Quick Start Guide

RENESAS AI Model Deployer

Quick Start Guide

Renesas RZ & RA Family
Using RZ/V2L-EVK

Renesas Electronics
www.renesas.com

Rev.1.00 Jun 2025

Notice

1. Descriptions of circuits, software and other related information in this document are provided only to illustrate the operation of semiconductor products and application examples. You are fully responsible for the incorporation or any other use of the circuits, software, and information in the design of your product or system. Renesas Electronics disclaims any and all liability for any losses and damages incurred by you or third parties arising from the use of these circuits, software, or information.
2. Renesas Electronics hereby expressly disclaims any warranties against and liability for infringement or any other claims involving patents, copyrights, or other intellectual property rights of third parties, by or arising from the use of Renesas Electronics products or technical information described in this document, including but not limited to, the product data, drawings, charts, programs, algorithms, and application examples.
3. No license, express, implied or otherwise, is granted hereby under any patents, copyrights or other intellectual property rights of Renesas Electronics or others.
4. You shall be responsible for determining what licenses are required from any third parties, and obtaining such licenses for the lawful import, export, manufacture, sales, utilization, distribution or other disposal of any products incorporating Renesas Electronics products, if required.
5. You shall not alter, modify, copy, or reverse engineer any Renesas Electronics product, whether in whole or in part. Renesas Electronics disclaims any and all liability for any losses or damages incurred by you or third parties arising from such alteration, modification, copying or reverse engineering.
6. Renesas Electronics products are classified according to the following two quality grades: “Standard” and “High Quality”. The intended applications for each Renesas Electronics product depends on the product’s quality grade, as indicated below.
   "Standard": Computers; office equipment; communications equipment; test and measurement equipment; audio and visual equipment; home electronic appliances; machine tools; personal electronic equipment; industrial robots; etc.
   "High Quality": Transportation equipment (automobiles, trains, ships, etc.); traffic control (traffic lights); large-scale communication equipment; key financial terminal systems; safety control equipment; etc.

   Unless expressly designated as a high reliability product or a product for harsh environments in a Renesas Electronics data sheet or other Renesas Electronics document, Renesas Electronics products are not intended or authorized for use in products or systems that may pose a direct threat to human life or bodily injury (artificial life support devices or systems; surgical implantations; etc.), or may cause serious property damage (space system; undersea repeaters; nuclear power control systems; aircraft control systems; key plant systems; military equipment; etc.). Renesas Electronics disclaims any and all liability for any damages or losses incurred by you or any third parties arising from the use of any Renesas Electronics product that is inconsistent with any Renesas Electronics data sheet, user’s manual or other Renesas Electronics document.
7. No semiconductor product is absolutely secure. Notwithstanding any security measures or features that may be implemented in Renesas Electronics hardware or software products, Renesas Electronics shall have absolutely no liability arising out of any vulnerability or security breach, including but not limited to any unauthorized access to or use of a Renesas Electronics product or a system that uses a Renesas Electronics product. RENESAS ELECTRONICS DOES NOT WARRANT OR GUARANTEE THAT RENESAS ELECTRONICS PRODUCTS, OR ANY SYSTEMS CREATED USING RENESAS ELECTRONICS PRODUCTS WILL BE INVULNERABLE OR FREE FROM CORRUPTION, ATTACK, VIRUSES, INTERFERENCE, HACKING, DATA LOSS OR THEFT, OR OTHER SECURITY INTRUSION (“Vulnerability Issues”). RENESAS ELECTRONICS DISCLAIMS ANY AND ALL RESPONSIBILITY OR LIABILITY ARISING FROM OR RELATED TO ANY VULNERABILITY ISSUES. FURTHERMORE, TO THE EXTENT PERMITTED BY APPLICABLE LAW, RENESAS ELECTRONICS DISCLAIMS ANY AND ALL WARRANTIES, EXPRESS OR IMPLIED, WITH RESPECT TO THIS DOCUMENT AND ANY RELATED OR ACCOMPANYING SOFTWARE OR HARDWARE, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY, OR FITNESS FOR A PARTICULAR PURPOSE.
8. When using Renesas Electronics products, refer to the latest product information (data sheets, user’s manuals, application notes, “General Notes for Handling and Using Semiconductor Devices” in the reliability handbook, etc.), and ensure that usage conditions are within the ranges specified by Renesas Electronics with respect to maximum ratings, operating power supply voltage range, heat dissipation characteristics, installation, etc. Renesas Electronics disclaims any and all liability for any malfunctions, failure or accident arising out of the use of Renesas Electronics products outside of such specified ranges.
9. Although Renesas Electronics endeavors to improve the quality and reliability of Renesas Electronics products, semiconductor products have specific characteristics, such as the occurrence of failure at a certain rate and malfunctions under certain use conditions. Unless designated as a high reliability product or a product for harsh environments in a Renesas Electronics data sheet or other Renesas Electronics document, Renesas Electronics products are not subject to radiation resistance design. You are responsible for implementing safety measures to guard against the possibility of bodily injury, injury or damage caused by fire, and/or danger to the public in the event of a failure or malfunction of Renesas Electronics products, such as safety design for hardware and software, including but not limited to redundancy, fire control and malfunction prevention, appropriate treatment for aging degradation or any other appropriate measures. Because the evaluation of microcomputer software alone is very difficult and impractical, you are responsible for evaluating the safety of the final products or systems manufactured by you.
10. Please contact a Renesas Electronics sales office for details as to environmental matters such as the environmental compatibility of each Renesas Electronics product. You are responsible for carefully and sufficiently investigating applicable laws and regulations that regulate the inclusion or use of controlled substances, including without limitation, the EU RoHS Directive, and using Renesas Electronics products in compliance with all these applicable laws and regulations. Renesas Electronics disclaims any and all liability for damages or losses occurring as a result of your noncompliance with applicable laws and regulations.
11. Renesas Electronics products and technologies shall not be used for or incorporated into any products or systems whose manufacture, use, or sale is prohibited under any applicable domestic or foreign laws or regulations. You shall comply with any applicable export control laws and regulations promulgated and administered by the governments of any countries asserting jurisdiction over the parties or transactions.
12. It is the responsibility of the buyer or distributor of Renesas Electronics products, or any other party who distributes, disposes of, or otherwise sells or transfers the product to a third party, to notify such third party in advance of the contents and conditions set forth in this document.
13. This document shall not be reprinted, reproduced or duplicated in any form, in whole or in part, without prior written consent of Renesas Electronics.
14. Please contact a Renesas Electronics sales office if you have any questions regarding the information contained in this document or Renesas Electronics products.

(Note1) “Renesas Electronics” as used in this document means Renesas Electronics Corporation and also includes its directly or indirectly controlled subsidiaries.

(Note2) “Renesas Electronics product(s)” means any product developed or manufactured by or for Renesas Electronics.

(Rev.5.0-1 October 2020)

Corporate Headquarters

TOYOSU FORESIA, 3-2-24 Toyosu,
Koto-ku, Tokyo 135-0061, Japan
www.renesas.com

Trademarks

Contact information

For further information on a product, technology, the most up-to-date version of a document, or your nearest sales office, please visit: www.renesas.com/contact/.

Renesas RZ/V2L EVKIT Disclaimer

By using this RZ/V2L EVKIT, the User accepts the following terms, which are in addition to, and control in the event of disagreement, with Renesas' General Terms and Conditions available at https://www.renesas.com/en-us/legal/disclaimer.html.

The RZ/V2L EVKIT is not guaranteed to be error free, and the entire risk as to the results and performance of the RZ/V2L EVKIT is assumed by the User. The RZ/V2L EVKIT is provided by Renesas on an "as is" basis without warranty of any kind whether express or implied, including but not limited to the implied warranties of good workmanship, fitness for a particular purpose, title, merchantability, and non-infringement of intellectual property rights. Renesas expressly disclaims any implied warranty.

Renesas does not consider the RZ/V2L EVKIT to be a finished product and therefore the RZ/V2L EVKIT may not comply with some requirements applicable to finished products, including, but not limited to recycling, restricted substances and electromagnetic compatibility regulations. Refer to Certifications section, for information about certifications and compliance information for the RZ/V2L EVKIT. It is the kit User's responsibility to make sure the kit meets any local requirements applicable to their region.

Renesas or its affiliates shall in no event be liable for any loss of profit, loss of data, loss of contract, loss of business, damage to reputation or goodwill, any economic loss, any reprogramming or recall costs (whether the foregoing losses are direct or indirect) nor shall Renesas or its affiliates be liable for any other direct or indirect special, incidental or consequential damages arising out of or in relation to the use of this RZ/V2L EVKIT, even if Renesas or its affiliates have been advised of the possibility of such damages.

Renesas has used reasonable care in preparing the information included in this document, but Renesas does not warrant that such information is error free nor does Renesas guarantee an exact match for every application or parameter to part numbers designated by other vendors listed herein. The information provided in this document is intended solely to enable the use of Renesas products. No express or implied license to any intellectual property right is granted by this document or in connection with the sale of Renesas products. Renesas reserves the right to make changes to specifications and product descriptions at any time without notice. Renesas assumes no liability for any damages incurred by you resulting from errors in or omissions from the information included herein. Renesas cannot verify, and assumes no liability for, the accuracy of information available on another company's website.

Precautions

This Evaluation Kit is only intended for use in a laboratory environment under ambient temperature and humidity conditions. A safe separation distance should be used between this and any sensitive equipment. Its use outside the laboratory, classroom, study area, or similar such area invalidates conformity with the protection requirements of the Electromagnetic Compatibility Directive and could lead to prosecution.
The product generates, uses, and can radiate radio frequency energy and may cause harmful interference to radio communications. There is no guarantee that interference will not occur in a particular installation. If this equipment causes harmful interference to radio or television reception, which can be determined by turning the equipment off or on, you are encouraged to try to correct the interference by one or more of the following measures:
• Ensure attached cables do not lie across the equipment.
• Reorient the receiving antenna.
• Increase the distance between the equipment and the receiver.
• Connect the equipment into an outlet on a circuit different from that which the receiver is connected.
• Power down the equipment when not in use.
• Consult the dealer or an experienced radio/TV technician for help.
Note: It is recommended that wherever possible shielded interface cables are used.
The product is potentially susceptible to certain EMC phenomena. To mitigate against them it is recommended that the following measures be undertaken:
• The user is advised that mobile phones should not be used within 10 m of the product when in use.
• The user is advised to take ESD precautions when handling the equipment.
The Evaluation Kit does not represent an ideal reference design for an end product and does not fulfill the regulatory standards for an end product.

General Precautions in the Handling of Microprocessing Unit and Microcontroller Unit Products

The following usage notes are applicable to all Microprocessing unit and Microcontroller unit products from Renesas. For detailed usage notes on the products covered by this document, refer to the relevant sections of the document as well as any technical updates that have been issued for the products.

1. Precaution against Electrostatic Discharge (ESD)
   A strong electrical field, when exposed to a CMOS device, can cause destruction of the gate oxide and ultimately degrade the device operation. Steps must be taken to stop the generation of static electricity as much as possible, and quickly dissipate it when it occurs. Environmental control must be adequate. When it is dry, a humidifier should be used. This is recommended to avoid using insulators that can easily build up static electricity. Semiconductor devices must be stored and transported in an anti-static container, static shielding bag or conductive material. All test and measurement tools including work benches and floors must be grounded. The operator must also be grounded using a wrist strap. Semiconductor devices must not be touched with bare hands. Similar precautions must be taken for printed circuit boards with mounted semiconductor devices.
2. Processing at power-on
   The state of the product is undefined at the time when power is supplied. The states of internal circuits in the LSI are indeterminate and the states of register settings and pins are undefined at the time when power is supplied. In a finished product where the reset signal is applied to the external reset pin, the states of pins are not guaranteed from the time when power is supplied until the reset process is completed. In a similar way, the states of pins in a product that is reset by an on-chip power-on reset function are not guaranteed from the time when power is supplied until the power reaches the level at which resetting is specified.
3. Input of signal during power-off state
   Do not input signals or an I/O pull-up power supply while the device is powered off. The current injection that results from input of such a signal or I/O pull-up power supply may cause malfunction and the abnormal current that passes in the device at this time may cause degradation of internal elements. Follow the guideline for input signal during power-off state as described in your product documentation.
4. Handling of unused pins
   Handle unused pins in accordance with the directions given under handling of unused pins in the manual. The input pins of CMOS products are generally in the high-impedance state. In operation with an unused pin in the open-circuit state, extra electromagnetic noise is induced in the vicinity of the LSI, an associated shoot-through current flows internally, and malfunctions occur due to the false recognition of the pin state as an input signal become possible.
5. Clock signals
   After applying a reset, only release the reset line after the operating clock signal becomes stable. When switching the clock signal during program execution, wait until the target clock signal is stabilized. When the clock signal is generated with an external resonator or from an external oscillator during a reset, ensure that the reset line is only released after full stabilization of the clock signal. Additionally, when switching to a clock signal produced with an external resonator or by an external oscillator while program execution is in progress, wait until the target clock signal is stable.
6. Voltage application waveform at input pin
   Waveform distortion due to input noise or a reflected wave may cause malfunction. If the input of the CMOS device stays in the area between V_{IL} (Max.) and V_{IH} (Min.) due to noise, for example, the device may malfunction. Take care to prevent chattering noise from entering the device when the input level is fixed, and also in the transition period when the input level passes through the area between V_{IL} (Max.) and V_{IH} (Min.).
7. Prohibition of access to reserved addresses
   Access to reserved addresses is prohibited. The reserved addresses are provided for possible future expansion of functions. Do not access these addresses as the correct operation of the LSI is not guaranteed.
8. Differences between products
   Before changing from one product to another, for example to a product with a different part number, confirm that the change will not lead to problems. The characteristics of a microprocessing unit or microcontroller unit products in the same group but having a different part number might differ in terms of internal memory capacity, layout pattern, and other factors, which can affect the ranges of electrical characteristics, such as characteristic values, operating margins, immunity to noise, and amount of radiated noise. When changing to a product with a different part number, implement a system-evaluation test for the given product.

Renesas AI Model Deployer

Quick Start Guide

Contents

1. Introduction

Renesas AI model deployer is a no-code, intuitive interface that streamlines the entire Vision AI development pipeline—from model selection and training to deployment—on Renesas MPUs and MCUs powered by NVIDIA TAO. It enables embedded developers to quickly build, optimize, and deploy AI models with guided workflows, hardware-aware tooling, and optional deep customization via Jupyter notebooks.

This QSG covers the following modules:

- Guidance for how to set up Renesas AI model deployer interface, and RZ/V2L-EVKIT board for model deployment.
- End-to-end tutorial from model selection to dataset curation to training and evaluation.
- Exporting the model, compiling for DRP-AI and deploying onto RZ/V2L-EVKIT board.

This QSG is based on Detectnetv2 demo, the GUI and the Jupyter notebook support various others, such as mobilenetv2 image classification on RA8 MCU, for more details on that, please visit our GitHub.

1.1 Assumptions and Advisory Notes

1. **Tool experience**:
   It is assumed that the user has some basic knowledge of AI. It is also assumed the user has knowledge in setting up and using RZ/V2 series MPU boards
2. **Subject knowledge**:
   It is assumed that the user has a basic understanding of microcontrollers, microprocessors, embedded systems to perform project customization, and setting up the RZ/V series MPU board with a Yocto-based environment.
3. **Hardware preparation**:
   Before running the Quick Start example or programming a target board (in this case, RZ/V2L-EVKIT), ensure that **default jumper/switch settings** are properly configured. Please refer to the respective board's **user manual** for detailed instructions.
4. **Version and appearance disclaimer**:
   The screenshots and UI content in this guide are provided for reference only. The actual interface and screen layouts may vary depending on the installed **tool versions and configurations**.

2. Hardware and Software requirements

This section lists the requirements necessary to have Renesas AI model deployer and RZ/V2L-EVKIT up and running.

- Please acquire RZ/V2L board and necessary equipment as described in section 4.2 of this QSG.

2.1 Operating System requirements

- The Toolkit supports Ubuntu 20.04.6 LTS and 22.04 LTS.

2.2 System Hardware Requirements

As the tool is using NVIDIA TAO, a NVIDIA GPU is necessary for training and evaluation.

Minimum System Configuration	Recommended System Configuration
8 GB system RAM	32 GB system RAM
4 GB of GPU RAM	32 GB of GPU RAM
8 core CPU	8 core CPU
1 NVIDIA GPU*	1 NVIDIA GPU*
100 GB of SSD space	100 GB of SSD space

- Note: TAO Toolkit is supported on discrete GPUs, such as H100, A100, A40, A30, A2, A16, A100x, A30x, V100, T4, Titan-RTX, and Quadro-RTX. TAO Toolkit is not supported on GPUs before the Pascal generation.

3. System Architecture

3.1.1 Overview

Renesas AI model deployer provides NVIDIA TAO features such as data augmentation, pruning and quantized aware training (QAT). The tool includes computer vision pipelines for object detection/image classification including dataset conversion, model training, model evaluation, inference and deployment.

3.1.2 High Level Architecture

The diagram illustrates the high-level architecture of the toolkit. Data management is performed locally using desktop data management. The toolkit comprises two workflows: MPU workflow and MCU workflow.

In the MPU workflow, once the model is trained on the toolkit, it is exported via ONNX and compiled by DRP-AI TVM for deployment on the board. In the MCU workflow, following model training on the toolkit, the model is exported to ONNX, translated to TFLite for quantization, and subsequently exported in .c format for MCU deployment. Real-time inference is achieved through web socket streaming from the board to the UI on MPU and via Segger Jlink for MCU.

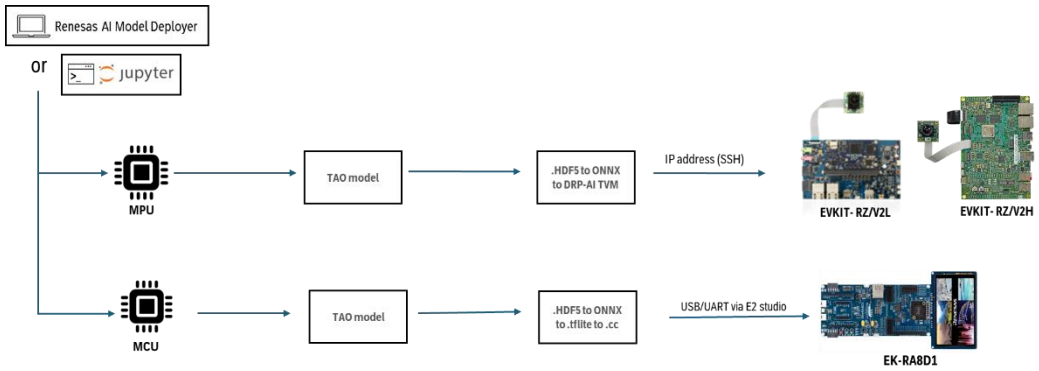


Figure 1: High Level Architecture

4. Installation Setup

4.1 Renesas AI model deployer

To get started by using the GUI please follow the installation procedure:

1. Visit Renesas NVIDIA TAO integration GitHub.
2. Navigate to releases, press the “Tags”
3. Click the latest vX.Y.Z tag
4. Under assets, click Renesas_AI_Model_Deployer_vX.Y.X.tar. It should download the .tar project with relevant files.
5. Then extract the file as follows:
   - `tar -xvf < name.tar >`
6. The project directory contains various folders and shell scripts such as:
   - Utils – Folder that stores all the files containing default values and configuration files for TAO including dataset analytics, conversion, training, evaluation and inference.
   - api.py – A python script that stores all information related to handling API calls.

7. You must then get your NVIDIA NGC API Key to be able to access NVIDIA TAO. Please refer to section **4.1.2 Generating a Personal NGC API Key**
8. Navigate to the toolkit directory, open a terminal and run the following commands inside the project folder to install the S/W (installation will take some time and request inputs such as NGC key and sudo rights):
   - `chmod ug+x *.sh`
   - `./docker_gpu_install.sh`
   - `./setup_tao_env.sh`
9. Start application, at times, Renesas TAO server may not start, please confirm **4.1.2 Prerequisites and start up error handling** section
   - `./gui_start.sh`
10. This should open a new terminal called TLT-server that outputs the logs of the various processes, whilst the main terminal will mention the status of the server.

4.1.1 Generating a Personal NGC API Key

1. Sign in to the NGC website. From a browser, go to **https://ngc.nvidia.com/signin** and then enter your email and password.
2. Click your user account icon in the top right corner and select **Setup**.

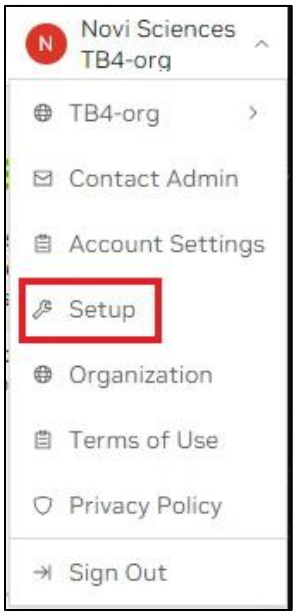


Figure 2: Side-panel of NGC website

3. Click Generate Personal Key from the available options. Personal Keys allow access to a set of NGC service APIs.

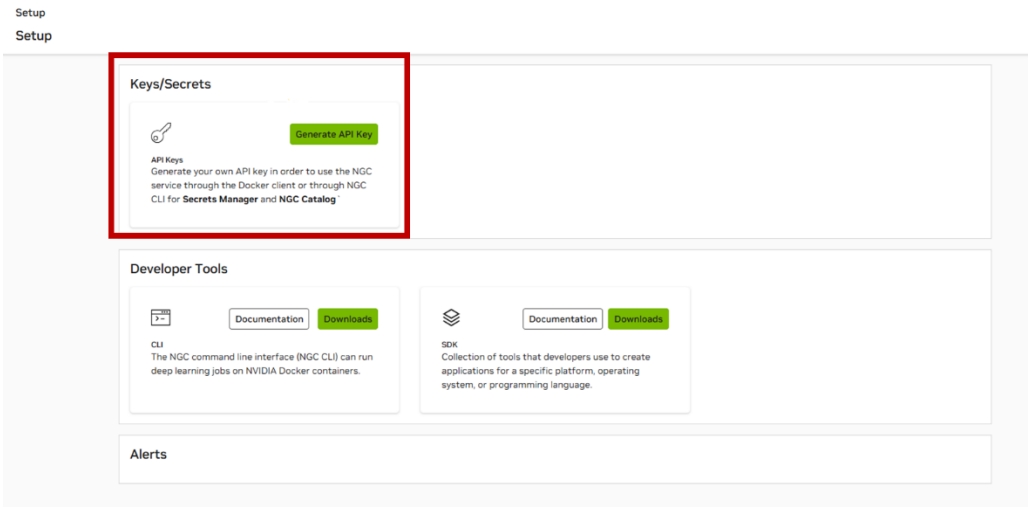


Figure 3: Personal Keys option

4. On the **Setup > Personal Keys** page, click **+ Generate Personal Key**, on the menu or the pane.

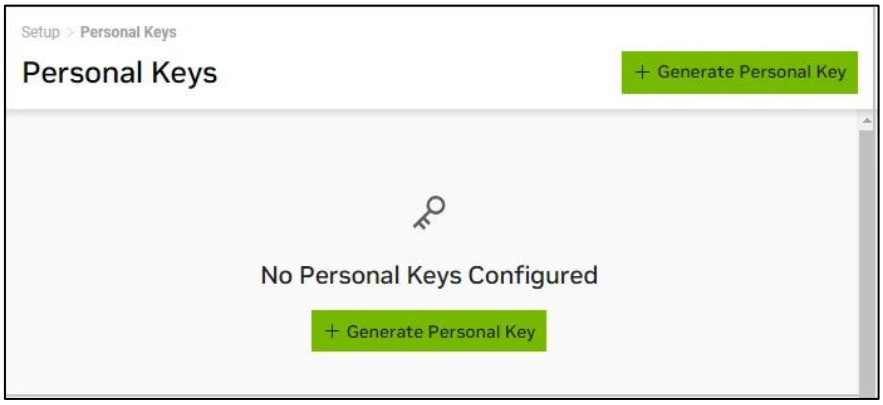


Figure 4: Personal Keys menu

5. In the **Generate Personal Key** dialog, fill in the required information for your key.

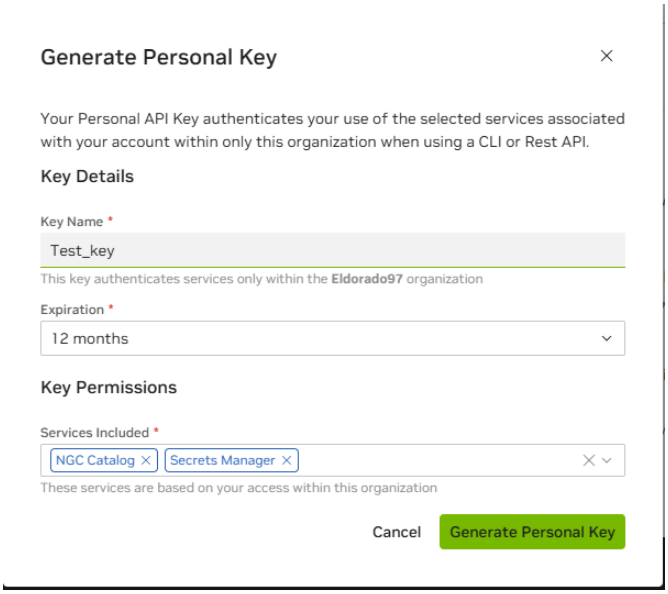


Figure 5: Personal Keys menu

- **Key Name**: Enter a unique name for your key.
- **Expiration**: Choose the expiration date for the key.

6. Click **Generate Personal Key** when finished. Your API key appears in the following dialog.
7. NGC does not save your key, so store it securely. You can copy your API Key to the clipboard by selecting **Copy Personal Key** or using the copy icon to the right of the API key.

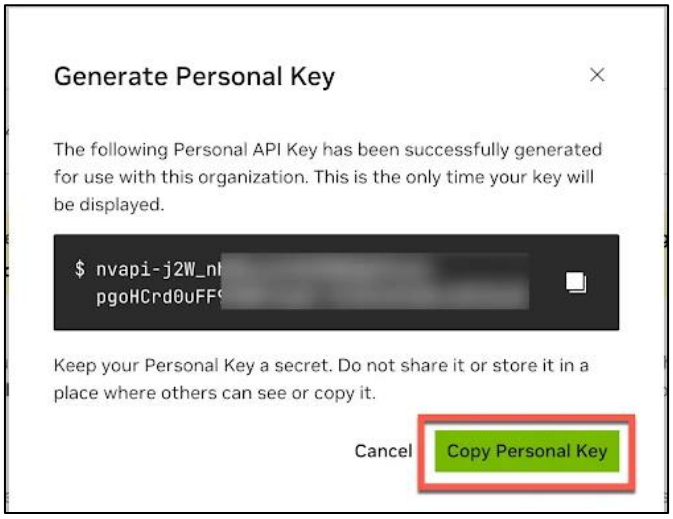


Figure 6: Personal Keys menu

4.1.2 Prerequisites and start up error handling

Below are the checklists to be taken into consideration before running the application.

1. Please check for Nvidia driver's version and CUDA versions by checking the commands below before starting the application in Ubuntu terminal.

- `$ nvidia-smi`
- `$ nvcc --version`

2. Please also check for backend applications running on GPU. Restart is recommended before running the application to avoid nvml error.
3. At times the port may be preoccupied when you try to `./gui_start.sh`, follow below steps to kill the process using port:8000.
   - Run the following command in the terminal to identify the Process-ID (PID).

     ```
     $ sudo lsof -i :8000
     ```

   - Run the following command to graceful / force kill all the PID identified.

     ```
     $ sudo kill < PID > or sudo kill -9 < PID >
     ```

   - Check to verify if the Port:8000 is free.

     ```
     $ sudo lsof -i :8000
     ```

4.2 RZ/V2L related setup

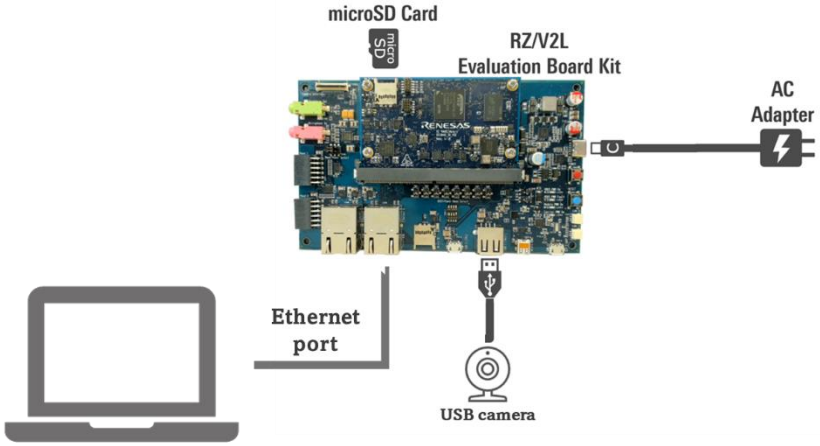
The set-up requirement for RZ/V2L is twofold, one is setting up AI SDK on the Ubuntu host machine to translate the model via GUI and second is to set up the RZ/V2L-EVKIT.

The detailed requirements can be followed on RZ/V AI GitHub [link], make sure to choose latest GitHub page from top left corner (labelled as “latest”) but at a glance:

1. Follow steps 1 & 2 that involve acquiring the board along with necessary components (listed in table below)

Equipment	**Details**
RZ/V2L EVKIT	Board to run demos
USB Cable Type-C	Connect AC adapter and the board.
AC adapter	Power supply to the board.
USB camera	Used to visualize the inference
microSD card	Must have over 4GB capacity of blank space.
	Operating Environment: Transcend USH-I microSD 300S 16GB
Linux PC (for initial setup)	Used for Setup microSD card and RZ/V2L AI SDK Setup.
	Operating Environment: Ubuntu 20.04
SD card reader	Used for setting up microSD card.
USB Hub	Used to connect USB Keyboard and USB Mouse.
USB Keyboard	Used to type strings on the terminal of board.
USB Mouse	Used to operate the mouse on the screen of board.
HDMI Monitor	Only if you want to run the end application code independent of GUI or host
Micro HDMI Cable Ethernet cable	Used to connect the HDMI Monitor and the board. To communicate with host PC

2. Step 3-5, download RZ/V2L AI SDK and setting up the docker container for your RZ/V board. Please note, for the best performance use **DRP-AI TVM v2.5.**
3. Step 7 provides an explanation of how to format the SD card to boot the board, bootloader and other Yocto necessary files. Please follow this step if you are starting with a new board or want to update the AI SDK.
4. Once the board is set up, please connect as shown in figure 7, do note the switch SW1 & SW11 orientation depend on which bootloader (eSD/eMMC) you opted for
5. Once you are done with testing, to safely shutdown the board, please check the following link.



.

Figure 7: Hardware wiring and connectivity

5. MPU DetectNet_V2 Objection detection

The demo uses **DetectNet v2** with a **ResNet-18 backbone**, a lightweight yet powerful architecture well-suited for real-time object detection tasks on embedded platforms. The model is trained using the **KITTI dataset**, which includes annotated images for cars, pedestrians, and cyclists, making it ideal for evaluating urban and autonomous perception scenarios.

5.1.1 Workflow Summary

In this guide, we will:

1. **Train** the DetectNet v2 model on the KITTI dataset using transfer learning.
2. **Prune** the model to reduce its size and remove redundant parameters, improving inference speed and efficiency for edge deployment.
3. **Retrain with Quantization-Aware Training (QAT)** to prepare the model for INT8 quantization, ensuring optimal performance on Renesas' DRP-AI hardware accelerators.

This workflow ensures a balance between accuracy, efficiency, and deploy-ability, making the model suitable for real-world embedded applications.

5.1.2 Model download

The model can be downloaded from Nvidia NGC here. Please store the .hdf5 model under <projectdirectory>/utils/config/detection/pretrained_models.

5.1.3 Dataset download

For this demo, KITTI formatted dataset separated to three classes (pedestrian, cyclists, cars) is used to trained the model and draw bounding boxes on the classes. The dataset can be downloaded from the following link. The labels are available here.

For more details on the dataset, please visit the KITTI vision benchmark suite webpage.

5.2 End-to-end GUI flow

As mentioned previously, you can start the GUI by inputting `./gui_start.sh` in terminal within project directory.

5.2.1 Landing page

When the user opens the application, Renesas AI model deployer landing page is displayed as follows

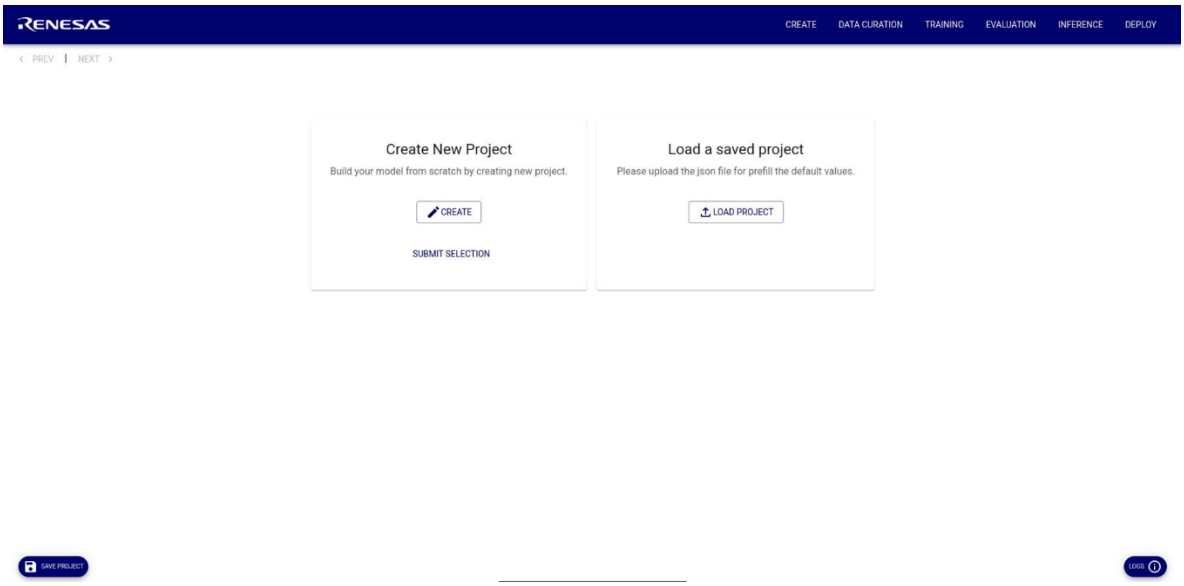


Figure 9: GUI Landing Page

On this page, you can load default settings from a previously worked-on project. The fields will be pre-filled based on your earlier work, allowing you to continue development seamlessly. You can also create a new project. For this tutorial, we will create a new project.

Step-1: Click on **CREATE** button to create a new project with the custom details.

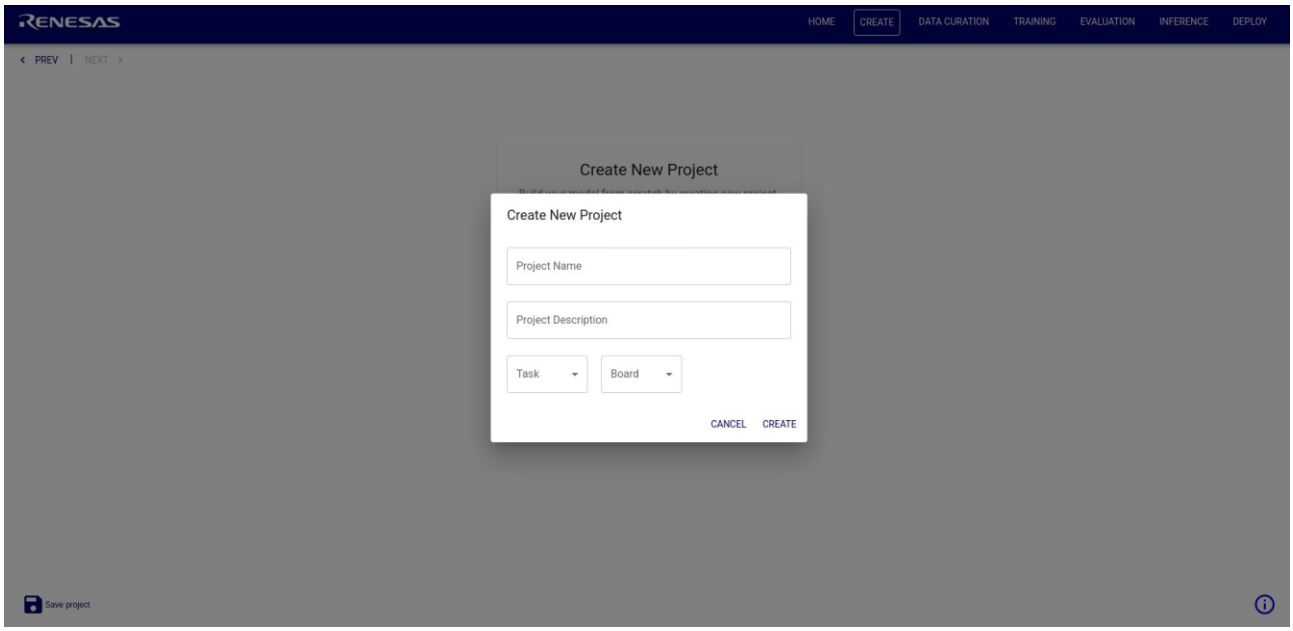


Figure 120: Create New Project pop-up

Step-2: Fill the **Project Name** (The Project name refers to a unique identifier that user assigns to specify the project)

Step-3: Fill the **Project Description** (A field where user can provide a brief explanation of the project)

Step-4: Select the **Object Detection** from the Task drop-down menu to configure the toolkit appropriately for the specific workflow, pre-trained models, and configurations required for the task.

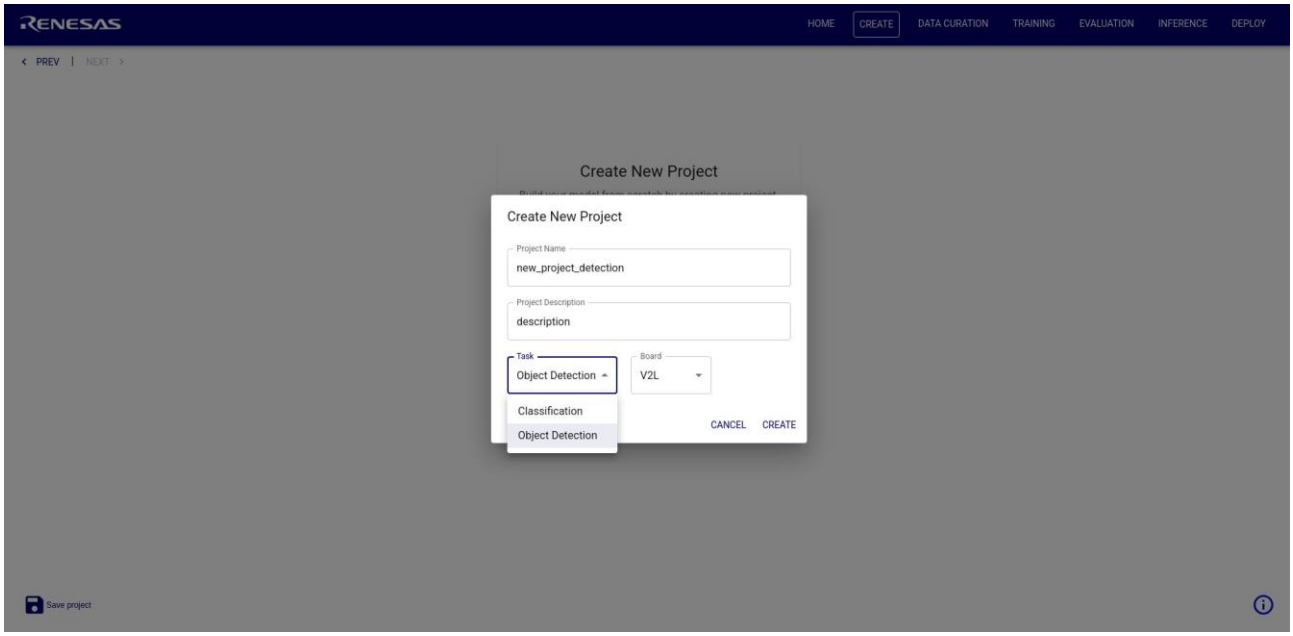


Figure 3]11: Object Detection selection from the Task drop-down menu

Step-5: Select either **V2H** or **V2L** option from the Board drop-down menu to choose the specific hardware platform or development board on which the model will run. For this demo, we will use **RZ/V2L.**

Step-6: Once the details are filled, click on **CREATE** button to create the project and wait till the **Successfully created the task!!** message is displayed like shown in the below image.

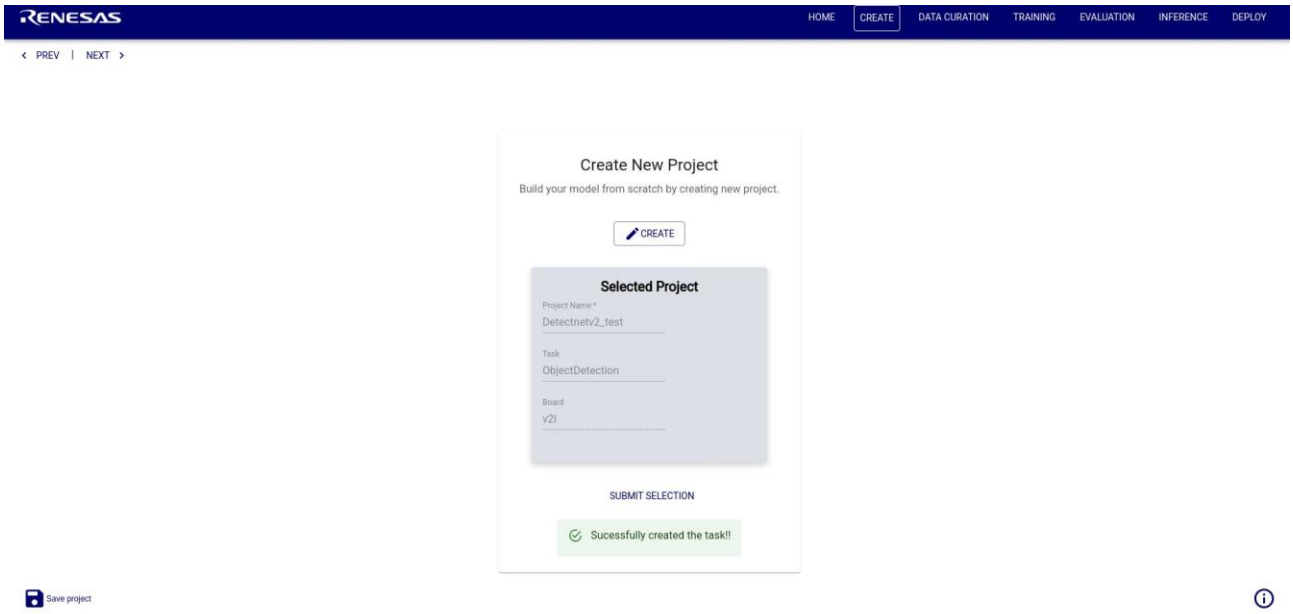


Figure 12: New Project Created

Step-7: Click on the **SUBMIT SELECTION** button to proceed to DATASET page.

5.2.2 DATASET Curation

In this page, the sample KITTI dataset will be selected, and the statistical analysis provided in graphical and tabular formats.

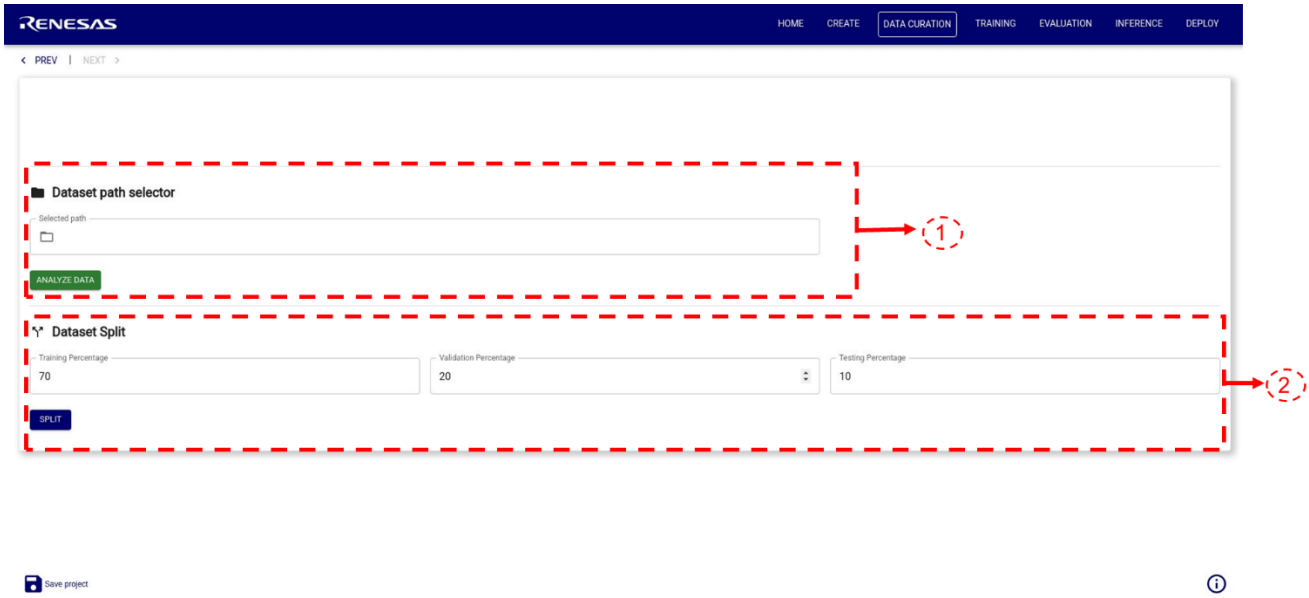


Figure 43: Dataset Page

Step-1: Choose the directory that contains **KITTI data** from the root folder by using the **Selected path** option from the **Dataset path selector** section and Click on **SUBMIT** button. It will take a few seconds to fetch the dataset to generate the analytics results are shown as in the image below. **Note: For this pipeline, only KITTI data is supported.** The path selector should point to a directory with two folders, image_2 and label_2 folders where the images are in .png format and the labels in .txt format with the same naming convention.

Figure 54: Display of Analysis of the Dataset Selection

Step-2: Adjust the values of **Training Percentage**, **Validation Percentage** and **Testing Percentage** from the **Dataset Split** section as per the requirement and Click on the **SUBMIT** button.

Wait until the "**Successfully completed dataset split**" message is displayed as shown in the below image, click on **GO FOR TRAINING** button to proceed to the **TRAINING** page.

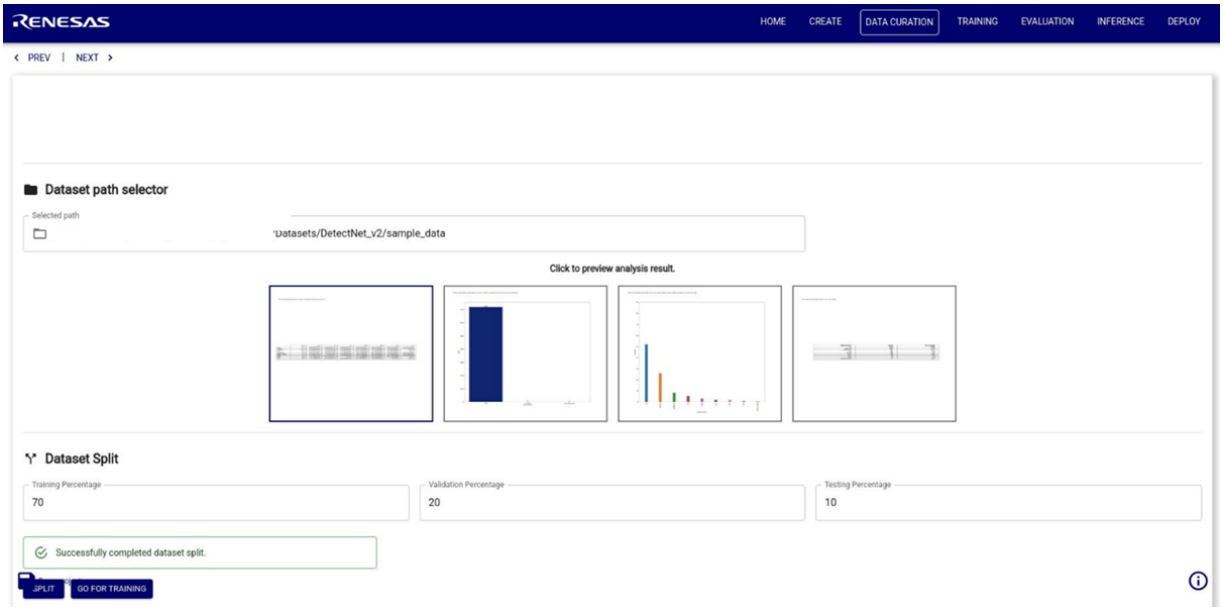


Figure 65: Display of GO FOR TRAINING button and successfully completed dataset split message.

5.2.3 TRAINING

Upon clicking the “GO FOR TRAINING” button from the following page is displayed.

In this page, user can train the detectnet_v2 model with predefined parameters.

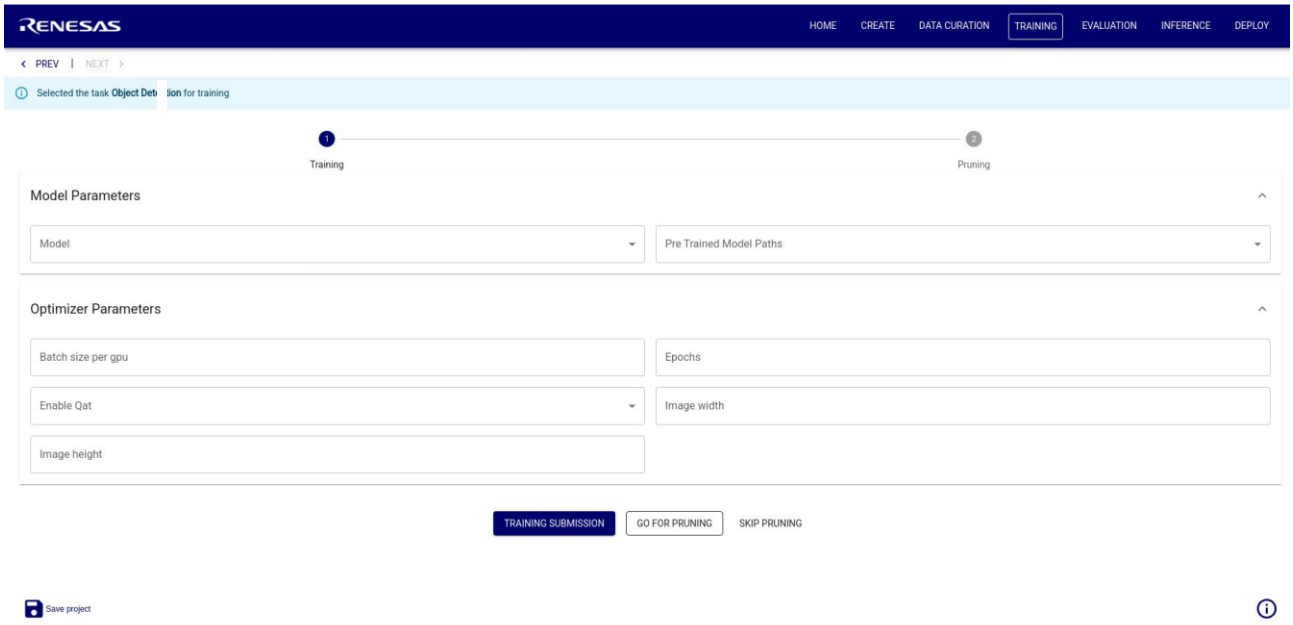


Figure 76: Model Training Page

Step 1: Choose **detectnet_v2** from the **Model** drop-down menu and **resnet_18** from the **Pre-Trained Model Paths** drop-down menu where the model is initially stored, as shown below.

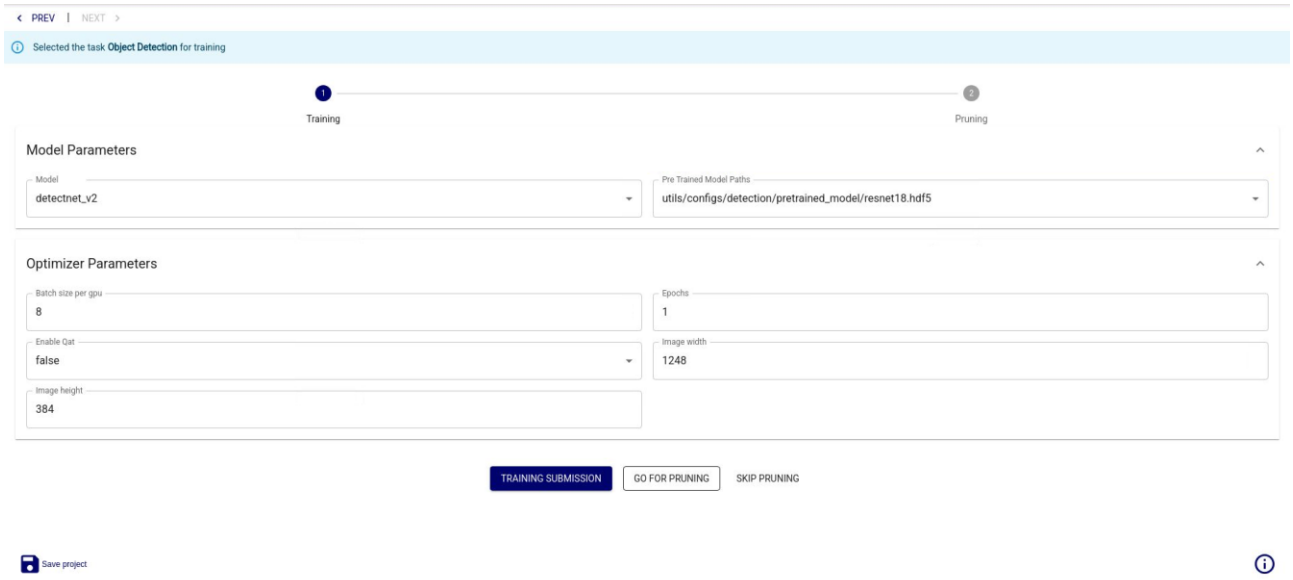


Figure17: Displaying the Model Parameters section details

Step-2: Adjust the predefined values in the **Optimizer Parameters** section as needed. Specifically, modify:

- **Batch Size per GPU**: The number of samples processed by each GPU in a single training iteration.
- **Epochs**: The number of complete passes through the training dataset. Do note, if you train for ~100 epochs, it will take several hours.
- **Enable QAT**: Turn on Quantization-Aware Training to prepare the model for INT8 quantization and improve inference performance during deployment.

Step-3: Click on **TRAINING SUBMISSION** button to start the training and "**Training started**" message is displayed to show the status as shown in the below image. By clicking the "logs" button in the right corner live logs can be viewed.

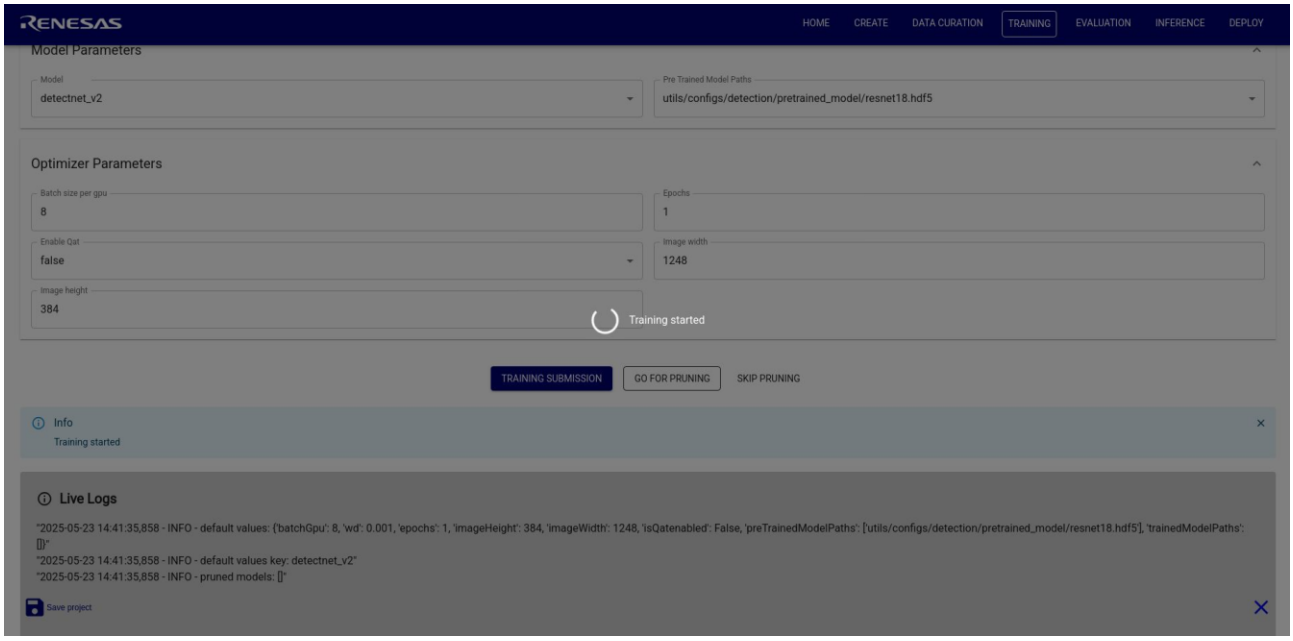


Figure 18: Displaying the Training started status

Step-4: Once training is complete, click the **GO FOR PRUNING** button to proceed to the next stage. For DetectNet v2 models, NVIDIA TAO supports **unstructured pruning**, which removes the lowest-magnitude weights based on a defined threshold. The model is then retrained to recover accuracy by learning better weight distributions, helping to reduce model size while maintaining or even improving performance.

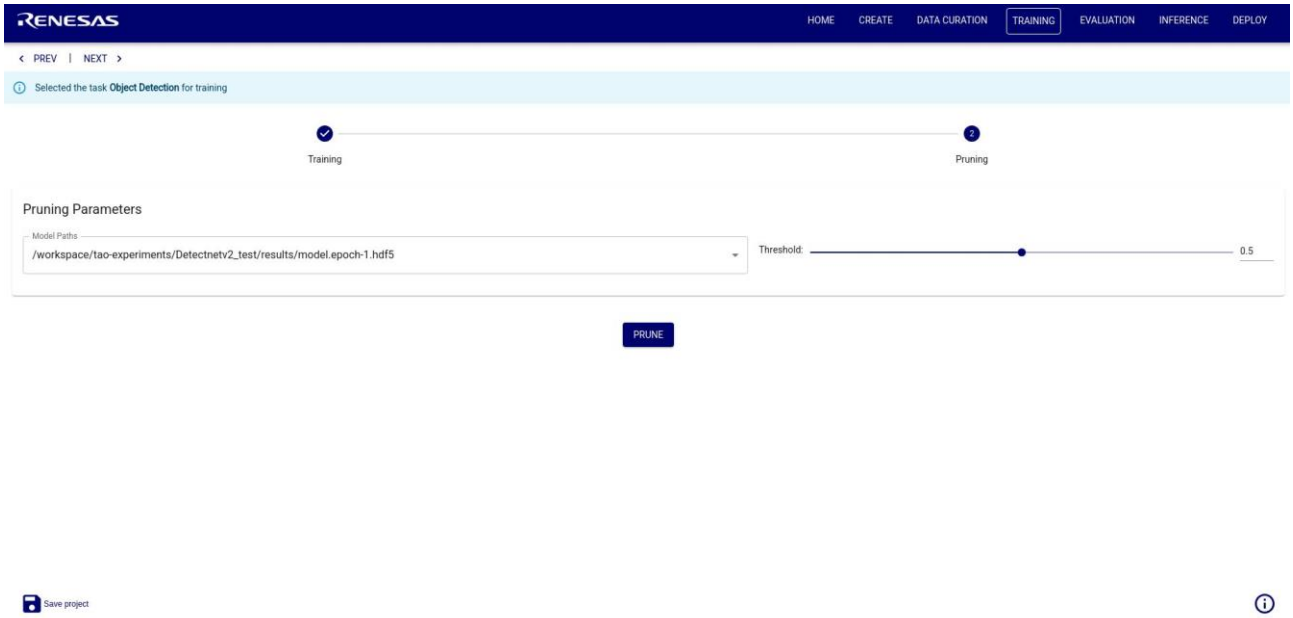


Figure 19: Pruning stage

Step-5: Select the model from the **Model Paths** drop-down menu and select/adjust the pruning **Threshold** slider as per the requirement. Click on the **SUBMIT SELECTION** button to start the pruning process. “Pruning started” message is displayed to show the status.

Step-6: Wait until the Pruning is completed; Retraining box would be visible with prefilled parameters based on the initial training selection. Click on **RE-TRAINING** button to re-train the model and the “**Re-Training started**” message is displayed as shown.

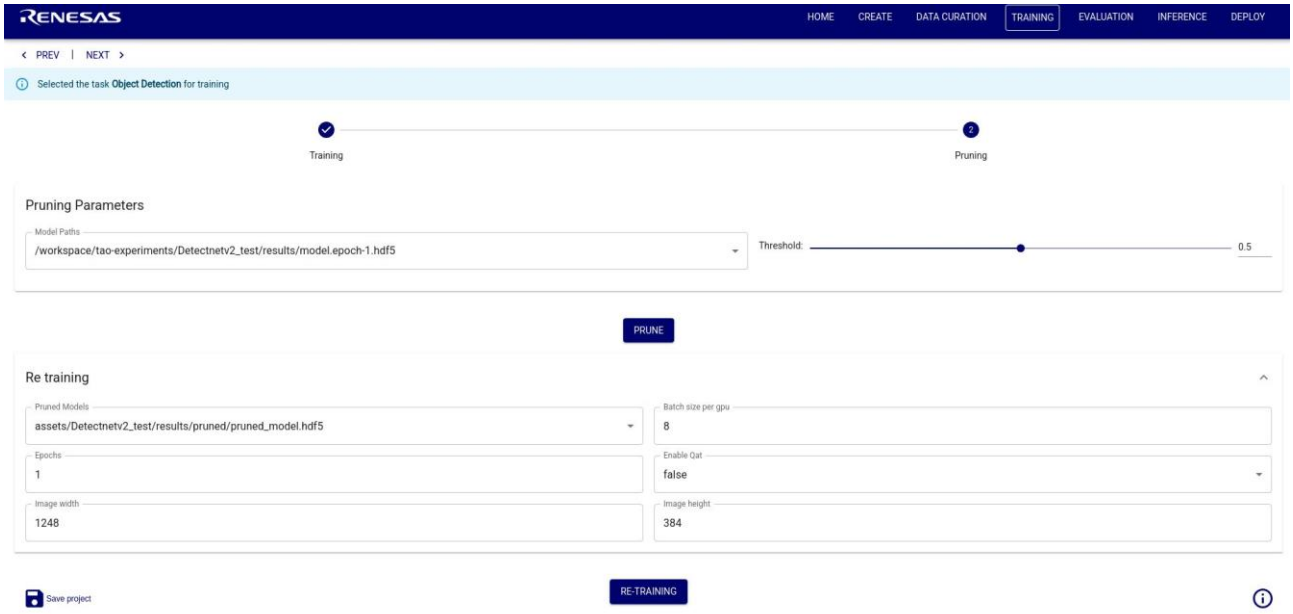


Figure 80: Displaying the status of the Re-Training process.

Step-7: Wait until the re-training is completed, click on next to go to **EVALUATION** page.

5.2.4 EVALUATION

In this page, it allows the users to assess the performance of the trained model.

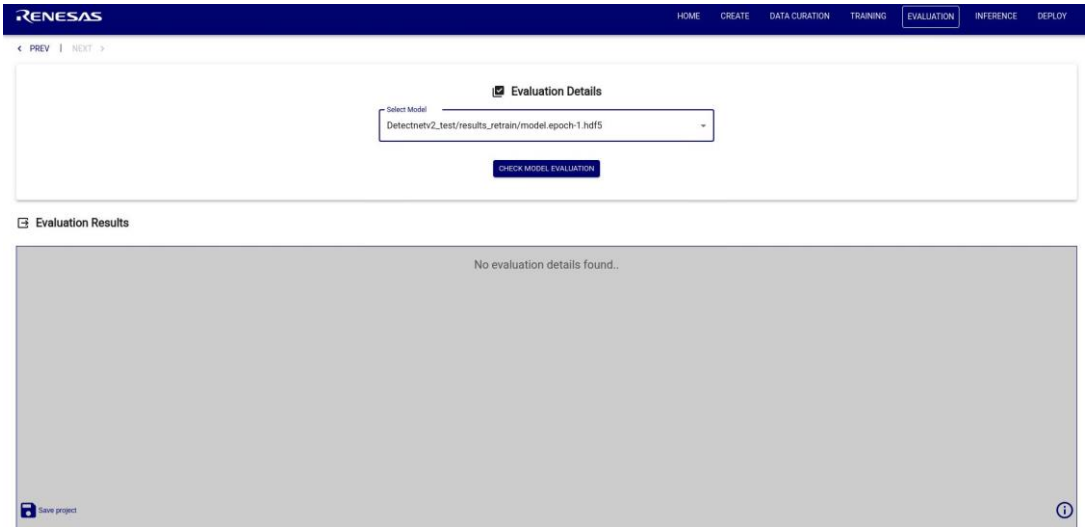


Figure 91: Model Evaluation Page Select

Step-1: Select the trained model from the dropdown from the **Select Model** drop-down menu and click on **CHECK MODEL EVALUATION** button to see the evaluation results in the image format. The "**Checking evaluation details**" message is shown as status.

Step-2: Wait until the Evaluation process is completed and click on the obtained evaluation image result to get the mAP values for **detectnet_v2 model** as shown in below image. Do note, the image below is based on a model that was trained for 120 epochs, pruned by 0.5 pth and retrained for another 120 epochs that resulted in the following mAP results. For further improvement, either train for more epochs, change the hyperparameters in the .yaml file or add more dataset for training.

Figure 10: Model Evaluation Results

Step-3: Close the image and go for the **INFERENCE** section.

5.2.5 INFERENCE

In this step, the user can check the inference results on an image and export the model (.hdf5) to ONNX format.

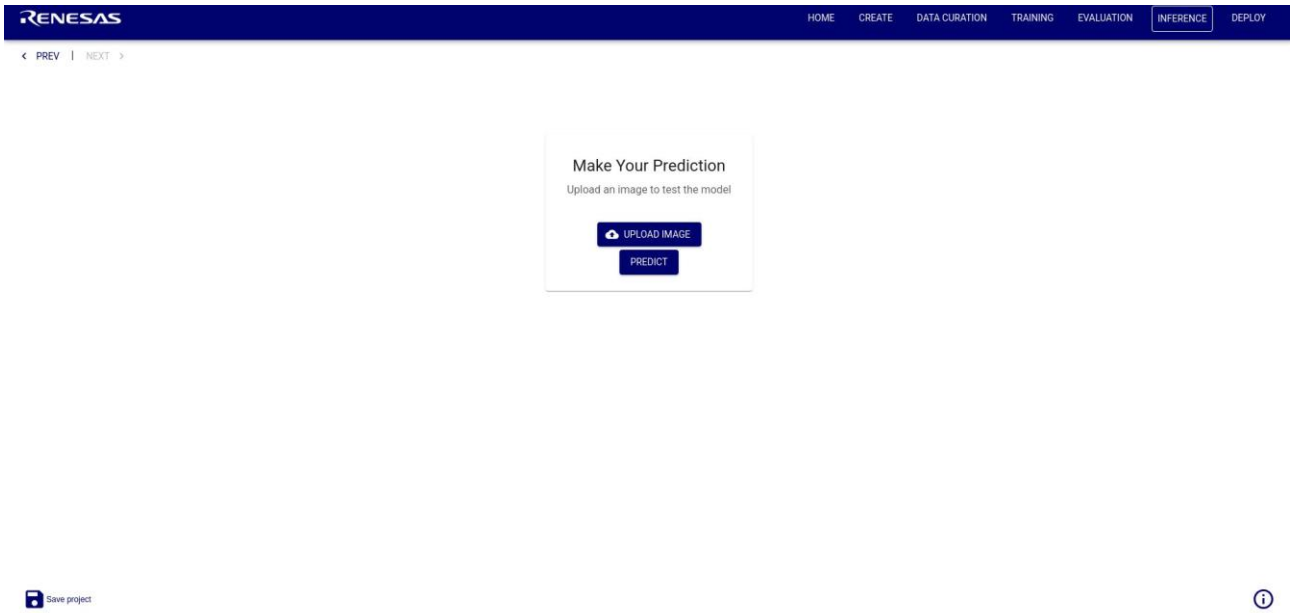


Figure 11: Inference Page

Step-1: Click on **UPLOAD IMAGE** button to select an image and Click on **PREDICT** button to get the inference results as shown below.

Figure 12: Prediction Results

You can then proceed to the deploy page by click on the next button.

Step-4: Select model form the **Model paths** drop-down menu and click on **EXPORT** button.

Step-5: Wait until the "**Model successfully exported**" message is displayed and then click on **DEPLOY** button in the header section to go to deploy page.

5.2.6 DEPLOY

The user first converts the .hdf5 model into. ONNX format which will then be compiled via DRP-AI TVM into binary that can be run on RZ/V2L board. The model can then be executed on the board by inputting board IP to connect to the board and deploy the model for checking the live inference. Do note, for RZ/V2H,during translation post training quantization is done in the backend to be runnable on DRP-AI 3. Similarly for RA8D1 use case.

Step-1: Select model from the **Model paths** drop-down menu and click on **EXPORT** button.

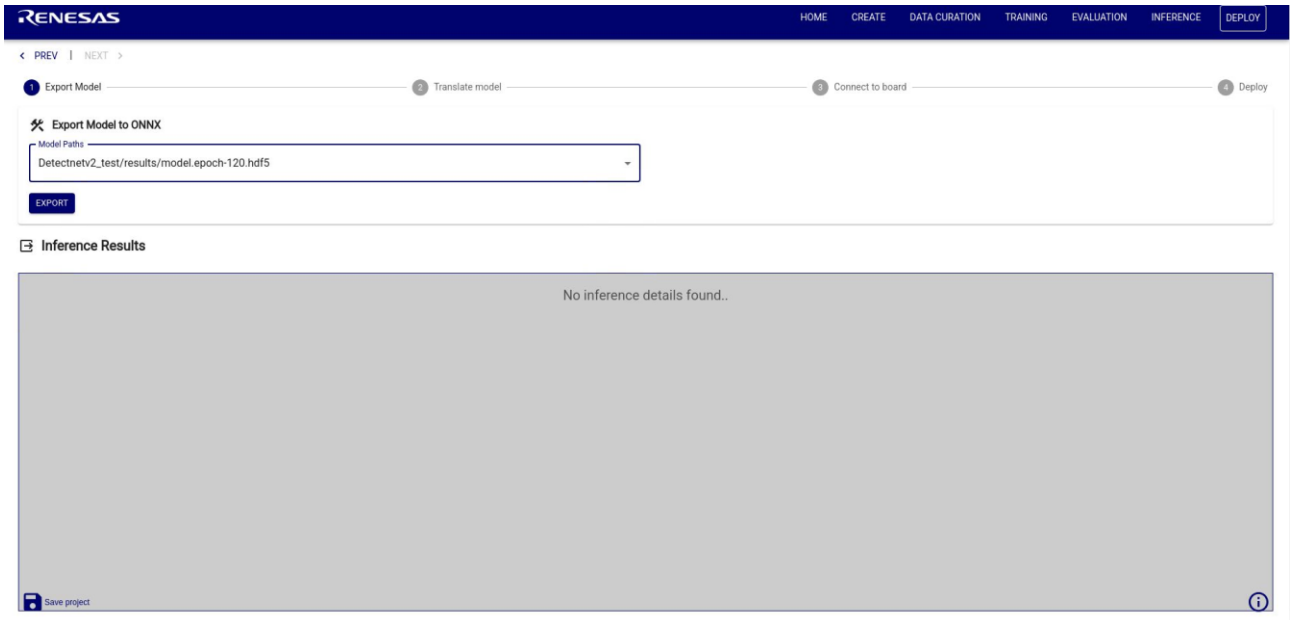


Figure 25. Exporting the model to ONNX

Step-2: Wait until the "**Model successfully exported**" and it will move you to the translation page, click on **TRANSLATE** button. In the translation page,

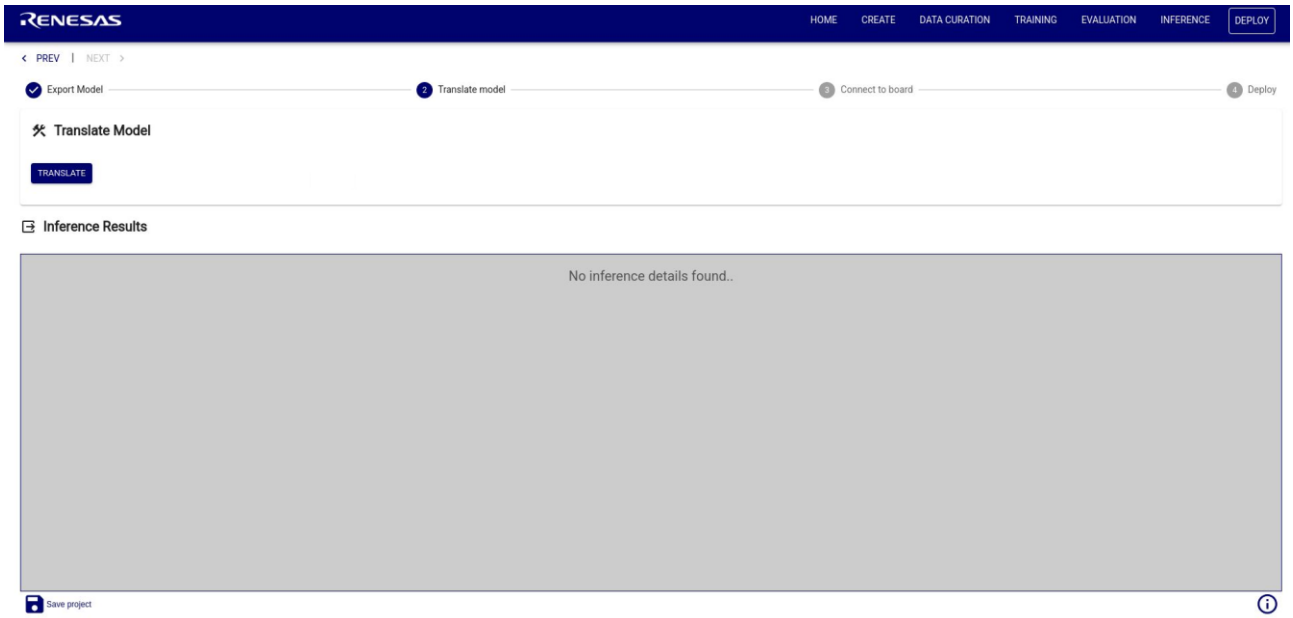


Figure 26. Translating the model

Step-3: Wait for the "**Model successfully translated**" message to be displayed and the user will be prompted to the board connection.

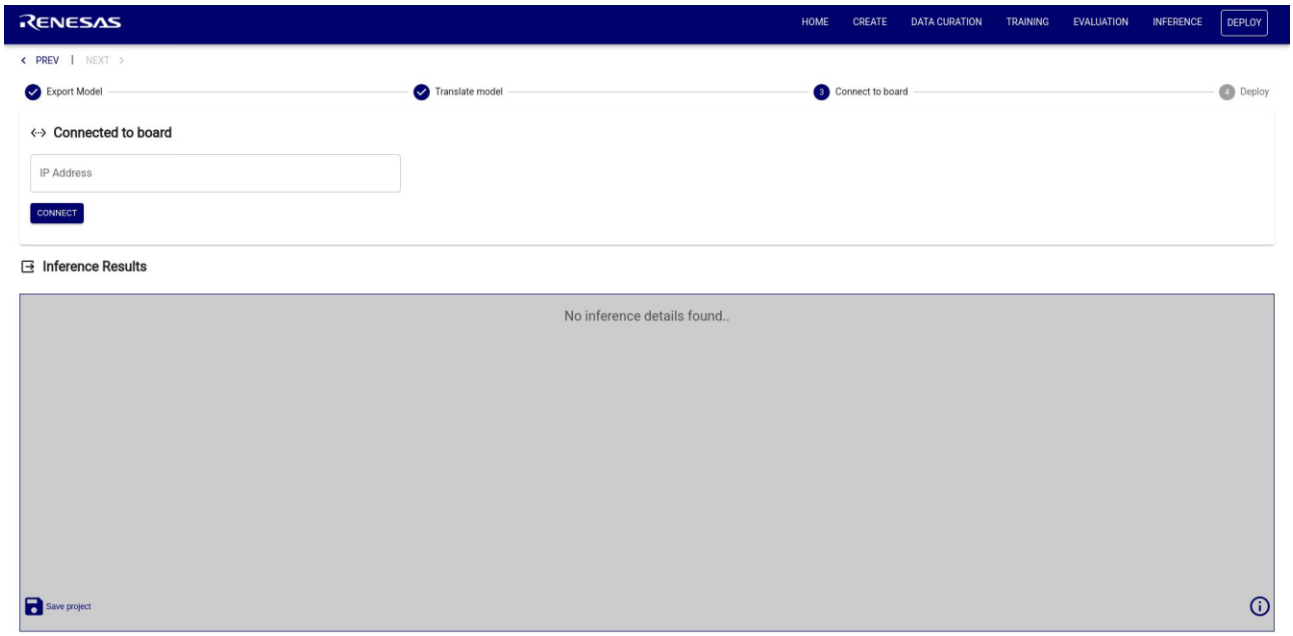


Figure 27: Input the board's IP address

Step-3: In the **Connect to board** section, enter the **IP address** and click on **CONNECT** button, wait for a response and the user will be prompted to deploy.

Figure 28: Connect to board section

Step-4: Wait for the "**Board Detected**" message display as shown in the below image. Wait for a couple of seconds and click **Deploy**, you should be able to see a live screen displayed as below.

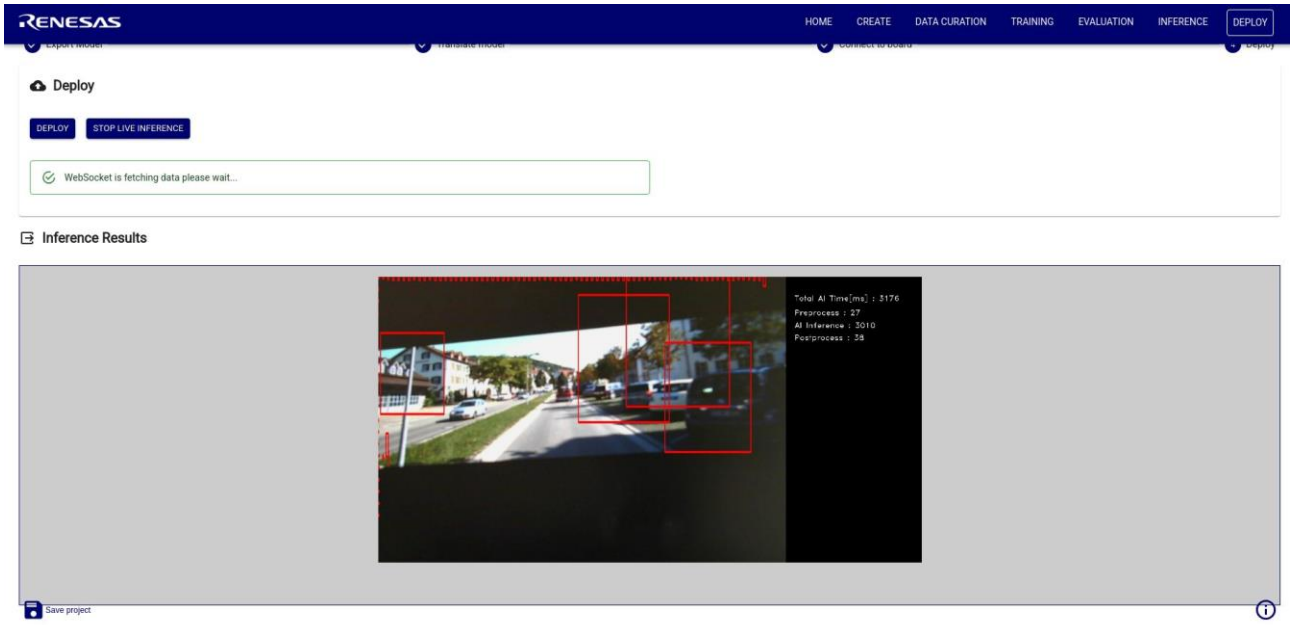


Figure 29: Board detected message display.

From Setup to Deploy, all the stages are completed

6. References

1. Renesas AI Model Deployer Github
2. TAO Toolkit | NVIDIA Developer
3. GitHub - NVIDIA/tao_tutorials: Quick start scripts and tutorial notebooks to get started with TAO Toolkit
4. TAO Toolkit Getting Started | NVIDIA NGC
5. TAO pretrained classification (Detectnetv2)
6. KITTI vision benchmark suite
7. DRP-AI TVM on RZ/V series
8. Release Notes - NVIDIA Docs.

9. RZ/V2L EVK Getting Started

7. Next Steps

After exploring the GUI-based Detectnetv2 workflow, users are encouraged to further explore the full capabilities of the TAO Toolkit through testing the other two pipelines, namely Segformer or Mobilenet v2. Alternatively, users are encouraged to explore provided Jupyter Notebooks for each supported AI model. These notebooks offer greater flexibility and control for advanced customization, training, evaluation, and deployment, enabling users to fully leverage TAO's potential beyond the GUI interface.

8. Website and Support

Visit the following URLs to learn about the kit and the RA family of microcontrollers, download tools and documentation, and get support.

Renesas Artificial Intelligence (AI)	renesas.com/ai
RA Product Information	renesas.com/ra
RA Product Support Forum	renesas.com/ra/forum
RZ/V2H Product Information	renesas.com/rzv2h
RZ/V2L Product Information	renesas.com/rzv2l
Renesas Support	renesas.com/support

Revision History

Rev.	Date	Description	
		Page	Summary
1.00	June.16.2025	—	Initial release

Renesas AI Model Deployer – Quick Start Guide

Publication Date: Jun.16.25

Published by: Renesas Electronics Corporation

RENESAS AI Model Deployer – Quick Start Guide

Renesas Electronics Corporation

R11QS0064EJ0100