



# eIQ Toolkit™ for i.MX RT1170 Model Runner Getting Started

---

Revision 4  
August 2023

---

# Contents

<b>1 Lab Overview .....</b>	<b>3</b>
<b>2 Hardware Requirements .....</b>	<b>3</b>
<b>3 Software Requirements.....</b>	<b>3</b>
<b>4 Model Runner with DeepViewRT.....</b>	<b>3</b>
4.1 Import Model Runner Project.....	3
4.2 Run the DeepViewRT Model Runner Example.....	4
4.3 Collect Profiling Information .....	5
<b>5 Model Runner with TFLM or Glow.....</b>	<b>9</b>
5.1 Import Model Runner Glow Project.....	9
5.2 Run the DeepViewRT Model Runner Example.....	10
<b>6 Conclusion.....</b>	<b>11</b>

## 1 Lab Overview

---

This document will cover how use Model Runner on an i.MX RT1170 EVK to get model profiling information on a target device. It will also cover how to use Model Runner with Glow or TFLM to send data to those inference engines running on an embedded device over an Ethernet connection. However profiling information is only available when using the DeepViewRT inference engine.

## 2 Hardware Requirements

---

This lab is written for the i.MX RT1170 evaluation board. This lab can also be used with the following evaluation boards that support DeepViewRT in the SDK by downloading their respective SDK packages:

- i.MX RT1050
- i.MX RT1060
- i.MX RT1064
- i.MX RT1160
- i.MX RT1170

## 3 Software Requirements

---

The following pieces of software are required:

- [MCUXpresso IDE](#)
- [MCUXpresso SDK for the board being used](#)
  - Make sure to unzip the package as some files inside of it will be used.
- [eIQ Toolkit](#)

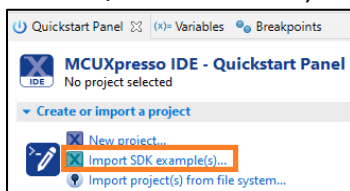
## 4 Model Runner with DeepViewRT

---

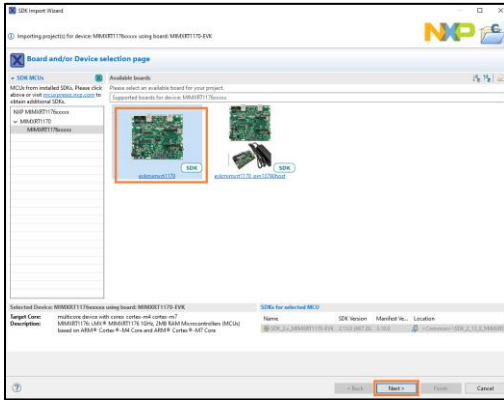
If you have a model in the DeepViewRT format, the first step to generating the model profiling information is to load and run the DeepViewRT Model Runner SDK project on the target device. This project can be found in the MCUXpresso SDK.

### 4.1 Import Model Runner Project

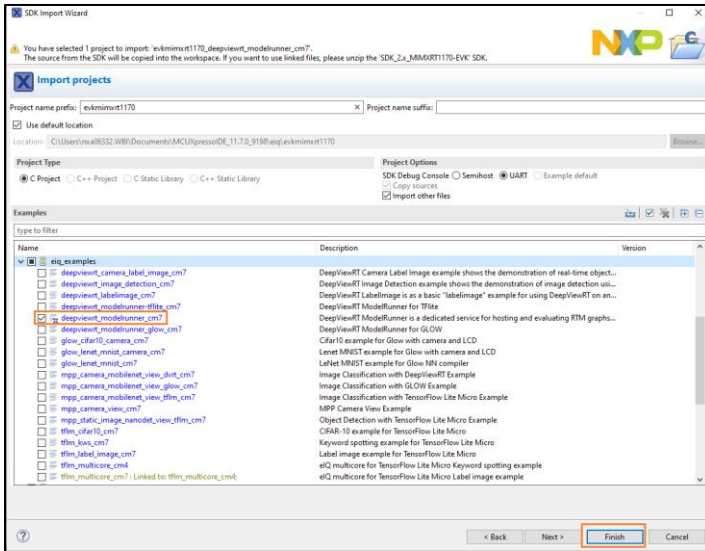
1. Open MCUXpresso IDE and select a workspace.
2. Add the RT1170 SDK zip file to the Installed SDKs window
3. In the QuickStart menu, select **Import SDK example(s)...**



4. Select the RT1170-EVK (**evkmimxrt1170**) and click on **Next**



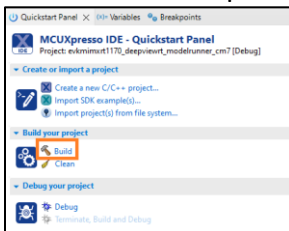
5. Expand the **eIQ** category and select the **deepviewrt\_modelrunner\_cm7** example. Click on the **Finish** button to import the project into your workspace.



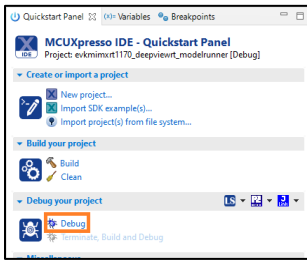
## 4.2 Run the DeepViewRT Model Runner Example

Now build and run the project that was just imported

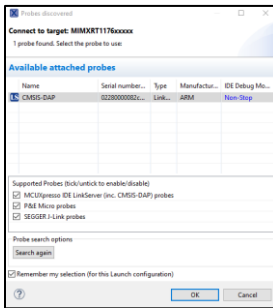
1. Connect an Ethernet cable to the **1G ENET** Ethernet port (J4) and connect the other end to a router that is on the same network as your PC. This is how the board will download the model and images from the eIQ Portal
2. Connect a micro-B USB cable to the board on J11
3. Open TeraTerm or other terminal program, and connect to the COM port that the board enumerated as. Use 115200 baud, 1 stop bit, no parity as the terminal settings.
4. In the Quickstart panel, click on **Build**. You should see no errors in the Console tab.



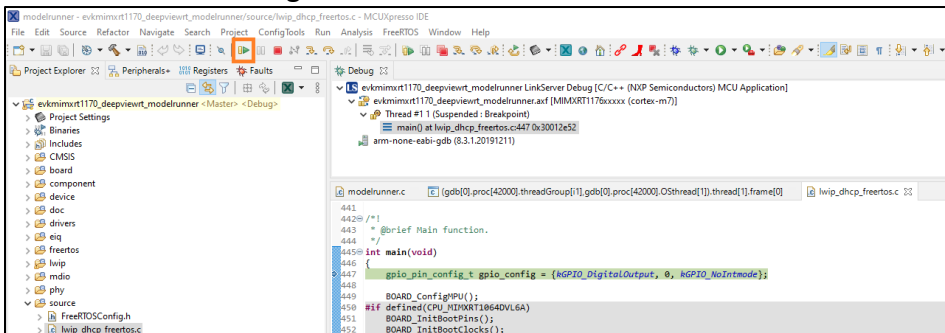
5. In the Quickstart pane select **Debug**



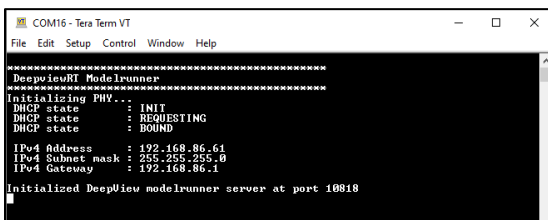
6. It will ask which interface to use. Select the CMSIS-DAP probe and click on OK.



7. The debugger will download the firmware and open up the debug view. Click on the Resume button to start running.



8. In the terminal you should see the board acquire an IP address from the router. Make note of this IP address as it'll be used in the next section.

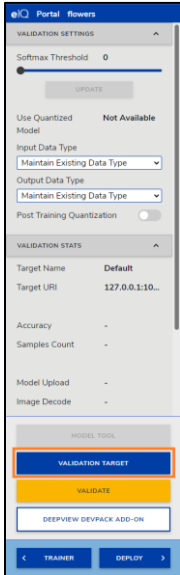


### 4.3 Collect Profiling Information

Now that firmware is loaded on the board, eIQ Toolkit can be used to collect the profiling information for that DeepViewRT model.

1. Open eIQ Toolkit
2. From the Home screen, click your existing eIQ Portal project. You can use the Flower project from the previous lab document.
3. Make your way to the Validation screen like done in the previous lab document.

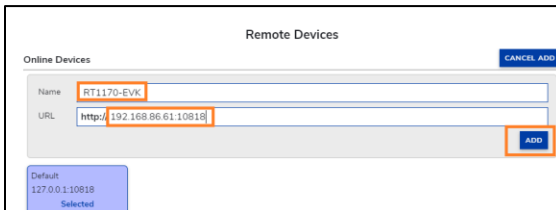
- By default the validation will run on the local PC. But you can run it on the EVK instead by click on the **Validation Target** button



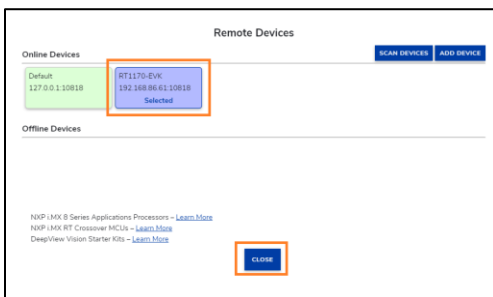
- Click on Add Device in the dialog box that comes up



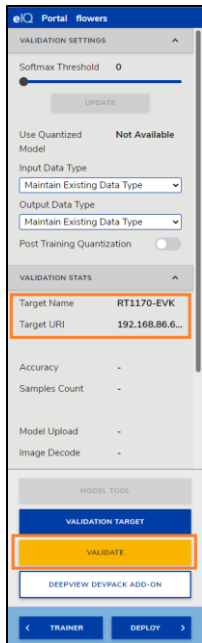
- Then add the board by giving it a name and the IP address that was printed out into the terminal, with the port **:10818** at the end. Then click on **ADD**



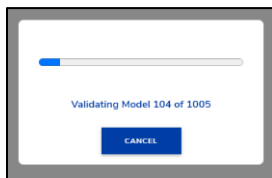
- Then click on the new board just added to that it says Selected underneath it and turns purple. And then click on Close at the bottom.



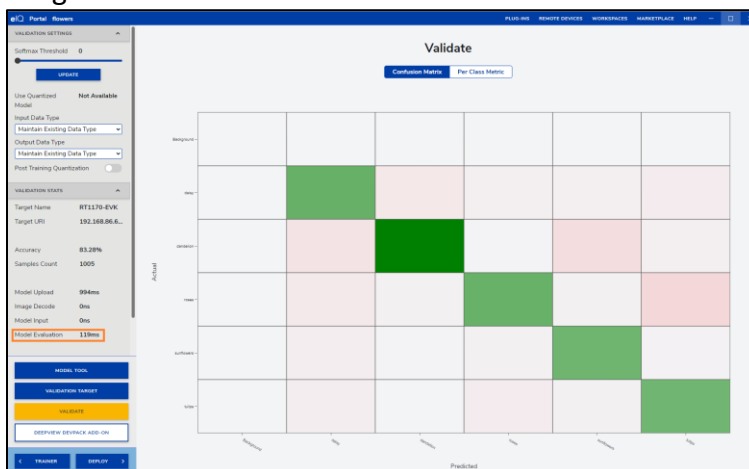
- Verify the **Target Name** and **Target URI** look correct for your board and then click on **Validate** to download the model to the RT1170-EVK over the Ethernet connection and start giving it images from your Dataset to do the validation on.



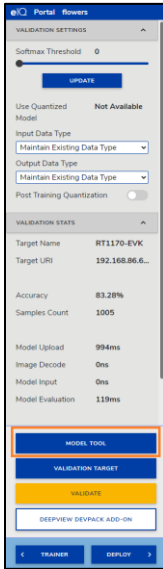
- eIQ Portal will begin sending over the Ethernet cable the images in the Validation category to the board for inferencing.



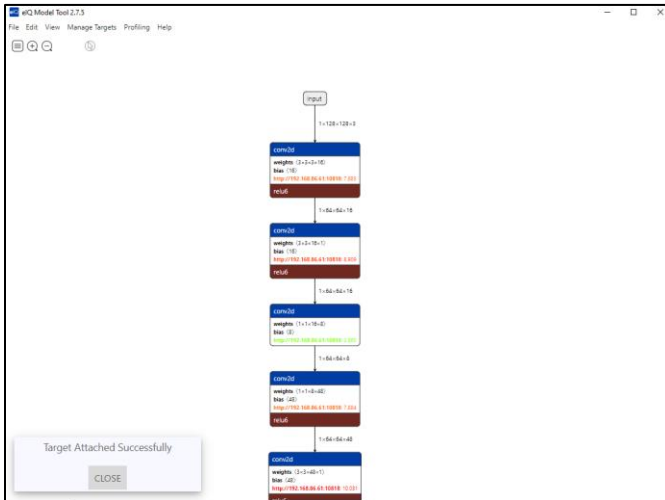
- You should see a confusion matrix come up that shows how accurate the model is for different images as well as the overall execution time for an inference. In this case it was 119ms.



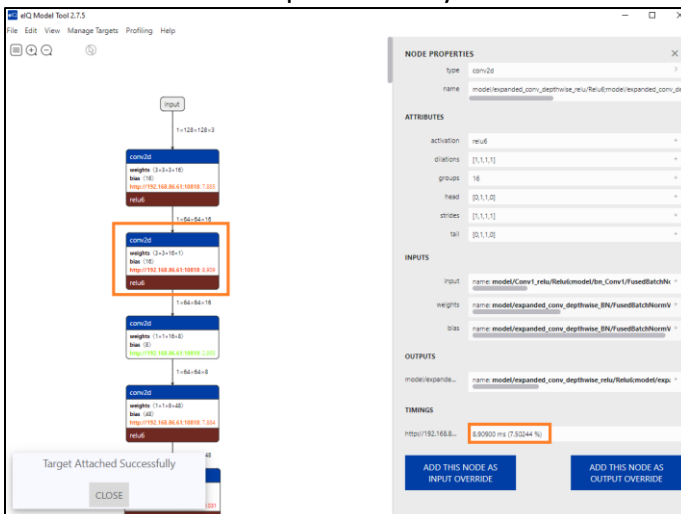
11. The profile information can be viewed by clicking on the **Model Tool** button



12. You should see the model diagram come up with the timings as recorded on the EVK:



13. You can click on a particular operation to see the time in milliseconds and the percentage of overall time that that particular layer used. This can be used to decide which layers to optimize.



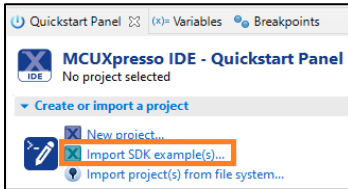


## 5 Model Runner with TFLM or Glow

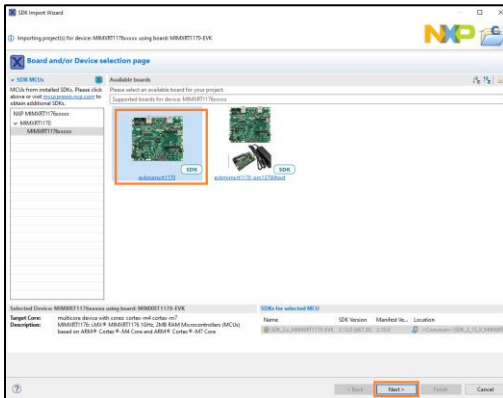
Model Runner can also be used to send images from a PC to a Glow model running on the board. This can be useful when verifying a Glow model on hardware. The same can also be done with TensorFlow Lite Micro (TFLM). Note that with these two inference engines that profiling information cannot be collected.

### 5.1 Import Model Runner Glow Project

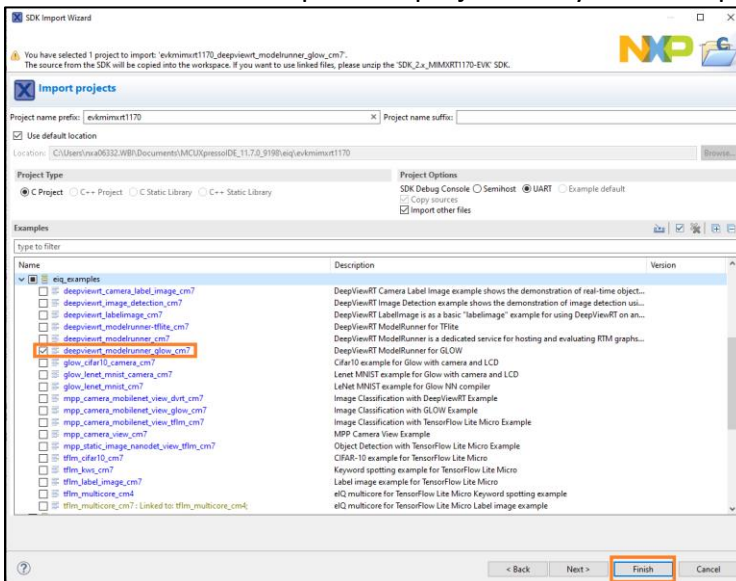
1. Open MCUXpresso IDE and select a workspace.
2. Add the RT1170 SDK zip file to the Installed SDKs window if not done in previous section.
3. In the QuickStart menu, select **Import SDK example(s)...**



4. Select the RT1170-EVK (**evkmimxrt1170**) and click on **Next**



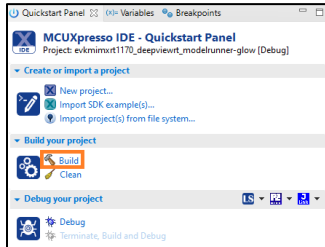
5. Expand the **eiQ** category and select the **deepviewrt\_modelrunner\_glow\_cm7** example. Click on the **Finish** button to import the project into your workspace.



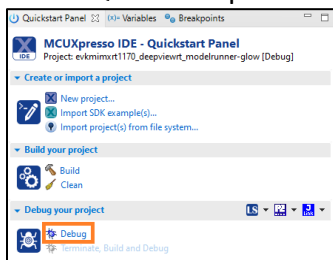
## 5.2 Run the DeepViewRT Model Runner Example

Now build and run the project that was just imported

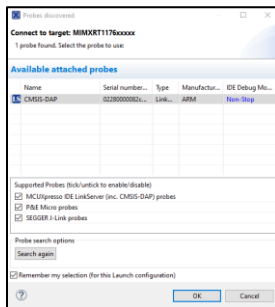
6. Connect an Ethernet cable to the **1G ENET** Ethernet port (J4) and connect the other end to a router that is on the same network as your PC. This is how the board will download the model and images from the eIQ Portal
7. Connect a micro-B USB cable to the board on J11
8. Open TeraTerm or other terminal program, and connect to the COM port that the board enumerated as. Use 115200 baud, 1 stop bit, no parity as the terminal settings.
9. In the Quickstart panel, click on **Build**. You should see no errors in the Console tab.



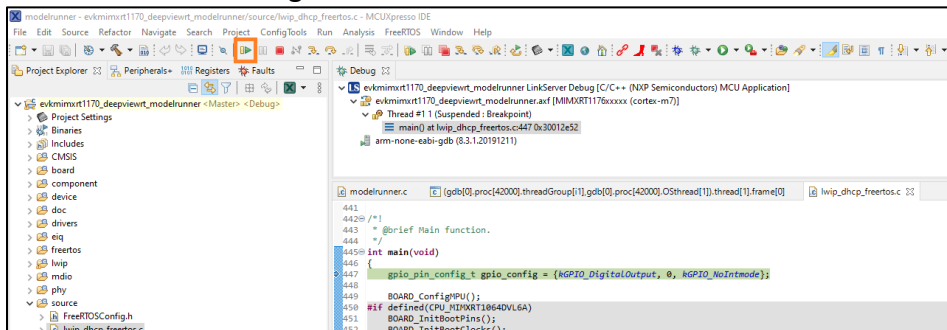
10. In the Quickstart pane select **Debug**



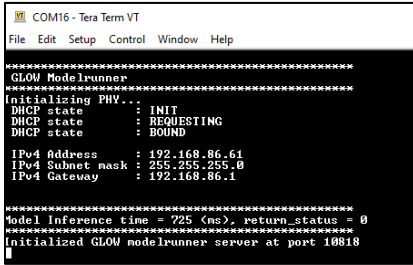
11. It will ask which interface to use. Select the CMSIS-DAP probe and click on OK.



12. The debugger will download the firmware and open up the debug view. Click on the Resume button to start running.



- In the terminal you should see the board acquire an IP address from the router. Make note of this IP address as it'll be used in the next section.



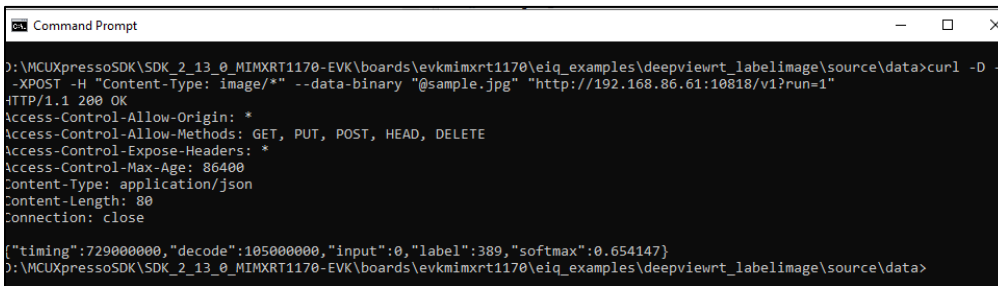
```

COM16 - Tera Term VT
File Edit Setup Control Window Help
*****
GLOW ModelRunner
*****
Initializing PHY...
DHCP state : INIT
DHCP state : REQUESTING
DHCP state : BOUND
IPv4 Address : 192.168.86.61
IPv4 Subnet mask : 255.255.255.0
IPv4 Gateway : 192.168.86.1
*****
Model Inference time = 725 (ns), return_status = 0
*****
Initialized GLOW modelrunner server at port 10818
  
```

- The project is running a Mobilenet image classification model which can categorize images into one of 1,000 different categories. You can then send images via Ethernet to the RT1170 board running that Glow model.
- Open up a Windows command prompt
- Navigate to an example panda image named **sample.jpg** in the RT1170 SDK at **<SDK unzip location>\boards\evkmimxrt1170\eiq\_examples\deepviewrt\_labelimage\source\data**
- In the Windows Command Prompt type the following which will use a program called curl to send the image to the board using its IP address from the previous step. You will get the inference time in nanoseconds and the result. This command should be one long continuous line:

```

curl -D - -XPOST -H "Content-Type: image/*" --data-binary "@sample.jpg"
"http://192.168.86.61:10818/v1?run=1"
  
```



```

Command Prompt
D:\MCUXpressoSDK\SDK_2_13_0_MIMXRT1170-EVK\boards\evkmimxrt1170\eiq_examples\deepviewrt_labelimage\source\data>curl -D -
-XPOST -H "Content-Type: image/*" --data-binary "@sample.jpg" "http://192.168.86.61:10818/v1?run=1"
HTTP/1.1 200 OK
Access-Control-Allow-Origin: *
Access-Control-Allow-Methods: GET, PUT, POST, HEAD, DELETE
Access-Control-Expose-Headers: *
Access-Control-Max-Age: 86400
Content-Type: application/json
Content-Length: 80
Connection: close

{"timing":729000000,"decode":105000000,"input":0,"label":389,"softmax":0.654147}
D:\MCUXpressoSDK\SDK_2_13_0_MIMXRT1170-EVK\boards\evkmimxrt1170\eiq_examples\deepviewrt_labelimage\source\data>
  
```

- The label number for this model is offset by 1 from this list of labels found [here](#). The model should correctly identified the image as a giant panda, which is item 388 in that list due to the offset.
- The default models that come with the SDK example can be replaced with your own custom model by replacing the model.c/.h/.o files in the example.
- If using TFLM make sure to use the Release target for high optimization compiler settings which will have a large impact on inference time for that inference engine.

## 6 Conclusion

This lab demonstrated how to use the Model Runner feature to do profiling of a model on an embedded system.