Reaching the Edge of the Edge: Image Analysis in Space

Robert Bayer, PhD Student, IT University of Copenhagen (ITU) (roba@itu.dk)
Supervisor: Pınar Tözün, ITU (pito@itu.dk) **Collaborator:** Julian Priest, ITU (jucp@itu.dk)

As the world increasingly relies on data, processing data efficiently is crucial. Depending solely on cloud-based solutions for efficient data processing encounters challenges related to latency, privacy, and network bandwidth. Firstly, real-time applications, such as autonomous driving, smart wearables and mobile devices have stringent latency requirements. This makes offloading all of the data processing to the cloud infeasible. Secondly, the data being processed has increasingly been our personal data, either at home through the use of personal assistants or in hospitals where doctors leverage medical data (e.g., image analysis) in order to provide us with the right treatment. The data-intensive applications are, therefore, met with not just ethical but also regulatory restrictions on data processing. Lastly, the constant rise in the amount of data collected could, in the future, lead to network congestion. While new networking technologies are being developed to target this problem, the transferred data is often redundant and discarded after simple pre-processing in the cloud. Additionally, certain data sources are not in easy-to-reach areas with good connectivity (e.g., space, underwater), making the transfer of all collected data infeasible. Processing data close to its source, *at the edge*, could alleviate all of these issues due to reduced off-device data movement. This, in turn, lowers latency, privacy risks, network bandwidth requirements, and carbon footprint and increases the value of hard-to-reach data sources.

While the amount of resources available on edge devices is growing, increasing these indefinitely is not a viable solution, especially in battery-powered deployments. Nowadays, medium-scale edge devices can reach 4-8 cores and 4-8 GBs of memory, which is highly limited compared to 100s of cores and TBs of memory found in the cloud. These devices are still, however, expected to perform highly data-intensive applications, which, on these devices, can only be achieved with high efficiency.

Satellites are a major data source at the edge. With the increase in accessibility, the number of new satellites constantly increases as their deployment becomes attainable for smaller companies and institutions. The reduction in component sizes saw satellites' overall size shrink, introducing a new standardized satellite format called CubeSat. These satellites comprise 10x10x10 cm units, allowing modular design and better space utilization of transport vehicles, decreasing the launch costs. The decrease in the size of the satellites leads to a decrease in the amount of resources. The reduced surface area of the satellite limits the power generation capabilities, while the volume limits the physical dimensions of the electronics.

Satellite imagery is one of the most widely spread data-intensive applications running on satellites. The limited power generation of CubeSats limits the choice of networking technologies used for communication with Earth, leading to reduced bandwidth of an already unstable connection. To better utilize this reduced bandwidth, developers increasingly rely on machine learning to perform smart filtering of data to increase the perceived value of the data sent back to Earth. This includes removing cloud-covered imagery, under- or over-exposed images, and more advanced methods such as super-resolution, which stacks multiple images in order to increase the spatial resolution of the imagery.

In this talk, I plan to present lessons learned in designing an image processing unit for the Danish Student CubeSat Organization's (DISCO)[1] first satellite DISCO1. We use this image processing unit to perform deep-learning-based image filtering on the imagery captured by the satellite. We identified the specific constraints of the DISCO1 satellite and three example use-case scenarios to assess the suitability of edge devices as image processing units on board of DISCO1. We investigated the performance of seven edge devices based on different architectures, showing that only the highest degree of parallelism found in GPUs or ASICs can fulfill the latency requirements of the real-time imaging inference scenario on large 4512x4512 pixel RGB images. Through this analysis, we identified the Coral Dev Board Mini as the most suitable candidate, satisfying all the requirements: latency, peak, and nominal power draw. We then show the steps necessary to integrate such a device into a satellite. As this satellite will serve as a platform for future student experiments, the integration of the image processing unit focuses on reliability, fault tolerance, and modularity. The satellite, together with the onboard image processing unit, was launched in April 2023[2] and currently orbits the Earth.

[1] `https://discosat.dk`

[2] `https://en.itu.dk/About-ITU/Press/News-from-ITU/2023/Students-launch-a-satellite-to-test-artificial-intelligence-in-space`