# Episode 218 – AI for EO, and Neural Network Supervisors and Overcoming the Clouds

Speaker: Aubrey Dunne, Co-Founder and CTO, Ubotica – 24 minutes

John Gilroy:

Welcome to Constellations, the podcast from Kratos. My name is John Gilroy and I'll be your moderator. Joining us today is Aubrey Dunne, co-founder and CTO at Ubotica, a company working to empower satellites with intelligence. With more than 50% of the earth covered in clouds at any given day, the ability to provide real-time insights about what is happening on the ground becomes complicated and nearly impossible. Aubrey is here to share his knowledge on AI-based cloud detection and removal for earth observation satellites that overcomes clouds and delivers insights about what's happening on earth right now. Aubrey ready to go?

Aubrey Dunne:

I certainly am, John.

John Gilroy:

Great, great, great. As I mentioned, Earth's average cloud cover severely limits usable earth observation imagery from satellites. So can you share with us some of the impacts that result when imagery is limited in this way?

Aubrey Dunne:

Absolutely, John. Yes, and thanks for having me here today, by the way. I'm delighted to talk to you on this topic. So the obvious way is that you cannot see the ground in the way that you would expect to or hope to when you schedule those acquisitions or when you take those images. So in the optical domain, you cannot see through cloud so that any cloud obscuration means you don't get the amount of data that you wanted. So obviously that's an inconvenience in itself, but then if you take the bigger system and look at it at a broader scope, it reduces the overall efficiency of an earth observation system or an earth observation satellite.

Aubrey Dunne:

So if you have a limited amount of bandwidth or a fixed amount of bandwidth, so a fixed amount of downlink capacity, and you are downlinking images that are full of pixels with clouds in them, well then your overall efficiency of valuable data is being reduced. And so, if you can in some way ensure that you can only downlink useful pixels, usable pixels, so the ground pixels, then your overall efficiency of the system goes up. So there's an efficiency play as well as the obvious. If you can't see the ground, it's, for most uses, not very useful.

John Gilroy:

Because 50% of the time then you can't use what you spend so much money getting up there, huh?

Aubrey Dunne:

Exactly. And what you'll see is some of the operators, they effectively have to spread this cost or amortize that cost across their imagery. So it actually ends up pushing up the cost to the end user of the valuable imagery because they have to cover the cost of the cloudy imagery as well.

John Gilroy:

So how can on-orbit artificial intelligence help serve as a solution to this cloud problem?

Aubrey Dunne:

I'm glad you asked that. So there are many different ways. We actually work on two ways ourselves within Ubotica, and maybe I can talk you through those two ways. So one way is to actually remove the cloud directly on satellite. So when the image is captured, to process it directly on satellite, detect where the cloudy pixels are, and then attempt to remove them and only downlink or only send down the non-cloudy pixels. And we do this using AI techniques on board. AI is very, very well suited to this type of application. And there are AI models that have been trained to detect clouds and to segment clouds, many different versions. We've trained some ourselves, but there are many different models out there. Being able to process directly on board means that you can effectively throw away or discard your unusable image. So this is one way of doing it effectively, reduce or only send down the useful information.

Aubrey Dunne:

Another way, which is something we demonstrated very recently actually in collaboration with NASA JPL, is dynamic targeting. And this takes a different approach. This is where you use some autonomy on the spacecraft to actually ensure that the spacecraft images where there is the least likelihood of cloud. And so this is something we supported JPL in demonstrating say about two weeks ago at this stage. And it works on a satellite that we co-own. We demonstrated this where we pitch forward, we look forward with the satellite about 500 kilometers ahead, we capture an image and we process that image with an AI algorithm to detect the cloud, and we find the least cloudy portion of that image, and then we use as a control signal to autonomously have the satellite pitch back to nadir. So it's looking straight back down again, and then it rolls to the area that had the least cloud and look ahead image.

Aubrey Dunne:

So it's this idea of dynamically only imaging or dynamically imaging where there's the highest value to be attained. So it takes a slightly different approach. In this case, you're actually actively controlling what and where you image in order to reduce the amount of cloud that you capture.

John Gilroy:

Aubrey, I've done a lot of radio and a radio phrase is, "Hey, the phrase that pays," and I think the phrase that pays in this discussion is I wrote down on satellite. So this is kind of a big deal. All it means is that don't have to worry about latency and when you get it, you're not wasting that latency. It's more efficient, isn't it?

Aubrey Dunne:

Absolutely. It's really an efficiency play at the end of the day. And in fact, there's been some publications published in the last couple of years which has shown that you can increase your efficiency of usable data or the value of the data you get by about 20 x by taking this approach of dynamically imaging, only imaging where you know there's a high probability of real useful data, in this case of non-cloudy pixels.

John Gilroy:

So Aubrey, can you explain a little about the cloud detection and cloud removal algorithms in onboard systems? So how does the AI identify, mask, I guess reconstruct, all these different regions to give, I guess what they call maybe surface fidelity? Is that the phrase?

Aubrey Dunne:

That works. I know what you mean by that. Sure. And I have a presentation on this tomorrow, so I should be familiar with this topic. So I'll talk about the case where we do what we call cloud removal and compression. So this is where we have a flow that runs on satellite. On board, and it has, if you like, three steps. So it has a cloud detection step, it has a cloud removal step, and it has a compression of what's left step. So these three steps run in a flow. So we have demonstrated this on-orbit at the end of last year. We validated and verified this on-orbit.

Aubrey Dunne:

And effectively you asked how does it detect the cloud? So we use a quite lightweight network called a segmentation network actually using an architecture called U-Net. It's an AI type type architecture. And this tries to identify almost per pixel, is the pixel cloudy or is it not? So the end of the results of applying the AI model to the images using segmentation is that you have a mask and that mask, it's a binary mask. Either it says, "There's a cloud in this pixel," or, "There's no cloud in this pixel."

Aubrey Dunne:

Now it's not that simple. We are up to threshold that mask and to allow us to adjust how cloudy we accept and how much cloud we say there has to be before we reject it. But this is principally how it works. So once we have that mask, we basically, we use tiles in the image and say, "If the cloud in this particular tile is above this threshold, we discard it." Otherwise, we don't discard the tile. We keep it and we compress it and we use what's called lossless compression or lossless JPEG compression. Again, we do this all this on satellite, and this is all done with hardware acceleration, so it's very efficient. So we compress the remaining tiles so that we get this double effect of both removing cloud and compressing what's left, and then we package it up into a single file and downlink it.

Aubrey Dunne:

On the ground side then, we reverse the process, we decompress the image and then reconstruct the tiles, put them into the locations that they came from in the original image using the metadata that's downlinked. So maybe that was a bit technical, but it's a flow that we have developed that allows us to fully do the compression, sorry, the cloud removal and compression on orbit in one go using very power efficient hardware on satellite as well.

John Gilroy:

Well, there's nothing wrong with technical because a lot of technical listeners, and when you said lossless, this is, technology that's been around for a while, it's not just invented last night, and so it's very reliable. Earlier you said you give a presentation tomorrow on this. Well, we are recording this from the floor of the Small Sat Conference in lovely downtown Salt Lake City, and you're presenting here on this specific topic. So if you can't afford to come to Salt Lake City, this is a place to catch up with Aubrey, huh?

Aubrey Dunne:

Absolutely more than welcome.

John Gilroy:

So Aubrey, what have you seen as some of the key challenges in achieving real-time inference for AI-based cloud detection and removal under the power and bandwidth constraints of spacecraft?

Aubrey Dunne:

It's a great question. So the power one is probably the easier one to address here. So the spacecraft that we target are typically relatively small, so they're CubeSats or small sats, they're maybe from a size of 6U and above. So about the size of a briefcase and larger, kind of size, they're quite small. This means that they have a limited area of solar panels, and therefore, they have a limited ability to generate energy for their batteries. So power becomes quite an issue on these spacecrafts. So we address the power side by using very power-efficient hardware. So this is hardware that specifically has blocks in it to accelerate the inference in a very power-efficient way. And similarly for the compression, and try to do this in a power envelope of about four watts is what we work on. So four watts is about one-tenth the power that your cell phone would use when you're watching Netflix. So it's very, very low power.

Aubrey Dunne:

We do this because the satellite has to use power for all the other systems on the satellite, and therefore, we only have a certain amount that we can use for this cloud compression approach. So that's on the power side. We also developed a network that is very lightweight, so it's a very small network if you like. I can't remember the exact size. It's about 10 megabytes, maybe slightly less. So because it's small, it means that it's relatively quick to run, and therefore, doesn't consume as much energy in its execution. On the bandwidth side, it's slightly different, because in the case of bandwidth, often what will happen is, the operator has a specific capacity, a specific amount of downlink that they have contracted and that's available. So we see the benefit here really being in, you're not necessarily reducing that, but what you're doing is, you're making sure that the images that you get down are more useful, or rather that you get more useful images down for the same amount of bandwidth that you have capacity for.

John Gilroy:

So Aubrey, I walked around the show floor here this morning, lots of people talking about reprogrammable satellites and good and bad news on this and some challenges too. So with satellites reprogrammable in orbit these days how are model upgrades, retraining and evolving mission profiles managed over the operational life of a satellite like this? This is challenging.

Aubrey Dunne:

It is. It's a great question because, from my perspective, this ability to reconfigure is really the key reason why you fly AI in many cases in the first place on a satellite. When you enable a satellite with AI, it's a capability that you're enabling. It's not application specific or not necessarily application specific, but you have a capability that you enable, which allows you to swap out the applications, to update the applications post-flight, but when the applications run on satellite, they're still hardware accelerated, they're still using hardware that's very, very efficient at performing inference. So it's a capability that you fly and then that you can address different applications with that capability. So in terms of updating the network, that's something that is very, very typical.

Aubrey Dunne:

So normally, you have a relatively limited amount of data to train a model on the ground. Once you fly the satellite, you capture more imagery on this imagery, then you downlink and you use that to retrain the model to update, to fine tune that model and then uplink a new model. We have done some work on making those uplinks as efficient as possible because, in some cases, or in many cases even, the communications are asymmetric. So the amount of data that you can downlink from the satellite is maybe 10 x the amount that you can uplink. Because that's the way these original satellites were designed. They were generally capturing data and downlinking it, but not so much into the direction. So when you want to update a model that's maybe 100 megabytes, that can be quite a challenge on a lot of these satellites.

Aubrey Dunne:

So what we tried to do is keep that model as small as possible and then have some ways to effectively compress the model update as well, meaning that we use less uplink bandwidth, that we can get the model update faster. But absolutely, fundamentally, the AI capability on all of the satellites, at least all the ones we're involved in, it's designed to be updatable, upgradable and swappable, and we can even hot swap models. So we can have a particular application that uses two models and ping pongs between the two models. It's really, really flexible capability, which is one of the reasons why we fly it.

John Gilroy:

Aubrey, earlier you mentioned lossless compression and downlink, and you just mentioned uplink. Let's go to the show-me-the-money part of this interview, the ROI on all this transmission. So in measurable terms now, how much can operators reduce downlink costs or processing latency after integrating onboard cloud removal, and what typical ROI scenarios are there? There's got to be something there.

Aubrey Dunne:

Absolutely. It's a fair question and a good question. So we did some analysis on this because obviously we developed this cloud removal and compression approach and solution. So the easy way to look at this is that you mentioned at the beginning that there's about 50% of the world covered by cloud. Depending on who you listen to, and I think NASA have reported it, it's about 67% is the number they give. So let's call it two-thirds on average at any one time. So if you can eliminate all of that by processing it on board and removing the cloud on board, you're effectively able to generate three times the amount, or even downlink, three times the amount of useful data than you would be if you didn't do the cloud compression and removal or cloud detection rather and removal. So if it's 67% cloudy, and if you remove all that cloud, you effectively can downlink three times more data.

Aubrey Dunne:

So this has real value to missions where, as I say, they're bandwidth limited or where they have a certain amount of data that they can downlink per day or per week, they can effectively triple that amount of useful data that they can get down. Another way of looking at it is on the ROI and the cost savings and cost benefits. And what we have analyzed here is that, on a typical CubeSat EO CubeSat, earth observation CubeSat, that by taking this cloud removal and compression solution over the course of a year, per satellite, you can save up to about $150,000 per year in downlink costs. So that's another way to look at this quantum.

John Gilroy:

You mentioned technical people. We have technical people that listen, and we have people who just want some stories. Tell me a story and maybe boil it down to something. So here's a story question for you now. Can you walk us through the first AI-driven ship detection event and what that meant for real-time earth observation?

Aubrey Dunne:

I would love to. Absolutely. So as I mentioned earlier, we have our own joint mission with a company called Open Cosmos in the U.K., which we launched last March. So this is a 6U CubeSat mission, about the size of a briefcase. And on board that, we have our AI software and hardware solution. So enables us to do power-efficient AI on board. And so in June of last year, June 2024, we captured the first operational image with the satellite over the port of Jeddah in Saudi Arabia. And when we captured that image, so it's acquired on satellites, it's saved, and then we started off our AI processing flow on that particular image. So we load the image back, we tile the image, so we cut it into smaller chunks, we process each chunk through our AI algorithm that performed a ship detection or actually ship segmentation.

Aubrey Dunne:

So this is where we're trying to identify the locations and the orientation and the size of all the vessels in each tile and then iterate through all the tiles over the whole image. So the image is about 800 square kilometers in total around the port of Jeddah. So in just under three minutes after starting the processing, we had detected all of the vessels in that image. That's the end of the AI part of the flow, but what we do with that data? So then what we did is, we take that data again on satellite in an automated flow and package it up into 200 byte packets. So really, really small packets of data, 200 bytes is a very, very small amount of data. And that just contains the location, the orientation, and the size of each of those vessels.

Aubrey Dunne:

And then we send those packets down over what's called an inter-satellite link, which is a radio link between our satellite and a constellation of communication satellites orbiting in low earth orbit. And it gets relayed from those satellites through the constellation back down to ground, to one of our servers. So in short, this is a way to transmit those 200-byte packets without having to rely on the traditional ground station overpass approach, which is how the images are downlinked. You have to wait up to 90 minutes to pass over a ground station. You may have contention for bandwidth at that ground station. So we bypass that by saying, "Let's condense this full image, 800 square kilometers, into a couple of hundred bytes, put them onto into an ISL packet and send it over this inter-satellite link down to ground." And so we got to about... Well, now we're down to about 10 minutes. I think in that case, we're about 20 minutes from start of acquisition to delivery.

Aubrey Dunne:

Most recently, again, looking at ports because we're interested in maritime situational awareness, and we've gone from 10 minutes from the start of acquisition to having that information about the vessels down on the ground. And so this allows you to look at that the latency is massively reduced here, so we're getting close to real time. It's near real time information. And this allows to action on that information in a very different way than if there was a two-hour or five-hour, 10-hour or 20-hour delay. For example, in the case of vessels, you can start interdicting those vessels, intercepting if there's some potential illegal activity that they're involved in. And by having that information down in a matter of minutes instead of hours or days, you know where the vessel is going to be because it's not going to move very far in a few minutes, and therefore, you can intercept and interdict. So that's an end-to-end story about having ship detection.

John Gilroy:

I think the headline for this interview, got to use the word efficiency. Every single answer you've given me, "Increase the efficiency, increase the efficiency," and corollary is the reduced costs. That makes sense. A word that's popular in the trade show today, I'm sure if you listen to people speaking, use the word "autonomous" and "satellites" because it's really important. So as these satellites become more, I guess, autonomous sensors, how do you address privacy, data sovereignty, AI decision governance, especially in maritime in some sensitive data sections?

Aubrey Dunne:

It's a great question. So there's a couple of ways that this is addressed and that we address it. So one is on encryption, so encrypting both at the payload level and at the comms level. So having a communications channel from satellite to ground, to the ground station, to the end server that is itself encrypted. And then also encrypting the data payload itself so that the packets of information that we're sending down, encrypting those. So you have this double layer of encryption. That's really important. Responsible management of data is really, really important as well. So some of this is a responsibility and a process element where you have to understand, once that data comes down to the ground, where does it get stored? How does it get saved? Who has access to that? How is that protected? How is that backed up? Et cetera. So that's I guess more general software engineering and enterprise system developments techniques or approaches.

Aubrey Dunne:

So they're two of the main ones. One other one on reliability, which is maybe a slightly different question, but I think related, is this concept of how can you trust the results of the AI model that you're running. And this has been something that has been talked about for many, many years on the ground as well, and even on space now. Because as people try to develop autonomous systems on satellite using AI, there is a question, "Well, if you don't really understand the black box that is AI, do you really want to trust it to control your satellite?" And it's a valid question. And one of the ways we've tried to address it and other people are taking a similar approach is by having something called, at a high level it's a neural network supervisor.

Aubrey Dunne:

So it's a separate software module that sits outside the network and monitors the performance of the network. And really simply what it does is, it looks at how the individual neurons or nodes in that network are firing, the patterns of firing, and it checks, did it fire that way during training time? Because if it didn't, then it's probably working in a domain for which it wasn't trained, and you should be more suspicious of that result and probably, either repeat processing, or at least don't trust that result to the same level that you trusted the other results. So it's a monitoring on-satellite approach that helps to, I guess, give more confidence and have a supervisory role over the performance of the AI model.

John Gilroy:

Aubrey, looking ahead, what other capabilities could be further enhanced by combining AI with other sources of data?

Aubrey Dunne:

Again, a very good question. So data fusion is another hot topic, and it's very often talked about in space and in non-space context as well. And this is really where you're pulling different data sources, what we might call orthogonal or data sources that are uncorrelated and combining them together in a fusion approach, which allows to develop or allows to generate a result that is of higher value than the results you would get from say treating any data source alone. So this is the general concept of data fusion. So in our work, we do some data fusion in the maritime situational awareness domain. So a simple example of this is where we run the flow I talked about earlier on board, where we detect the vessels on satellite, package up that information into small packets, send it down to ground really quickly, low latency, and then we fuse it with what's called AIS. So that's the automated identification system. This is a transponder information from the vessels themselves. This is terrestrial data. Well, it's transmitted terrestrially from the vessels.

Aubrey Dunne:

And this provides GPS locations or GNSS locations of the vessels. And by fusing that information with the insights that we pulled down from the satellite, we can generate a higher confidence in the locations of the vessels and also start identifying vessels where there is a mismatch between where we see they are and where they say they are effectively with their AIS. So this allows us to start detecting the likes of what are called dark vessels, which are vessels that turn off their transponder or spoofing where vessels, they transmit data indicating they're in a particular location, but actually they're in a separate location. That's one example of pulling this information together. There's other data, for example, like ORF intelligence, which is where, on satellites... And this is something we're looking at as well... Where you capture effectively the signature or the energy emitted by ship-borne radars. And this allows you, again, to try to pinpoint where the vessels are based on the radar that they are transmitting as they run their radar systems on ship. So it's another data source that you can also fuse in a similar way.

John Gilroy:

So much change, so much innovation. I got to ask another forward-looking question. See if you can handle this one. So from your perspective, what are the key remaining technical or regulatory hurdles for AI to become a standard feature across the whole EO satellite market?

Aubrey Dunne:

It's a great question, and there's probably quite a few answers to that, but maybe technically, I think we are quite far advanced. We first flew AI on a satellite in 2020, a PhiSat-1 and a ESA, European Space Agency, satellite in 2020. And that was the world's first hardware-accelerated AI inference on a CubeSat, at least outside of the defense industry. So that was only five years ago. So things are moving along very, very fast. And as you say, today, you walk around the conference here today, a lot of people are talking about onboard AI and AI autonomy. So the technology is progressing very, very fast. It's not easy, but I think that we have shown that it is definitely doable, and it's at the point where it is operational. And that's why, for some of our customers, we are operationally deploying AI for them on satellites. So the technology is definitely getting there.

Aubrey Dunne:

As time goes on, for example, in terms of earth observation sensors, the resolution increases, their spectral resolution increases so there's more data. So is this constant problem of more data, how do you process more data on-satellite? Because when we want to deploy AI on board, we have to be able to touch all those pixels. And so there is an element of new hardware coming online that's suitable for spaceflight and that's suitable for performing some of this AI inference on orbit. But I think that is coming, and we see it now with some of the large AI hardware providers that people are taking this new space approach where they qualify, or they characterize rather, hardware and then fly it even though it's not designed for space. So they take this new space approach of flying what's called commercial off-the-shelf hardware. So technically, we're definitely making progress.

Aubrey Dunne:

I think regulatory-wise, there's definitely some work that needs to be done probably, or some movement that needs to be had over the next couple of months and years, particularly on the autonomy side. Because, at the moment, and I'm not a regulatory expert, but there are some restrictions around what can be done autonomously on satellite. And that feeds into the licensing and the national regulatory bodies to provide licenses for some of these missions. And so, I certainly know for some of the missions I've been involved in, we were prevented from doing what we might call true autonomy because of the licensing, because they need human in the loop.

Aubrey Dunne:

And this is understandable, because it's looking at, if the AI doesn't work very well, if the control doesn't work very well and the satellite goes out of control, then we can have this knock-on effect of taking out or creating a lot of space debris and it can accelerate from there. So I understand the why, but we need to probably have that discussion a bit more openly about how do we progress the regulation to enable some of this automation in a controlled way and taking it step by step by step.

John Gilroy:

Aubrey, I'd like to thank you for your time today at the Small Sat Conference. You have been listening to the Constellations Podcast from Kratos. I'd like to thank my guest, Aubrey Dunne, co-founder and CTO at Ubotica.