

VSE REPORT: HUAWEI ATLAS 200DK IMAGE CLASSIFICATION SYSTEM

Ang Boon Yew, Tea Lee Seng, Yang Xiaoyan

Institute of Systems Science, National University of Singapore, Singapore 119615

ABSTRACT

Deep learning models can be trained to classify images and objects accurately using large datasets but need to be deployed and integrated with other devices in a local environment. This achieved by integrating pre-trained model with edge devices to perform tasks such as image classification and object detection in a particular location such as train station or convenience store. In this project, we demonstrate the set-up and deployment of an edge vision system using the Huawei Atlas 200DK developer kit and show how it can be used for an image classification application.

1 Introduction to Huawei Atlas 200DK

1.1 Overview

The Huawei Atlas 200 Developer Kit is a developer board product that uses the Atlas 200 AI accelerator module as the core and leverages on the Huawei Mind Studio development kit for users to quickly and flexibly develop Deep Learning applications [1]. By integrating the HiSilicon Ascend 310, a High-Computing, Low-Power A.I. processor, the Atlas 200 is ideal for research and rapid deployment in applications which require analysis and inferential computing on the edge, such as smart city applications using voice, images and video data. An example of a use of the Atlas 200 will be image classification, which can easily leverage on a well-trained model such as ImageNet. For this task, the Atlas 200DK can serve as a inference server to efficiently process images from locally connected cameras for follow-up actions. In this demonstration, we deploy the sample classification code from the Ascend GitHub repository [2]. The Atlas 200DK I/O interface and dimensions are as shown in Fig. 1.

1.2 Development and Deployment Environment

The Huawei Atlas 200DK requires the Ubuntu 16.04.3 version Linux OS, which 16.04.6 is the latest security updated. For this deployment, we set up Mind Studio in a Linux Virtual Machine (VM). As Mind Studio communicates with Atlas 200DK via TCP/IP, we also set up an ethernet link between

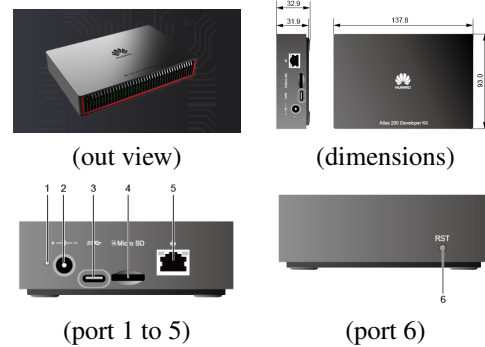


Fig. 1. Dimensions and ports of Huawei Atlas 200DK

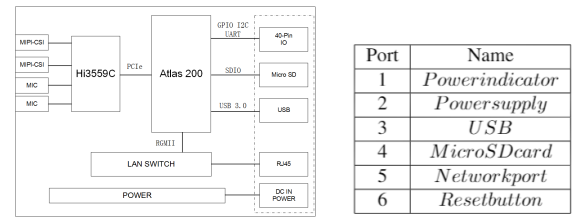


Fig. 2. Atlas 200DK System Architecture and Port description

the VM and Atlas 200DK and it used for development. Mind Studio is an AI full-stack development platform which allows the developer to construct the system engine layers by drag-and-drop nodes. It provides web services and also supports network migration, lowering the AI development threshold. Atlas 200DK uses the Da Vinci Model which is .om file for its AI models that applies for pre-trained models under the Caffe and TensorFlow frameworks and can be converted using Mind Studio.

2 System Implementation

2.1 Project and Device Setup

After setting up Mind Studio, we created an empty project with target 'Atlas DK' in Mind Studio and uploaded the classification sample code into the empty project folder. We then added the Atlas DK configuration by giving a name to the device, using the 'NIC' communication mode and setting the IP

Address to: 192.168.0.2. If the 'USB' mode is used, the IP Address is set as 192.168.1.2 instead.

2.2 Model Conversion

To convert pre-trained models to the specified Da Vinci model, the 'Convert Model' under 'Tool' menu in Mind Studio was selected. The pre-trained model type 'Caffe' was selected and the prototxt and model weight files were uploaded. In our classification system we used the ResNet-50 pre-trained model, hence two files (ResNet-50-deploy.prototxt and resnet50.caffemodel) were used for the model conversion. Mind studio can also convert 'TensorFlow' models to Da Vinci models and is saved in a default location under the 'Che/model-zoo/my-model' directory.

2.3 Project Deployment

To prepare the project environment, the environment variables 'DDK_HOME' and 'LD_LIBRARY_PATH' had to be added. The deployment was then done by accessing the project location and running the deploy script by typing 'bash deploy.sh 192.168.0.2 internet' command on local Linux host.

2.4 Running Model Application

The Da Vinci offline model and sample images to be inferred were uploaded to the directory of the HwHiAiUser user on the Atlas 200DK edge device via SFTP. The command 'ssh HwHiAiUser@192.168.0.2' was used to login to Atlas 200DK edge device and directed to the run_classification.py script path. To run the image classification task, the following command was used: 'python3 run_classification.py -m /resnet50.om -w 224 -h 224 -i /img1.jpg -n 5' using 224x224 sizes specified for ResNet50 Fig 3. shows the classification results for top 5 confidences scores for some images.

```
HwHiAiUser@devinct-mint:~/AI_PROJECTS/ascend_workspace/classification/out$ python3 run_classification.py -m /resnet50.om -w 224 -h 224 -i /img1.jpg -n 5
[INFO][general.post.cpp:164] Success to deal file=/home/HwHiAiUser/img1.jpg.
[INFO][general.post.cpp:165] Top index and confidence:293:0.99927,253:0.99487,253:0.00487,350:0.00213,340:0.00187
HwHiAiUser@devinct-mint:~/AI_PROJECTS/ascend_workspace/classification/out$ python3 run_classification.py -m /resnet50.om -w 224 -h 224 -i /img2.jpg -n 5
[INFO][general.post.cpp:164] Success to deal file=/home/HwHiAiUser/img2.jpg.
[INFO][general.post.cpp:165] Top index and confidence:293:0.99923,288:0.000904,290:0.00007,353:0.00005,340:0.00004
HwHiAiUser@devinct-mint:~/AI_PROJECTS/ascend_workspace/classification/out$ python3 run_classification.py -m /resnet50.om -w 224 -h 224 -i /img3.jpg -n 5
[INFO][general.post.cpp:164] Success to deal file=/home/HwHiAiUser/img3.jpg.
[INFO][general.post.cpp:165] Top index and confidence:331:0.940289,280:0.03406,332:0.00071,335:0.00012,322:0.00011
HwHiAiUser@devinct-mint:~/AI_PROJECTS/ascend_workspace/classification/out$ python3 run_classification.py -m /resnet50.om -w 224 -h 224 -i /img4.jpg -n 5
[INFO][general.post.cpp:164] Success to deal file=/home/HwHiAiUser/img4.jpg.
[INFO][general.post.cpp:165] Top index and confidence:281:0.493652,285:0.482422,282:0.021194,287:0.002216,292:0.000592
HwHiAiUser@devinct-mint:~/AI_PROJECTS/ascend_workspace/classification/out$ python3 run_classification.py -m /resnet50.om -w 224 -h 224 -i /img5.png -n 5
[INFO][general.post.cpp:164] Success to deal file=/home/HwHiAiUser/img5.png.
[INFO][general.post.cpp:165] Top index and confidence:259:0.921875,261:0.854535,676:0.006115,260:0.004108,283:0.002510
HwHiAiUser@devinct-mint:~/AI_PROJECTS/ascend_workspace/classification/out$
```

Fig. 3. Classification results for five images

3 System Limitations

3.1 Outdated Development Operation System

Currently, the Atlas 200DK only supports the Ubuntu 16.04.3 version, although Ubuntu has 2 long term support versions: 18.04.3 and 16.04.6 at this moment. In addition, the Huawei

support video also advised that other versions might give unexpected errors during deployment. As the 16.04.6 version has the latest security updates with minimum architectural changes, the Atlas 200DK can support at least this version or other versions for more flexible usage.

3.2 Project Upload Function

The upload project function only supports downloaded projects from the same Mind Studio platform, although a more extensive workaround is needed such as creating an empty project and copying the projects files to the directory. The upload project feature may not function as well as previously downloaded projects cannot be uploaded at times.

3.3 Model Conversion Errors

We also attempted to use our pre-trained TensorFlow model as a .pb file with the Atlas 200. However we were unable to do so as we encountered the unknown model conversion error as shown in Fig.4.

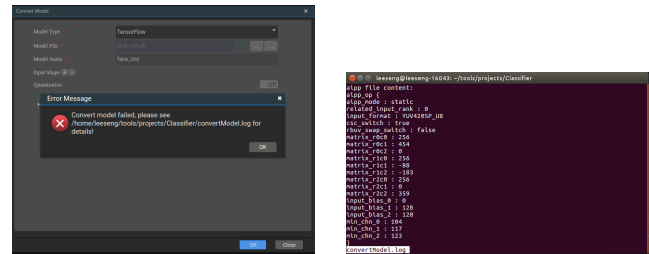


Fig. 4. Model Conversion Error

4 Conclusion

We are aware ResNet50 inference is about 2images/second on a Personal Computer. Huawei has achieved engineering prowess by releasing edge/IOT device that performs same task with much lower power and better performance without 3rd party A.I. technologies. We are grateful to experience the device and hopefully having chance to experience Ascend 910 platform in future.

5 References

- [1] Huawei, "Huawei atlas200dk documentation," <https://ascend.huawei.com/doc/Atlas200DK/1.3.0.0/en/en-us/topic/0178961794.html>.
- [2] Huawei, "Atlas 200dk classification sample code," <https://github.com/Ascend/sample/classification>.