Artificial Intelligence for High-performance Human Space Flight Avionics Systems

Carles Hernàndez, *Universitat Politècnica de València (Spain), carherlu@upv.es, +34963875738*
Nick Hauptvogel, *Airbus Defence and Space (Germany), nick.hauptvogel@airbus.com, +494215393221*
Laura Medina, *Universitat Politècnica de València (Spain), laumecha@upv.es, +34963875745*
Charles-Alexis Lefebvre, *Ikerlan Technology Research Centre (Spain), calefebvre@ikerlan.es, +34943712400*
Jose Flich, *Universitat Politècnica de València (Spain), jflich@disca.upv.es, +34963875753*
Martin Rönnbäck, *Cobham Gaisler (Sweeden), martin.ronnback@gaisler.com, +46317758670*
Dierk Lüdemann, *Airbus Defence and Space (Germany), dierk.luedemann@airbus.com, +494215395103*

Speakers Laura Medina, *Universitat Politècnica de València (Spain), laumecha@upv.es, +34963875745*
Nick Hauptvogel, *Airbus Defence and Space (Germany), nick.hauptvogel@airbus.com, +494215393221*

Topics 3.1 Data management aspects for remote sensing missions - on-board processing, AI, ML, cloud based processing, data storage...
5.1 High performance and High reliability
5.3 Next generation processing: new technologies, new products
5.4 System specific OBC solutions
7.2 Avionics solutions and architectures for reference autonomy applications (launcher, rovers/robotic systems, exploration missions, complex applications...)

I. Introduction

The space environment is known for its extremely hazardous conditions for astronauts and resulting from this, requires the highest demands for safety and security. Current designs for human space flight avionics systems are traditionally very conservative. This is generally not an issue as astronauts are present all the time to monitor and control the spacecraft. However, new types of spacecraft (e.g. Deep Space Gateway, Missions to Mars) require much higher levels of autonomy, which often cannot be reached with traditional designs. Furthermore, much computing power and power consumption is needed to run the computationally expensive tasks required to reach such high levels of autonomy, which is a major challenge for on-board systems in contrast to on-ground systems.

This challenge also applies to upcoming space missions that will additionally rely on the support of robotic control and artificial intelligence, for instance to explore the moon with the European Large Logistics Lander EL3 [1]. Existing projects like the International Space Station ISS are subject to constant evolution as well, as seen in the Bartolomeo [2] extension of the European Columbus Module, which provides additional opportunities to explore automated in-space assembly via robotic factories performed in the PERIOD [3] project.

For all of these next generation space missions, multi-core platforms and software systems are the key enabler. A safe and secure architecture relies on strong separation of concerns in this environment with mixed criticality tasks consisting of vital bare metal real time spacecraft controls, telemetry and ground communication tasks, up to computationally intensive image processing by artificial intelligence algorithms and neural networks.

In this paper we present an Airbus-designed system architecture that builds on the multi-core platform which is being developed in the project SELENE [4]. The platform is a RISC-V multicore system with FPGA-based hardware acceleration of artificial intelligence inference processes and built-in support for mixed-criticality tasks exploiting both the hardware partitioning capabilities of the platform for timing interference and the spatial partitioning and isolation capabilities of the Jailhouse hypervisor. Based on this, the Airbus system architecture places a particular focus on mixed criticality and artificial intelligence tasks for robotic use cases, which are seen to be representative for a broad range of potential applications in space.

II. SELENE Computing Platform

The SELENE System-on-Chip (SoC) is depicted in Figure 1. This SoC comprises a NOEL-V multicore RISC-V system. Each NOEL-V core has private L1 instruction and data caches and is connected to an Advanced High-performance Bus (AHB) to a shared L2 cache, forming a General Purpose Processing (GPP) element. The GPP and the AI acceleration subsystems are interconnected by an AXI high-speed interconnect. The SELENE SoC is highly configurable and for the demonstration platform, a versatile six-core system has been chosen.

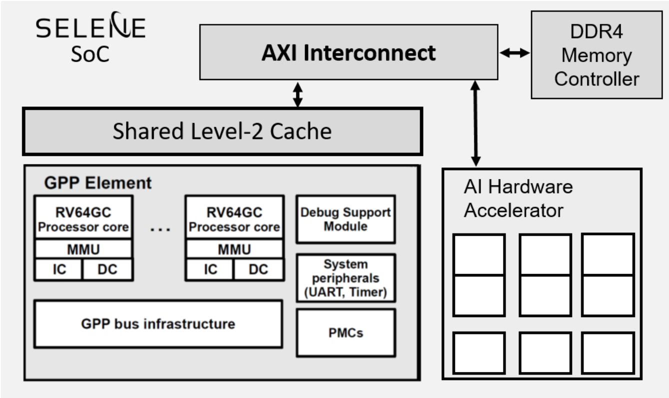


Fig. 1: Baseline SELENE SoC architecture.

With a single GPP element, coherency between all cores is handled by the L1 caches in the GPP element. Furthermore, the IO subsystem is connected through an IOMMU directly to the internal bus of the GPP element, thereby providing coherency between the IO peripherals and the processor cores. This structure can be changed at a later stage in the project.

The SELENE platform relies on Linux as the default operating system. In particular, it is using a Debian-based RISC-V Linux adaptation to NOEL-V that can be found at https://github.com/siemens/isar-riscv/. The integration of the artificial intelligence software toolchain is built on top of this Linux distribution.

III. AI FRAMEWORK

The SELENE AI framework uses the European Distributed Deep Learning Library (EDDL) library [5] to train neural network models and to perform the inference process of already trained neural network models. Below, we describe the main elements of this framework in the context of the SELENE platform.

A. *AI Models deployment*

The EDDL is a general-purpose open-source deep-learning library [5] and it is the library we use to deploy AI models on the SELENE platform. The EDDL can also be used to train the neural network models. However, we support the deployment of pre-trained models that were exported using the ONNX format to ensure the compatibility with existing workflows and other frameworks.

In the SELENE platform, the EDDL runs on top of the Linux OS deployed in the NOEL-V processor, as presented in Figure 2. Thus, the inference process can be executed entirely in the NOEL-V multicore system. To speed up the inference process the SELENE AI framework allows offloading heavy computations to the SELENE AI hardware accelerators. A low-level runtime achieves this task and guarantees that the SELENE AI hardware accelerators can communicate with the EDDL. In particular, the runtime ensures that low-level constraints are met, such as memory allocation, translation, and interrupt handling. The SELENE Linux image integrates the EDDL and the low-level runtime. The final application running on the NOEL-V processor infers the AI algorithms and allows the user a transparent use of the accelerators.

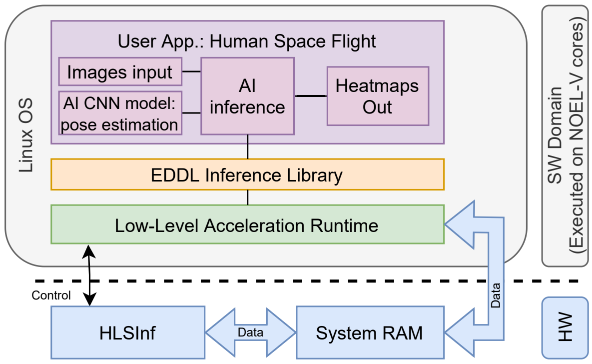


Fig. 2: The SELENE AI framework

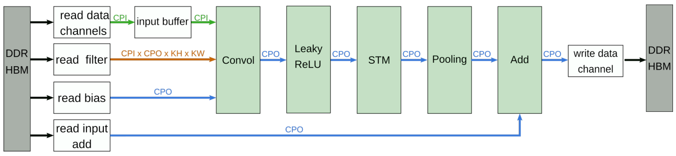


Fig. 3: HLSinf High-level Architecture

B. *HLSinf Accelerator*

The default AI hardware accelerator of SELENE is HLSinf. This accelerator is an open-source project designed in High-Level Synthesis for inference processes of neural network models based on convolutions [6]. Thus, this accelerator is especially well suited for SoCs including embedded FPGAs or systems including FPGA devices. However, the use of tools like Catapult HLS [7] also enables using this accelerator for ASIC targets. HLSinf implements the channel slicing concept since the accelerator takes a set of channels as input in parallel and produces a set of channels as output in parallel. The channels per input (CPI) define the input speed up, and the channels per output (CPO) define the output speed up. The accelerator allows configuring the value parameters to adapt the accelerator size and parallelism to the desired use case.

Figure 3 shows the current design of the HLSinf accelerator. As we can see, HLSinf currently supports several functions such as convolution, ReLU, leakyReLU, STM (an aggregation of softmax, hyperbolic tangent, and element-wise multiplication), element-wise addition as well as pooling operations. HLSinf can be customized in several dimensions. First, we can define the operations to support at design time. More specifically, the HLSinf allows defining the convolution algorithm among three alternatives: direct convolution, Winograd algorithm, and depthwise separable convolution. In the second dimension, the accelerator data type and precision format can be customized. Currently, the accelerator supports single-precision floating-point format (FP32), fixed-point formats and integer formats. Finally, in the third dimension, the input

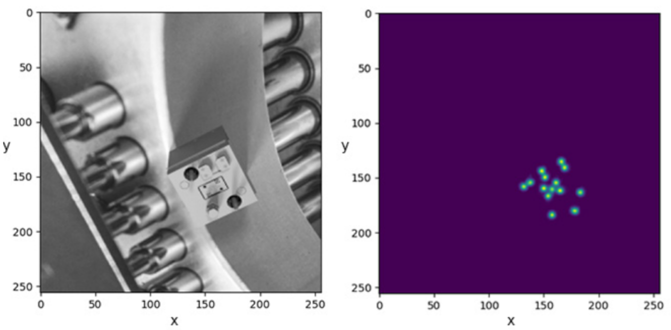


Fig. 4: Input image and key point prediction heatmap from Convolutional Neural Network.

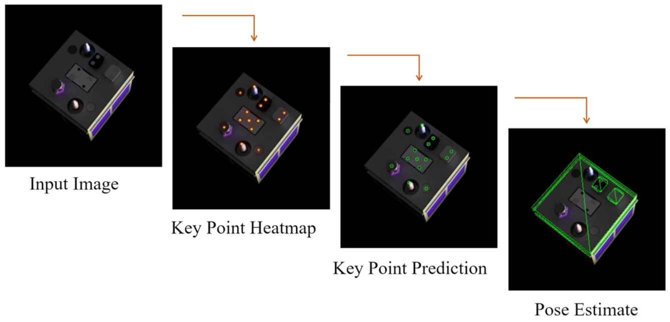


Fig. 5: Multi-stage process to estimate the pose of a given input image.

speedup and the output speedup can be defined to adapt the accelerator size and parallelism.

IV. Human Space Flight

Airbus is part of Europes long-lasting history of empowering humankind to explore space. The architecture described in this paper represents a generic Robotic Servicing Module (RSM), which forms the basis for multiple space-related domains, including intravehicular robotics, on-orbit servicing [8], in-space assembly [9] [10] and space-debris removal [11]. The platform contains four mixed-criticality tasks on its separate cores - management, robotic arm control, LIDAR based navigation and robotic visual pose estimation with the dedicated FPGA AI acceleration unit. SpaceWire and CAN Bus interfaces are used to access simulated camera and LIDAR sensor data, a TCP/IP Ethernet connection for each core serves the exchange of telemetry and telecommand data to the Test Bench, which is used for evaluation of the platform.

A. *AI-based Visual Pose Estimation*

To quantify if the SELENE platform is capable of increasing the computational power compared to current space-grade equipment, particular emphasis is placed on the AI-based visual pose estimation as benchmarking application. In this task, the platform will receive camera images containing a representative payload in the form of a CubeSat via the SpaceWire interface from the Test Bench. They are fed into the pose estimation algorithm, which applies a hybrid AI approach, comprising a two-stage pipeline to estimate the satellites pose. First, it derives key point locations on the satellites surface from the images via the inference of a Convolutional Neural Network, which predicts the key point location via probability distributions, so-called heatmaps (Figure 4).

Afterwards, a non-AI Perspective-n-Point algorithm maps the detected two-dimensional key point locations onto a priori known positions of the key points from the 3D satellite model, resulting in a 6D pose estimate, translation and rotation. The approach is visualized and described in Figure 5 and Figure 7.

B. *Convolutional Neural Network*

The Convolutional Neural Network is responsible for the extraction of key point locations via heatmaps. It makes

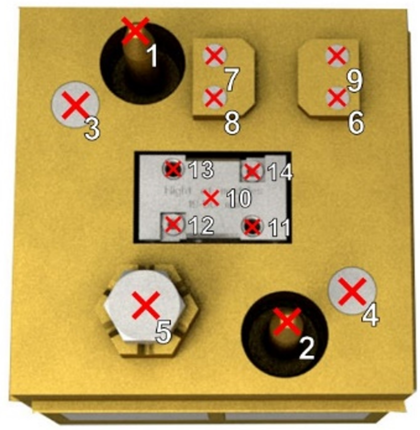


Fig. 6: Key point locations on virtual payload satellite structure

predictions on the AI accelerator of the multi-core SELENE platform. The neural networks architectures for the use case originate in human pose estimation but were demonstrated to be applicable to objects as well [12] [13] [14]. Also, the hybrid approach via heatmaps and a subsequent classical algorithm has been widely explored [15] [16] and was often found more accurate than directly regressing a pose from images [17] [18].

In the project Manipulation and Tool Operations for On-Orbit Servicing (MANTOS), Airbus identified two potential architectures that are suited for the use case of satellite pose estimation - the HRNet [13] and Stacked Hourglass network [14]. Their common characteristic is a encoder-decoder structure in the network which extracts semantically rich features before upsampling to a higher resolution for the final heatmaps. To retain spatial accuracy in the output, the Stacked Hourglass uses direct transfers of the input to the output data (skip connections), while the HRNet keeps high resolution branches throughout the network. Both architectures were investigated with an artificial dataset that was created using a virtual payload model in the form of a CubeSat shown in Figure 6 and rendering techniques described in [19]. The SELENE benchmarking task makes use of the dataset and model architectures to apply the use case to the multi-core

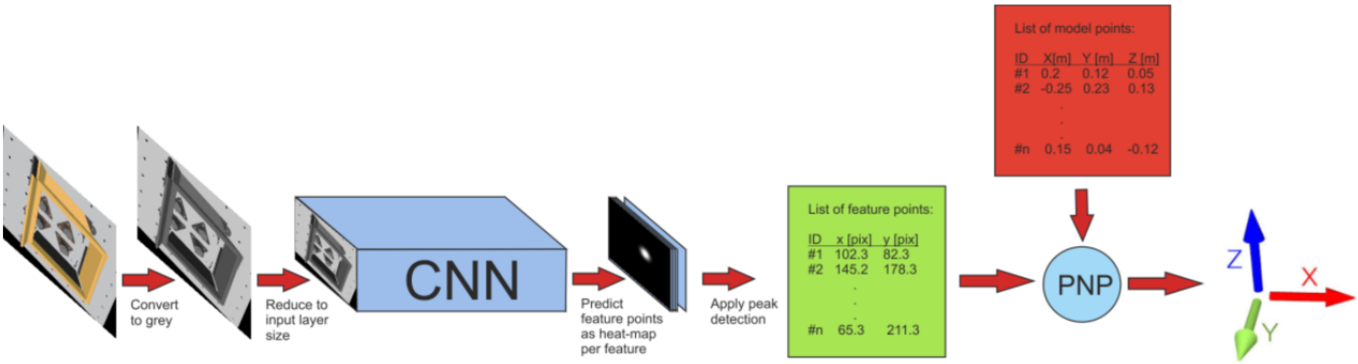


Fig. 7: Hybrid AI approach with Convolutional Neural Network and Perspective-n-Point Algorithm.

RISC-V platform.

For inference, the European Distributed Deep Learning Library EDDL was chosen as the main hardware acceleration library for inference on the RISC-V SELENE platform by the consortium [20]. A big advantage of the framework is the support for the Open Neural Network Exchange (ONNX) format, which enables framework interoperability, for instance, between training and inference of models [21].

C. Evaluation

The benchmarking of the SELENE platform with AI-based pose estimation requires an evaluation strategy to generate comparable results. Therefore, test cases have been identified by Airbus that ascertain repeatability of test runs and their results.

After testing the general feasibility, a test of effectivity, i.e. a stress test, is foreseen. The platform receives images as fast as it can process them and reports the results to the Test Bench. Execution time-related metrics will be collected and documented. Next to the speed of inference, the accuracy is monitored closely to ascertain results with integrity. In a preliminary study, it was shown that an EDDL-trained model is able to achieve a similar accuracy compared to a Keras-trained network with the same architecture, however at the expense of being slower in training and inference in the same GPU environment [21]. An important accuracy metric that focuses on the first stage in the hybrid AI approach is the percentage of key point predictions that have an Euclidean distance of smaller than 1px to the true location, called predictions with sub-pixel accuracy.

The test runs are planned to be repeated in different hardware configurations of the platform. Moreover, modifications to the neural network (in terms of network size, quantization etc.) will reveal if significant changes in performance can be achieved.

V. Conclusion

As a mixed-criticality system, the SELENE multi-core platform provides the opportunity to maximize its efficiency while respecting and meeting safety- and security-critical requirements. It uses strong isolation capabilities at hardware and software level for an optimal separation of concerns. Also, it provides the infrastrucutre to speed up the inference process of machine learning models significantly using the custom library EDDL, a low-level acceleration runtime and the HLSinf open-source FPGA accelerator. Those components will provide the foundation of a multi-core system capable of handling computationally expensive tasks, especially the inference of neural networks.

Using the SELENE platform, Airbus presents a system architecture that will serve as a basis for future mission on-board data processing with a special focus on mixed-criticality tasks and robotic automation with the help of artificial intelligence. As next step, Airbus will continue the integration of the use case to the SELENE platform to be able to validate the platform in practice and obtain evaluation results. For this, the detailed evaluation approach will offer a comprehensive benchmarking against current space-grade equipment and will indicate with its results how compute-intensive tasks like neural network inference for image processing will be feasible on spacecrafts in the future, broadening the range of potential on-board applications significantly. It can therefore be seen as a step towards highly autonomous spacecraft operations in future human space flight missions.

References

[1] ESA, "*ESA - European Large Logistics Lander*," 2022, https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/Exploration/European_Large_Logistics_Lander [Accessed: 20.01.2022].

[2] Airbus, "*Bartolomeo - Airbus*," 2022, https://www.airbus.com/en/products-services/space/in-space-infrastructure/bartolomeo [Accessed: 20.01.2022].

[3] ——, "*Airbus pioneers first satellite factory in space - Airbus*," 2022, https://www.airbus.com/newsroom/press-releases/en/2021/03/airbus-pioneers-first-satellite-factory-in-space.html [Accessed: 20.01.2022].

[4] C. H. et. al, "SELENE: self-monitored dependable platform for high-performance safety-critical systems," in *23rd Euromicro Conference on Digital System Design, DSD 2020, Kranj, Slovenia, August 26-28, 2020.* IEEE, 2020, pp. 370–377. [Online]. Available: https://doi.org/10.1109/DSD51259.2020.00066

[5] H. D. project, "European distributed deep learning (eddl) library," 2020, https://github.com/deephealthproject/eddl.

[6] U. P. de València, "Hlsinf neural network inference accelerator for fpgas," https://github.com/PEAK-UPV/HLSinf, 2021.

[7] "https://eda.sw.siemens.com/en-us/ic/ic-design/high-level-synthesis-and-verification-platform/."

[8] Airbus, "*On-Orbit-Services - Services - Airbus*," 2021, https://www.airbus.com/space/Services/on-orbit-services.html [Accessed: 15.03.2021].
[9] M. A. R. Garzon, K. Nottensteiner, A. Wedler, and G. Grunwald, "Robotic technologies for in-space assembly operations," in *Advanced Space Technologies in Robotics and Automation (ASTRA)*, Juni 2017. [Online]. Available: https://elib.dlr.de/116805/
[10] A. Beyer, G. Grunwald, M. Heumos, M. Schedl, R. Bayer, W. Bertleff, B. Brunner, R. Burger, J. Butterfaß, R. Gruber, T. Gumpert, F. Hacker, E. Krämer, M. Maier, S. Moser, J. Reill, M. A. R. Garzon, H.-J. Sedlmayr, N. Seitz, M. Stelzer, A. Stemmer, G. T. Manteiga, T. Wimmer, M. Grebenstein, C. Ott, and A. O. Albu-Schäffer, "Caesar: Space robotics technology for assembly, maintenance, and repair," in *69th International Astronautical Congress (IAC)*, Oktober 2018. [Online]. Available: https://elib.dlr.de/122123/
[11] CORDIS, "*A Low Cost Active Debris Removal Demonstration Mission*," 2017, https://cordis.europa.eu/project/id/607099/de [Accessed: 14.04.2021].
[12] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," 2014.
[13] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," 2020.
[14] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," 2016.
[15] G. Pavlakos, X. Zhou, A. Chan, K. G. Derpanis, and K. Daniilidis, "6-dof object pose from semantic keypoints," 2017.
[16] C. J. Liang, K. M. Lundeen, W. McGee, C. C. Menassa, S. Lee, and V. R. Kamat, "Stacked hourglass networks for markerless pose estimation of articulated construction robots," *35th International Symposium on Automation and Robotics in Construction*, 2018. [Online]. Available: https://par.nsf.gov/biblio/10072688
[17] A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," 2016.
[18] D. Hirano, H. Kato, and T. Saito, "Deep learning based pose estimation in space," in *Proceedings of the International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*, 2018.
[19] K. Foerster, "Deep learning for keypoint localization and its application to object pose estimation," Master's thesis, Institute of Space Systems, University of Stuttgart, 2020.
[20] M. Sainz, U. Muñoz, and F. Eizaguirre, "Hardware accelerated machine learning: A literature review," 7 2020.
[21] N. Hauptvogel, D. Luedemann, and J. Matevska, "Neural networks for robotic pose estimation in safety-critical space systems," in *Proceedings of the SPACE&AI Online Conference by ESA and CLAIRE*, September 2021.

Artificial Intelligence for High-performance Human Space Flight Avionics Systems

DASIA 2022

Carles Hernàndez, Nick Hauptvogel, Laura Medina, Charles-Alexis Lefebvre, Jose Flich,
Martin Rönnbäck, Dierk Lüdemann

H2020 SELENE project grant No. 871467

19th May 2022

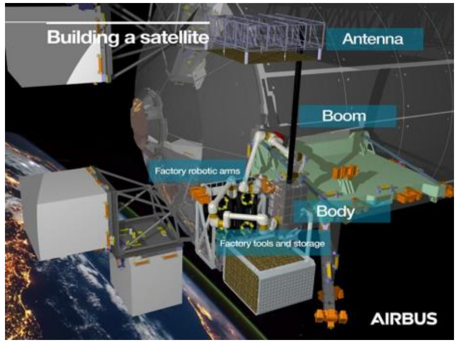
Agenda

- Introduction
- SELENE Computing Platform
  - AI Framework
  - HLSinf Accelerator
- Human Space Flight Use Case
- Conclusion

Introduction

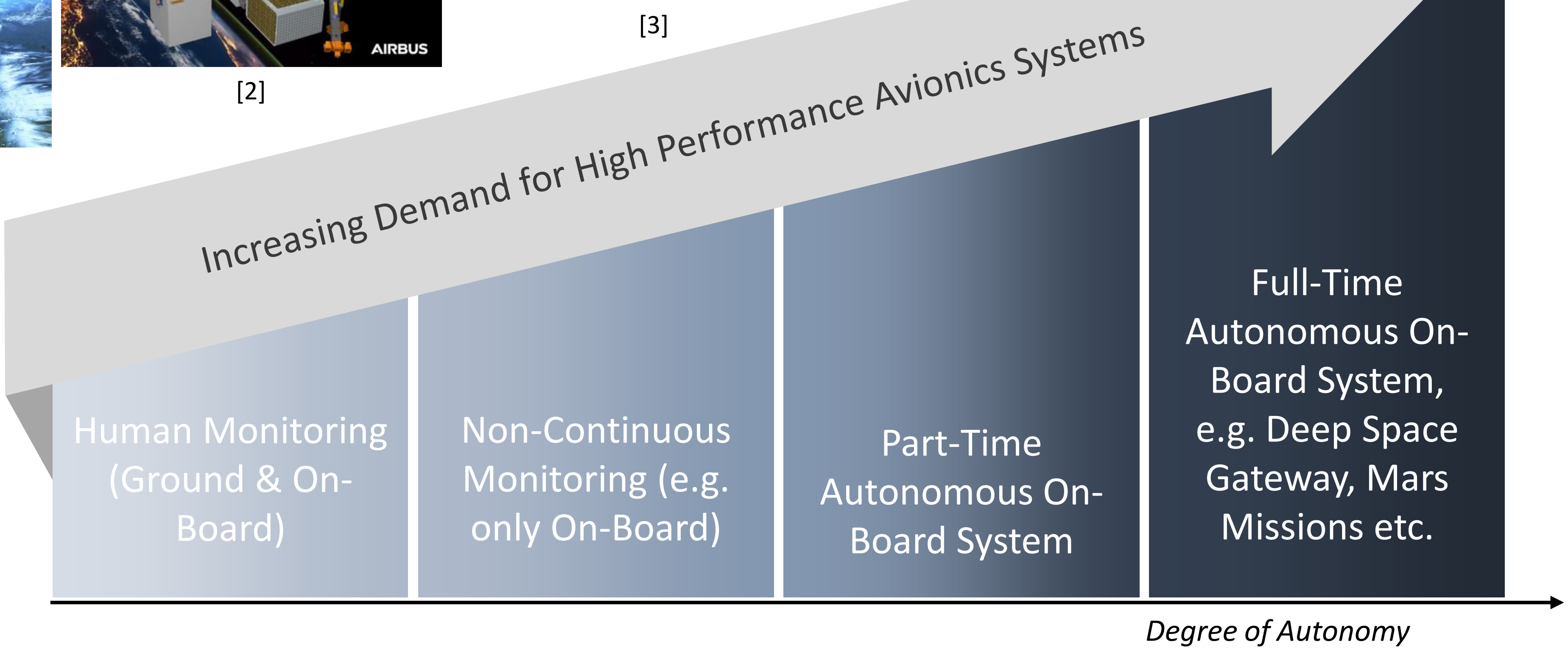
SELENE

[1]



[2]

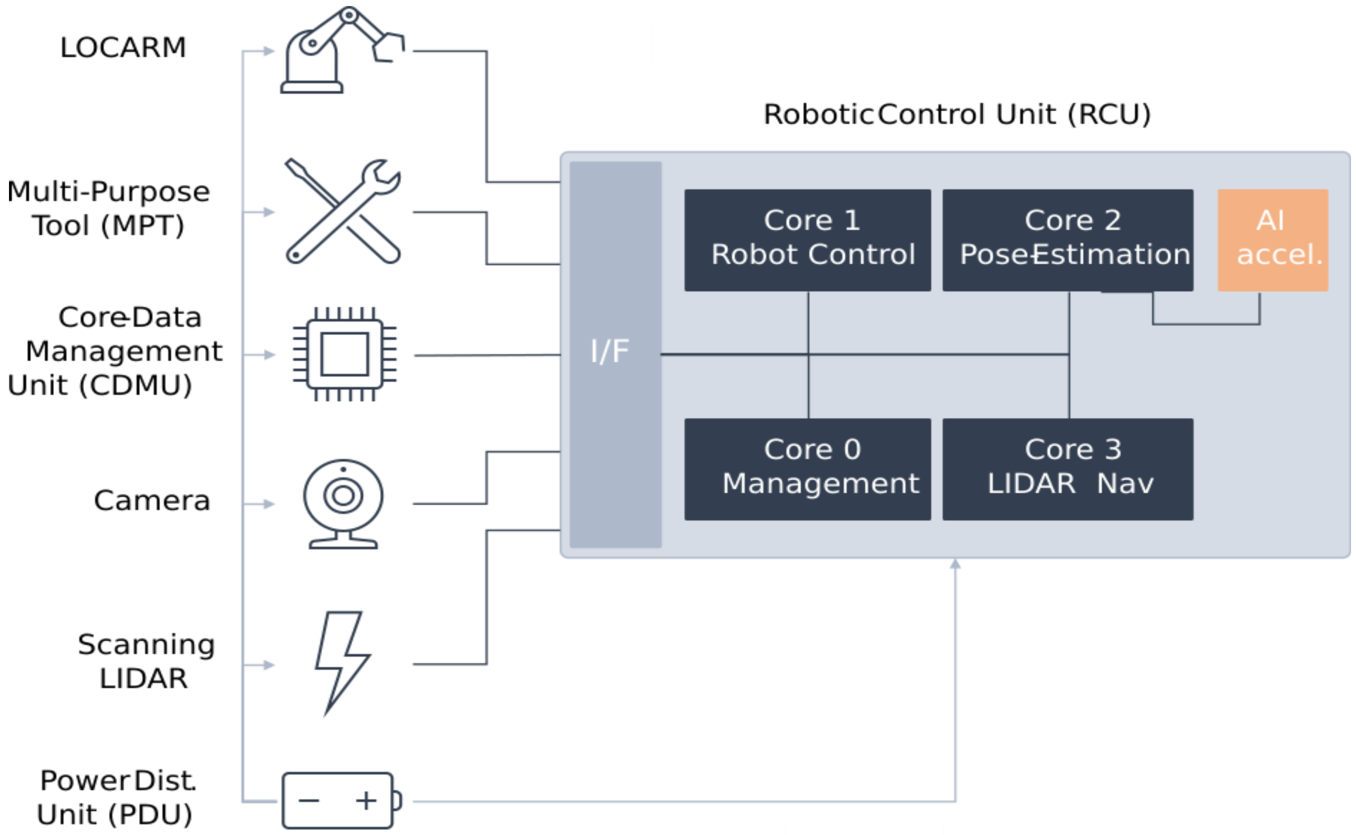
[3]



Introduction

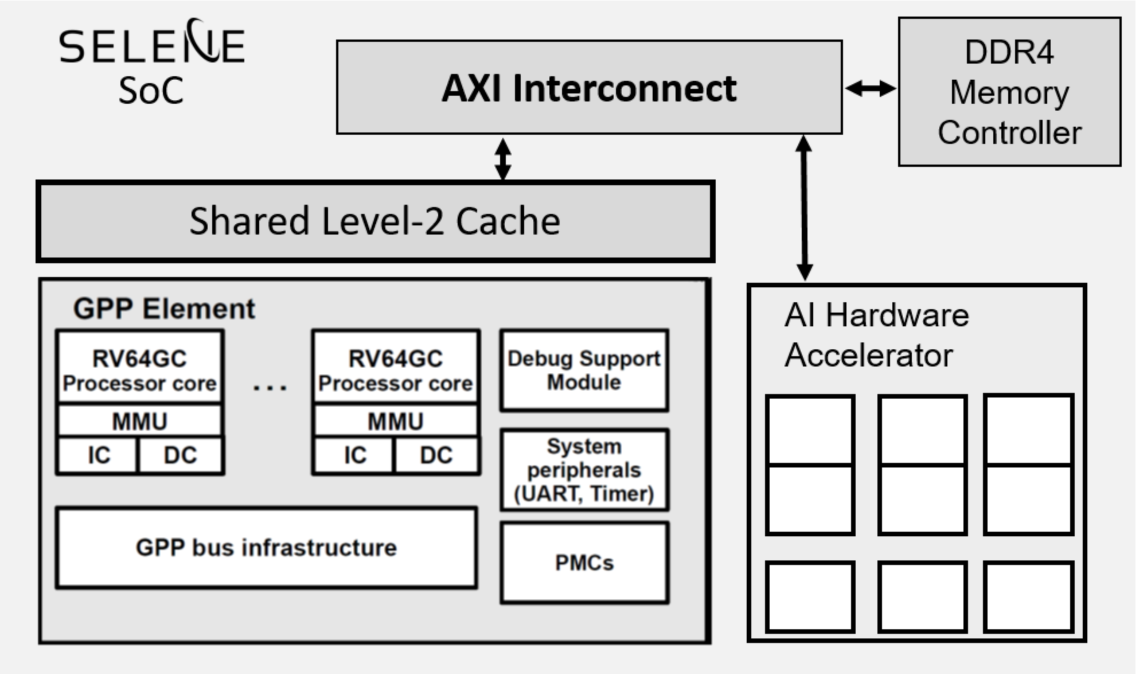
SELENE

- Multi-Core platforms as key enabler
- Mixed-criticality execution
- Safety and Security + High Performance Capabilities
- SELENE: RISC-V Multi-Core platform w. AI hardware acceleration and hypervisor



Computing Platform

- 6 NOEL-V RISC-V cores
- AI Hardware accelerator
- Linux as the default OS



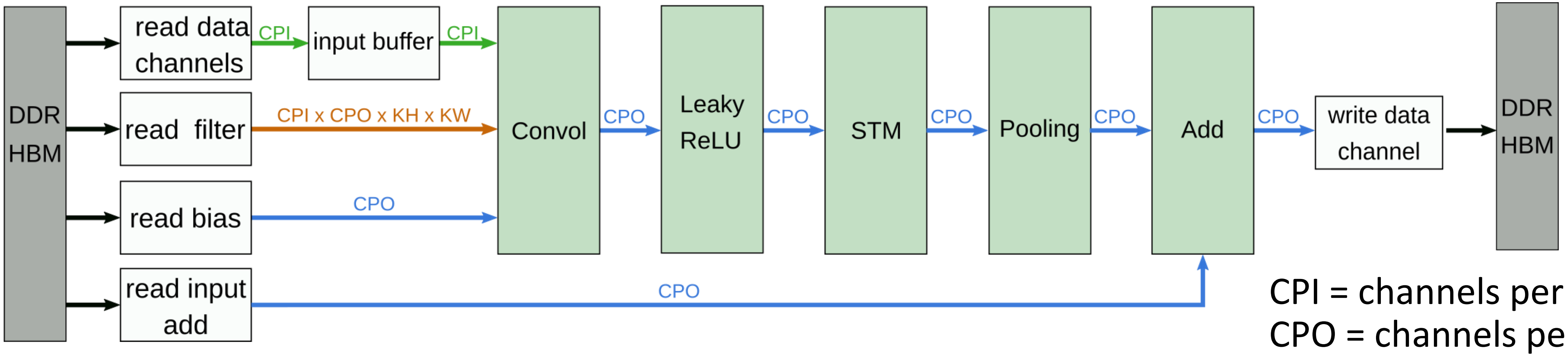
SELENE SoC

AI Framework

- Use of EDDL (European Distributed Deep Learning Library):
  - General-purpose, open-source deep-learning library
  - Used for **training** and **inference** processes
  - Offloads heavy computations to the **accelerator**
  - Executed in the SELENE **64bit RISC-V** cores
  - Supports ONNX format

HLSinf Accelerator

SELENE

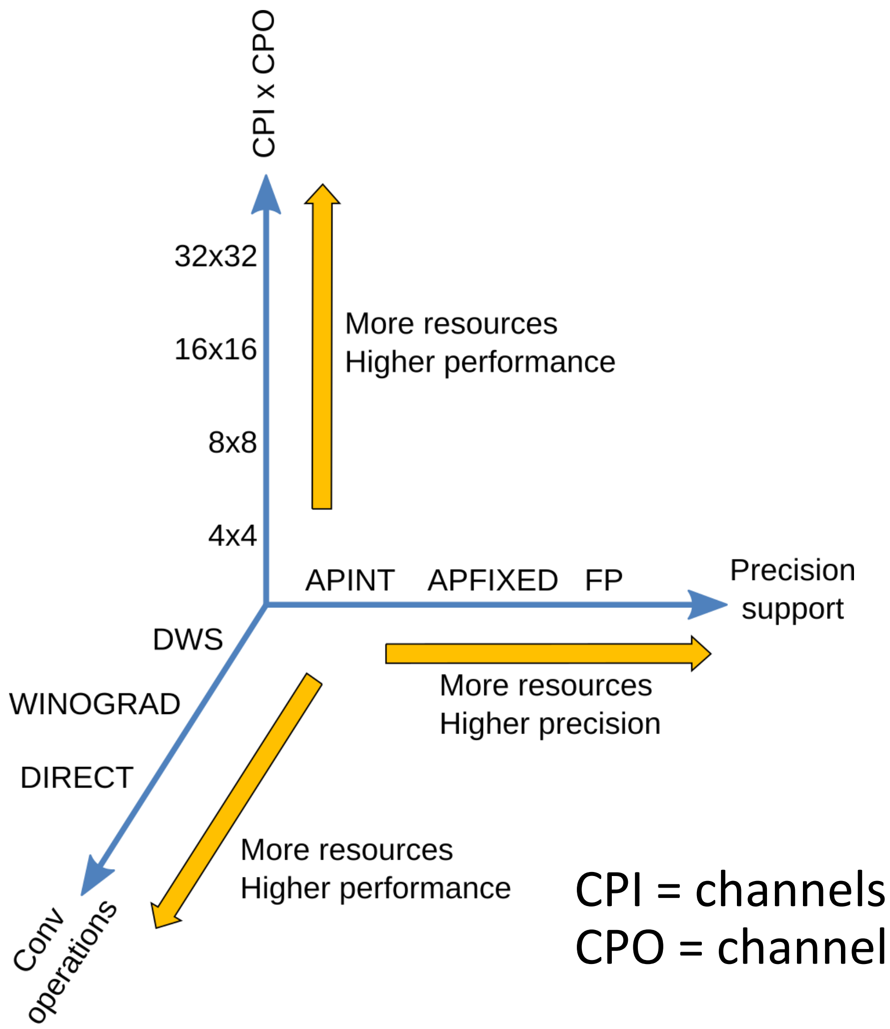


CPI = channels per input
CPO = channels per output

- **High-Level Synthesis** open-source project
- Uses Channel slicing
- Modules connected with streams building a **dataflow model**

HLSinf Flexibility

SELENE



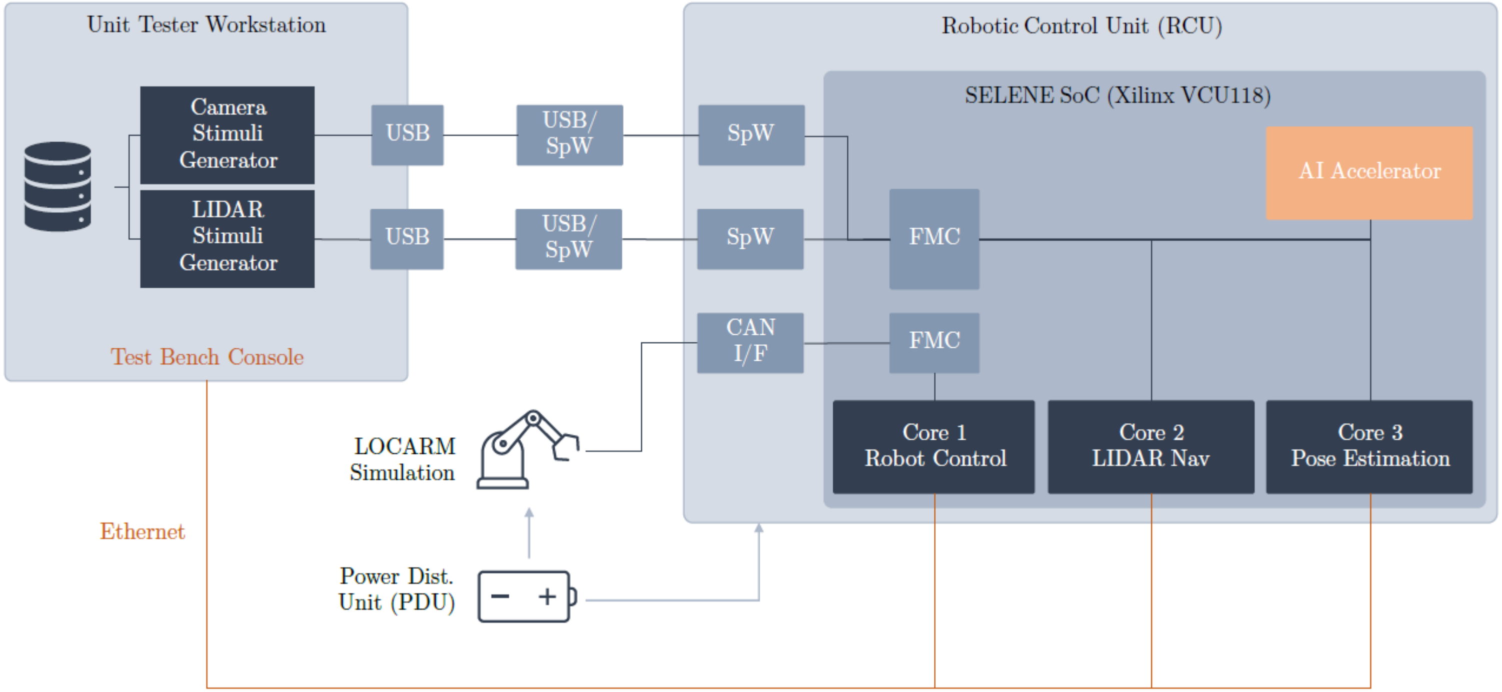
CPI = channels per input
CPO = channels per output

- **First** dimension: type of operations to be supported
- **Second** dimension: data types and precision formats
- **Third** dimension: input speedup and output speedup.

Human Space Flight

SELENE

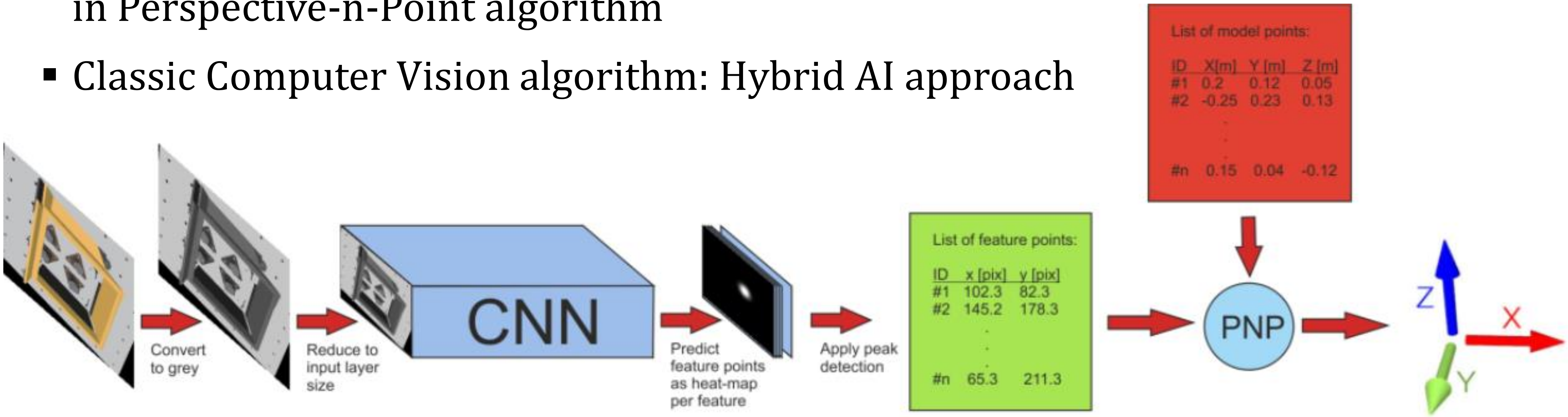
Intravehicular Robotics	On-Orbit Servicing
In-Space Manufacturing	Space-Debris Removal



- Housekeeping
- Robotic Control of simulation
- LIDAR Navigation
- **AI Pose Estimation**

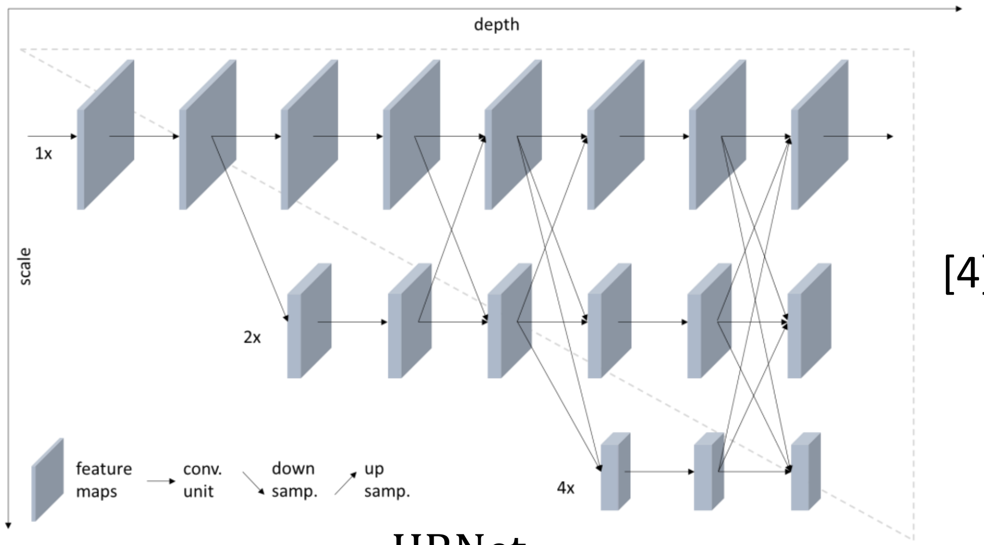
AI-based Pose Estimation

- Input: Camera Images of representative payload (e.g. CubeSat)
- Convolutional Neural Network detects key points via heatmaps, i.e. probability distributions
- Detected key points combined with 3D payload model in Perspective-n-Point algorithm
- Classic Computer Vision algorithm: Hybrid AI approach



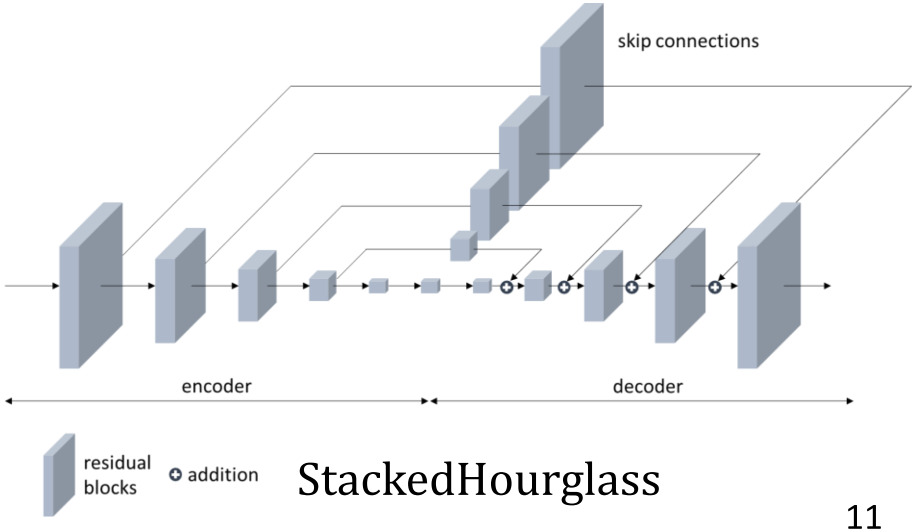
Convolutional Neural Network

- CNN prediction: Key point location
- Architectures from Human Pose Estimation
- Encoder-Decoder architectures
- High-Resolution branch vs. Skip connections
- European Distributed Deep Learning (EDDL) library as inference framework
- Support for .ONNX Neural Networks



[4]

HRNet

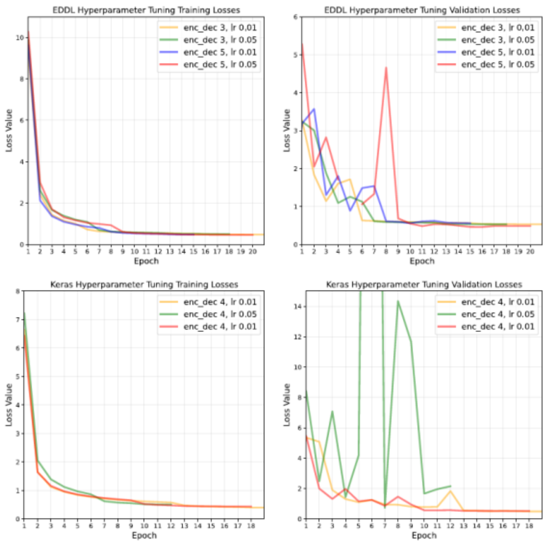


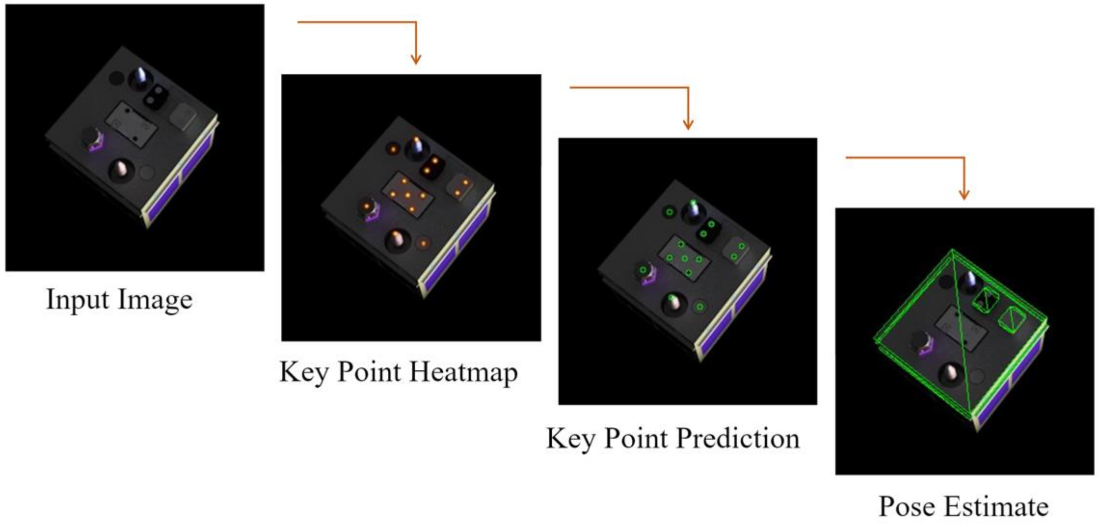
[5]

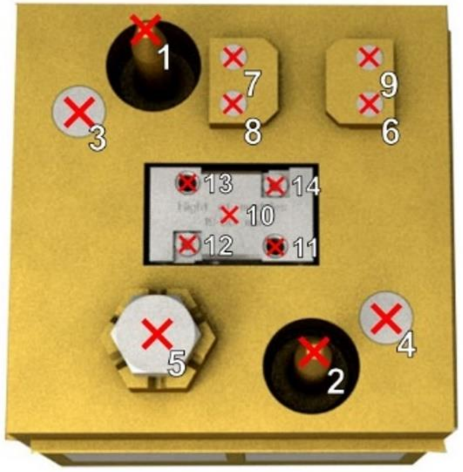
StackedHourglass

Evaluation

- Artifically rendered data set for Benchmarking
- Accuracy metric: Euclidean distance of true and predicted key points
- Execution time comparison between SELENE configurations, network types, sizes etc.







Conclusion

SELENE

- SELENE: Performance & Safety/Security
- AI acceleration with EDDL and HLSinf FPGA accelerator
- Airbus system design as basis for future on-board mission scenarios in robotic use cases

- Next steps
  - Integration continued in SELENE
  - Completed with detailed benchmarking against current space-grade equipment

Sources

SELENE

[1] https://www.airbus.com/en/newsroom/press-releases/2021-02-dlr-to-use-airbus-bartolomeo-service-for-first-ever-in-orbit

[2] https://period-h2020.eu/sites/default/files/PERIOD_Demonstrator.jpg

[3] https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/Exploration/European_Large_Logistics_Lander

[4] Image based on J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," 2020

[5] Image based on A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," 2016.

Contact

www.selene-project.eu