

A Dual Camera System with AI-enabled Imaging Control for Earth Observation Applications: A Feasibility Study

Simon Vellas
OHb Hellas
Athens, Greece
simon.vellas@ohb-hellas.gr

Grigorios Tsagakatakis
Institute of Computer Science, FORTH
Heraklion, Greece
greg@ics.forth.gr

Nikolaos Nikolopoulos - Anastasiadis
OHb Hellas
Athens, Greece
nikolaos.nikolopoulos@ohb-hellas.gr

Mathieu Bernou
OHb Hellas
Athens, Greece
Mathieu.Bernou@ohb-hellas.gr

Panagiotis Tsakalides
Institute of Computer Science, FORTH
Heraklion, Greece
tsakalid@ics.forth.gr

Abstract—Given a mission profile, using very similar components but different configurations can allow an Earth Observation satellite to acquire large swath observations, offering very high revisit frequencies, at the expense of lower spatial resolution or vice versa. In the feasibility study summarized herein, the objective is to have a wide-swath optical camera for the maximization of the coverage, combined with a high-resolution, narrow-swath camera to achieve the highest required spatial resolution and get the best possible image of an Area-of-Interest (AoI), with the help of on-board AI decision making. The performed analysis concludes with a case study demonstration on active fire detection.

Keywords—artificial intelligence, earth observation, cognitive, constellation, on-board, fine-pointing, active fire detection

I. INTRODUCTION

Optical Earth Observation satellites strive to find a balance between spatial and temporal resolution [1], i.e., ground sampling distance and revisit period. In the mission concept proposed for the “Dual-camera satellite with on-board AI-based decision-making capabilities” study, in the context of the ESA Cognitive Cloud Computing in Space campaign, a first camera is placed on a satellite platform as forward-looking and is characterized by a large swath but relatively low spatial resolution (LR camera). The LR camera is responsible for capturing images with a large field of view. The second camera, placed on a second satellite platform, is designed to acquire images of specific Areas-of-Interest (AoI) within the field of view of the first camera but at a significantly higher spatial resolution (HR camera). The challenge, in this case, is how to select which region to image. The extremely tight timing constraints prohibit more traditional approaches which rely on high latency space-to-ground communications. As a result, this challenge can only be met by introducing AI systems on-board the satellite, following an edge-computing paradigm. In this concept, the decision on which AoI the second camera will focus, is performed by an intelligent AI-enabled control module that analyses the observations from the first camera (LR), in order to provide the necessary signals to the control of the second camera (HR). The AI-enabled AoI selection takes the coarse observations from the first camera (LR) and executes a process that assigns a specific value to the different regions, based on whether there is an event of interest taking place in the region. The AI-enabled imaging control system is realized through on-board Deep Reinforcement Learning. Thus, in the

proposed satellite configuration for the space segment of the study, a leading satellite is equipped with the low-resolution imager, identifying regions of interest, based on the AI decision-making algorithm and broadcasts the information to a trailing satellite carrying the high-resolution imaging device.

The approach is further studied with a use case, selected to highlight the merits of the dual-camera system and to identify any potential limiting points. In coordination with ESA, the selected application was agreed to be active fire detection, while demonstration was performed on two commercially available embedded devices which are of interest to the space community for future cost-effective high-performance on-board processing solutions. Finally, the study is concluded with a future outlook and elaboration on the potential benefits of such an architecture for European small satellite missions.

II. MISSION ANALYSIS

A two-satellite system can be utilized to implement the dual camera concept. One satellite, the “scout”, is equipped with the low – resolution camera and an AI accelerator to detect the AoI, while the second one, the “fine-pointing”, is equipped with the high-resolution camera. Thus, the scout will capture the wide field of view image, locate the AoI and transmit the required data to the fine-pointing satellite. The latter will receive the data, perform the necessary attitude adjustment (attitude change and settle) and capture high-resolution images. The satellites will be in the same orbit with a predefined distance between them. In this scenario, the total time available between the LR and the HR image capturing is defined by the said distance and must include the time required for the AoI identification to take place, as well as the satellite attitude change time and settle time. Regarding the satellite platform, it is possible to realize the dual camera concept with Cubesats (12 – 16U), but a microsatellite platform in the range of 100 kg will ensure redundancy and a better lifetime for the mission, as it will be easier to use multiple components, as well as a propulsion system for station keeping, collision avoidance and deorbiting, thus constituting a more robust solution in general. In this case, a Control Moment Gyro (CMG) actuator would be preferable to provide the necessary torque for attitude adjustment compared to reaction wheels, due to the torque-to-watt ratio the first solution can offer compared to the latter for agile spacecraft [2].

The relative distance between the two satellites is an important factor for the realization of the concept because it is the parameter that will set the limit for the AI processing time and the available time that the fine-pointing satellite (HR) will have to correct its attitude and settle, in order to capture the high-resolution image. This value will be constrained by the range of the Inter-Satellite Link (ISL) solution that will be used in the final implementation.

From (1), a satellite on an assumed circular orbit of 500 km has an orbital speed of 7.61 km/s, where v is the linear velocity, the product GM is the gravitational parameter and r is the radius of the orbit.

$$v = \sqrt{GM/r} \quad (1)$$

From (2), the angular velocity of the satellite is 0.0011 rad/s.

$$\omega = v/r \quad (2)$$

The angle θ that the satellite covers in 22 seconds, which is the total processing and attitude change and settle time assumed for the concept, is 0.0242 rad (3).

$$\theta = \omega \cdot t \quad (3)$$

The cord length, or in other words the linear distance between the satellites, for the given angle, will be 166.44 km (4).

$$l = 2 \cdot r \cdot \sin(\theta/2) \quad (4)$$

These calculations can be made in reverse, starting from the linear distance of the two satellites based on an existing ISL solution to define the maximum time available for the AI processing and for the fine-pointing satellite to perform the attitude adjustment.

Several ISL solutions were assessed during the study. Using a realistic one, such as the Antwerp Space LEO – LEO Ka-band ISL for small satellites [3], providing an assumed satellite distance capability of 4700 km which is within its specification, the maximum time available for AI processing and attitude adjustment (attitude change and settle) will be 630 seconds. This is considered viable for the sum of typical AI processing in this context and microsatellite attitude change and settling times. Also, the angle between the two satellites will be 0.7 radians or 40°. In terms of Keplerian orbital elements, this translates to a True Anomaly difference of 40°. True Anomaly is the angle between the argument of perigee direction and the current position of the body, that is used to define its position along its orbit. Conclusively, with this approach, the time constraint is lifted but depends on the capabilities of the ISL solution.

An important parameter that requires further investigation in the two-satellite scenario is the “downtime” that the system experiences during the AI processing for the identification of the AoI. If a mission scenario required continuous monitoring with the first satellite, for the assumed characteristics of 500km orbit altitude, 120 km swath for the LR camera and the 22 seconds of available time for the AI processing and the attitude adjustment of the fine-pointing satellite, near-real-

time processing would not be possible for sequential imaging, as new data would be generated every 16 seconds.

For every acquisition/processing/acquisition cycle, the system would not be able to proceed until the processing would be over. Only after this time would it be possible to initiate the new cycle. Depending on the implementation of the system and the target image resolution, it might be the case that continuous monitoring of the ground is not possible. A constellation of multiple satellites would be able to lift this constraint. Another solution would be to broadcast the information to a third satellite (e.g. relay to LEO) that would then acquire the high-resolution image of the AoI. To summarize, the two-satellite scenario is feasible and could be implemented in a space mission for an AI on-board application execution, however, the AI processing time is a decisive factor for different mission scenarios.

III. ACTIVE FIRE DETECTION DEMONSTRATION

The dual camera concept as previously described provides great benefits in different aspects of EO and related application scenarios. By implementing AI on board, which can dynamically assess the current conditions and select which action to take, the cost of data is reduced since the platform will only downstream data that the algorithm has denoted as of interest. Subsequently, the platform’s operation is flexible in nature as it can calculate and execute orbital adjustments by itself, given the current environmental conditions (e.g. cloud detection like in [4]). As an added benefit from its autonomous capabilities to self-orient, the operational cost can be further decreased by reducing the need for pre-planning activities compared to traditional operational approaches. The platform can also achieve reduced latency in capturing data, made possible due to its autonomous operations and on-board data processing implementing AI, thus paving the way for new applications focusing on emergency response, disaster monitoring, and the time-critical domain in general, especially for capturing high-resolution images of unplanned and sudden events. Finally, the concept has the potential to address applications like Land Cover and Land Use Classification [5].

A. Case Study on Active Fire Detection

AI algorithms can be executed on-board satellites in order to extract information that is much smaller in size and therefore can be transferred to the ground by low-capacity telemetry channels. This approach is considered in scenarios for rapid detection of events like fires [6]. To demonstrate the merits of the cognitive cloud dual camera concept, the case of Active Fire Detection (AFD) was explored. Given an image of a scene acquired by a satellite, the objective in AFD is to classify each pixel (i.e. spatial location) as part of an active fire front or not. An example is shown in Fig. 1.

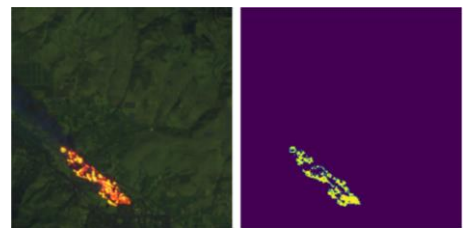


Fig.1. Example of satellite image (left) and active fire mask (right) used in the study.

B. Dataset Generation

For the implementation, a publicly available dataset that includes 10 bands from the Landsat 8 MSI instrument at specific resolutions, as well as associated manual annotation is considered [7]. Each image in the dataset is 256x256 pixels. In the conducted study, these images were subsampled to 32x32 pixels, in order to approximate the behavior of the first, LR camera. Each LR image is then divided into 4x4 patches which correspond to potential AoI that can be acquired by the second, HR camera. The associated images from the second camera are 64x64 pixels, corresponding to $\frac{1}{4} \times \frac{1}{4}$ of 256x256 which is the size of the original image. Only the manually annotated dataset which consists of 118 pairs of remote sensing image and manual annotation pairs is considered.

The theoretical model studied here is composed of the Deep Policy Network (DPN), which is responsible for selecting the AoI from the first LR camera and the Active Fire Detection network which performs pixel-level active fire front segmentation on the HR AoI acquired by the second camera.

C. The Deep Policy Network

The DPN is responsible to selecting the AoI from the LR image acquired by the first camera that needs to be sampled at higher resolution by the second, HR camera. A graphical overview of the operations executed by the DPN is shown in Fig. 2.

The network accepts at inputs LR images of size 64x64 pixels at 3 bands and produces a vector of 16+1 elements. The predicted vector corresponds to a binary vector where all elements but one are zero. The location of the non-zero element indicates the patch where the AoI has been selected. The extra element corresponds to the non-sampling case.

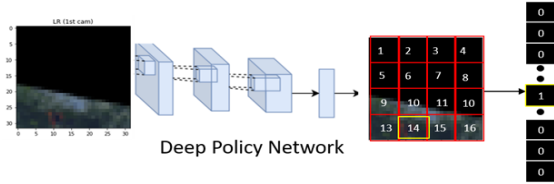


Fig. 2. AoI selection by the DPN.

D. The Active Fire Detection Network

Once the DPN selects the AoI, this information is communicated to the attitude adjustment module which orients the second HR camera towards the selected AoI. The selected HR patch is then introduced to the Active Fire Detection Network which assigns a value of 0 or 1 to each pixel, based on the absence or presence of an active fire located on the area associated with this pixel.

E. Operational Behavior

To demonstrate the potential of the proposed scheme, Fig. 3 and Fig. 4 present indicative examples of the system operation. Each figure is composed of six panels.

The top left panel shows the image acquired by the first (LR) camera. The top middle panel shows the ideal HR fire map. This image is not available to the platform since it assumes an AoI as large as the first camera's FoV (Field of View). This map is thus shown for reference. The top right panel shows the selected AoI. The image is 4x4 corresponding to the 16 potential AoI by the second camera. The bottom left

panel shows the HR AoI acquired by the second camera, given the output of the DPN network.

The bottom middle panel presents the active fire front segmentation by the AFD network. This network assumes the HR AoI as input and can be optionally executed on the platform. The bottom left panel presents the manually annotated active fire front for the HR AoI patch.

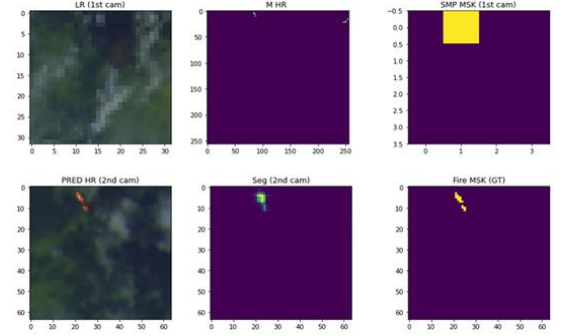


Fig. 3. A case where the fire is almost indistinguishable in the LR image, yet the DPN correctly identifies the AoI. The AFD network performs reasonably well.

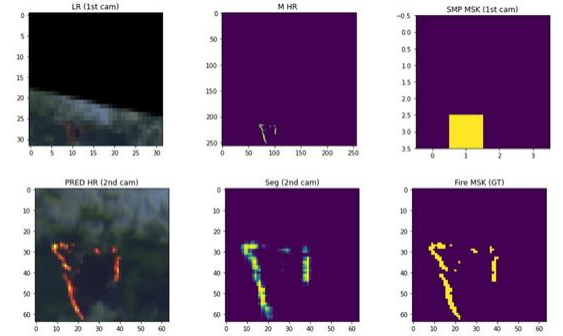


Fig. 4. A case where the fire is visible in the LR image and the DPN selects the correct AoI. The AFD performs very close to the target behavior.

IV. USE CASE DEMONSTRATION SETUP

In this section, the selected demonstration hardware and the implementation approach followed for the demonstration of the two AI networks (HR and LR network) on the two devices is presented. For this demo, it was decided that the two selected devices will represent the two satellites, where each of the two computers will be the main payload processing unit.

The device selected to represent the computing platform of the scout satellite is the Xilinx Kria KV260 SoM (System-On-Module) [8], featuring the Xilinx's DPU AI specialized soft-core IP. Xilinx Kria is a series of SoM devices designed for high performance and low time-to-market edge applications. The K26 SoMs are integrated into application-focused starter kits to be used for initial evaluation and early development. The Xilinx Kria KV260 Vision AI Starter Kit that was used for this study includes a non-production version of the K26 SoM, a carrier card, and a thermal solution. It combines components such as a Zynq UltraScale+ MPSoC based silicon device, 4GB of RAM and the OnSemi AP1302 ISP image sensor processor.

The role of the first, scout satellite will be to select an AoI, i.e. a fire tile, in a LR image. This essentially entails a classification among the tiles of the original LR image and is

performed using a ResNet model, the LR network. Following the development flow for embedded applications on the Xilinx's Kria SoM and DPU, the Vitis-AI toolchain is used to prepare the model for execution on the target platform. This preparation includes the quantization and compilation steps. The original float model is quantized into INT8 precision to be executed on the DPU using a representative, calibrating dataset. Then, the quantized version of the model is compiled into an XMODEL file that uses the DPU's highly efficient instruction set and dataflow model. The final step is the development of the host application. Python and the Vitis AI unified Python high-level API were used to execute the quantized XMODEL. This application reads the input LR image, runs the inference task and returns the predicted index that corresponds to the HR tile of the original LR input image.

The board selected to represent the fine-pointing satellite is the Google Coral Dev Board [9], featuring the Edge TPU AI accelerator. The Coral Dev Board is a single-board computer with a removable SoM, containing an Edge TPU co-processor designed by Google which is a small Application Specific Integrated Circuit (ASIC). It provides a complete system with a System on Chip (SoC) which runs a derivative of Debian Linux called Mendel. The Coral Dev Board is a fully integrated system which is equipped with NXP's i.MX 8M SoC, 8GB eMMC memory, 1GB of LPDDR4 RAM and a GC7000 Lite GPU. The Coral Dev Board is capable of performing high-speed machine learning (ML) inferencing, executing state-of-the-art TensorFlow Lite models in a power efficient manner. The TensorFlow Lite is a lightweight version of the ML library TensorFlow, which is designed for mobile and embedded devices. The Coral Dev Board was selected for the specific use case, since it is allowing to quickly prototype and deploy low-power embedded systems with on-device ML inferencing in a small form factor.

The role of the second, fine-pointing satellite is to find the exact mask of the active fire, that is, to identify the pixels which depict fire on the HR image of the fire tile location, as it is received from the scout satellite. To conduct the fire segmentation, a Unet model is used. The development flow for embedded applications on the Coral Dev Board includes the utilization of the Mendel Development Tool (MDT) which is a command-line tool allowing the communication with the Coral device running Mendel Linux. The Edge TPU supports only TensorFlow Lite models that are fully 8-bit quantized and then compiled specifically. The model cannot be trained directly with TensorFlow Lite. Therefore, the TensorFlow Lite Converter was used to convert the TensorFlow model to a TensorFlow Lite file and the quantization and compiling steps were applied. The quantization of a model implies the estimation of the range of all the floating-point numbers in it, including the input, intermediate layers and the model output. The estimation of the range is usually calculated by utilizing a small batch of a few hundred images from the training or the evaluation dataset, which constitutes the representative dataset. The demonstration is completed by executing the host application and presenting the segmentation mask which depicts the fire location on the HR image.

These two devices were selected based on a previously performed market analysis of available COTS devices for use in space missions as well as an analysis of the interest of the space community to use these devices in space (both from the industry and the Agency) and are depicted in Fig. 5.



Fig. 5. Use case demo setup. On the left the Kria KV260 board and on the right the Coral TPU Board.

In the selected use case, in the total time available the producing of the high-resolution fire detection mask is included. Thus, the following equation can express the total time required for finishing the whole operations pipeline for a single identification:

$$t_{tot} = t_1 + t_2 + t_3 \quad (5)$$

Where t_1 is the time required for AoI identification, t_2 is the time required for high resolution mask production and t_3 is the time required for attitude change and settle.

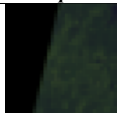
Initially, the first script executing on Kria, reads a LR image stored in the device and performs a classification task to label the image's tiles as <fire> or <not fire>. After that, the location of the tile that contains fire is transmitted to Coral. In this setup, the transmission of information is done through SCP, a Linux secure copy command. Then, the script executing on the Coral device receives this info and runs the segmentation algorithm. In this case the information exchanged is the selected image tile itself. However, this has only been done for demo purposes as it has no impact on the overall system performance. Finally, it is noted that the attitude adjustment is not simulated in this demo.

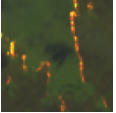
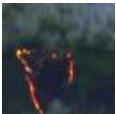
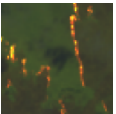
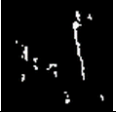
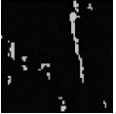
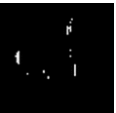
V. EVALUATION

In this section, the results presented in Table I from running the dual camera demo on Landsat-8 active fire satellite imagery will be evaluated. The initially set objective and requirements have been met and will be discussed in the following paragraphs.

Table I contains the LR input image obtained from Landsat 8 imagery and artificially degraded in quality to simulate a picture captured by the low-resolution camera from the scout satellite. Three sample images are chosen to assess the models' performance and the 2 versions of the AI models, the float and the quantized one, are also considered. The first model (LR), a ResNet, decides the position of the AoI. Additionally, the ground truth mask and the mask as it is predicted by the Unet models (HR) are being presented.

TABLE I. MODEL'S ACCURACY

	Sample 1	Sample 2	Sample 3
LR-Input			
Ground Truth Fire Patch		No Fire	

	<i>Sample 1</i>	<i>Sample 2</i>	<i>Sample 3</i>
Fire Patch Predicted – Float Model		No Fire	
Fire Patch Predicted – Quantized Model		No Fire	
Ground Truth Fire Mask		N/A	
Fire Mask Predicted – Float Model		N/A	
Fire Mask Predicted – Quantized Model		N/A	

The results indicate that the quantized LR model is performing well, as it has a 91% match on the AoI patches prediction in relevance to the float model. Regarding the HR model, the float model is capable of capturing the main features of an active fire, resulting in the mask extraction which depicts the fire fronts. On the other hand, the absence of previously predicted fire pixels in the quantized model predictions, indicates that some information loss occurred during the quantization process, as expected.

In Table II the latency and throughput measurements for the inference and the pre-processing stages are presented:

TABLE II. LATENCY AND THROUGHPUT MEASUREMENTS

	Latency (Seconds)	Throughput (FPS)
<i>Inference</i>		
Kria KV260	0.0031	319.91
Coral Dev Board	0.0675	14.81
<i>Pre-Processing</i>		
Kria KV260	0.0645	15.50
Coral Dev Board	0.0528	18.94

The near-real-time objectives of the mission are also met through this demo. The overall execution of the demo is about 16 seconds. This time includes the inference task as well as the pre-processing stages in both devices. Moreover, the time for the SCP commands that perform the information transmission between the two devices and the communication overhead to connect from the local environment to these devices are also included. The AI processing (inference) takes a few milliseconds as indicated in Table II.

It is evident from this demo that an implementation of AI algorithms on specialized HW-accelerator platforms is highly efficient, especially when comparing the achieved performance to power consumption ratio. The relevant results are captured in Table III.

TABLE III. POWER CONSUMPTION

Power (Watt)	
Kria KV260	6.6 (5.3 Idle)
Coral Dev Board	5

This demonstration clearly highlights the feasibility of implementing complex AI algorithms in a power-efficient manner and achieving very satisfactory performance results. Of course, the size and type of the algorithm that can be efficiently implemented on these devices, as well as the limitations imposed by them must be taken into account when designing such a system. However, it is evident that for a light to medium-complexity algorithm it is possible to achieve very good performance, addressing the needs of modern time-critical applications.

VI. DISCUSSION

This paper presents a study that explores the feasibility of a two-satellite system for Earth Observation, with on-board decision making enabled by AI, that is capable of providing information about events of interest with low latency. An end-to-end system for active fire detection is emulated, with the results combined with the orbital analysis performed showing in this preliminary stage that the concept is feasible.

The focus of the study lies on an EO use case, but the concept is also applicable to other missions like observations of moons and asteroids. The key idea can also be appreciated in the context of autonomous robotic platforms for space exploration. Furthermore, the proposed approach can be extended to scenarios involving satellite constellations, which offer better revisit times as well as instrument versatility and low latency due to the use of constellation-wide Inter-Satellite Links.

The proposed AI-enabled technology is an initial stepping stone in developing more complex and advanced missions, with greater scope and capabilities. The investigated idea could enable the realization of novel EO applications on reconfigurable satellites, targeting national agencies and commercial customers through a flexible Satellite-as-a-Service business model, that can serve end user needs with efficiency based on intelligent platform tasking and scheduling.

ACKNOWLEDGEMENT

This work was funded by the ESA OSIP program and the Cognitive Cloud Computing in Space campaign.

REFERENCES

- [1] D. Kasampalis et al., "Contribution of remote sensing on crop models: a review," *Journal of Imaging*, vol. 4, no. 4. MDPI AG, p. 52, March 2018.
- [2] R. Votel and D. Sinclair, "Comparison of control moment gyros and reaction wheels for small earth-observing satellites," 2012.
- [3] "Ka band LEO-LEO ISL for small satellites," Gaussteam, <https://www.gaussteam.com/wordpress/wp-content/uploads/2020/03/IAA-AAS-CU-20-07-15-Ka-Band-ISL-and-Transceiver-for-Small-Satellites-min.pdf> (accessed Aug.9, 2023).

- [4] G. Giuffrida et al., "CloudScout: A deep neural network for on-board cloud detection on hyperspectral images," *Remote Sensing*, vol. 12, no. 14, p.2205, July 2020.
- [5] A. S. Kucik, and G. Meoni, "Investigating spiking neural networks for energy-efficient on-board AI applications. A case study in land cover and land use classification." 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, June 2021.
- [6] B. Hally, L. Wallace, K. Reinke, S. Jones, and A. Skidmore, "Advances in active fire detection using a multi-temporal method for next-generation geostationary satellite data," *International Journal of Digital Earth*, vol. 12, no. 9. Informa UK Limited, pp. 1030–1045, July 2018.
- [7] G. H. de Almeida Pereira, A. M. Fusioka, B. T. Nassu, and R. Minetto, "Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 178. Elsevier BV, pp. 171–186, August 2021.
- [8] "Kria KV260 Vision AI starter kit," Xilinx, <https://www.xilinx.com/products/som/kria/kv260-vision-starter-kit.html> (accessed Aug. 9, 2023).
- [9] "Dev Board," Coral, <https://coral.ai/products/dev-board/#tech-specs> (accessed Aug. 9, 2023).