                                     ➡ Image original width
    param_isp_fd[tile_no].width = width;
    param_isp_fd[tile_no].height = height/3;                 ➡ Set 1/3 of original height
    R_MMU_VAtoPA((uint32_t)&ave_result[tile_no], &(param_isp_fd[tile_no].accumulate));   ➡ Set luma accumulation output address
                                                                                          The address is in uncatchable RAM
    clip_rect(0, tile_no*param_isp_fd[tile_no].height, width, param_isp_fd[tile_no].height,
            area1_offsetx, area1_offsety, area1_w, area1_h,
            (uint16_t*)&param_isp_fd[tile_no].area1_offset_x, (uint16_t*)&param_isp_fd[tile_no].area1_offset_y, (


    clip_rect(0, tile_no*param_isp_fd[tile_no].height, width, param_isp_fd[tile_no].height,   ➡ Divide accumulation area to 3 segments
                area2_offsetx, area2_offsety, area2_w, area2_h,
                (uint16_t*)&param_isp_fd[tile_no].area2_offset_x, (uint16_t*)&param_isp_fd[tile_no].area2_off

    clip_rect(0, tile_no*param_isp_fd[tile_no].height, width, param_isp_fd[tile_no].height,
                area3_offsetx, area3_offsety, area3_w, area3_h,
                (uint16_t*)&param_isp_fd[tile_no].area3_offset_x, (uint16_t*)&param_isp_fd[tile_no].area3_off

    /* Initialize variables to be used in termination judgment of the DRP library */
    drp_lib_status[tile_no] = DRP_NOT_FINISH;            ➡ Reset the status of each tile,
                                                           Tile status will be updated in interrupt
    /**********************/
    /* Start DRP Library */
    /**********************/
    ret_val = R_DK2_Start(drp_lib_id[2*tile_no], (void *)&param_isp_fd[tile_no], sizeof(param_isp_fd[tile_no]));   ➡
    DRP_DRV_ASSERT(ret_val);                                                                                        Start DRP Processing
}
```

Users can decide how to load the DRP library according to the Number of tiles and Segmented Processing attributes in the application documentation of the DRP library.

- Number of tiles: Indicates that the DRP library needs to occupy how many hardware tiles
- Segmented processing: Indicates that the DRP task can be split into multiple tiles for parallel execution

Number of tiles	1
Segmented processing	Supported

There are 11 ways to place the DRP library in the Tile. You can choose which loading method to use according to the Number of tiles and Segmented attributes of the DRP library in a flexible way. Below are some examples:



Tile Pattern	Macro Setting of Argument tile_pattern of R_DK2_Load Function
	R_DK2_TILE_PATTERN_1_1_1_1_1_1
	R_DK2_TILE_PATTERN_2_1_1_1_1
	R_DK2_TILE_PATTERN_2_2_1_1
	R_DK2_TILE_PATTERN_2_2_2
	R_DK2_TILE_PATTERN_3_1_1_1
	R_DK2_TILE_PATTERN_3_2_1
	R_DK2_TILE_PATTERN_3_3
	R_DK2_TILE_PATTERN_4_1_1
	R_DK2_TILE_PATTERN_4_2
	R_DK2_TILE_PATTERN_5_1
	R_DK2_TILE_PATTERN_6

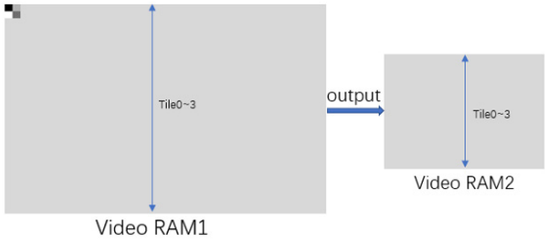
RZ/A2M DRP Realizes Efficient Masked Face Detection CONTINUED

Step 2

After getting a frame of a grayscale image, we load a DRP library by resize bilinear_ fixed to zoom this frame of image. Key features of the DRP library:

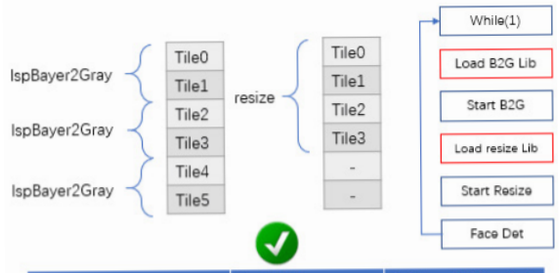
- Input 8bpp grayscale image
- Support ⅛ ¼ ½ 1x 2x 4x 8x 16x fixed zoom ratio
- Separate control of horizontal and vertical scaling
- Input width range 128~1280, input height range 8~960
- Occupies 4 tiles hardware resources without supporting segmented

After processing Step 2, we read the grayscale image from Video RAM1, reduce the width and height to the original ½, and write the image to Video RAM2 for the next face detection.



The execution time of these two steps is about 4.6ms and 8.2ms. The parallel processing and the loading speed of the DRP library of less than 1ms have greatly optimized the execution speed of image preprocessing before face recognition.

File name	Param1	Param2	Param3	Param4	Param5	Param6	Param7	Param8
Jpg	x-coordinate of the left eye	y-coordinate of the left eye	x-coordinate of the right eye	y-coordinate of the right eye	x-coordinate below the nose	y-coordinate below the nose	x-coordinate below the mouth	y-coordinate below the mouth
pic05358.jpg	2.873178e+02	8.731138e+01	3.043895e+02	9.148446e+01	2.941465e+02	1.022966e+02	2.916806e+02	1.098840e+02



Process	Total time consumption (ms)	DRP Load Time(ms)
Bayer To GrayScale	4.63	0.44
Resize	8.14	0.34
Total	**12.77**	

Step 3

We used a light open source face recognition algorithm https://github.com/nenadmarkus/pico

A data set of masked faces was added based on it, and the new data set was trained. After testing, its recognition speed and recognition accuracy are relatively ideal.

First, you need to prepare a data set, which can be directly trained through the data set in our sample package or downloaded from the network.

This data set contains 7092 face pictures and 4283 face mask pictures, and the key features of the face in the pictures are calibrated.

The AI model is not a commonly used neuron network model, but a decision tree model, which has the characteristics of fast execution, and its model size is only tens of KB to 200KB.

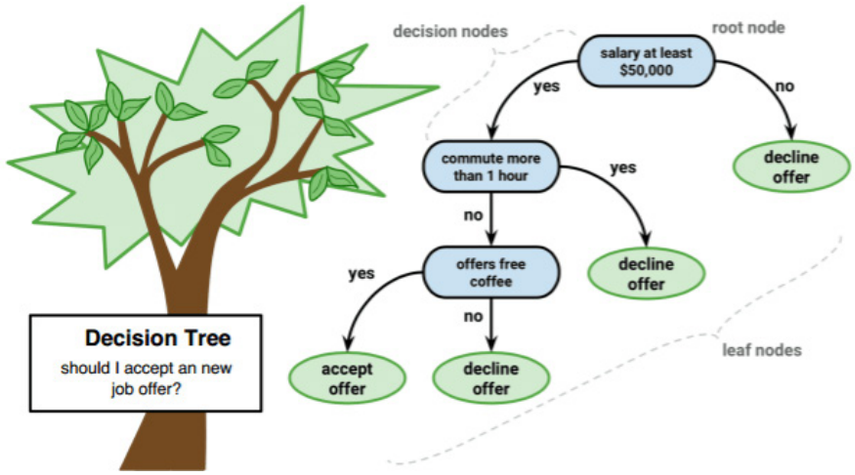
```
mkdir -p caltechfacescd caltechfaces
wget http://www.vision.caltech.edu/Image_Datasets/Caltech_10K_WebFaces/Caltech_WebFaces.tar
7za x Caltech_WebFaces.tar
rm Caltech_WebFaces.tar
wget http://www.vision.caltech.edu/Image_Datasets/Caltech_10K_WebFaces/WebFaces_GroundThruth.txt
```

RZ/A2M DRP Realizes Efficient Masked Face Detection CONTINUED

The following are the more commonly used algorithms in the AI field:

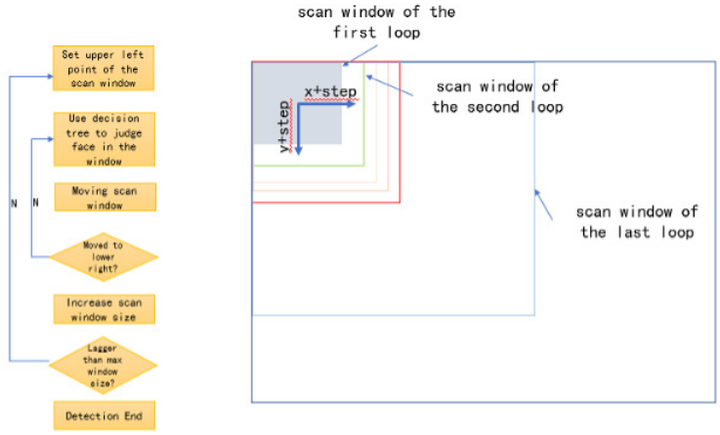
- Linear regression
- Logistic regression
- Neural network
- SVM
- KNN
- K-Means
- Decision Tree
- Random forest
- AdaBoost
- Naive Bayesian
- Gradient descent
- Principal Component Analysis

A decision tree is a non-parametric supervised learning model that allows you to follow the results of the tree-like decision branch step-by-step from the root node to the top leaf node, thereby predicting the target value based on the result of the top leaf node. It is often used for target classification and regression



Resource: [Machine Learning Notes by Heart of Machine]

In the face detection process, we use a sliding window to scan the image generated in Step 2 step-by-step, first use the smallest sliding window, and then gradually increase the size of the window. Use the decision tree model to detect whether there is a face in each sliding window.



The following is the configuration of some key parameters, we can balance between detection accuracy and performance through parameter adjustment.

```
typedef struct
{
int minsize;
int maxsize;
float angle;
float scalefactor;
float stridefactor;
float qthreshold;
int usepyr;
int noclustering;
int verbose;
}R_DETECTION_PARAM;
```

- Minimal scan window, default ¼ of height 720/4)
- Maximal scan window, default 2/3 of height (720 * 2/3)

Bigger rage, longer time

- whether to rotate by a certain Angle, default is 0, range 0.0~1.0, Represents 0 degrees ~360 degrees
- The magnification of the scan window each time it increases, default is 1.1, That is, the window increases by 10% per cycle
- Step size of scan window moving, default is 0.1, That is the 10% of minimalwindow size
- Confidence threshold, default is 5.0, larger value means more reliable
- Whether to enable zoom function, default is 0, After enabling, do one more detection after 1/2 down scaling
- Output all possible regions and integrate the regions that are close to each other. The current setting is 0, that is, integration is required
- Whether to print the results through the Console

RZ/A2M DRP Realizes Efficient Masked Face Detection CONTINUED

By implementing all the steps, we could generate the following 1280x720 resolution input, which detects 1 face, 3 faces and 7 faces in the screen respectively. In the mask face mode, the detection speed can be above 15FPS.

Delve into the product details of the RZ/A2M MPU here.

Scan Window Size: 180x180 ~ 480x480
Face number: 1
Model: facefinder2.dat
facefinder_mask.dat

Items	Performance
Bayer2Gray	4.63 ms
Resize	8.14 ms
Detect Mode (face + mask)	26~44 ms
Time/Frame Rate	39~58 ms / 17~24 fps
Detect Mode (face)	14~22 ms
Time/Frame Rate	27~34 ms / 29~35 fps
Detect Mode (mask)	14~24 ms
Time/Frame Rate	27~38 ms / 26~35 fps

Scan Window Size : 180x180 ~ 480x480
Face number : 3
Model : facefinder2.dat
facefinder_mask.dat

Items	Performance
Bayer2Gray	4.63 ms
Resize	8.14 ms
Detect Mode (face + mask)	42~60 ms
Time/Frame Rate	56~72 ms / 13~17 fps
Detect Mode (face)	21~30 ms
Time/Frame Rate	36~43 ms / 23~28 fps
Detect Mode (mask)	23~30 ms
Time/Frame Rate	36~43 ms / 22~26 fps

Scan Window Size : 180x180 ~ 480x480
Face number : 7
Model : facefinder2.dat
facefinder_mask.dat

Items	Performance
Bayer2Gray	4.63 ms
Resize	8.14 ms
Detect Mode (face + mask)	51 ms
Time/Frame Rate	64 ms / 15 fps
Detect Mode (face)	24 ms
Time/Frame Rate	33 ms / 26 fps
Detect Mode (mask)	35 ms
Time/Frame Rate	48 ms / 19 fps

RENESAS

At Renesas we continuously strive to drive innovation with a comprehensive portfolio of microcontrollers, analog and power devices. Our mission is to develop a safer, healthier, greener, and smarter world by providing intelligence to our four focus growth segments: Automotive, Industrial, Infrastructure, and IoT that are all vital to our daily lives, meaning our products and solutions are embedded everywhere.

STAY CONNECTED

CLICK HERE TO VISIT OUR **WEBSITE**

© 2024 Renesas Electronics Corporation. All rights reserved.
Document No.: R19PM0004EU0100