Searching for extraterrestrial artefacts on the moon and in the solar system: detection strategies and techniques

Birkbeck
UNIVERSITY OF LONDON

Searching for Extraterrestrial Artefacts on the Moon and in the Solar System: Detection Strategies and Techniques

Lewis James Pinault

A thesis submitted to

Birkbeck College, University of London
for the degree of Doctor of Philosophy (PhD).

Primary supervisor: Ian A. Crawford

Thesis submission date: 7 January 2024

Re-Submitted with Minor Corrections 3 November 2024

Abstract

Building on 1980s hypotheses developed by Alexey Arkhipov, this Thesis presents new Machine-Learning based strategies, methods and techniques for detecting extraterrestrial technological artefacts. The focus is on ways to find evidence of submicron to micron debris that could have travelled across interstellar space via natural forces to the Solar System. It also applies these techniques to the object detection of spacecraft hardware remnants >0.5m on the Moon: exploring the author's hypothesis that such dust-sized artefacts could include pre-programmed material designed to construct exploratory probes from local resources, much as submicron-scale DNA carries instructions for one of the most complex constructions we know so far in the Galaxy, humans.

Imminent robotic and human activities on the Moon and other planetary bodies would benefit from advanced *in situ* Computer Vision and Machine Learning capabilities to identify and quantify microparticle terrestrial contaminants, lunar regolith disturbances, the flux of interplanetary dust particles, possible interstellar dust, β-meteoroids, and secondary impact ejecta. The YOLO (You-Only-Look-Once-ExtraTerrestrial) algorithm fine-tunes Tiny-YOLO to specifically address these challenges as well. Designed for coreML model transference to mobile devices, the algorithm facilitates edge computing in space environment conditions. In collaboration with JAXA, training on images from the Tanpopo aerogel panels returned from the International Space Station, YOLO-ET demonstrates a 90% detection rate for surface contaminant microparticles, and demonstrates promising early results for detection of both microparticle contaminants on the Moon and for evaluating asteroid return samples.

YOLO-ET demonstrates an 80% detection rate for Apollo lunar landing modules, correctly identifying a known Luna 16 as a landing module. YOLO-ET also detects two potential candidates for Luna 9, with confidence levels of 61% and 43%, a spacecraft whose exact location has thus far remained undetermined. The light computing resource demands of YOLO-ET suggest that is well suited to continuous video object detection over the Moon's surface.

Acknowledgements

We are at the cusp of a paradigm shift from SETI to SETA, the Search for ExtraTerrestrial Artefacts, and it's a privilege to submit one of the first PhDs staking out this new and strangely wonderful terrain.

We don't know, yet, whether we are alone, but I know that *I* am not alone. First and foremost, I wish to hugely thank and humbly acknowledge Professor Ian A. Crawford for his unstinting support and confidence, generous insights, patient guidance, and quiet, steady enthusiasm.

From the outset he's had a perhaps more conservative view of near-term detection outcomes, but never wavered in his belief that we could set out assumptions, tools and techniques for *how* extraterrestrial artefact detection might best be undertaken in our Solar System, for decades ahead. His uniquely twinned astronomy and lunar geology careers have challenged me on all key fronts, he was first to recognise Arkhipov as the pioneer in this field, and he accepted with wisdom and healthy scepticism my testable assumptions that programmed matter might be the real prize in artefact detection. His introductions to Japan's Institute for Space and Astronautical Science (ISAS) at the Japan Aerospace Exploration Agency JAXA were invaluable for the experimental phases of the work, especially while constrained by Covid-19 travel limitations.

I must thank Ian and Birkbeck College/ UCL both for allowing breaks in studies in the course of this research. Emergency eye surgery and subsequent successive operations left me in the end with better-than-ever vision and a multifocal technological artefact of my own implanted in my eye, but the time outs for recovery and the day job at Airbus Ventures were essential to this project's completion. Birkbeck College enjoys a deserved reputation as a pioneer in the working person's aspirations to academic learning and achievement, and it's a privilege to be part of this tradition to this day.

By dint of my early (teenage!) interest in Japan's potential for outer space development, I studied Japanese while an undergraduate at MIT, and enjoyed the chance to work for an engineering company in Kawasaki straight after, mining the seabed in anticipation of future lunar and

asteroid mining; little did I know that decades later those hard-earned mineralogical and language skills would serve me well as a collaborating researcher with JAXA's ISAS team. Professor Haruyama Junichi, kindly introduced by Ian, stimulated great insights into both how Machine Learning can be applied to the discovery of pit craters on the Moon, and to how they might provide a better preserved environment for extralunar artefacts. Professor Haruyama in turn introduced me to the Tanpopo team, where I was warmly welcomed by Professors Yamagishi Akihiko and Mita Hajime.

Their introduction in turn to Professor Yano Hajime was instrumental to the development and execution of my experimental work in the laboratory. I owe deep thanks to Hajime for his tireless collaboration in adapting his ingenious CLOXS system and for our hours together in the clean rooms. I am confident that his inventiveness and collaborative spirit will create many new opportunities for spacecraft exploration and sample collection and analysis, and I hope I may continue to be part of them even with the completion of this project. Professor Okudaira Kyoko and her University of Aizu students Honoka Kiryu and Matsuo Yuhi were excellent companions and collaborators on the Machine Learning side of this project.

By happy coincidence Professors Crawford and Yano had both worked with Dr Penny Wozniakiewicz at the University of Kent; she and her colleague Matthias van Ginneken introduced me to the adventures of micrometeorite collection on Earth, and importantly helped kick off the planning of experiments to use Kent's two-stage hypervelocity gas gun to test spacecraft hardware particle impacts in mono-crystalline aluminium and regolith simulant. This has provided valuable inspiration for experiments begun here training my Machine Learning model on Earth spacecraft particles atop lunar regolith simulants.

As they learned of my Machine Learning modelling work and interest in distinguishing extraterrestrial microstructures, Professor Yano and this team also invited me to participate in exciting work analysing samples of the Ryugu asteroid returned by the Hayabusa2 mission. The Model's ability to find correlations between Ryugu and micrometeorites of the TransAntarctic Mountains collection, potentially paving the way for a recalibration of Earth's wider meteoritic classifications and assumptions about origins of the Solar Nebula, provided important demonstration in this project of future pathways for potential spacecraft-borne analyses. My great thanks to Matthew Genge of Imperial College and Luigi Folco of Università di Pisa for their support in these efforts.

I wish to thank all my colleagues at Airbus Ventures, a sponsor of this project, particularly my Managing Partner Thomas d'Halluin, who has been generous with both time and encouragement. Our own Limited Partner investors have also been supportive, particularly the Aerospace team at Development Bank of Japan. Dozens of our portfolio company teams have helped inspire and inform this project, and several may be the first to carry experiments based on this work to the Moon and the wider realms of the Solar System. Whether it's Catalog DNA, compacting advanced computing power of such high density to the micrometre scale that near future space missions can carry on-board processing power equivalent to even the largest systems on Earth today – and perhaps one day even see powerful Artificial Intelligence contained in star-journeying dust particles; or ispace bringing networked rovers and miners ready to sift and scan the lunar surface carrying Machine Learning scanners based on the developments of this project, all these innovative startups and entrepreneurs have had their own role in creating artefacts I'd be proud for any future civilisation to discover us by.

I owe particular thanks to startup companies where I've served as board director, including satellite communications pioneer Infostellar; telerobotics developer Telexistence; Q-CTRL using its fine quantum sensing capabilities to ready high resolution surveys of the Moon like never before; and Atlas AI, whose platform for applying Machine Learning to satellite imagery combined with other data sets is providing extraordinary predictive capabilities for many of our planet's greatest challenges. For computer vision and Machine Learning, I've had patient teachers: I especially appreciate Professor Stefano Ermon's allowing me to audit his Stanford University, Machine Learning, CNN for Visual Recognition, and Deep Generative Models courses.

I also thank startups CesiumAstro for providing spacecraft hardware for me to destroy and particulate for this project's experiments, and Manna Electric for providing additional lunar simulant.

Startup company energies and prize competitions characteristic of Silicon Valley have their place in the Search for Extraterrestrial Artefacts, but steady patience and careful data handling have been cornerstones of this project, for all the benefits of speed and efficiency that Computer Vision and Machine Learning afford; I learned these skills both as a meteoriticist and observer of countless Martian images while at the University of Hawai'i Manoa, under the patient guidance and sharp-eyed instruction of the late and dearly missed Dr Ed Scott, and Professor Pete Mouginis-Mark. There have been naysayers, actively opposing this project from the outset – you know who you are – but I thank you too, for

helping ensure that I kept tight academic discipline in building a foundation that I hope many others will choose to brave and enjoy in future.

I wish to thank all the members of the UK SETI Research Network, who welcomed me and heard me out as the black SETA sheep of the family, including Lewis Dartnell, Anders Sandberg, Stephen Baxter and of course Ian Crawford once again. I'm particularly grateful to Charles Cockell for including my chapters in his seminal *Extraterrestrial Liberty* series: while he's not a likely champion of the world government I might advocate to tackle our planetary challenges, if we do meet these challenges it will be thanks in part to the vigorous forum Charles created.

There are many others where the encounters may have been brief, a conversation over one of my posters or Q&A sessions after one of my APEX sessions at Birkbeck. Two in particular stand out – Mark Robinson, head of the Lunar Reconnaissance Orbiter Camera project, welcoming my interest in automating searches of LROC images, while at a 2015 conference in lovely Frascati, Italy; and the team from ESA's solar system exploration programme, suggesting that I should put in a proposal to ESA to make space dust probes a reality – the best way I can thank you is to say I'm working on it!

To all those who knew and loved Carl Sagan and helped make him the inspiration he was to me and millions of others, my special lasting thanks; the trajectory of my life would not have been the same without him.

At MIT it was my privilege to work as Dr Harold E ('Doc') Edgerton's lab assistant, do my summer work at his companies, and act as his general aide-de-camp, diver and equipment handler on unforgettable marine archaeological adventures. From the strobe lamp to side-scanning sonar, radar advances, and high-speed photography, filming the first atomic explosions and working as Jacques-Yves Cousteau's underwater cameraman, Doc was a pioneering inventor who inspired me to never stop imagining new adventures and discovering the means to undertake them. Doc introduced me to Arthur C Clarke, then running a diving school in Colombo; though ours would be a relationship of correspondence, Sir Arthur was also unstinting in his lifelong energy and support for new discovery, taking up my invitation to join a Caltech summer programme on Mars exploration by video in his 80s. To Doc and Sir Arthur and their loved ones, my many thanks for their many inspirations.

To everyone who read my first (business!) book *Consulting Demons* through to the end you knew I would wind up here – thanks for the faith.

To the teams at iDu Optics, Turi Create, Apple, ThorLabs, and many other patient suppliers and advisers – thanks for helping me equip and tinker together my own bat cave.

For all five of the brothers and sisters who raised me and set me on stranger paths, I of course so owe much to you – one need not look further than my brother David's *Providence Blue* for evidence – and Steve those days and nights helping you grind your own reflecting telescope mirror together and later watching with reverence for all the mysteries of the night sky that you brought to life for me, my great thanks indeed. Mogs, a Princess of Mars, is amongst the wisest here on Earth.

To Sasha, and Julien and Ariane - and Alfie and Mizu - much of our time together was sacrificed in making this project possible, my thanks for the love and generosity you shared throughout.

Only from the stars could it have come, and no thing of chance was it. It was a creation of artifice and mind . . . a child of intelligences, remote and unguessable, working corporally in metals, it indubitably was. He stared at it in amaze, his brain a racing wild-fire of hypotheses to account for this far-journeyer who had adventured the night of space, threaded the stars, and now rose before him.

The Red One, Jack London, 1914

. . . I had assumed without question that this crystalline apparition had been built by some race belonging to the Moon's remote past, but suddenly, and with overwhelming force, the belief came to me that it was as alien to the Moon as I myself.

The Sentinel, Arthur C. Clarke, 1951

. . . optics at the tether-tip automatically magnified a glitter, just ahead . . . probably some piece of space junk, left here by an earlier, wastrel generation . . . By now, Gerald knew how all kinds of normal junk reflected sunlight – from archaic launch vehicles and satellites to lost gloves and tools – each playing peekaboo tricks of shadow. But this thing . . .

Existence, David Brin, 2012

And it's worth looking for artefacts within our solar system; maybe we can rule out visits by human-scale aliens, but if extraterrestrial civilisation had mastered nanotechnology and transferred its intelligence to machines, the 'invasion' might consist of a swarm of microscopic probes that could have evaded notice. It's even worth keeping an eye open for especially shiny or oddly shaped objects lurking amongst the asteroids.

On the Future, Martin Rees, 2018

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 2

Chapter 3

Chapter 1

Introduction

1.1 Overview and Background

This Chapter highlights my motivation for developing the means to detect what are perhaps the least glamorous but potentially most numerous supply of extraterrestrial artefacts in the Solar System, the particulate waste of others' technologies, and the more exciting if less likely prospect of detecting probes and devices that may have been intentionally programmed to develop from particulate matter, taking advantage of natural interstellar transport mechanisms to proliferate throughout the Galaxy. I describe in this Chapter the background to the research problem and the programme I've undertaken, my thesis aims and how I set out to solve them.

In my work as a Partner at Airbus Ventures, it's been a privilege over these last nine years to have invested in dozens of spacefaring startups around the world, many readying and now deploying advanced technologies and capabilities beyond Earth orbit. As diverse as suites of autonomous rovers to the Moon, novel plasma fusion propulsion systems, and quantum sensors for lunar orbital assays, here on Earth this portfolio of startups is also packing intelligence into smaller and smaller bits of matter – in one case not only storing massive data amounts in synthetic DNA, but using enzymes that synthesise DNA without a DNA template. Developments demonstrate the molecule's extreme information density, low-energy requirements and ability to be preserved and remain accessible for at least millennia in silica particles for later information, read, write and compute functions suggests greater intelligence in smaller matter seems a reasonable expectation [1].

Yet other startups are sweeping across vast areas of our planet with heretofore unseen artificial intelligence capabilities, using fleets of AI- driven drones to plant millions of tree seedlings [2], and in yet another case, deploying Machine Learning applications to transform Earth Observation data sets into practical predictive economic capabilities for crops, infrastructure, electrification and more [3]. In this work assessing and funding promising entrepreneurs, I have become increasingly optimistic that tapping into the resources of the Earth-Moon system in particular – be it from clean energy derived from lunar materials, rare Earth elements found on the Moon, platinum group minerals extracted from metallic

asteroids (which may have also crashed onto the Moon) and shipped to Earth, or the expansion of off-world manufacturing likely to be pioneered in the Earth-Moon system – we have a chance to douse the fires beginning to consume our planet.

And even if we don't douse the fires and do not survive as a technological species, we are already leaving our trash behind [4, 5]. One portfolio startup is already observing hundreds of thousands of particles of debris in Earth orbit at 1 - 2 cm resolution [6], and there are undoubtedly millions more at smaller, sub-micrometre levels; soon sensing and processing capabilities will extend to objects throughout the Earth-Moon System, anticipating more trash to follow its exploration and industrial development. I can't help but wonder what a future alien archaeologist would make of our technological remnants. Trained as a meteoriticist before beginning my work with Airbus, I knew that powerful tools already at our disposal could likely reveal the products of technological artefacts in extraterrestrial materials at the sub-micrometre level.

Discovering non-terrestrial artefacts on the Moon is a vision at least as old as Sir Arthur C. Clarke's *Sentinel* and *2001: A Space Odyssey* vivid imaginings [7, 8]; see Figure 1.1. In 2012, I was captivated by David Brin's novel *Existence* [9], in which alien AIs reside in small crystals, ready to lure unwitting species into uploaded existences, the better to save galactic resources for themselves, and further reduce the rise and spread of potential inimical species, a theme well developed in the *Remembrance of Earth's Past* trilogy [10, 11, 12].

Could we, and possibly many others, I also can't help but speculate, indeed pack artificial intelligences into small bits of matter, perhaps with the ability to use local materials, to grow, DNA-like, into complex exploratory probes?

In the present day, this project sets out to demonstrate that we have the tools to constrain the actual possibilities.

Figure 1.1: Humans encounter a Monolith artefact on the Moon in 2001:A Space Odyssey. Creator: Movie Poster Image Art. Credit: Getty Images.

1.2. From Arkhipov Particles to Arkhipov-Bracewell Probes

An early discussion with Professor Ian Crawford, directly leading to his becoming my research supervisor for this project, explosively illuminated and connected all these threads, by introducing me to the papers of Dr Alexsey Arkhipov, a Ukrainian Search for Extraterrestrial Intelligence (SETI) specialist turned Search for Extraterrestrial Artefacts (SETA) advocate beginning in the 1980s. Through my previous meteoritics studies and research I was familiar with the principles of the interstellar transport of dust grains,[1] and in Arkhipov I discovered these principles applied to the problems of manufactured dust – their transport, survivability, and possible quantities intersecting natural collecting plates like the Moon [16, 17, 18, 19, 20]; see Figures 1.2 and 1.3.

Figure 1.2: Interstellar Dust Grains is an illustration by Mark Garlick/ Science Photo Library which was uploaded on November 4th, 2019.[1]

[1] See Sterken et al. [13, 14] and Totani [15] for a current thorough overview on estimates of the flux of Interstellar Dust Particles.

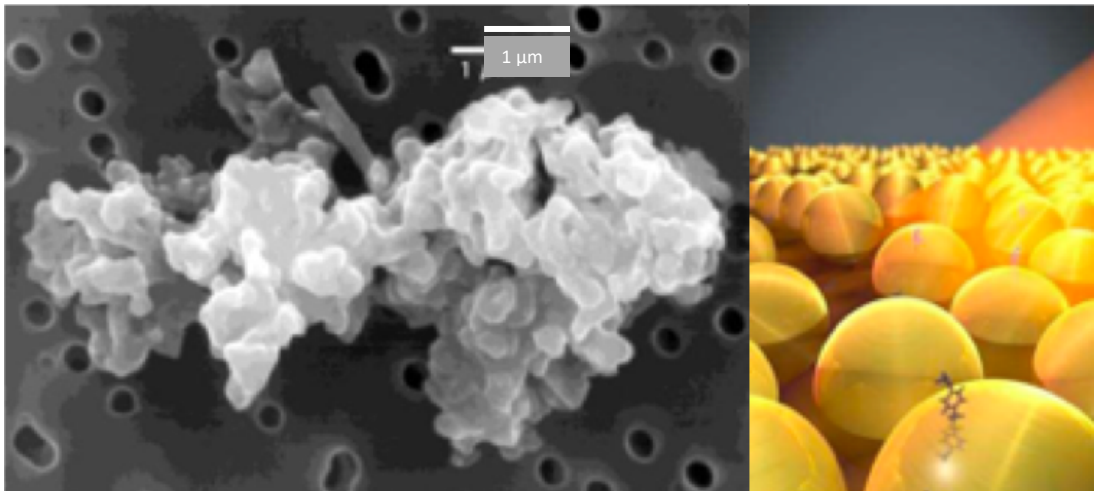


Figure 1.3: Left: Porous chondrite interplanetary dust particle. Right: London Nanotechnology Centre researchers have 'developed a system to quickly detect trace amounts of chemicals like pollutants, explosives or illegal drugs. The new system can pick out a single target molecule from 10,000 trillion water molecules within milliseconds, by trapping it on a self-assembling single layer of gold nanoparticles.' Credits: Left, Donald E. Brownlee/Elemar Jessberger Wikipedia Commons [21]; Right, Benjamin Miles and Nanowerk Newsletter.[2]

This I saw was an exciting key to significantly constraining the presence and quantity of other technological civilisations in our Galaxy. Since the time of Arkhipov's work, exoplanets have been discovered to be a commonplace in the Galaxy, and ancient planets with conditions potentially amenable to life now add to the taxonomy, multiplying the plausibility that any technological remnants from the Milky Way's first few Gigayears should by now have passed our way. More enticingly, advances in Artificial Intelligence, Computer Vision and Machine Learning, as well as nanotechnologies themselves, now suggest ways and means of detection Arkhipov would not have had at his disposal.

[2]https://www.imperial.ac.uk/news/116589/researchers-improve-technology-detect-hazardous-chemicals

In Chapter 4 I explore these ideas further, as a fitting reflection on the spacecraft object detector I've successfully trained on Earth spacecraft hardware on the Moon. I understand that Dr Alexsey Arkhipov is now retired in Kharkiv Ukraine (F. Graham, personal communication, November 16. 2021) and still chooses to remain there amongst all the strife, but I hope I may look forward to one day soon sharing the results of this project with him: based on this project work I firmly believe we have in our hands the means to deeply constrain the uniqueness of our existence in the Galaxy, right here, right now – or at least on the Moon, beginning as soon as the late 2020s with the first Artemis landings.[3] In Chapter 5, I highlight some of my recent proposal work for bringing particle detectors to the Moon's surface.

[3] Objectives highlighted in the Artemis III Science Definition Report for astronaut-emplaced experiments offer key opportunities for constraining e.g. anthropogenic contaminants and non-lunar extraterrestrial materials on the Moon; see https://www. nasa.gov/wp-content/uploads/2015/01/artemis-iii-science-definition-report-12042020c.pdf

1.3. Trends and Accelerating Progress in the Field of SETA

Beginning in April 2014, together with Professor Crawford I began building out a frame of work that would focus on detecting Arkhipov Particles[4] as well as Arkhipov-Bracewell Probes.[5] For the former, we began considering a variety of sorting and beneficiation processes that might differentiate Arkhipov particles from lunar regolith materials; we also began to consider what such material might look like, given advances in nanotechnologies on Earth. Somewhat elegantly, I began to understand, Earth nanotechnologies themselves, designed for detection of impurities, contaminants, explosives etc., could be the right answer for detecting extraterrestrial nanotechnologies and other advanced materials. For the latter, we recognised that a good start would be the work of UCL Centre for Planetary Sciences colleagues Katherine Joy [23] and Roberto Bugiolacchi [24], who had developed in Moon Zoo[6] a crowd-sourced tool for distinguishing features on the lunar surface, including spacecraft hardware, based on Lunar Reconnaissance Orbiter Camera (LROC) images [25].

---

[4] As I dub them and define them, an Arkhipov Particle is a fragment of extraterrestrial technology, from angstroms to a micrometre in length, that can be borne through the Galaxy by radiation pressure, stellar winds and other natural phenomena, from comets to incorporation into meteorites, much like naturally occurring interstellar dust particles.

[5] Arkhipov-Bracewell Probe is a double-barrel denomination. Arkhipov did not conceive of his Particles, as I've introduced, as being pre-programmed to become natural Galactic travellers that could grow into Probes upon contact with resources; Bracewell [22] identified Probes as a superior civilisation's preferred means of exploration and communication, over radio technologies, but though he considered that they might shed debris or crash into natural bodies, Bracewell did not envisage them as designed for travel at microscopic scales.

[6] https://www.moonzoo.org/

I saw that the image-recognition software being actively developed amongst startups in Silicon Valley and elsewhere, particularly for autonomous driving systems and facial recognition, might helpfully incorporate human crowd-sorted Moon Zoo into their data training models. Davies and Wagner [26] had also identified LROC images as a potential data base for automated searching for alien artefacts, and in May 2015, when I presented a poster on initial results with primitive model runs using face recognition software at the 3rd European Lunar Symposium in Frascati (Figure 1.4), I also had the opportunity to discuss the evolving work in person with Dr Mark Robinson, leading the LROC mission out of Arizona State University. Mark relayed (M. Robinson, personal communication, May 14, 2015) that his team was still at the stage where he had graduate students searching with magnifying glasses on enlarged LROC images, for two as yet unidentified Apollo Lunar Module ascent stage crash sites, and that he would welcome collaboration on any automated approaches.

Figure 1.4: An illustration of my early work on Object Recognition for spacecraft hardware, from a poster I presented at the 3rd European Lunar Symposium in Frascati, Advancing a Search for Non-Terrestrial Artefacts on the Moon. [27]

Three important trends then began to converge around this Project work: 1) a major shift in the general interest and funding for SETI; 2) the first practical and methodical use of automated search systems for interplanetary and interstellar dust particles; and 3) improvements in Machine Learning that in this Project are already shifting the directions of the first two trends.

The shift for SETI came with the Breakthrough Initiatives[7] under venture capitalist Yuri Milner's sponsorship, injecting new life and significant funding into SETI research [28, 29] with additional provisions for "Messaging to Extraterrestrial Intelligences," as well as a small lightsail project to Alpha Centauri, Breakthrough Starshot [30], thereby expanding SETI beyond the bounds of radio astronomy. By 2020, James Benford, a specialist in microwave source physics and electromagnetic power-beaming for space propulsion and a project leader for Breakthrough Starshot offered the candid observation that SETI signal-detection results had not been forthcoming (J. Benford, personal communications, June and July, 2020), and that it was time for the SETI community to pivot to SETA, the Search for Extraterrestrial Artefacts, as coined by Freitas *et al.* [31].

In my communications with Dr Benford, I shared my work and ideas around Arkhipov and Machine Learning using Apollo landing sites as a training base. Professor Crawford and I were then invited to a formal virtual workshop organised by Vishal Gajjar, which included a number of Breakthrough-sponsored participants, entitled 'SETI in the Solar System' (password SETA)[8], to help explore

[7] https://breakthroughinitiatives.org/

[8] SETI in the Solar System, a virtual workshop convened by Prof. Vishal Gajjar, University of California at Berkeley, 28 July 2020

and facilitate the extension of SETI work to SETA. Notably in the workshop a pre-print paper was introduced featuring work on lunar anomaly detection using Machine Learning trained on Apollo Landing site data sets by Lesnikowski *et al.* [32] – it features a distinctly different (and less scalable) approach than the methods introduced in this Thesis (Chapter 4), but nonetheless I could not help but feel the race was on!

Meantime, the Breakthrough initiatives have had another important influence in the shift from SETI to SETA. While to my knowledge no other researcher is practically pursuing ideas and methods for detecting Arkhipov Particles, Tomonori Totani and others have considered 1μm grains arriving on Earth from exoplanets via natural processes [33]. There is considerable growing focus on the detection of small probes of about a 100g mass – notably detecting the very type of lightsail probe that Breakthrough proposes to launch – a bias that may overlook the wider and deeper-in-time opportunities for Arkhipov Particle detection, but one that can only add to the possibilities for Arkhipov-Bracewell Probe detection.

The second main of area of convergent developments for this Thesis has been in the application of automatic detection systems, and most recently, Computer Vision and Machine Learning, for identifying and classifying interplanetary and interstellar dust particles from sample return missions. As examined in Chapter 2, these are the first practical applications of Computer Vision with supporting algorithmic systems applied at the microscopic scale in the space exploration

sciences, and the most advanced work, conducted in Japan under the auspices of JAXA and the Institute for Space and Astronautical Sciences, formed the base for the experiments and advances in methods undertaken in this Thesis and detailed in Chapter 3. This is the first approach well suited for adaptation to Arkhipov Particle detection on future missions and *in situ* sampling of the Moon and asteroids. The resulting core Machine Learning models also directly inform the experiments detailed in Chapters 3 and 4.

Finally the pace of developments in Machine Learning itself have far outstripped any other developments in this field. As described in Chapter 2 the advances and range of applications in Machine Learning has exponentially expanded with parallel developments in computing speed and capacity, just in the years since this project began. Fortunately I have been able to adapt some of its most powerful capabilities into the heart of this Project, pursuing my hypotheses that Arkhipov Particle and Arkhipov-Bracewell Probe detection are best pursued by designing systems for autonomous operations in small spacecraft.

I fully expect that all of these investigations will be fruitful, combining new-found emphasis on SETA, advances in collection and examination of microscopic particles in space, and the inexorably advancing powers of AI Machine Learning. The Moon seems a particular ripe opportunity for investigation. As Professor Crawford noted to James Benford in 2020, per Basilevsky if the Moon's rocks survive on the order of 100s of Ma (I. A. Crawford, personal communication, July 2020; see Figure 1.5. and [34]) then spacecraft hardware made of presumably sturdier stuff should still be detectable.

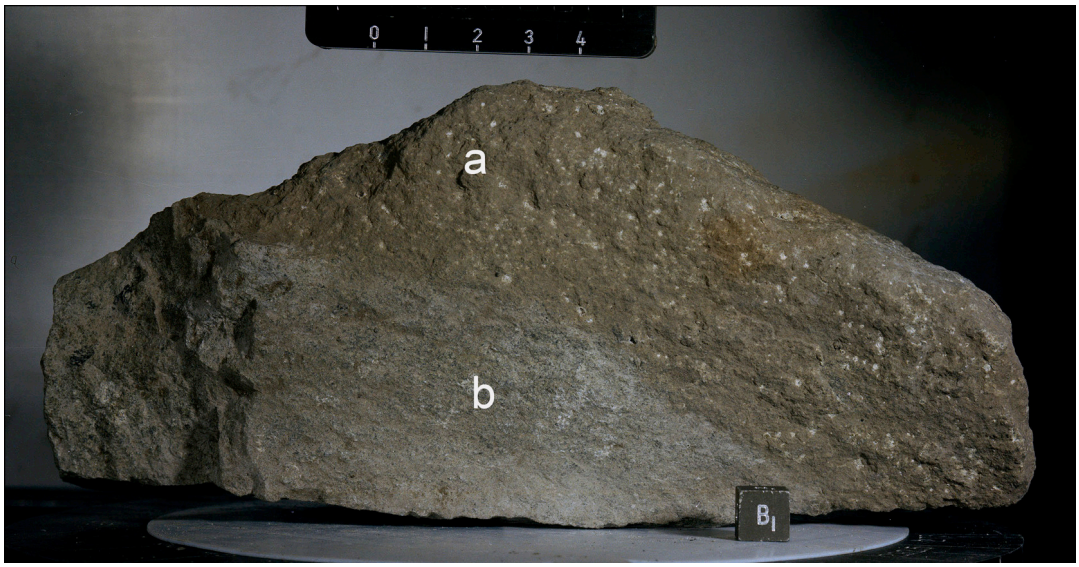


Figure 1.5: Apollo 16 sample 65015 showing two types of surfaces: (a) rounded and abraded surface exposed to space, covered with brownish patina and pitted with micrometeorites (b) fresh and very jagged fracture surface of the sample part that was buried in the soil.' Scale cube is 1cm x 1cm x 1cm. From Basilevsky *et al.* [34]

1.4. Expanding the Drake Equation and Constraining Fermi's Paradox

Recently, James Benford has been focused on the thousands of stars that have passed within a light year of Earth, advocating for a survey of the Moon for alien artefacts. To quote from Benford [35]

> Stars come very close to our solar system frequently. About two stars per million years come within a light year. An extraterrestrial civilisation that passes nearby can see there is an ecosystem here, due to the out-of-equilibrium atmosphere. They could send interstellar probes to investigate. We estimate how many probes could have come here from passing stars. And where would they be now? The Moon and the Earth Trojans have the greatest probability of success. Close inspection of bodies in these regions, which may hold primordial remnants of our early solar system, yields concrete astronomical research. This argues for a Search for Extraterrestrial Artifacts (SETA) strategy of exploring for alien artifacts near Earth.
>
> I propose a version of the Drake equation to include searching for alien artifacts, which may be located on the Moon, Earth Trojans, and Earth co-orbital objects. The virtue of searching for artifacts is their lingering endurance in space, long after they go dead. I compare a search for extraterrestrial artifacts (SETA) strategy with the existing listening to stars search for extraterrestrial intelligence (SETI) strategy. I construct a ratio of a SETA Drake equation for artifacts to the conventional Drake equation so that most terms cancel out. This ratio is a good way to debate the efficacy of SETI versus SETA.

As described below, one of the great attractions of the search for waste-material Arkhipov particles is that indeed discovery is not contingent on specific solutions to Fermi's 'where-are-all-the-aliens' Paradox [36], and at orders of magnitude greater possibilities for extraterrestrial presence than Benford [35] proposes. We can still find evidence of civilisations that existed much as we do, as technological species exploring off-planet possibilities, whether or not: other civilisations characteristically succumb to the planetary dangers unleashed by the very technological capabilities that enable their first spacefaring ventures; they have long since transcended existence as we know it; they aim to keep their own existence and whereabouts secret; choose to communicate using unknown advanced technologies; are surreptitiously observing us; or are simply too far or too few or too unmotivated to reach out and find us. By searching for debris generated by past civilisations, we may be able to falsifiably demonstrate that we are not the only intelligence this Galaxy has known, or at least significantly constrain the problem.

With the work undertaken in this Thesis we will have the tools to begin to constrain the 'Drake Equation.' Originally conceived as a means to calculate the number of presently radio-communicating civilisations [37], it has stimulated decades of debate which I do not treat here in the historical detail, a topic generously covered by others in the literature, with an especially useful review and exposition introduced by Vakoch & Dowd [38].

The artistic representation of the Drake Equation by Bagia[9] (see Figure 1.6) to me beautifully captures the depth, wonder and uncertainty embodied in the Equation itself. In this project I sense that the art form of the Equation is raised a notch, through Machine Learning.

Figure 1.6: The Drake Equation: original art by Antonio Bagia created in 2013

While the 'Equation' (as I refer to it here throughout) has been amply criticised on its mathematical and predictive merits – see e.g. Denning [39] – as an heuristic, lending itself to loosely defined solutions by trial and error, it is a natural for Machine Learning, as further discussed in Chapter 4.

[9] https://www.saatchiart.com/art/Painting-The-Drake-Equation/972743/3469873/view

The original eponymous Drake Equation, proffered by the extraordinary SETI pioneer Frank Drake, runs as follows:

$$N = R \times f_p \times n_e \times f_l \times f_i \times f_c \times L, \qquad (1.4.1)$$

where N is the number of detectable civilizations in space, R denotes the rate of formation, f_p is the fraction of stars that form planets, n_e factors for the number of planets hospitable to life, f_l indicates the fraction of planets where life actually emerges, f_i is the fraction of planets where life evolves into intelligent beings, f_c is the fraction of planets with intelligent creatures capable of interstellar communication, and L is the length of time that such a civilisation remains detectable [40].

With new-found evidence for determining rates of star formation and the on-going discovery of exoplanets that appear to be as numerous as the stars in our Galaxy [41], useful approximations for more of the Drake Equation's terms' values can be determined. Arkhipov's work, as I demonstrate in this thesis, is the beginning of a key to further significantly constrain these parameters. To begin with, we needn't concern ourselves with the last two factors, (f_c) the wherewithal and presumed interest in interstellar communications, and (L) longevity since as a starting point we are only looking for trash, and similarly for (R), we needn't be concerned with rates of star formation, rather just however many stars there have ever been in our Galaxy, particularly in the first 3 to 4 Gyrs, allowing

ample opportunity for waste transport to our corner of the Milky Way[10].

Fermi drew attention to the possibly ample abilities of a single alien intelligence's simple, self-replicating probes to cover the whole of our Galaxy many times over in its cosmic lifetime; yet we have no sign of such coverage in our purview, though we can conceive of such coverage ourselves, begging the question oughtn't they already have found us? – Fermi's core conjecture. With the work in this Thesis we can shed new light on Fermi's Paradox – we can add to the reasoning and premises for how evidence of intelligent life may have been communicated to the Solar System – via Arkhipov Particles and Arkhipov-Bracewell Probes – and address its seeming contradiction, their apparent absence, by seeking them with more powerful tools, including Computer Vision, Machine Learning, and high-resolution imagery. Chapters 2 and 3 detail the development and application of these tools to microscopic particles in space, and Chapter 4 begins to address this further with practical searches for alien artefacts on the Moon. Contemporaneously, Bracewell introduced the idea of the practicality of probe deployment by 'superior' civilisations [22]. There is nothing paradoxical *per se* about this – our ability to conceive of probes does not necessarily entrain others' actions, and interactions could be consciously muted for many reasons – the notion of a paradox in fact can lead to reckless thinking: i.e. because we don't see something, it must be true it isn't 'there.' Paradoxical standing aside, Fermi nonetheless raises an important question: if it's so easy to permeate the Galaxy, where is the evidence of such permeation in our own neighbourhood?

[10] The SETI Institute, a not-for-profit organisation founded by Drake, has recently modified his Equation to constrain (N) to the Milky Galaxy, and expanded (f_c) to include civilisations that 'develop a technology that produces detectable signs of their existence;' making the Equation specifically relevant to technosignature detection in our Galaxy as well as to electromagnetic communications. See https://www.seti.org/drake-equation-index

1.5 SETI vs SETA – And the Merits of Searching for Arkhipov Particles and Arkhipov-Bracewell Probes

SETI is an express aim of the massive, multinational collaborative Square Kilometre Array (SKA) [11] antenna project (see Figure 1.7); alongside the Break-through Listen initiative, it represents a major, hopeful undertaking with as yet no measurable prospect of anticipated returns – a formula for impatience that is now opening the doors wider for SETA.

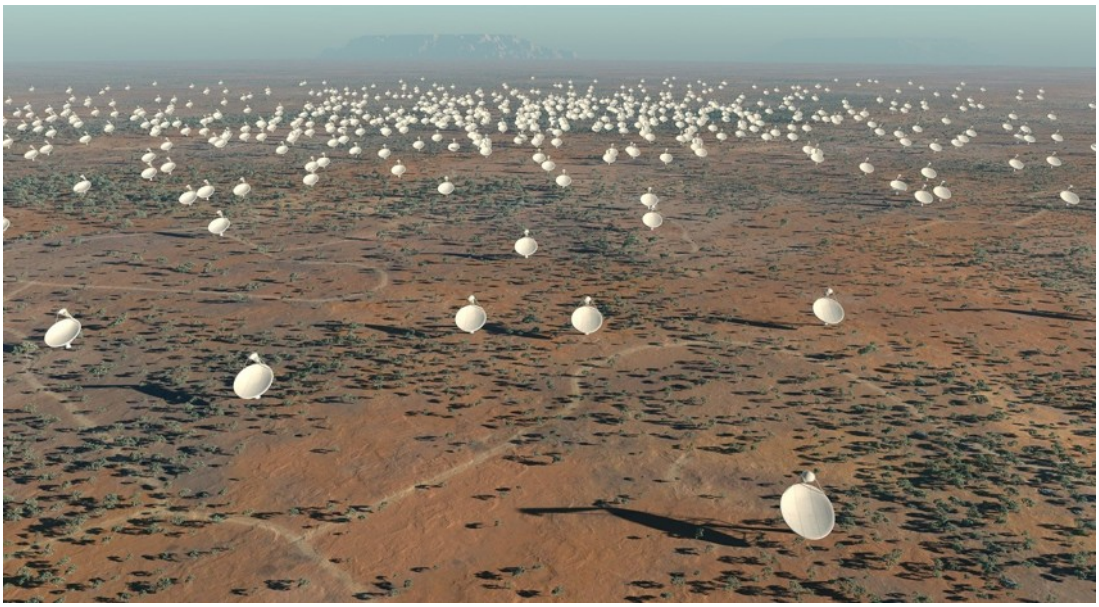


Figure 1.7: Artist's impression of the central core of Square Kilometre Array (SKA) deployed in Australia. Credit: SKA Project Development Office and Swinburne Astronomy Productions.

[11] SKA's 'Cradle of Life' objectives include: From Pebbles to Planets, observing planetary formation processes in detail beginning at the scale of micron-sized Inter Stellar Dust; Zooming in on the Solar System's Edge, locating the precursors of amino acids, ribonucleotides, sugars, and lipids to better understand how life on Earth formed and better predict the likelihood of it arising on other planets; Examining Exoplanets, specifically studying the signals detected from exoplanet systems for planetary magnetospheres and their interactions with their host stars; and Locating Intelligent Life: looking for indicators of technology as a proxy for the intelligence that produced them, whether messages being broadcast deliberately or just the electromagnetic noise that technology creates which can leak out into space.

.

Jason Wright, an accomplished specialist in the detection of exoplanets, and somewhat despairing of the usefulness of the Drake Equation going forward, published an article three years ago on Strategies and Advice for SETI [29] which helps frame and differentiate how the work I undertake here carves out new territory in the SETI and SETA domains, and offers the kind of practical immediacy he and others have advocated for SETA – see especially Freitas [31, 42]; Arkhipov [16, 17, 43, 44, 45]; Crawford [19, 20, 36, 46]; Davies & Wagner [15]; Loeb & Turner [48]; and Lingam & Loeb [49] – by my now using a combination of newly powerful Machine Learning software, hardware, and experimental access to dust and spacecraft hardware materials.

Wright [29] traces the early modern history of radio SETI: from Cocconi & Morrison [50] to Schwartz & Townes [51], the first SETI programmes Drake [52], and the first proposals for searching in our Solar System by Bracewell [22] and early suggestions for searching for technosignatures around other stars by Dyson [53]. Wright then identifies three main factors in a 'resurgence' in the field of SETI today:

> One is the discovery of exoplanets, and the determination that the n_e term in the Drake Equation (i.e. the average number of Earth-like planets per star)[12] is on the optimistic end of estimates, which significantly increases pessimists' estimates of the number of potential signals there exist to find.

[12] In strictly canonical usage, (N_e) is number of planets per star with environments suitable for life.

> Second is the Breakthrough Listen Initiative founded by Yuri Milner [30] and executed by UC Berkeley, which has greatly increased the amount searching done, the number of people trained in the field, the visibility of the field, and the opportunities for practitioners to collaborate and contribute to the effort.
>
> Third has been a broadening of the focus of the field in the 2010's beyond the laser and radio projects that could find support in the limited funding environment of the '80's, '90's, and naughties [*sic*], inspired by the flourishing of the field of astrobiology under NASA's aegis. Tarter et al. [54] argued, rightly, that SETI belonged under the same astrobiology 'umbrella' as other ways to search for life, and coined the term 'technosignatures' by analogy to 'biosignatures' to emphasize the parallel and complementary approaches of the fields.

Wright [29] then goes on to highlight 'three complementary fronts' for progress in SETI, namely theory, instrumentation, and observation. Based on his extensive experience in the field on each front, he offers 'personal observations and recommendations for which sorts of projects are ripe for new work (and which are not), and how to practice SETI generally.'

I believe readers of Wright's survey of the field and his tracing of the 1) resurgent shift to more optimistic estimates of potential signals, 2) wider participation, and 3) broadening of focus to include technosignatures to be a

succinct yet comprehensive accounting of the state of play today.[13] But by my coming from another line of experience outside astronomy – in planetary geophysics, investments in spacecraft hardware development, and use of Machine Learning – I find several assertions that bear challenging, stimulate new opportunities for exploration, or point to new gaps for further exploration – much as I believe Wright in fact intends.

Wright [29] concurs with Denning [39] that the 'now-elaborate and extensive discourse concerning the Fermi Paradox [is] quite literally, a substantial body of analysis about nothing, which is now evolving into metaanalysis of nothing'[14] and Wright then observes that 'as with work on the Fermi Paradox, I suspect that further work on the Drake Equation will probably not provide significant new insight into how to search for technosignatures.' While admiring its utility as a guide to the problem to date, Wright concludes that 'in the end, the Drake Equation's terms are fundamentally too uncertain and its assumptions too narrow for it to truly estimate the number of technological species in the Galaxy.'

Here in this thesis I aim to demonstrate that the door should in fact not be closed on further work on the Fermi Paradox and the Drake Equation. As discussed in Chapter 4, and based on the experimental findings of this project, I believe

[13] For those with appetite for more exhaustive coverage, the doorstopper of the tome *Life in the Cosmos* [55] lacks for neither breadth nor detail, and offers a remarkably accessible and overdue integration of bio- and techno-signature detection prospects, well informed by developments in evolutionary biology, human-engineered interstellar travel prospects, and a wide array of possible detection strategies.

[14] Crawford and Schulze-Makuch's substantive advance on the Paradox – 'Zoo Hypothesis or Nothing' [47] is notable evidence in counterpoint.

that both 1) new data sets encompassing the search for Arkhipov Particles and Arkhipov-Bracewell Probes in our Solar System, and 2) a new, Machine-Learning, task-oriented advance beyond the Drake Equation that is able to improve task performance as it experiences massive new data sets, together comprise a new and powerful means to ultimately constrain an answer to the Drake Equation's most stubborn uncertainty, namely how often in our Galaxy spacefaring technologies have arisen from earliest life – and moreover, that through the methods introduced by this project that this uncertainty can be usefully constrained before we ourselves ever leave our Solar System. And as I describe in Chapter 4, this new data-driven and Machine-Learning driven set of constraints on the probabilities can also help suggest the direction of resolution(s) of the Fermi Paradox, and indeed lend itself well to not only 'significant new insights on how to search for technosignatures,' but possibly where to search.

	Spatial Scale / Distance			
	Solar System	Exoplanetary	Circumstellar	Extragalactic
Structures	•Free floating •Surface / subsurface •Lights / collectors	•Satellite belts •Lagrange point megastructures •Surface features	•In transit •Directly imaged	•Direct energy propulsion
Environmental changes	•Martian ice cores •Venusian atmosphere •Ancient Earth	•Atmospheric spectroscopy •Surface changes •Commonalities among exoplanets	•Stellar atmosphere pollution •Circumstellar gases •Stellar motions	•Stellar population management •Stellar motions
Excess Heat	•Exothermic "asteroids" •Heat islands	•Rotationally modulated surface hotspots	•Dyson spheres	•Dyson spheres

Figure 1.8: From Wright (2021): 'Some Potential Technologies and Artefact TechnoSignatures Grouped According to Their Scale and Their Kind.' This Thesis as indicated in red focuses on detection of structures – from nanometres to a metre – that can be detected on the surface or in the subsurface materials of natural bodies in our Solar System

These are extraordinary claims for a SETA approach that treats only one corner of the overall opportunities for SETI. Indeed as laid out by Wright [29] my experimental work in this project begins addressing only but one part of one corner, micro particles and their impact evidence and macro structures on solar system bodies (Figure 1.8).

Zooming in further, Wright [29] builds from developments led by Sofia Sheikh at a NASA workshop (TechnoSignatures Workshop Participants 2018), which is more specific to the search for Solar System artefacts, and useful here for comparing the merits of the methods introduced in this project for detecting Arkhipov particles and Arkhipov-Bracewell probes. Sheikh [56, 57] formalises the axes of merit for technosignatures in nine categories. I introduce and briefly summarise each here, and compare and contrast her assessment of the merits of searching for Solar System artefacts with the merits of the search for Arkhipov Particles and Arkhipov-Bracewell probes as introduced in this Thesis (See Figure 1.9). Using her nine-axes analyses, Sheikh builds a credible case for the comparative benefits of the search for Solar System artefacts over searching for radio/optical technosignatures and waste heat from megastructures – cases which I touch on only lightly in this work and will allow others to judge; suffice it to say that there seems to be a widening acceptance of, and arguably a continuing convergence to, the merits of Solar System artefact searches over those for radio, optical, and megastructure waste heat.

My focus in this work however is on the relative merits of searching for Arkhipov particles and Arkhipov-Bracewell probes as artefacts found on the surface and subsurface of terrestrial planets and airless bodies in our Solar System. With Sheik's Nine Axes of Merit as a reference I created the panel on the right-hand side of Figure 1.9 with an assessment of the relative Merits of a Search for Arkhipov Particles and Arkhipov-Bracewell Probes, placing yellow circles representing Arkhipov Particles and red circles representing Arkhipov-Bracewell Probes along the 9 Axes.

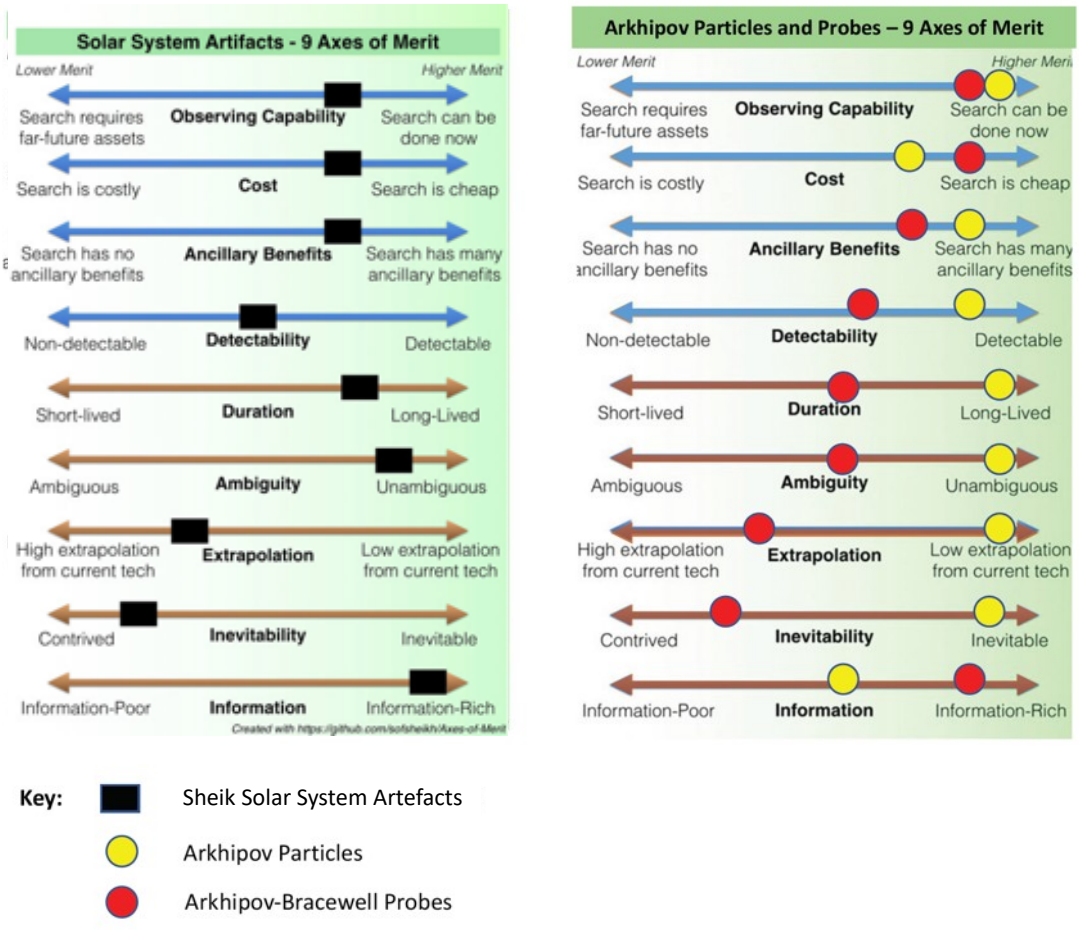


Figure 1.9: A comparison of the relative merits of the Search for Arkhipov Particles and Arkhipov-Bracewell Probes: in the left-hand panel, Sheik's original assessment of the 9 Axes of Merit for searching for Solar System Artefacts (Lower Merit is to left of the sliding scales, Higher Merit to the right) [56, 57]. I have created the right-hand panel, adapting Sheikh's methodology to draw comparison with the relative Merits of Searching for Arkhipov Particles and Arkhipov-Bracewell Probes.

I explain the rationale for my placements of the yellow circles representing Arkhipov Particles and the red circles representing Arkhipov-Bracewell Probes along the 9 Axes as follows:

1. ***Observational Capability*** is defined by Sheikh as the 'technological ability of astronomy as a whole at the time a search for the technosignature is proposed.' She refers to the difficulties of developing and deploying new technologies with the efficiencies and thoroughness required to match the goals of the task, if not the actual extraterrestrial technology itself, per Klein and Gulkis [58]. My experience and biases as a venture capitalist investor in aerospace technologies inclines me to lead the curve of technological development, and anticipate fast-accelerating advances in detection and deployment technologies, particularly with respect to the newer but increasingly self-learning technologies applicable from Machine Learning developments. So based on experience, I feel compelled to give higher scores to Sheikh's assessment of our observing capabilities for APs and A-BPs.

2. As Sheikh notes, following Davies and Wagner [26], the ***Cost*** of Searching for solar system artefacts generally is relatively low because, in many cases, it relies on existing instrumentation and resources; the more so for Arkhipov Particles and Probes, I contend here, with the amplifying power of Machine Learning applied to existing instrumentation and sensors.

3. The ***Ancillary Benefits*** of an artefact search is less a focus for the purposes of this Project – this is not a pitch for funding nor an attempted rationalisation for what I sense to be an urgent and purposeful priority: the timely discovery of our comparative fragility, potential new sources of knowledge, or contact in some form with other intelligences. Nonetheless I recognise that other synergies, such as the reliable identification of non-artefact exotic materials such as micrometeorites in the lunar regolith can offer significant additional scientific benefits. In the later chapters of this thesis I explore how the benefits of the work in this project extend to: imminent robotic and human activities on the Moon and their prioritisation of Computer Vison/ Machine Learning capabilities to identify and quantify microparticle terrestrial contaminants, lunar dust disturbances, the flux of Interplanetary and Interstellar Dust Particles, *ß*-meteoroids, and secondary impact ejecta. Missions to other celestial bodies also prioritise not only large-scale analyses of sample return microstructures and their correlates found on Earth, but abilities to conduct and report microscopic analyses *in situ* e.g. on asteroid surfaces. Therefore I have placed the marker for Arkhipov Particles significantly to the right of the Merit scale, since searching for them can bring so many Ancillary Benefits to other planetary science studies of microscopic phenomena. The marker for searching for

Arkhipov-Bracewell Probes I have placed as having similar Ancillary Benefits to Sheikh's, with synergies to both existing missions and research.

4. ***Detectability*** of Solar System artefacts is the very essence of this Thesis's strengths, and is thus placed firmly to the right amongst the axes in both cases.

5. Similarly ***Duration*** of Arkhipov Particles and Arkhipov-Bracewell Probes again brings us to the right of the axes, much as Sheik recognises the longer-lived nature of physical artefacts generally.

6. ***Ambiguity*** is still a factor in Solar System Artefacts with Arkhipov-type origins. But even if discovered materials overlap with existing Earth technologies (e.g. traces of heat resistant tungsten alloys, a logical choice for spacecraft materials) isotopic analyses should still be able to determine their age. Thus I have scored discovered to-hand Arkhipov Particles high on the Axis of Merit for Ambiguity, while Arkhipov-Bracewell Probes, which may first be revealed by more remote processes such as Computer Vision and Machine Learning techniques as applied in this Thesis to images of the Moon, may require further examination and future *in situ* analyses to distinguish them from Earth spacecraft hardware and determine their unambiguous extraterrestrial origin.

7. As for ***Extrapolation***, the fact that we can conceive of constructing our own Arkhipov-Bracewell Probes, as discussed in Chapters 6 and 7, indicates that we may be closer in both space and time to postulated extraterrestrial technology than currently generally anticipated. Extrapolation for Arkhipov Particles scores highest in my assessments on this Axis, since we are already producing detectably manufactured particles escaping our Solar System, while Arkhipov-Bracewell Probes seem a likely Extrapolation so long as our technological civilisation survives to produce them (a mitigating factor in my assigning a lower score for their Extrapolation).

8. Sheik argues that there is no 'physically motivated' compulsion for an extraterrestrial civilisation to introduce Bracewell Probes of any kind, and scores Artefacts low for ***Inevitability***. But as I argue throughout this Thesis, the most compelling case for discoverable Arkhipov Particles in our Galaxy is that any technological spacefaring civilisation of any kind will inevitably produce microscopic trash, and I have scored Arkhipov Particles high for Inevitability - barring motivations Sheik excluded from the Axes, notably Concealment, i.e. ETs preventing the creation of technosignatures or attempting to hide them. Sheik notes that there are overlaps amongst the 9 Axes of Merit, and with regard to Arkhipov-Bracewell Probes, the Extrapolation factor comes into

play: we can conceive of A-BPs, and as I demonstrate in this work their seeming absence in our Solar System may but await more exhaustive searches using finer-scale detection techniques.

9. Finally Richness of ***Information*** may feature across a wider space of the Axes of Merit – an Arkhipov Particle of mere trash may yet shed light on advanced materials engineering; and even a fragment of a long-defunct Arkhipov-Bracewell Probe could provide insight to stunning feats of engineering – or house a veritable Encyclopaedia Galactica, if not an artificial or transcendent intelligence. When considering in this Thesis work various mechanical, magnetic, conductive and other sifting approaches to search for manufactured particles in lunar regolith simulants, it appears that some of the more promising methods may on depend on some of the very technologies that may have been used in the making of Arkhipov particles and Arkhipov-Bracewell Probes; our own technologies hint that even more might be in store for future discoverers, from an Encyclopaedia Galactica to intelligences themselves (see Figure 1.10). Accordingly, as indicated in my right-hand panel in Figure 1.9, the yellow-circled marker for Arkhipov Particles is placed mid-range on the Axis of Merit for Richness of Information, and I have placed the red-circled marker for Arkhipov-Bracewell Probes toward the highest-scoring end of the Axis.

Figure 1.10: Above, the startup company CatalogDNA demonstrates its data storage capabilities with the whole of present-day Wikipedia contained in DNA. Credit: CatalogDNA. https://catalogdna.com/

1.6 Summary and New Directions

Theoretical consideration of extra-terrestrial intelligence saw its first practical experiments arise and grow with the acceleration of humanity's own space technologies and capabilities. Refinements in radio and more recently optical astronomy have enabled wider searching across multiple parts of the electromagnetic spectrum to the extent that Fermi's 'where are they' challenge has led to increasing speculation that technological, space-exploring civilisations like Earth's are rare, short-lived, evolved beyond recognition, or just plain hiding, whether in fear or as a policy of non-interference.[15] It has been well noted, beginning with Bracewell in 1960 [22][16] that purpose-built, possibly self-replicating automata, not far beyond Earth's own technical capabilities and launched by just one motivated civilisation in our Galaxy, should have had time to span and connect across our Galaxy several times over, suggesting that 'they' are not there, or not yet here, or are consciously and carefully avoiding us or biding their time for reasons unknown.

As discussed above, the Drake equation has been the subject of numerous papers and indeed at least one extensive monograph on its history and development. It has also drawn numerous critiques – in its original incarnation it has attracted much criticism as being neither a delimiting equation nor comprising much more than a compilation of currently unknowable factors – but it has

[15] See especially Frank J. Tipler's 1980 critique 'Extraterrestrial Intelligent Beings do not Exist' [59]

[16] Hungarian-American mathematician and physicist John von Neumann introduced ideas for self-replicating automata in the 1940s, summarized and published posthumously in 1966 [60]; Bracewell adapted the principles of von Neumann's Universal Assembler into his concepts for self-replicating interstellar space probes.

regardless been a reliable source of stimulation and debate for more than half a century. Now, with recent interpretations rendering it susceptible to the disciplines of probability and statistics, it has become a more durable tool, notably under treatments by Maccone [61] adapted and further refined by Lingam & Loeb [55].

Most recently Benford [35], as seen above, specifically adapted the Drake equation for artefacts, albeit without a probability and statistics driven analysis. Both Lingam and Benford cite Arkhipov in their recent work, but leave out the wider implications for the Drake Equation, addressed in the experimental work of this Thesis. Based on these experiments and the Machine Learning methods I employ, I can envisage additional terms to base assumptions for future testing using the experimental methods developed here, to incorporate the constraints of detection/ non- detection of Arkhipov Particles on the Moon. In addition, I suggest how the probability and statistics driven approach pioneered by Maccone [61] not only adds to robustness for adding additional data, as demonstrated by Lingam & Loeb [55], Benford [35] and here by the present author. In the next few decades, Machine Learning may begin to not only populate key elements of the Drake Equation, but render powerful predictions to inform it, e.g. regions of interest in the Galaxy, sources of Arkhipov particles and probes, distribution of asteroidal technological artefact materials etc., using autonomous and inter-networked spacecraft to cumulate and refine discovered data for iterative learning experiences.

1.7 Aims of This Work and Thesis Structure

This thesis aims to create and demonstrate new tools to automatically identify and classify 1) particles captured on Solar System dust collection missions, including intercepted particles that may be of technological origin (i.e. Arkhipov particles); 2) spacecraft hardware remnants (including those of possible Arkhipov-Bracewell probes) captured in images of Solar System objects by deep space missions; and 3) exotic materials that may be uncovered *in situ* in conjunction with lunar and asteroid mining and other operations, including Arkhipov particles and traces of Arkhipov-Bracewell probes.

This first chapter introduces the background and context for SETA, and how this work can better address the need to operationalise programmatic searches for extraterrestrial artefacts in our Solar System – and indeed begins to do so. In Chapter 2, I describe the methods employed that are common to all of the experiments undertaken, introducing general concepts of Machine Learning, and expanding on the specific developments in Object Identification and Classification used in this work, beginning with the Japanese Aerospace Exploration Agency (JAXA's) Tanpopo Astrobiology Project.

In Chapter 3, I describe the camera and hardware elements common to my experiments newly imaging Tanpopo panels and my novel approaches examining Earth spacecraft hardware particles atop lunar simulants, and highlight procedures used for all the experiments described in Chapters 3, 4 and 5. Chapter 3 describes experiments in Object Identification, Localisation and Classification I conducted on samples returned from the International Space Station as part of the Tanpopo Astrobiology Project (See Figures 1.11 and 1.12) [62].

The Tanpopo project itself – focused on panspermia, microbiology exposure and contamination experiments – is introduced, as well as the receiving laboratory's set up and original means of identifying and classifying objects.

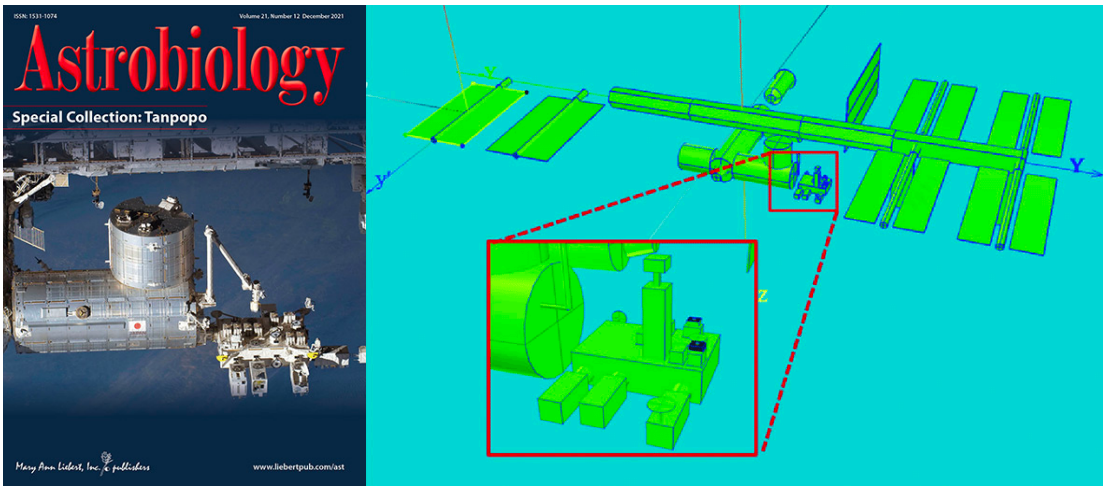


Figure 1.11. Left: The Tanpopo Project featured on the cover of Astrobiology Volume 21, No. 12, December 2021; Right: as detailed in that issue, location schematic of the Aerogel Collection Plate Module on the outside of the International Space Station. Credit: Astrobiology and Astrobiology Project Tanpopo team.

In Chapter 3 I also explain the changes I made to the laboratory's imaging systems, and introduce the algorithms that I developed for speeding the Object Identification and Classification process, improving its accuracy, and 'containerising' the system into a form factor and power and computing resource requirements suitable for spacecraft deployment. The results are described and implications for future interplanetary and interstellar dust particle collection and analyses are discussed.

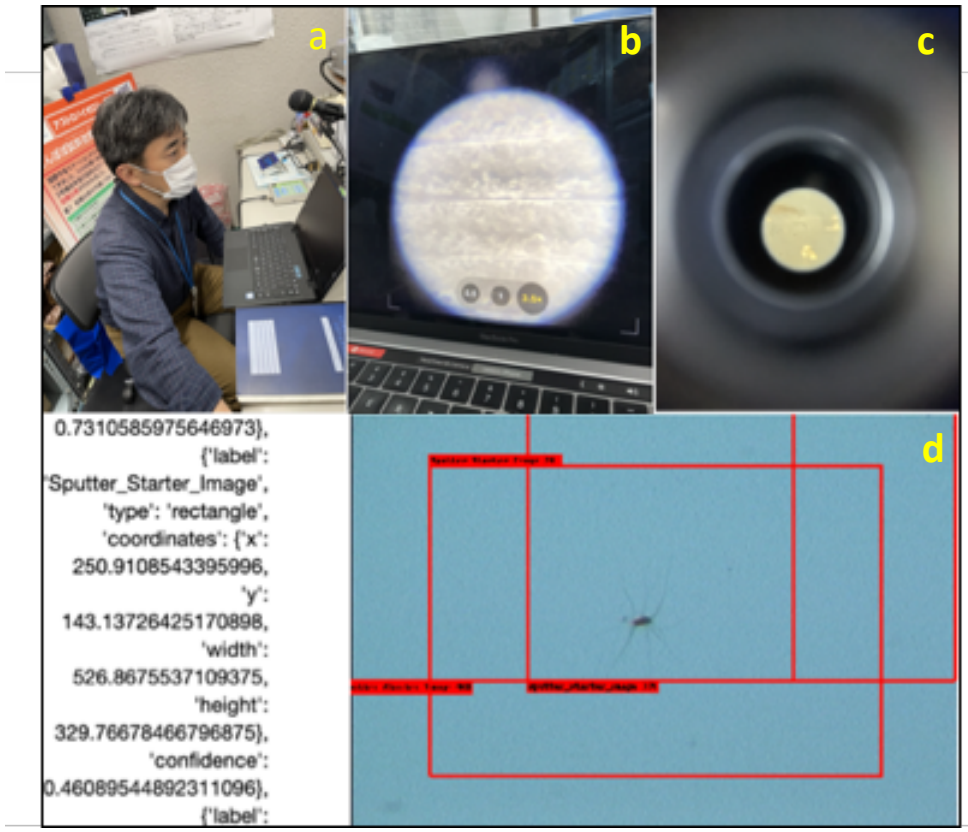


Figure 1.12: Above from left: (a) Hajime Yano reading out results of one of this Project's Tanpopo aerogel searches; (b) a magnified image of the aerogel aluminium frame edge, as captured by the iLabCam® set up; (c) looking remotely 'down the hole' of the iPhone 12MaxPro at a magnified block particle fragment in the aerogel; and (d) Machine Learning search algorithm result readouts, text and images generated and displayed simultaneously, correctly identifying and bounding a 'Sputter' in the aerogel using the algorithm core Machine Learning model developed in this thesis.

Chapter 3 additionally includes the rational for undertaking experiments using a two-stage gas gun to inject spacecraft hardware particles into monocrystalline aluminium and frozen lunar simulant targets. Recognising that Arkhipov particles would likely arrive at the Moon at hypervelocities, I have arranged and designed these experiments to determine what types of spacecraft hardware particles would most likely survive, vaporise or deform, at what rates,

and to use the resulting micro-crater morphologies to determine whether natural (e.g. glass beads) versus manufactured (e.g. titanium alloy) impactors produce characteristic morphologies that could be recognised by Computer Vision/ Machine Learning systems. In conjunction with the design of the two-stage gas gun experiment, I have consequently prepared, developed and trained Machine Learning models on un-shot particle samples of Earth spacecraft materials set atop and mixed into lunar regolith simulants. There is a modest chance that some Arkhipov particles arriving at the lunar surface from a trailing position, at lower velocities, may survive relatively intact, and that such particles that are shielded by regolith churning processes, say those deposited in pit craters, may be reasonably modelled for Machine Learning training purposes even without the hypervelocity impact experiment results that could infer their presence through characteristic crater morphologies.

In Chapter 4, I describe experiments conducted on images captured by the Lunar Reconnaissance Orbiter. I introduce the algorithms that I developed for detecting spacecraft hardware, and how I trained the Machine Learning model on images of the Apollo spacecraft landing sites and other spacecraft hardware. I compare the results (See Figure 1.13) to other approaches recently undertaken for detecting spacecraft hardware on the Moon, beginning with the Moon Zoo project and some early work I did on image processing there, and including recent attempts by other researchers to process LROC images using Virtual Auto Encoders, a different, and as I demonstrate, less scalable Machine Learning approach. I introduce possible new spacecraft hardware detections on the Moon

using the Machine Learning model I developed, and discuss designs based on my model for incorporating real-time Machine Learning Object Detection and Classification into future lunar and asteroid missions.

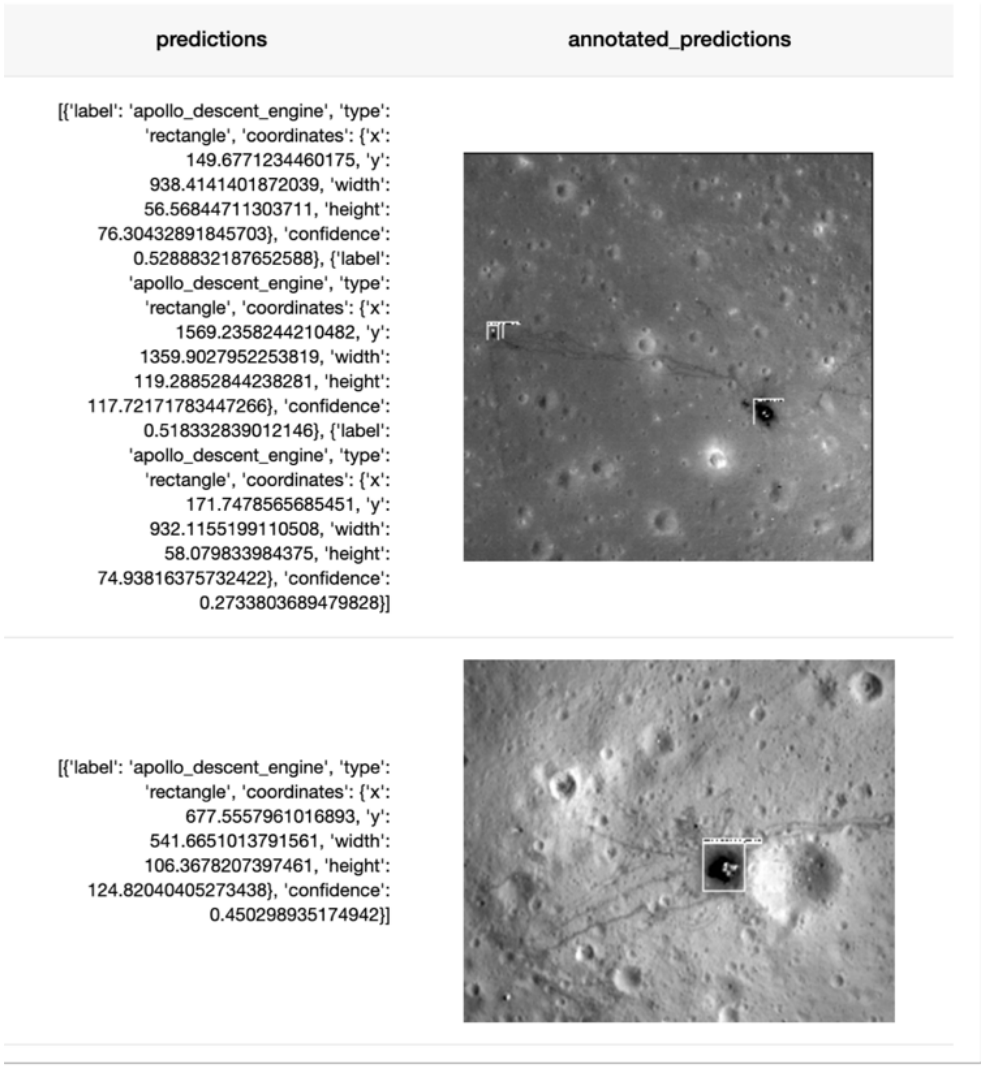


Figure 1.13: You-Only-Look-Once (YOLO) Predictions for Earth Spacecraft Hardware using this work's core Machine Learning model.

Chapter 5 features future work now in various stages of proposal. This includes potential mission payloads to be carried to the Moon by startups in my company portfolio and others (See Figures Figure 1.14), and ideas for particular locales and future Space Agency missions, including inside pit craters on the Moon, and sample return missions from the moons of Mars. I also suggest how recent accelerating advances in materials science, sensing and computing might make construction of Arkhipov-Bracewell probes for our own exploration use practical. In Chapter 5 I also summarise the Conclusions of this project, and I offer a reflection on space policy and initiatives that can support the Moon as a natural resource for surviving our planetary crises, and how in exploiting it, we can also come to better understand and value our own uniqueness in the Galaxy.

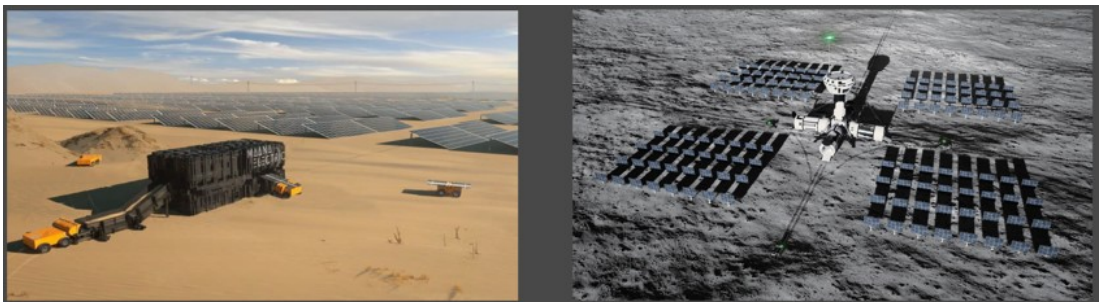


Figure 1.14: Future lunar mining and manufacturing processes could produce significant volumes of material to enable practical searches for Arkhipov Particles and Arkhipov-Bracewell Probe remnants, using the Computer Vision/ Machine Learning algorithms, models, and scanning designs developed in this thesis. Current startup projects to convert terrestrial sands to solar panels are planned for example as the basis for solar panel manufacture on the Moon. Left, artist's rendering of autonomous vehicles bringing sand to a terrestrial desert solar panel plant; solar panels in the background. Right, artist's rendering of a solar panel manufacturing Moon base, surrounded by solar panels. Credit: Maana Electric. https://maanaelectric.com.

Chapter 2

Developing YOLO-ET on Tanpopo Samples

2.1. Introduction[17]

The search for Arkhipov Particles and Arkhipov-Bracewell Probes as introduced in Chapter 1 is a timely one for driving space science advances in the study of microparticles generally. Extraterrestrial microparticles though millimetres or less in size, bear wide-ranging significance for understanding the interstellar medium, planetary system origins, delivery of water and life precursor materials to Earth and other planetary bodies, developing planetary protection measures for microcontaminants, and identifying the distribution of potential resources in the Solar System. Earth observation and missions to low Earth orbit, the Moon, asteroids and deep space destinations have already created a substantive inventory of these particles including:

[17] Much of this Chapter is based on the full-length article I published April 2024 in *Astronomy and Computing*, attached here in this Thesis at the Special Appendix.

1. Micrometeorites (MMs) — With their small mass, low deceleration through the atmosphere and gentle rain upon the Earth, some surviving micrometeorites are found to be relatively unaltered, with unmelted portions giving direct evidence of their precursor bodies and evolutionary sequence. MMs are generally categorised as meteoroids reaching the Earth's surface, and recovered like meteorites, with sizes in the 10s to 100s of μm [63, 64, 65]

2. Interplanetary Dust Particles (IDPs) — Finer grained than MMs and captured in the stratosphere, with sizes up to 10 μm, IDPs are effectively a category of MMs, and are also presumed to be of asteroidal and cometary origin, like Antarctic Micrometeorites (AAMs) and Cosmic Spherules (CSs), as well as fully melted and recondensated meteoroids recovered from the deepsea floor [66, 67];

3. Interstellar Dust Particles (ISPs) — Originating from outside our Solar system and owing to the Sun's Galactocentric orbit and other influences, these particles can travel at Earth encounter speeds from ~5 km s-1 to ~100 km s−1 or even greater hypervelocities [68];

4. Lunar, cometary and asteroidal micromaterials — The Apollo and Luna missions of 1969–1976 and the Chang 'e 5 mission of 2020 returned dust particles from the Moon to Earth; the Stardust mission intercepted

the coma of comet Wild 2 and returned cometary and interstellar dust materials back to Earth in 2006 [69]; and near-Earth asteroid sample returns by Hayabusa in 2010 (from S-type asteroid Itokawa) [70], Hayabusa2 in 2020 (from C-type asteroid Ryugu) [71], and OSIRIS-REx in 2023 (from B-type asteroid Bennu) [72] also successfully added to Earth's inventory; and

5. Anthropogenic contaminants — Fragments from spacecraft exteriors, engines, spacesuit microfibres and outgassed materials from extravehicular activities are produced in both normal operations and as a result of material degradation and microparticle impacts in space. These are likely to accompany human and robotic activities on the Moon and in the Solar System e.g. [73, 74, 75].

Each of these categories of microparticles in and from the space environment has its own significance, but many are also interrelated, and in practice, on lunar and asteroid surfaces they may be mixed or amalgamated together. Current micrometeorites may offer direct comparison to past asteroid and lunar sample returns for example, affording a recalibration of terrestrial micrometeorite collections by overcoming the selective biases of atmospheric entry, an important step toward better understanding Solar system formation processes see e.g. [76].

Almost all microscopic analyses of extraterrestrial samples involve a detailed examination of their petrological features, textures, mineralogy, and chemical composition, drawing on a depth of research expertise, judgement and experience to offer classification suggestions and understand origins and implications for the early Solar system and more. Besides a heavy experience requirement, the equipment required for these undertakings can be their own burden. Even the most recent automated micro-scanning systems in extraterrestrial sample labs, while increasingly powerful, are substantially sized and practically immobile [77]. Their power-intensive requirements, slow speeds of operation and high consumption of computing resources can lead to lengthy processing times particularly for the characteristically high volumes of data presented by these applications.[18] Here I introduce a novel approach using Computer Vision and AI Machine Learning combined with advanced on-device optical and computing technologies, that can serve as an important complement to researchers' experience and a companion to their field efforts. These advancements can overcome the limitations of current systems to rapidly and accurately identify, localise, and classify microparticles making it a more robust and practical solution for *in situ* anthropogenic contaminant and extraterrestrial sample analyses.

For this part of the Thesis, I collected data from 'F' (false) samples which are not captured micrometeoroids but rather anthropogenic contaminants

[18] The field of view versus area to be scanned can also drive higher data volumes. The heretofore intensive time requirements for data processing for extraterrestrial micromaterials is well illustrated by the case of the million fields of view from the Stardust mission samples, examined by upwards of 30,000 citizen scientists over six years [78].

identified on the surfaces of the Tanpopo aerogel panels exposed outside of the JAXA Kibo module of the International Space Station. The data were prepared and imaged using a digital optical microscope, recording, and processing techniques designed for rapid and automated identification followed by initial morphological classification of the identified features by experienced space scientists (Section 2.2). Following data acquisition and archiving, I then introduced YOLO-ET (You Only Look Once ExtraTerrestrial), a modified YOLO deep learning algorithm trained on the aerogel panel images to provide an optimised pipeline for detecting these 'F' sample features (Section 2.3). I analysed the performance of the trained model on unseen data in (Section subsection 2.4.2), and in Chapter 3, I discuss the applications to anthropogenic contaminants introduced to lunar simulants, as well as micro-scaled features within asteroid Ryugu samples, and their potential correlations to micrometeorites from the TransAntarctic Mountains in (Section 3.2) and (Section 3.3) respectively.

2.2. Data Acquisition and Archiving

2.2.1. Astrobiology Project Tanpopo

Tanpopo is Japan's first space experiment for astrobiology utilising the Exposure Facility of the Japan Experimental Module (JEM) of the International Space Station (ISS), designed for the exposure of extremophile microbes and astronomical organic analogues, and for the collection of potentially organic-bearing micrometeoroids impacting the ISS before entering the Earth's atmosphere, in order to explore the potential for and any

evidence of two-way interplanetary transport of life precursors and life [79, 75, 80]. For impacting microparticles, including micrometeoroids, space debris, and possible terrestrial particles that might carry microbes as bioaerosols, the capture of these particles was achieved using silica aerogel capture panels [81]. These were first placed on the Exposed Experiment Handrail Attachment Mechanism (ExHAM) unit on the ISS in 2015–2019 for the Tanpopo mission and followed by the Tanpopo2 mission in 2019–2020 (Figure 2.1.).

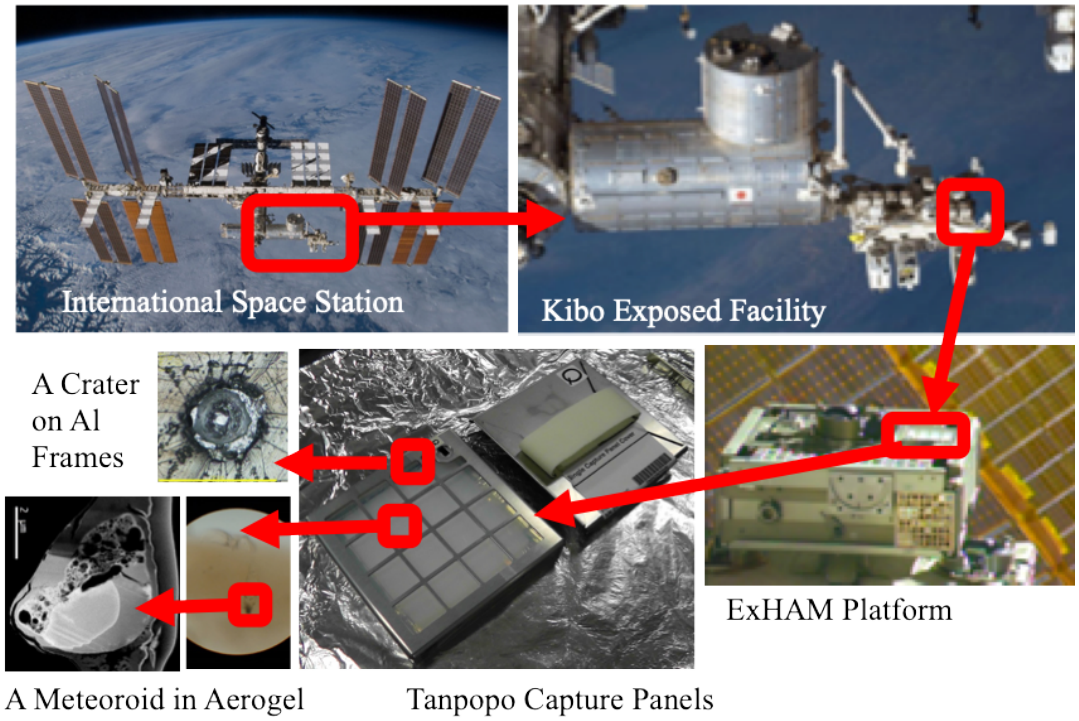


Figure 2.1: Progressive clockwise zoom-in sequence of the Tanpopo astrobiology mission onboard the International Space Station (ISS), showing the placement of silica aerogel panels on the Kibo Exposed Facility for capturing impacting microparticles [75]. The sequence depicts the ISS with highlighted Kibo module, the ExHAM unit where aerogel panels are mounted, and close-up views of an impact crater on the aluminium frames of the panel and an aerogel panel in which an intact captured micrometeroid (diameter ~ 0.10mm) was discovered in subsequent analyses.

The first set of 10x10cm area silica aerogel panels was exposed for one year before being returned to Earth [82] and these aerogel panels were examined under the microscope in clean room facilities at the Institute of Space and Astronautical Sciences (ISAS) in Sagamihara Japan [80]. CLOXS, which stands for 'Captured particles Locating Observation and eXtracting System' [83] is a specialised, compact (approx. $1m^3$) processing machine designed for the Tanpopo mission (Figure 2.2.)

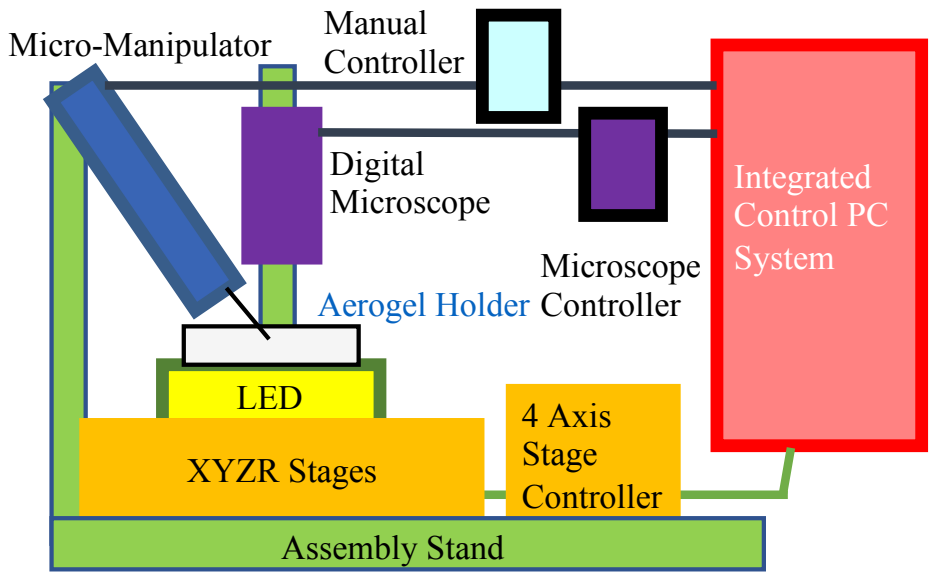


Figure 2.2: Schematic representation of the approximately 1 m^3 CLOXS system, illustrating the precise arrangement of the micro-manipulator, 10cm x 10cm aerogel holder, LED lighting, and XYZ stages mounted on an assembly stand, all coordinated by manipulator and microscope controllers, and integrated into a central control PC system for meticulous particle extraction and analysis.

It processes the returned aerogel from space, placing samples on an X-Y-Z rotatable coordinate stage, and automatically scans and images them under the microscope at 100 x to 245 x magnification to integrate a

microscale map of the entire aerogel panel by moving the stage in micrometre increments (Figure 2.3).[19]

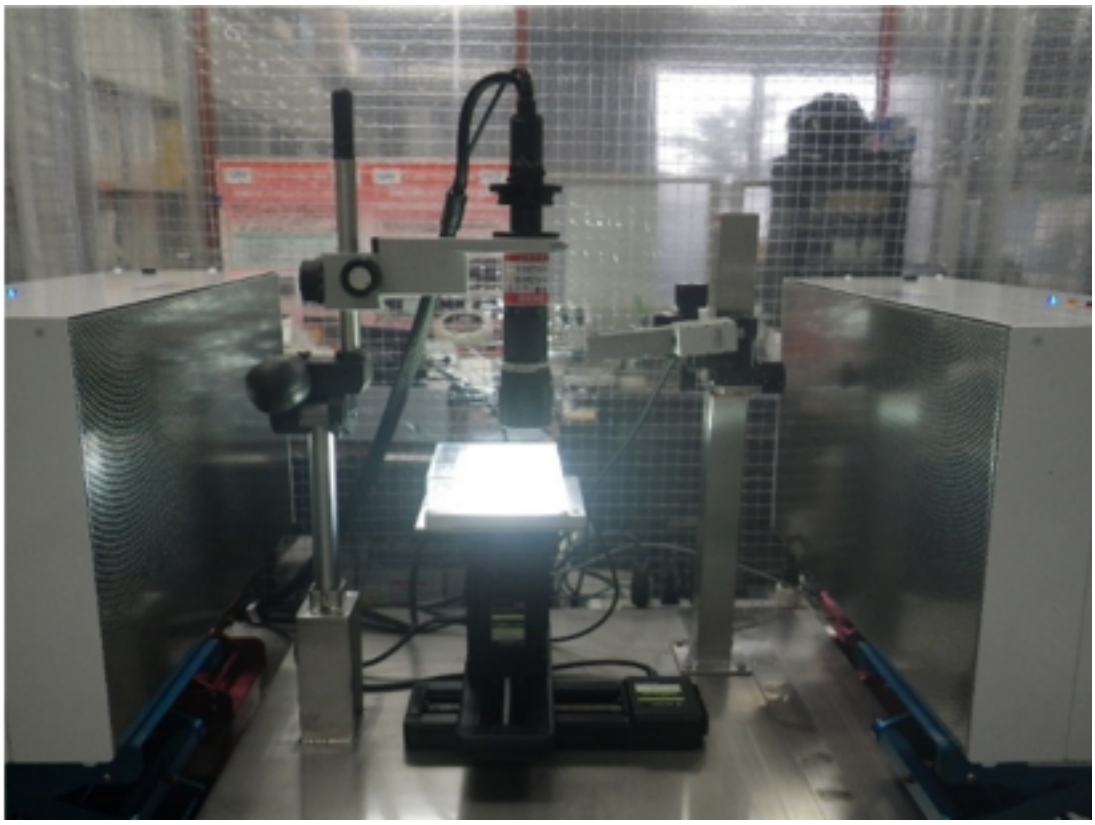


Figure 2.3: The approximately 1m^3 CLOXS system set-up in the ISAS clean room.

When objects of interest are identified from the integrated mapping image at 100x magnification by a scientist, the coordinates of the region of interest of the panel are recorded and the X-Y-Z stage can be automatically moved for revisiting the location for higher magnification investigation. The stack of the revisited images at different focal length depths may contain true penetration tracks and surface objects.

[19] A single integrated map of one 10cm x 10cm aerogel panel may comprise ~ 10^6 unique images with up to a megabyte of data per image.

The Tanpopo mission's classification of surface impacts from hypervelocity impactors is pivotal to understanding not only the impact process, which can lead to penetration of the aerogel by the impactor and its possible vaporisation, but also to glean information about the impactor's composition and origin from the remnants it leaves behind. Not unlike examination of Stardust Mission samples [84], for Tanpopo this involved laborious microscopic examination and imaging of 1000s of samples, with inherent human errors, and earlier efforts by the ISAS team focused on track types—carrot, pit crater, straight, and teardrop.[20] The identified particles and particle impact tracks (Figures 2.4. and 2.5.) in the aerogel that are of interest are cut

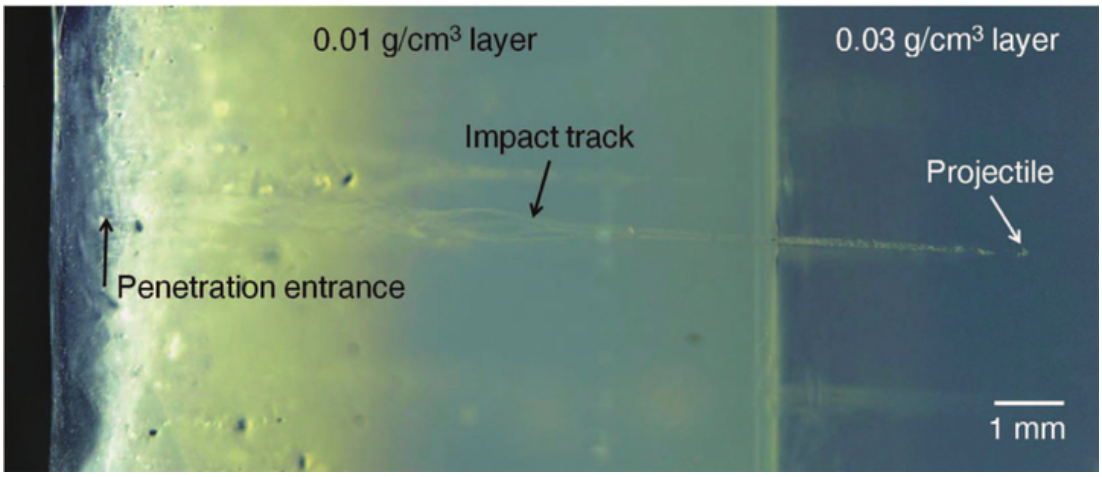


Figure 2.4: Silica aerogel sample post-impact from a hypervelocity particle experiment, simulating the conditions for the Tanpopo project. Side view. This experiment conducted on Earth tests the resilience of the aerogel panels designed for microparticle collection onboard the ISS. Credit: Tabata *et al.* [81]

[20] Track morphologies were classified as 1) pit type – a hemispherical or bulbous shape whose particle likely burst or sublimated; 2) carrot type – a wider-shouldered elongated shape with a single particle at its terminus; 3) straight type – a shape whose entrance diameter typically matches particle diameter along its length and at its terminus; and 4) teardrop type – a shape whose diameter blooms beyond the diameter of the entrance hole, with scattered materials along both wall and terminus of the penetration track [85].

out into suitable-sized chips that contain impact tracks of particles captured in space, where a needle is then used to cut the aerogel without contamination; this is then distributed to research groups worldwide for detailed biological and chemical analyses of the captured microparticles [86].

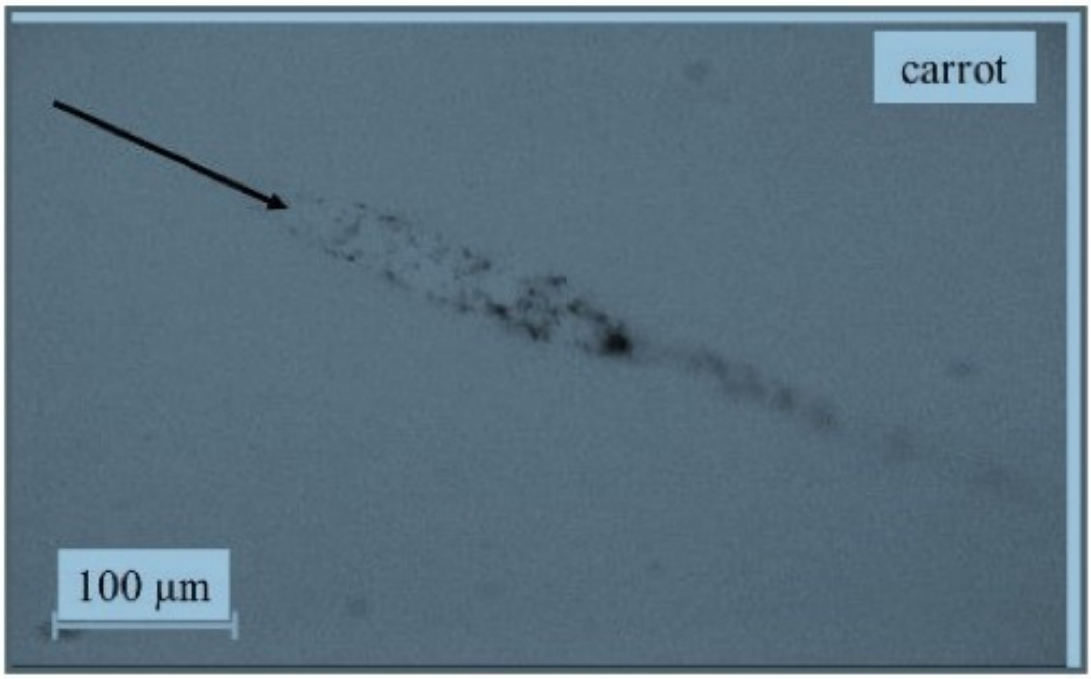


Figure 2.5: A 'carrot' shaped track of a hypervelocity impactor in the Tanpopo silica aerogel panel returned from the International Space Station. The arrow indicates the impact direction. Viewed from above, i.e. the track is roughly parallel to the surface.

As for microparticles collected on the surface of the Tanpopo aerogel panels, these are presumed not to be 'True' hypervelocity impactors associated with morphological features (i.e. carrot tracks, pit craters, straight tracks, and teardrop tracks), but rather 'False' incidentally collected particles impacted at much slower velocities such as material released by ISS docking and undocking activities, venting materials, secondary impacts from primary

impact ejecta, possible spacecraft component fibres, and fragments of the aerogel itself. This study has prioritised accurately classifying surface residual effects of these 'F' samples, such as sputter, fibre, block, bar, and aerogel fragments. Semi-automated methods have been employed to enhance classification but until the work of this Thesis, matching the expertise of human scientists has remained a challenge, requiring a series of manual re-sizings, whitening, and contrast adjustments to secure even modest levels of confidence. Given the abundance of samples, continuous improvement in automated techniques is essential to accurately assess the microparticle remnants.

2.2.2. Machine Learning Data Set

This research focuses on the Tanpopo 1-2 missions 2015–2020 collection of aerogel surface features larger than ~100 μm in the 10 cm × 10 cm aerogel panels captured at typically 100x to 245x magnifications.[21] The total number of 'F' sample images in the collection is nearly 5000. In consideration for the computational power and memory limitations required to train machine learning models with large image input sizes, my data sample consists of 395 images, which is less than 10% of the total 'F' samples, in .jpeg format each 480 × 704 pixels in size. My dataset was necessarily limited by the lack of annotated data (image-label pairs). With the open-source Python widget Bbox,

[21] The 10cm x 10cm silica aerogel panel (6 at a time are deployed on the International Space Station, exposed for one year; five ISS cargo returns of 6 sets each have been made) consists of two layers, an ~8.5mm thickness top layer of 0.01g/cm^3 density and base layer of 7mm thickness with 0.03 g/cm3 density, designed to slow and capture particles at hypervelocities of up to ~ 10km s^{-1} [87].

I additionally manually annotate each image with a bounding box around each object (some images contain multiple particles and others none), and a class label of either ‘sputter’, ‘block’, ‘fibre’, ‘bar’, and ‘aerogel fragment’ (Figure 2.6).

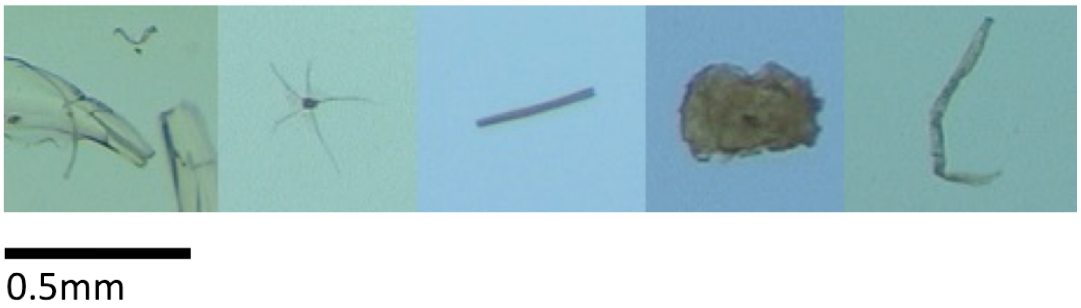


Figure 2.6: Examples of Tanpopo surface objects. From left to right, aerogel fragment, sputter, bar, block and fibre.

Note that the images have a variety of hues and brightnesses as they were captured under different lighting conditions, however in the interest of human time-saving I did not preprocess the images to account for this. Similarly I did not threshold, convert to grey scale or remove noise in the manner typically used to enhance the images for human inspection, as this would defeat the gains offered through the speed of Machine Learning. I was however inspired by the challenges and building on prior work using these more laborious methods [88, 89].

I split the data sample into 80% for training (316 images) and 20% for testing (79 images). These are randomly sampled whilst maintaining the baseline ratio between different classes. The training data are the images used

to optimise the model weights during the training phase of the model. The test data are not seen during the training of the model. Note that whilst it is common to additionally set aside a validation dataset for informing when the model is sufficiently trained, i.e. not under- or over- fit, my relatively limited data sample hampered my ability to reserve additional samples for validation, which could weaken the model's performance. Machine Learning methods are data greedy and perform better with more data, so to compensate for the relative data paucity, I applied automatic augmentation in the form of flipping the training images vertically, horizontally or both (corresponding to 180° rotation), resulting in a factor of 4 increase to the training sample (Figure 7).[22]

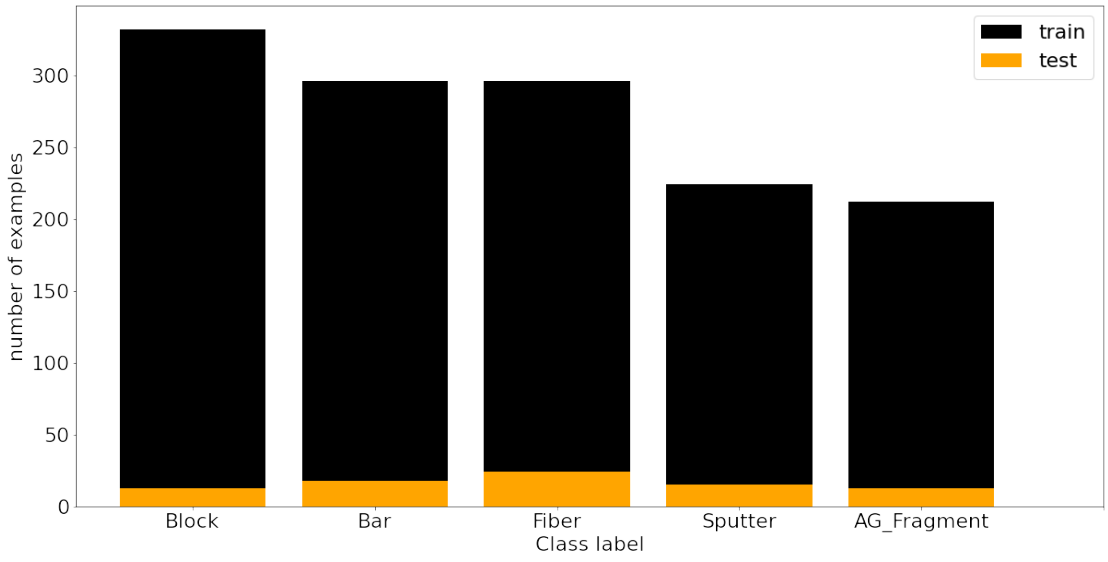


Figure 2.7: Distribution of training and test data over the different classes. Note that that these do not directly correspond to the number of images, as images can contain multiple objects of different classes or no objects altogether. Additionally there is significantly more training data as the augmentation is applied only to the training sample.

[22] Note that 'blank' images (aerogel images with no apparent particles present) are not included in the training or test cycles, to conservatively avoid their boosting model performance; in practice however, as demonstrated by Stardust mission analyses [84], the deselection of null results by the YOLO-ET model can itself be a powerful savings in human observer time consumption.

2.3. YOLO-ET: A Highly Efficient Convolutional Neural Network for Extraterrestrial Microparticle Detection and Classification

2.3.1. Machine Learning

Machine Learning (ML) (see [90] for a review) involves constructing layered architectures where each layer performs specific operations on data. These layers, particularly in neural networks, are composed of nodes or neurons with associated weights. During training, ML algorithms process input data through these layers, where each operation transforms the data based on the current weights. The goal is to optimise the weights to minimise a predefined cost or loss function (Section 3.3.2), which measures the difference between the algorithm's predictions and the actual outcomes. The optimisation is typically done using techniques like gradient descent, where the algorithm iteratively adjusts the weights based on the gradient of the loss, improving the model's predictions over time [91]. Supervised learning, a subset of Machine Learning, involves algorithms that improve at tasks over time by learning from labelled data. This Thesis applies supervised learning to object detection, training models to recognise and categorise microscopic particles on aerogel panels. The data consists of pairs of images and their corresponding labels that are the bounding box coordinates (x and y), height (h), width (w), and class.

2.3.2. YOLO

Machine Learning (ML) algorithms employing Deep Learning techniques have been gaining traction in the astronomical sciences for nearly a decade, with applications ranging from galactic surveys [92], dark matter mapping [93], and

notably in regard to this work, galactic cluster detection [94], using a streamlined YOLO technique. YOLO [95], an acronym for 'You Only Look Once', is a supervised learning approach to real-time object detection in computer vision. YOLO's novel architecture enables it to process images in a single pass, predicting both the bounding boxes and class probabilities (confidence scores) for objects within the image simultaneously. This contrasts with earlier two-step detection systems e.g. [96], which would first propose regions and then classify them. The efficiency of YOLO allows it to detect objects rapidly with a high degree of accuracy, making it ideal for applications that require real-time processing.

Unlike more conventional techniques, YOLO is a type of convolutional neural network (CNN) consisting of a series of convolutional layers and pooling layers rather than neurons [97]. Jaeger *et al.* [98] use a 16-convolutional layer Visual Geometry Group CNN, a tailored version of VGG-16 [99] with 20,000 training images to classify impact craters on aluminium foils from the Stardust interstellar dust collector, which are typically less than one micrometre in size and sparse, making them difficult to find. While this method excels in accuracy for small objects, its deep architectures lacks YOLO's speed, limiting its use in real-time scenarios. Additionally, it primarily assesses the probability of crater presence in an image[23] without pinpointing exact locations, and is not optimised for images containing multiple objects of different classes.

[23] The Stardust system flagged the absence of an impactor crater, allowing the rejection of images not bearing further examination. As a key measure performance, this facilitated automated scans of one hour, where a year of human scanning effort was previously required.

2.3.3. YOLO-ET

YOLO-ET, is a modification of YOLO optimised for the detection of extraterrestrial microparticles. Specifically, I employ Tiny YOLOv2 (Figure 2.8) which is a smaller, simplified version of the original YOLOv2 [100] with a Darknet-19 base network (a 19-layer network inspired by the VGG-16 model).

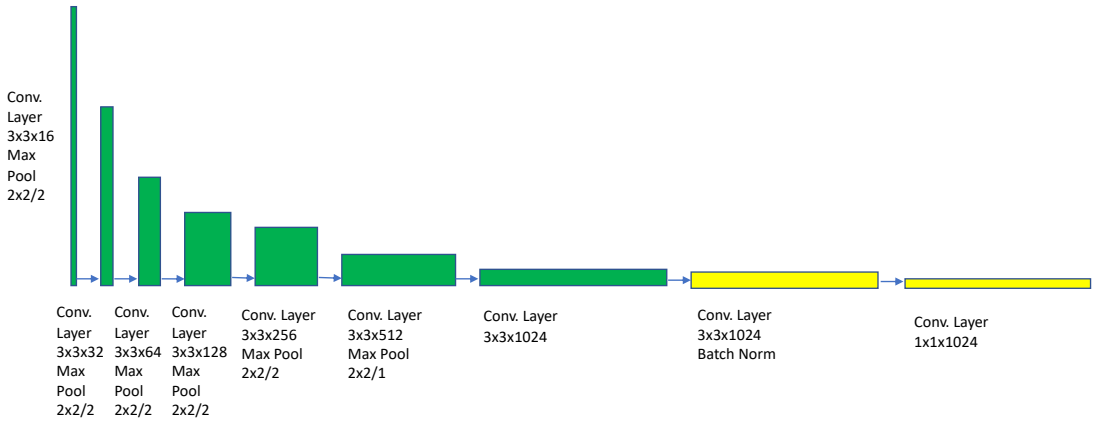


Figure 2.8 TinyYOLOv2 architecture showing the series of convolutional and max pooling layers, with Batch Normalisation marked in yellow. The numbers represent the filter size and number of filters in each layer and the/represents the stride in the max pooling layers. Batch Normalisation is introduced after convolutional operations and before the activation functions, and leads to faster convergence during training by reducing internal covariate shift, i.e. the natural tendency to change the mean and variance of the inputs with each layer. Aside from helping to stabilise the training process by ensuring that the distribution of inputs to each layer remains more consistent during training, Batch Normalisation also helps regularise the model and reduce overfitting, so that the model can generalise better to unseen data.

For an overview on the development of YOLO see [101]. YOLOv2 is designed to be more compact and faster than YOLO, making it suitable for applications with limited computational resources, such as mobile devices or real-time systems (see e.g. [102]). While maintaining the core principles of YOLO's

single-pass detection, YOLOv2 simplifies the architecture with fewer convolutional layers and filters. In the original YOLO architecture, bounding box[24] predictions were made relative to the dimensions of a grid cell; this approach had some limitations in terms of accuracy, particularly around predicting the correct size and location of objects. YOLOv2 improved upon this by predicting bounding box coordinates directly. Instead of the network learning offsets relative to a grid cell, YOLOv2 learns to predict bounding box coordinates relative to the location of the grid cell, along with anchor box dimensions, which makes predictions more precise (Figure 2.9).

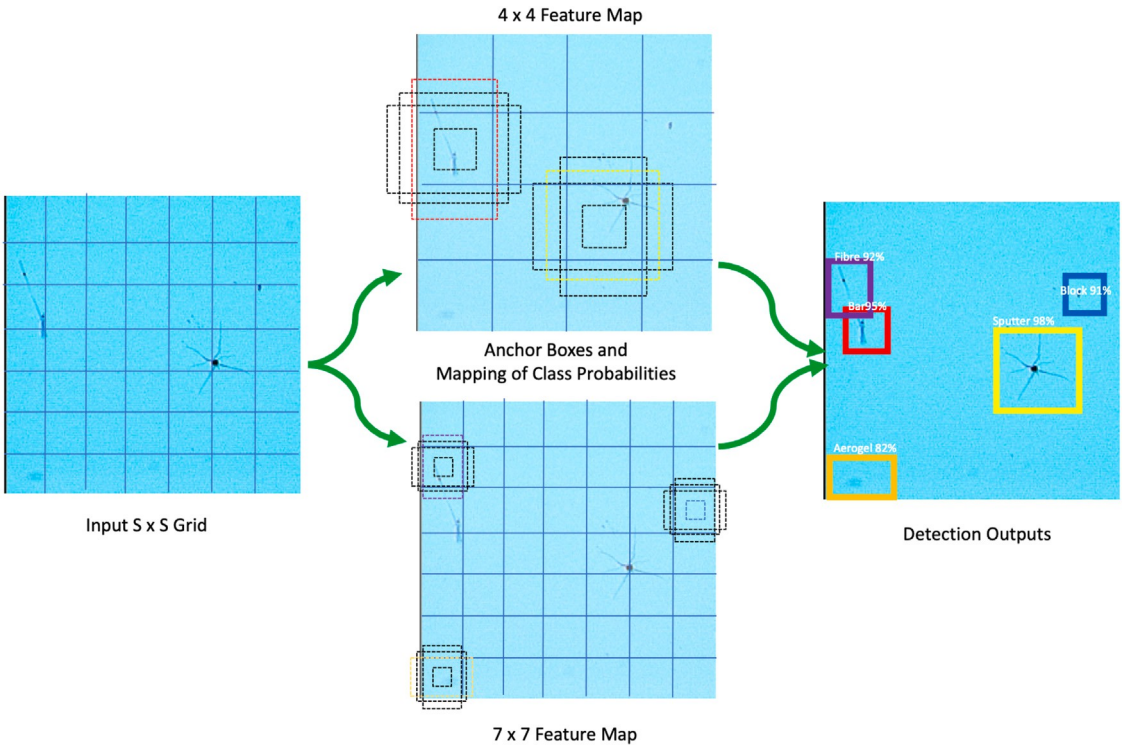


Figure 2.9: Diagram illustrating the concept of anchor boxes in TinyYOLOv2,showcasing various predefined box shapes and sizes strategically positioned across a Tanpopo aerogel image in YOLO-ET implementation. These anchor boxes enable the efficient and accurate prediction of object boundaries and classifications within a single pass of the network.

[24] A bounding box is used to identify the location of an object within an image or video frame in object detection tasks; it is usually defined by centre, width and height coordinates.

This reduction in complexity results in faster processing speeds but typically at the cost of some detection accuracy compared to the full YOLO model. The 'Tiny' version of YOLOv2 is specifically optimised to be more lightweight and faster, sacrificing some accuracy for the sake of speed and smaller model size. This makes Tiny YOLOv2 particularly well-suited for deployment in environments with limited computational resources, such as mobile devices, embedded systems, or applications where real-time performance is crucial.

I underscore the suitability of the TinyYOLOv2 architecture for mobile use, which represents a significant advancement in deploying deep learning models on devices without the need for high-powered computing resources. YOLO-ET thus embraces Tiny YOLOv2's trade-off between speed and precision, optimised for scenarios where real-time performance is crucial. For a deeper understanding of the efficiency and effectiveness of the TinyYOLOv2 architecture in mobile environments, I direct readers to the comprehensive study detailed in [102]. This study provides empirical evidence supporting my choice of architecture, demonstrating its superior performance in scenarios demanding high efficiency and reliability on mobile devices.

2.3.3.1. Non-Maximum Suppression

Non - Maximum Suppression (NMS) [103] is a post-processing technique commonly used in object detection algorithms and is a key feature in YOLO and YOLOv2, that ensures each detected object is only recognised once. When an object detection model predicts multiple bounding boxes for the same object,

NMS helps in selecting the most probable bounding box and discarding the rest. It does this by comparing the overlap between two boxes (A and B) using a metric called Intersection Over Union (IOU),

$$IOU = \left| \frac{A \cap B}{A \cup B} \right|, \tag{2.1}$$

and retaining only the boxes below the defined IOU threshold, and above the defined confidence score threshold while suppressing the others. This reduces redundancy and increases both detection interpretability and accuracy. After trialling different values, the default IOU threshold and confidence threshold in my work are 0.5 and 0.3, respectively, providing both a limited clutter of redundant bounding boxes and a practical level of accuracy for distinguishing microparticle types.

2.3.3.2. Loss Function

The loss function quantifies the difference between the values predicted by the model and the actual values in the training data. A key goal in Machine Learning is to find the set of parameters, the weights and biases in the context of convolutional neural networks, that can optimise toward actual values by iteratively moving toward the minimum value of the loss function. The slope or derivative of the loss function with respect to its parameters is defined as its gradient, and moves in the direction of the steepest increase of the function. Moving in the opposite direction of the gradient, the algorithm iteratively adjusts the parameters to reduce loss, referred to as the gradient's descent. The learning

rate is a hyperparameter that determines the size of the steps taken toward the minimum — too large, and the algorithm might overshoot the minimum, too small, and it will converge very slowly, consuming additional computing resource. Batch sizes determine the amount of data used to calculate the gradient at each step.

The YOLO loss function specifically combines terms for bounding box prediction accuracy, object presence confidence, and class prediction, ensuring the model is well-tuned across all aspects of object detection. In contrast to Grishin *et al.*'s [94]'s work on galaxy clusters, YOLO-ET retains the comprehensive YOLO loss function, exploiting the full power of YOLO to simultaneously tackle the presence of multiple objects of different classes in an image. This is also useful for aerogel-captured particles, where multiple particles may overlap or appear at different depths in the aerogel panel, and it is essential for real world observational tasks in the planetary sciences like searching for microparticles and tell-tale microcraters *in situ* on the surface of the Moon.

The relevant loss function is defined as follows,

$$
\begin{aligned}
L = {} & \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\
& + \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\
& + \sum_{i=0}^{s^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} (C_i - \hat{C}_i)^2
\end{aligned}
$$

$$+\lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{noobj} (C_i - \hat{C}_i)^2$$
$$+ \sum_{i=0}^{s^2} \mathbb{1}_{ij}^{obj} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2, \quad (2.2)$$

where, the first two terms, weighted by λ_{coord}, penalise errors in the position (x, y) and size (w, h) of predicted bounding boxes compared to the ground truth. These are crucial for precise localisation. The third term penalise errors in object scores C_i, distinguishing between object presence and absence. The fourth term, scaled by λ_{noobj}, specifically penalises false detections and the final term assesses the classification error for each class c across the objects detected, ensuring accurate class predictions [95].

2.3.3.3. Turi Create

I deployed YOLOv2 through Turi Create,[25] an open-source machine learning library developed by Apple. It provides a simplified approach to creating machine learning models, especially for developers interested in practical field application. Turi Create supports various types of models, including classifiers, recommender systems, and image classifiers, and is particularly known for its ease of use in creating models for iOS apps. The library is optimised for scalability and performance, enabling the development and deployment of models on both Macs and mobile iOS devices. Using Turi Create for object

[25] https://github.com/apple/turicreate

identification, localisation, and image classification is remarkably straightforward, allowing more user development time for focusing on the customisation of the learning model itself. AI Machine Learning is becoming increasingly accessible and user-friendly with applications such as Turi Create and Microsoft Lobe [26] providing highly accessible implementation of AI including in educational settings.[27]

This user-friendly entry point into object detection provides a streamlined experience, but at the cost of customisation depth. This abstraction means users are not able to fine-tune all model hyperparameters, i.e. the configuration settings of the network defined before training begins. In the case of YOLO-ET, these include: the learning rate — i.e. the magnitude by which the weights are updated during training, the anchor box dimensions, the NMS (confidence) threshold and the IOU threshold. It is also not trivial to employ a validation set directly within the framework. However, Turi Create still affords some degree of control, allowing for the adjustment of hyperparameters such as batch size and maximum iterations, which can significantly influence model performance and training time.

During training, I experimented with various batch sizes. While larger batches demand more memory due to the increased number of images loaded simultaneously, they tend to smooth out the loss curve, leading to a more stable model. Conversely, smaller batches, although more memory-efficient, can

[26] https://www.lobe.ai

[27] Lobe requires neither Machine Learning nor coding experience and should enable a wide range of user engagements. At the time of this research however its templates are only set up for Image Classification tasks.

result in a noisier gradient descent trajectory. High-resolution inputs restricted batch capacity, thereby decelerating the training convergence. Nevertheless, larger batches expedited convergence toward the global minimum of the loss function. For this Thesis, a batch size of 32 was identified as the most effective, balancing computational resource demands and learning stability.

In the training of my model, an epoch is defined as one complete pass through the entire dataset, whereas an iteration is one update of the model's weights, which occurs after processing a batch of samples. Recall my model uses 1264 training images with a batch size of 32, thus each epoch consists of 1264/32 = 39.5, or approximately 40 iterations. Setting max iterations to 2000 means the training process involved roughly 2000/40 = 50 epochs. After experimenting with various numbers of max_iterations, 2000 was found to be optimal, striking a balance between model performance and computational efficiency. This choice was partly influenced by Turi Create's constraints on validation set usage, which limited my ability to employ traditional validation techniques to fine-tune the number of iterations. Instead, I relied on trial and error in the creation of the model, along with runtime considerations, to determine the most effective training duration. Since my goal is automated on-device deployment for laboratory and *in situ* use in space environments, the trial-and-error component in the model creation phase to achieve this goal is a practical trade off: the balance of convenience with a good modicum of configurability makes Turi Create a practical tool for rapid development, while

recognising the limitations for more advanced experimentation and nuanced model optimisation.

2.3.3.4. Transfer learning

My dataset is relatively modest in size even with augmentation, but training machine learning models effectively requires extensive datasets and prolonged computational training times. Transfer learning (see e.g. Tan *et al.* [104] offer a practical solution to this challenge by utilising a pre-trained model – a model initially trained on a specific task and dataset – and adapting it to a different, yet related, task or dataset. This approach can take two forms:

- Direct Application: If the new task closely aligns with the original training task, the pre-trained model may be used as-is, leveraging its existing knowledge; and
- Modification and Retraining: More commonly, the latter layers of the network are modified and retrained, while the initial layer weights are kept fixed. This tailors the model more closely to the new task and data.

The latter method is advantageous as it significantly reduces the volume of training data required and shortens the training time compared to training a model from scratch. This efficiency stems from the model's ability to build upon the knowledge already acquired during its initial training phase.

For further refinement, fine-tuning comes into play. This process involves making minor adjustments to the model's weights, already pre-trained on a large dataset, to achieve a more precise adaptation to the new

task. Turi Create's implementation of TinyYOLO executes this process in a user-friendly manner. Initially, the model is pre-trained on the standard ImageNet dataset [105], which comprises over 1 million images of 224 × 224 resolution, spanning 1000 classes. This foundational training equips the model with a broad understanding of various visual features. Subsequently, it undergoes fine-tuning to adapt to higher-resolution images, specifically to a resolution of 448. This pre-trained model is further refined using my specialised Tanpopo dataset to create YOLO-ET. Using a desktop AMRadeon Pro Vega 64X 16 GB GPU, training time takes 0.02, 0.17 and 0.37 s per iteration for batch sizes of 1, 32 and 64 respectively. The runtime for prediction on 79 test images takes 1.14 s, demonstrating the efficient and practical application of transfer learning and fine-tuning in customising models for specific tasks in planetary and astronomical sciences, and opening the door to tasks that could be readily implemented on-device in laboratory and spacecraft environments.

2.4. Evaluation and results

Figure 2.10 illustrates the model's training loss over time. While it may seem tempting to continue training until the loss approaches zero, it is critical to halt the training process beforehand to avoid overfitting. Overfitting occurs when a model becomes excessively attuned to the training data, to the extent that it perfectly predicts the classes and localisations. Such hyper-specific learning compromises the model's ability to generalise and perform

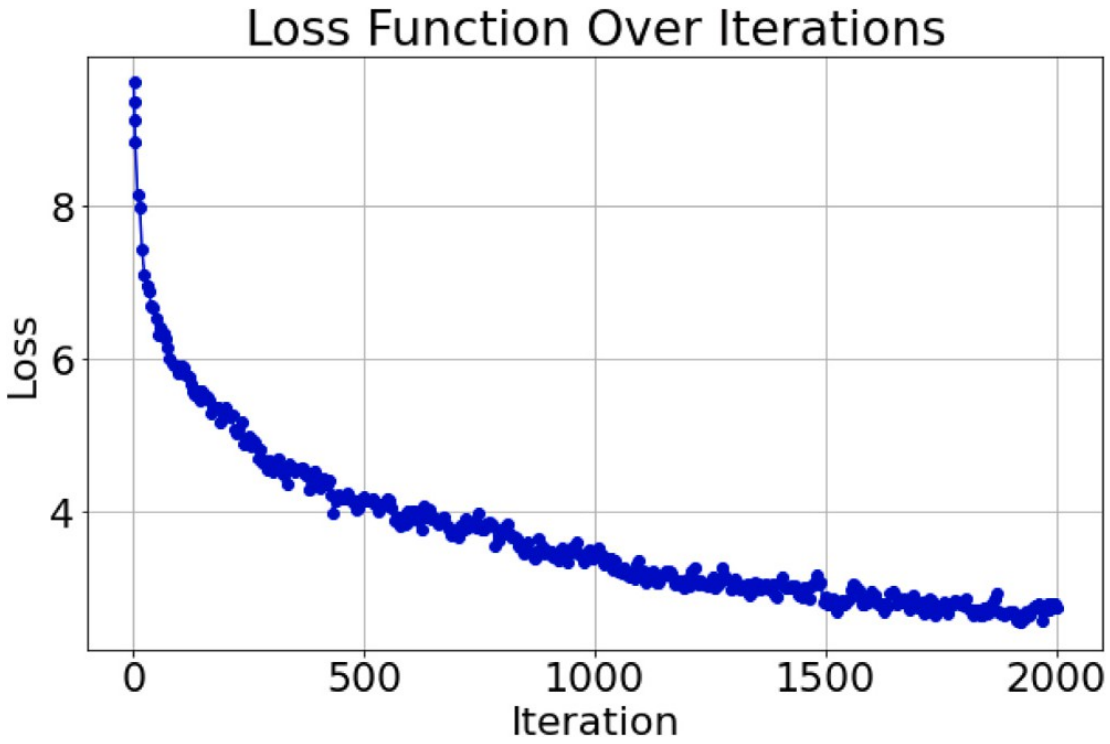


Figure 2.10: YOLO-ET training loss of the network over time.

accurately on new, unseen data. Therefore, identifying the right moment to stop training is essential for maintaining the model's effectiveness on diverse datasets. Evaluating the model on the test set that the model has not encountered during training, serves as a proxy for real-world, unseen data and provides a more accurate measure of how well it will perform in the real world in comparison to the training set (Figure 2.11). Figure 2.12 shows an example of the model applied to test data. Note that the NMS/IOU thresholds have failed at removing the duplicate detection, as both boxes are above the 30% confidence level and the overlap of the boxes are less than 50%.

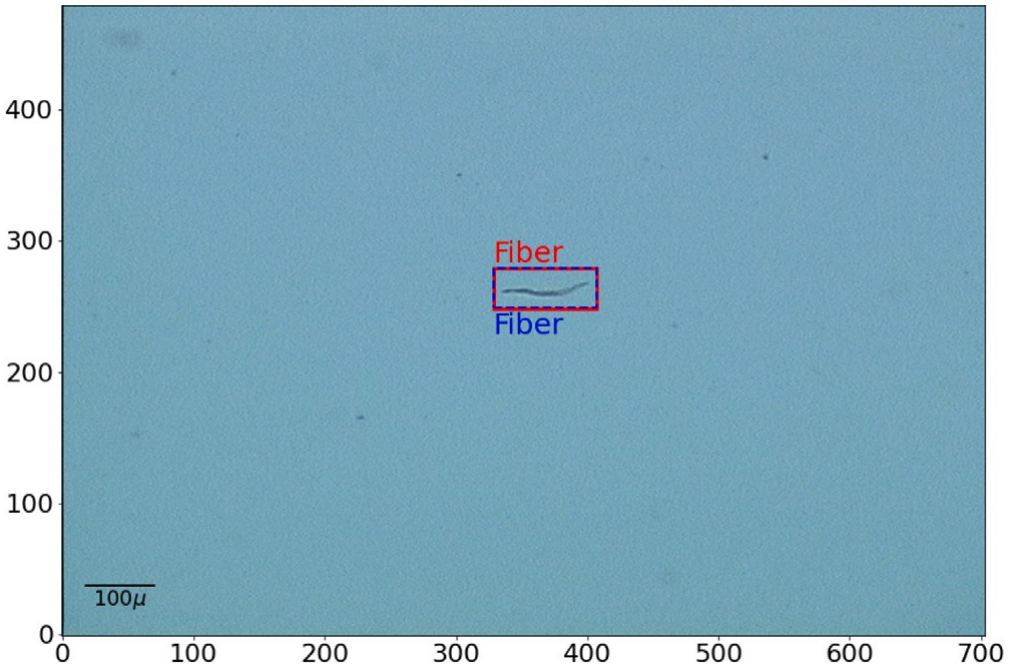


Figure 2.11: Training example image depicting a fibre with the ground truth bounding box in dashed blue and the prediction from the network in red. Units are pixel coördinates.

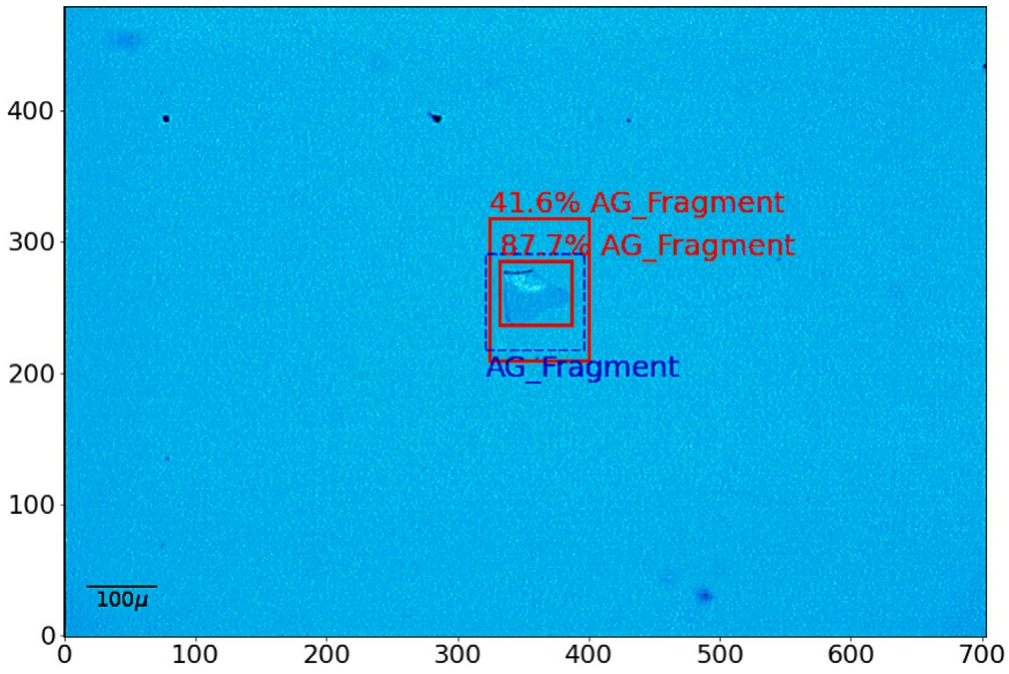


Figure 2.12: Test data example depicting an aerogel fragment with ground truth bounding box in dashed blue and the network predictions in red. The confidence score of the detected object is also shown at 42% and 88%. Units are pixel coördinates.

Figure 2.13 shows another test data example with a single detection. Note the variations in the hues, brightness and contrast of the images that typically make such classifications challenging.

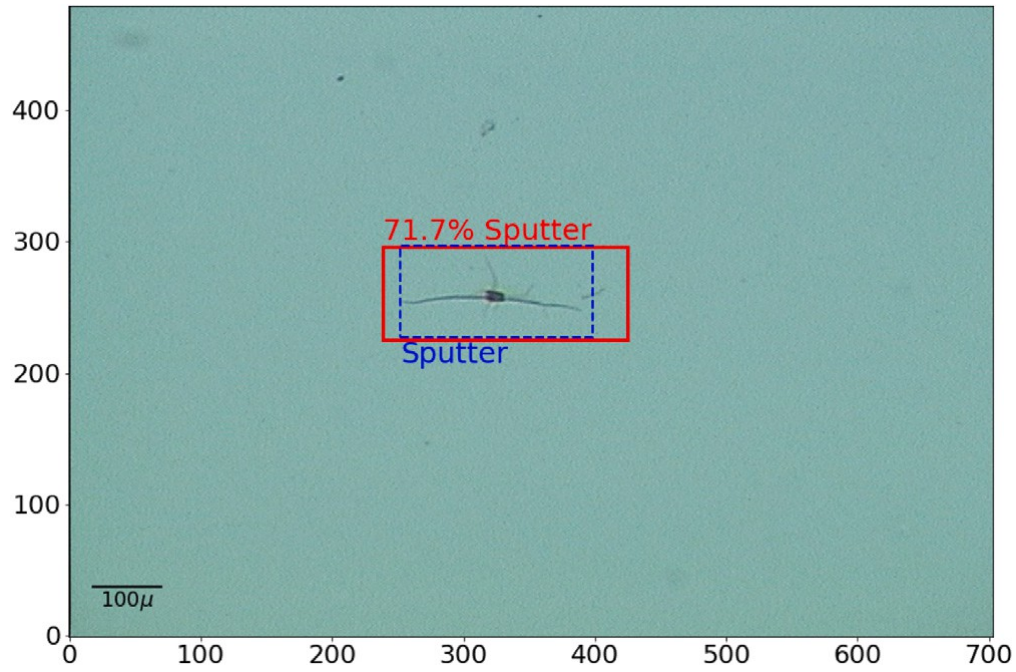


Figure 2.13: Another output prediction of YOLO-ET applied to test image. The red box shows the predicted bounding box, the annotated classification and confidence probability at 72%. Units are pixel coördinates.

2.4.1. Evaluation Metrics

To accurately assess the performance of our network, I utilised a range of evaluation metrics, each chosen for its specific relevance and effectiveness in addressing the unique aspects of our problem. This diverse set of metrics ensures a comprehensive and nuanced understanding of the network's capabilities and weaknesses, allowing for a more targeted and effective optimisation.

2.4.1.1. Precision and Recall

Precision and recall are two fundamental measures used in machine learning for evaluating the performance of classification models, especially in scenarios where the classes are imbalanced. Precision measures the accuracy of the positive predictions made by the model. It is the ratio of true positives (TP, correct detections) to the total predicted positives (both true positives and false positives FP).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2.3}$$

High precision indicates a low false positive rate but does not consider false negatives (FN, missed detections).

Recall measures the ability of the model to find all the relevant cases within a dataset. It is defined as,

$$\text{Recall} = \frac{TP.}{TP + FN} \tag{2.4}$$

High recall indicates that the model is good at finding the positive instances but does not indicate how many negative instances were incorrectly labelled as positive.

The trade-off between precision and recall often depends on the specific requirements of the task. For example with the Tanpopo aerogel-captured surface samples above, if minimising the mis-identification of terrestrial debris particles as extraterrestrial in origin were paramount, then a high precision would help avoid making incorrect categorisations. On the other hand high recall is essential when

the goal is to ensure no potential extraterrestrial particle is missed. This might be prioritised in space environment studies where noting every possible particle is more critical than the occasional false identification. The priority, for Tanpopo surface samples, much like in ML applications for cell pathology, is not to miss anything.

In practice, ML modellers often look at both precision and recall together, sometimes combining them into a single measure called the F1 score, which is the harmonic mean of precision and recall:

$$\text{F1 Score} = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (2.5)$$

The F1 score provides a balance between precision and recall, considering both false positives and false negatives.

2.4.2.2. Average Precision

Note that the definition of a TP also depends on the IOU threshold with respect to the ground truth box and confidence threshold. Average Precision quantifies the model's performance across different levels of precision and recall, which are typically varied by adjusting the threshold for classifying a detection as a true positive. It is calculated by plotting a Precision–Recall curve, which shows the trade-off between precision and recall for different thresholds. The area under this curve (AUC) represents the AP. Essentially, it is the average of precision values at different recall levels and is specified at a particular IOU threshold.

Mean Average Precision (mAP) is an extension of AP that is used when there are multiple classes to be detected. mAP is the mean of the APs

calculated for each class individually. It is computed by first calculating the AP for each class independently. Then these AP values are averaged across all classes. This gives a single metric that summarises the performance across all classes. This is particularly important here where we need to detect multiple types of objects as it gives a holistic view of the model's performance across all these different classes, making it a more comprehensive and balanced metric.

2.4.2. Results

The final loss of my network was 0.8605. YOLO-ET correctly detects 90% of the test data with over 50% overlap (IoU) with the ground truth box. A summary of the results is shown in the confusion matrix (Table 2.1).

		Predicted	
		Positive	Negative
Actual	Positive	71	15
	Negative	17	N/A

Table 2.1: Confusion Matrix based on the test data set.

Of the False positives, 47% are incorrectly identified as block, 35% as bar and only 1 each of fibre and sputter and aerogel fragment. The FPs are less of a concern as the confidence levels of all the detections are below 50% with the exception of the

AG fragment with a confidence of 88%. On inspection this detection is a duplicate detection where 2 bounding boxes are picking up the same object with high confidence. It is also notable that blocks tend to be detected but mis-classified with 3 incorrectly classified as Fibre and 2 incorrectly classified as AG fragment. Figure 2.14 shows the distribution of FPs and FNs over their respective class labels.

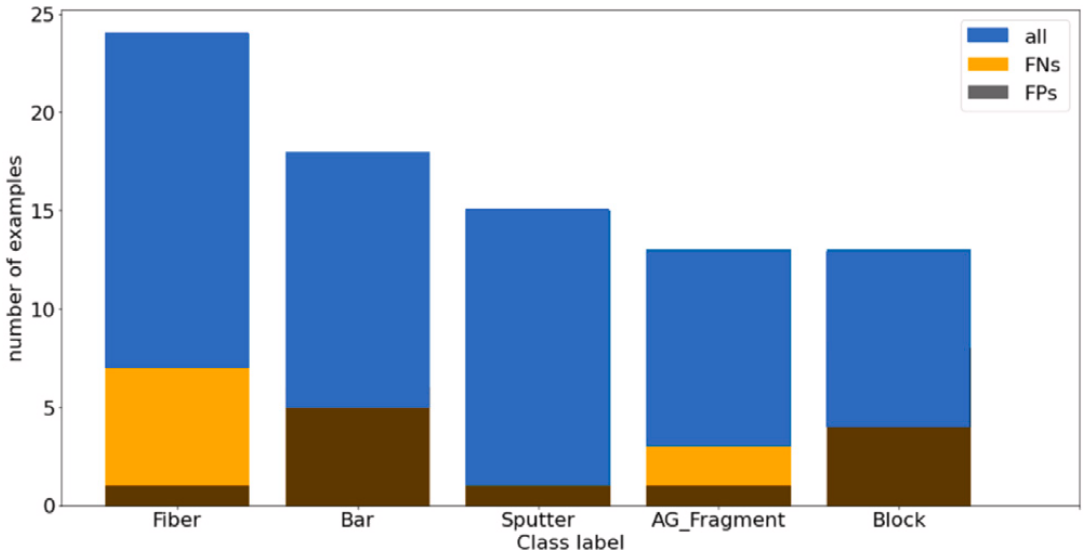


Figure 2.14: Distribution of FPs and FNs over their classes.

It is evident that the model performs best on detecting and classifying sputter which has the fewest FPs and no FNs. From these values I compute the evaluation metrics, with the model attaining a precision, recall and F1 score of 81%, 83% and 82% respectively. Additionally, the Turi Create environment provides tools to easily compute the average precision and mean average precision of the model. These are summarised in Table 2.2.

Class	AG fragment	Bar	Block	Fibre	Sputter	mAP
Train	0.852	0.789	0.938	0.943	0.996	0.903
Test	0.629	0.726	0.698	0.819	0.992	0.773

Table 2.2: Average precision 50 (IOU > 0.5) and mean average precision from the Turi Create environment of the trained YOLO-ET applied to the train and test data from Tanpopo Project aerogel panels. The prioritised classes are grouped morphologically as aerogel fragments, bar, block, fibre and sputter, and likely represent surface residual effects of material released by ISS docking and undocking activities, venting materials, secondary impacts from primary impact ejecta, possible spacecraft component fibres, and fragments of the aerogel itself.

These results well surpass the targets set by the Tanpopo Astrobiology Project team of a relatively modest 70% or better recall for surface objects detected on the aerogel panels returned from the ISS, to help ensure that objects of interest and/or entry points to greater depths in the silica aerogel were not being missed, enabling hypervelocity microparticle tracks to be flagged and further scrutinised at different magnifications, points of view, and focal depths. In practice human observers of inanimate objects outperform most convolutional neural networks unless fine-tuned, with overall accuracies of 90% [106], and YOLO-ET shows performance at these human-comparable levels. At least as notable however is the savings in human labour and computing resource with the implementation of YOLO-ET, and its on-device capabilities for ready real-time use in both laboratory and field environments.

The F-sample dataset is proprietary to the Tanpopo Project team and currently no other published work exists for comparison, however, in prior work, before my development and implementation of the YOLO-ET system, a VGG-16 network was used but for object classification not detection, and Tanpopo aerogel surface object

images at 245x magnification obtained by the CLOXS system were cropped from 704 × 480 pixels to 224 × 224 pixels and used to train a model with highest possible recall.

On the cropped images for most of the surface object categories the project goal of 70% or greater recall was met using the VGG-16 network: 93% for Blocks, 92% for Fibres, and 87% and 83% for Sputter and Bars respectively. Recall performance on Aerogel Fragments was poor however, at 29%. To improve performance a series of image pre-processing tasks were conducted: first a Zero Count Analysis (ZCA) whitening transform [88, 89] was employed to accommodate the different colour hues of the aerogel panels, and next a thresholding sequence, where images were converted to grey scale, noise-reduced, and pixels re-filled against various thresholds, to help create more distinct object boundaries. Both ZCA and thresholding techniques brought Sputter recall to 95%, but there was a somewhat poorer recall performance with thresholding: in the 72%–79% range for Fibres, Blocks and Bars, and still under 70% for Aerogel Fragments, at 62%; ZCA for Aerogel Fragments achieved still only 33% recall.

The low recall rate using the VGG-16 model even with intensive pre-processing, for just one of the five Tanpopo surface object categories necessitated continuing need for human inspection of all panels. The YOLO-ET model and system were thus developed to achieve Tanpopo project recall performance across all categories, without the need for labour-intensive cropping and pre-processing techniques, implemented on a faster network and model that could allow real-time automatic capture on mobile devices.

To validate these results it is best compare the YOLO-ET model to the performance of VGG-16 on unprocessed data (i.e. no thresholding/ ZCA). Turi Create does not natively support K-fold cross-validation (CV) as a built-in feature however, but for purposes of direct comparison with the prior study, I manually created the folds for 4-fold CV and calculated the average Precision and Recall across all folds.

For 4-fold CV, the dataset was divided into four equal segments. I then constructed and evaluated four distinct models, each trained on a unique combination of three segments for training purposes, thereby ensuring that every segment is utilised once as a testing set. Note that this results in slightly less data available for training the model (75% versus 80%), but the implementation of CV allows checking for robustness in the model, which is important when no validation dataset is available.

With no adjustment to Intersection over Union (bounding box overlap thresholds), YOLO-ET trained with this slightly smaller dataset demonstrated comparable if not better performance on Precision but slightly worse on Recall. Note however the models are not directly comparable as the VGG-16 study was an image classification task, where the objects were perfectly centred and cropped down. My model introduces the additional complexity of object localisation, where images can contain more than 1 object and are not necessarily centred (see Figure 15).

The performance accuracy of Image Classification-only models is generally higher than with Object Detection (see e.g. Lin *et al.* [107]). But my augmented

training dataset is more diverse, and furthermore I note that the false negatives in the image classification model VGG-16 are defined as the number of objects incorrectly classified, whereas the false negatives in my model are both the number of objects that are incorrectly classified and those that are not detected.

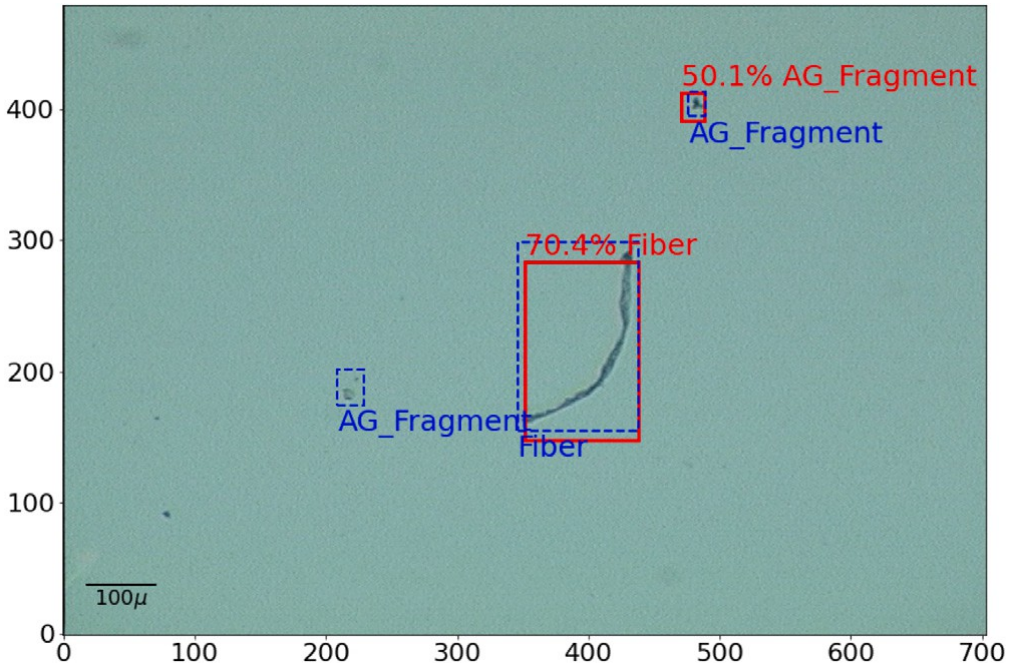


Figure 2.15: Test image with multiple objects. True labels are shown as dashed blue boxes and model predicted by model CV-3 shown in red with confidence scores.

The balance between Precision and Recall is a trade-off, and setting for example a lower confidence threshold and IoU score would typically result in a higher Recall rate whilst reducing Precision. Despite the bigger challenges faced by object detection compared to image classification, by dropping the IoU threshold to 0.3 and the confidence score threshold to 0, the performance of YOLO-ET evaluated

on Precision and Recall exceeds that of VGG-16 across all folds (See Figure 16 and Table 2.3).

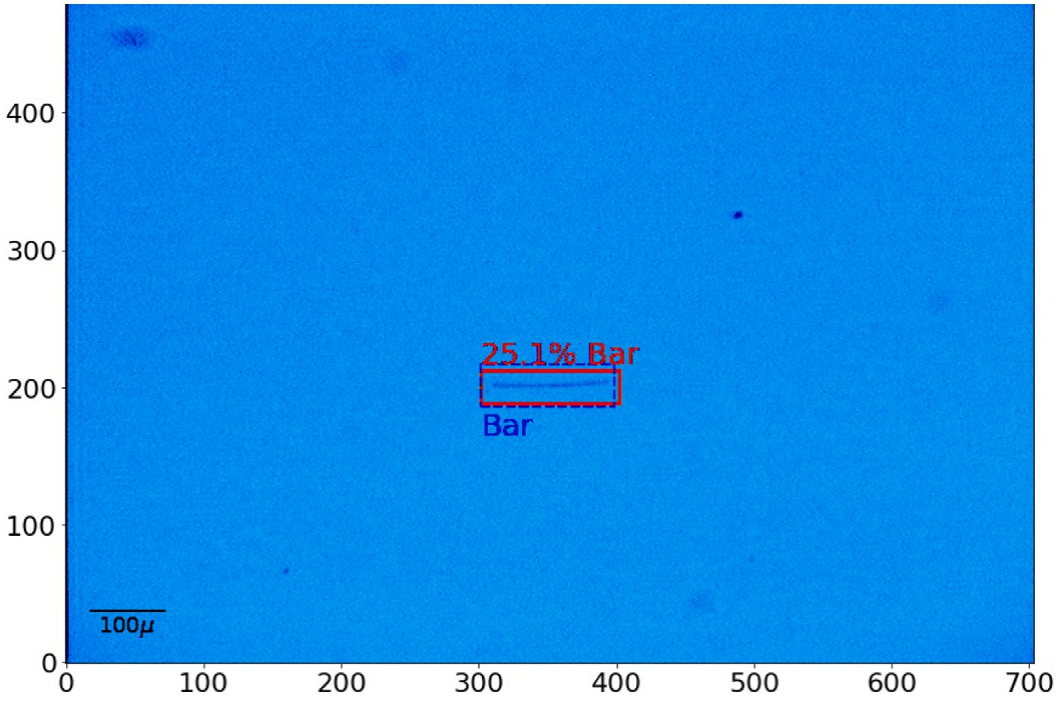


Figure 2.16: Test image of an object with a low IoU threshold and confidence score of 25.1% that would not typically have made the threshold for positive detection. True labels are shown as dashed blue boxes and model prediction shown in red with confidence score.

Jaeger *et al.* [98] also offer a useful contrast between Precision and Recall automatic detection of impact craters on aluminium foils utilising a Convolutional Neural Network (CNN), for Image Classification purposes. Their approach simplifies the problem to binary classification, focusing solely on distinguishing circular craters. Despite this simplification, Jaeger *et al.* employ synthetic data to train their model. This reliance on synthetically generated craters facilitates their

model in achieving an impressive precision rate performance of 99.8%. However, it is crucial to note that the model's Recall rate stands at 66.7%. This discrepancy between high Precision and relatively lower Recall underscores the challenges inherent in balancing these metrics, particularly when training AI models on synthetic versus authentic datasets.

Class	AG frag	Bar	Block	Fibre	Sputter
VGG-16	0.857	0.811	0.818	0.710	1.000
CV-0	0.929	0.938	1.000	0.944	1.000
CV-1	0.917	1.000	0.933	0.900	1.000
CV-2	1.000	1.000	1.000	1.000	1.000
CV-3	0.824	0.857	0.963	0.947	1.000
CV-avg	0.918	0.949	0.974	0.948	1.000
CV-std	0.063	0.059	0.028	0.035	0.000

Comparison of Precision using a VGG-16 image classification model and my 4-fold cross-validation runs. CV-avg and CV-std are the average and standard deviations of the cross-validation runs.

Class	AG frag	Bar	Block	Fibre	Sputter
VGG-16	0.286	0.833	0.931	0.917	0.872
CV-0	0.812	1.000	1.000	0.850	1.000
CV-1	0.957	1.000	0.933	0.818	1.000
CV-2	0.875	0.500	1.000	0.833	1.000
CV-3	0.933	1.000	0.963	1.000	0.950
CV-avg	0.894	0.875	0.974	0.875	0.988
CV-std	0.056	0.217	0.028	0.072	0.02

Comparison of Recall using a VGG-16 image classification model and my 4-fold cross-validation runs. CV-avg and CV-std are the average and standard deviations of the cross-validation runs.

Table 2.3: Precision and Recall comparisons using a VGG-16 image classification model and my 4-fold cross validation runs.

2.4 Chapter 2 Conclusions

The use of Computer Vision and Machine Learning in the search for Arkhipov Particles strongly complements how imminent robotic and human activities on the Moon and other planetary bodies could also benefit from these advanced capabilities *in situ* to identify and quantify microparticle terrestrial contaminants, lunar regolith disturbances, the flux of interplanetary dust particles, possible interstellar dust, β-meteoroids, and secondary impact ejecta, as further highlighted in Chapter 5 ahead. The YOLO-ET algorithm, an innovation in this field described in this Chapter, fine-tunes Tiny-YOLO that can potentially help to specifically address these challenges. Designed for coreML model transference to mobile devices, the algorithm facilitates edge computing in space environment conditions. YOLO-ET is deployable as an app on an iPhone ProMax with LabCam® optical enhancement, ready for lunar ruggedisation. Training on images from the Tanpopo aerogel panels returned from Japan's Kibo module of the International Space Station, YOLO-ET demonstrates a 90% detection rate for surface contaminant microparticles on the aerogels, and as described in Chapter 3 shows promising early results for detection of both microparticle contaminants on the Moon and for evaluating asteroid return samples, a first step toward searching for traces of Arkhipov particles or their impact signatures. YOLO-ET is now ready to be adapted to real-time analyses in the JAXA ISAS Sagamihara receiving labs, and its tasking to track hypervelocity impacts through silica aerogels can be directly useful to deploying similar capture panels on the Moon and throughout the Solar System, a practical means to capture and detect Arkhipov particle impacts as described in Chapter 5.

Chapter 3

Deploying YOLO-ET

3.1. Deploying the model[28]

Using the methods developed for this Thesis, I have demonstrated that applying AI Machine Learning to 2D aerogel images with YOLO-ET greatly speeds up and simplifies the identification, localisation, and classification of Tanpopo aerogel-captured surface particles. I believe this is an important first step toward identifying potential Arkhipov Particle impacts, particularly anticipating that such collecting panels may become widespread in use on all future spacecraft deployed

[28] Like Chapter 2 much of this Chapter is based on the full-length article I published April 2024 in *Astronomy and Computing* attached here in this Thesis at the Special Appendix.

throughout the Solar System: a rapid and effective means of detection and classification will be key. As described in Chapter 2, the advances made with YOLO-ET and the Tanpopo images have been demonstrated to speed the object detection training process, improve accuracy, and consume fewer computing resources, and as will be seen in this Chapter, all while taking advantage of the in-built optics and compact form factor of a mobile device. The model requires no pre-processing of the data. However to truly realise the potential of this model for space missions, it is imperative to address the dependency on traditional microscopy. While the CLOXS system represents significant capabilities in that it can automatically relocate the coordinates of objects of interest and centre the stages accordingly, the YOLO-ET system can now greatly augment these capabilities by identifying, localising and classifying objects on the first pass in real time. The YOLO-ET core Machine Learning models can be readily translated to a mobile device, in this project as an App, allowing the iPhone's camera, enhanced by a LabCam® adaptor, to act as the object detector for untrained, real-world images, with the self-contained iPhone and App able to bound and classify new images based on the core Machine Learning models developed. I propose the integration of this model into a mobile application for both laboratory and space environments, harnessing the capabilities of widely accessible technology like smartphones. This integration marks a significant step towards edge computing, where data processing is performed at or near the source of data generation.

3.1.1 Adapting and Integrating LabCam® to CLOXS

Adapting from the most recent developments in field research [e.g. 108, 109], I selected the iPhone ProMax 12 (Table 3.1) and the LabCam®[29], a user-friendly combination, for the initial deployment (Figure 3.1).

Camera	iPhone Pro Max 12	
Ultra-Wide	12 MP	f/2.4
Wide	12 MP	f/1.6
Telephoto	12 MP	f/2.2

Table 3.1 iPhone ProMax 12 Specifications

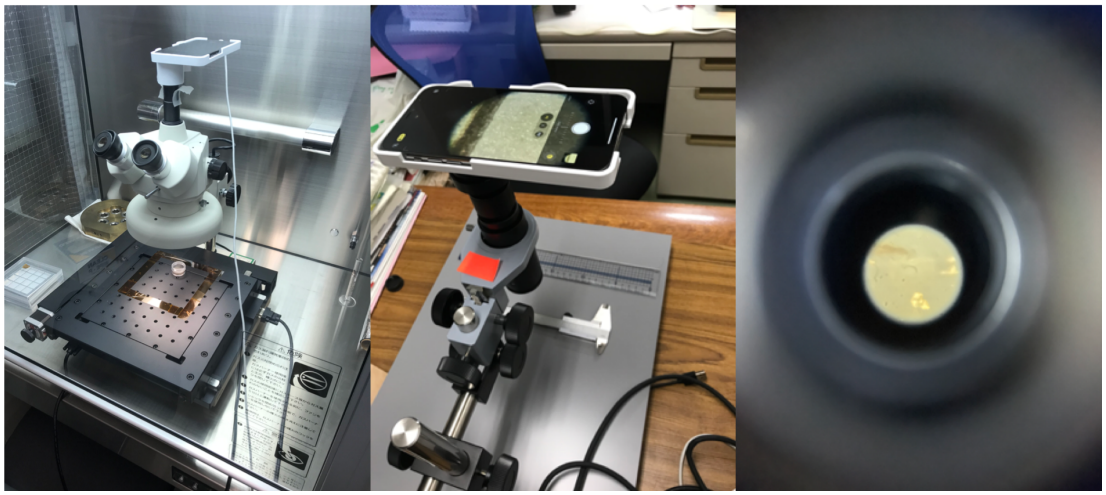


Figure 3.1: Adapting the LabCam® and iPhone Pro Max 12 into the ISAS CLOXS system. *Left*: Clean bench set up to start single-operator imaging and analysis. Note the LabCam® mount, and 50mm traverse motorised xy stage, 100mm traverse stage and joystick controls. *Centre:* Hozon Co. Ltd. microscope and LabCam® mounted with iPhone Pro Max 12 for calibration. *Right:* Looking remotely 'down the hole' of the iPhone Pro Max 12 at a magnified block particle fragment in the aerogel.

The iPhone's advanced camera system, processing power and sophisticated autofocus technology make it an ideal choice for capturing high-quality images of

[29] https://www.ilabcam.com/

microparticles. This autofocus feature is critical for my application, as it ensures that images are sharp and highly detailed, facilitating accurate Object Detection without the need for manual focus adjustments. Moreover, the convenience, portability and widespread availability of iPhones offer practical advantages for replicating my methodology across diverse settings, particularly fieldwork and applications in resource-constrained environments. While alternatives such as small PCs and specialised AI cameras exist, the iPhone's integrated ecosystem and the availability of sophisticated development tools in Turi Create make it an attractive choice for implementing advanced AI-driven object detection tasks directly on the device.

Meanwhile, the LabCam® attachment enhances the iPhone's capability to function as a makeshift microscope. In effect, it is a portable microscope that can be easily taken to the sample, rather than the other way around. This makes it ideal for real-world and real-time *in situ* detection of say micro-particle contaminants on the lunar surface. With the ability to capture images with up to 100x magnification in integration with an iPhone alone, the LabCam® provides an easy to use, ready system for microparticle detection. I am now preparing these capabilities for translation to real-time laboratory examination of aerogel panels at Japan's Institute of Astronautical and Space Sciences (ISAS), to identify, localise and classify 3D tracks and hypervelocity impact particles candidates across the inventory of existing space-flown aerogel panels as well as in future Astrobiology Project Tanpopo aerogel panel returns from the International Space Station, using my mobile on-device Machine Learning models.

The combination of the iPhone Pro Max 12, LabCam® mount and integrated magnification, coupled to the CLOXS system at ISAS allows real-time processing to a GPU-equipped desktop running my YOLO-ET algorithms and core Machine Learning model – a full surrogate of what could be packed and space-hardened into *in situ* and in-spaceflight missions. With additional optics added to the system with LabCam®, the CoreML models developed and experiments described in this Thesis demonstrate both real-world laboratory identification and classification of extraterrestrial microparticles and autonomous edge-computing capabilities for future spacecraft missions to detect, localise and classify them.

Thus a unique advantage of the methods employed in this Thesis is that the YOLO-ET imaging and Machine Learning processing and classification is self-contained, with 'on-board' GPU processing whose form factor and computing power can be readily incorporated into small spacecraft. With these applications in mind, further experiments were conducted on identifying and classifying granulated microscopic spacecraft materials distributed atop lunar regolith simulants, as a surrogate for *in situ* detection of anthropogenic contaminants on the lunar surface. Finally, based on the newly returned 'ground truth' of Scanning Electron Microprobe (SEM) images obtained from the asteroid Ryugu samples returned by Hayabusa2, as a further demonstration the model was trained and tested to establish potential correlations with SEM images from the suite of micrometeorites obtained from the TransAntarctic Mountains.

3.2. Experiments with Spacecraft Microparticles on Lunar Simulants

An anticipated future use of the CLOXS imaging system as adapted and coupled with YOLO-ET in this work is examination of Tanpopo-like aerogel panels deployed and retrieved from the lunar surface. Together with Professor Yano, I am developing a mission concept for aerogel panel deployment to the Moon as early as the Artemis lunar landing missions (see Chapter 5 Future Work), with important opportunities to advance not only the core Tanpopo astrobiology objectives, but to collect more information and contribute to studies on microparticle anthropogenic contaminants, lunar regolith disturbances by human activities on the Moon, the flux of interplanetary dust particles, β-meteoroids, possible interstellar dust, and secondary impact ejecta (see e.g. [110, 111, 112, 113].

Ultimately a 'Mini-CLOXS' could support both post-retrieval examination of aerogel panels returned from the lunar surface and other missions to Earth laboratories, as well as *in situ* examination on lunar and other planetary surfaces. A timely factor in bringing these capabilities to the Moon is to help establish a baseline for forward contamination caused by robotic and human activities there. A key element of potential forward contamination are particles of spacecraft, experimental packages, communications equipment etc. that may be deposited and distributed around the Moon by (i) routine operations, including outgassing of propellants and spacesuits, mechanical interfaces, vehicle track and wheel

movements, etc.; (ii) natural material degradations from micrometeoroid bombardment, day/night temperature cycles, interaction with the solar wind etc. and (iii) larger scale de-orbited and hard-landed material.

With these in mind I have taken the YOLO-ET algorithm developed and trained on surface particles captured on the Tanpopo aerogels, and used the same Turi Create, iPhone, and LabCam® system described above to test its capabilities for imaging and identifying spacecraft remnants mixed into lunar regolith simulants. Samples of both JSC-1 (see Appendix 6.1 for properties and composition) and Manna Electric lunar simulant (Appendix 6.2) of 0.05 g each were prepared and evenly and separately deposited into 2 cm diameter plastic vials. Test model portions of the CesiumAstro Nightingale satellite (Figure 3.2) whose

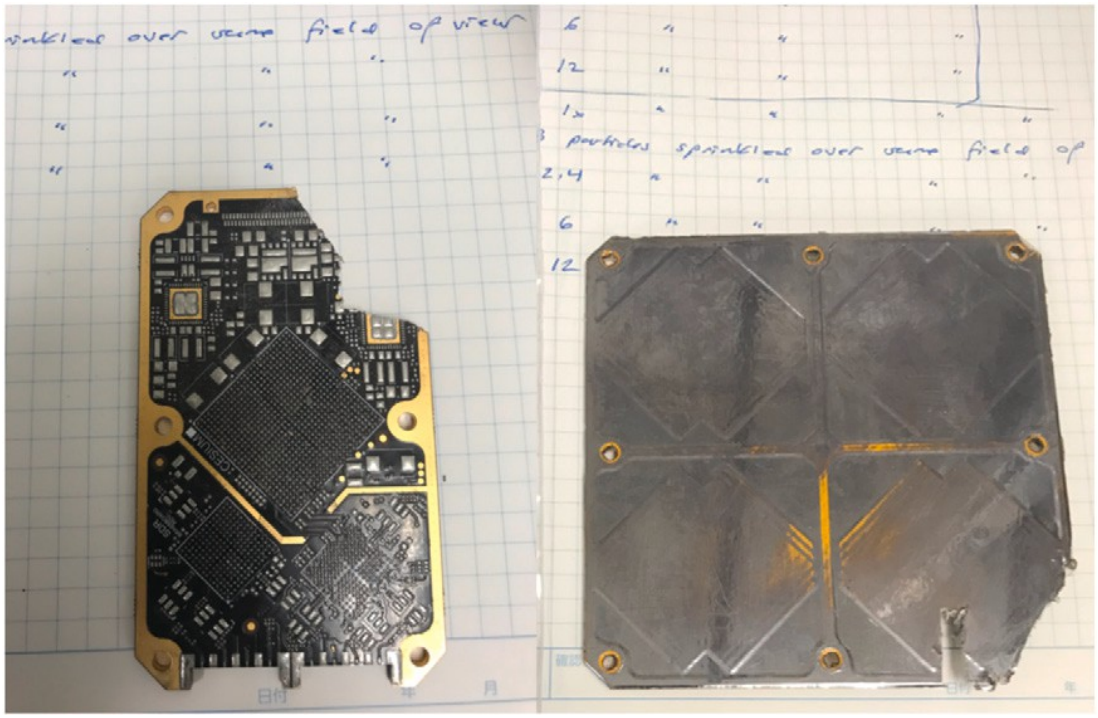


Figure 3.2: *Left:* Software-Defined Radio (SDR) board, Polyimide Arlon 85 with SAC 305 solder and copper layers, with impregnation material. *Right:* Nightingale Antenna board, Rogers material with copper and silver coating and STAMET radome.

compositions are detailed in Appendix 6.3 were particulated with a band saw, producing spiralled fragments that were then sieved to 80 μm or less to approximate average JSC-1 grain sizes, to set the challenge to distinguishing similar sized fragments amongst similar sized grains (Figure 3.3).

YOLO-ET first trained on a limited data set of images of 80 μm diameter glass beads, and then drilled fragments of spacecraft grade Al, Ti, and CFRP, also sieved to 80 μm or less, in various lighting conditions and magnifications, set atop the lunar simulant deposits. Utilising the specially combined optics of a 4x Hozon lab microscope, a 10xLabCam®, and an iPhone Pro Max 12 with 1x-12x magnification, the overall range is 40x to 480x. Optimal experimental resolution for early training and ground truth experiments in YOLO-ET are 100x; with next-generation LabCam® and iPhone Pro Max 15, 225x magnification can be anticipated using purely analogue optics, highly suitable for spacecraft deployment.

At the optimum magnification of 100x, lighting conditions and depth, for which an *in situ* Mini-CLOXS would be designed, early YOLO-ET training and ground truth experiments have been demonstrated to show ready identification of Nightingale antenna particles and Software Defined Radio board microcircuitry particles (Figure 3.4), and once optimised the resulting model can be directly exported to CoreML format, for streamlining integration into the on-device application. The initial training and detection returns results

3.2. Experiments with Spacecraft Microparticles on Lunar Simulants

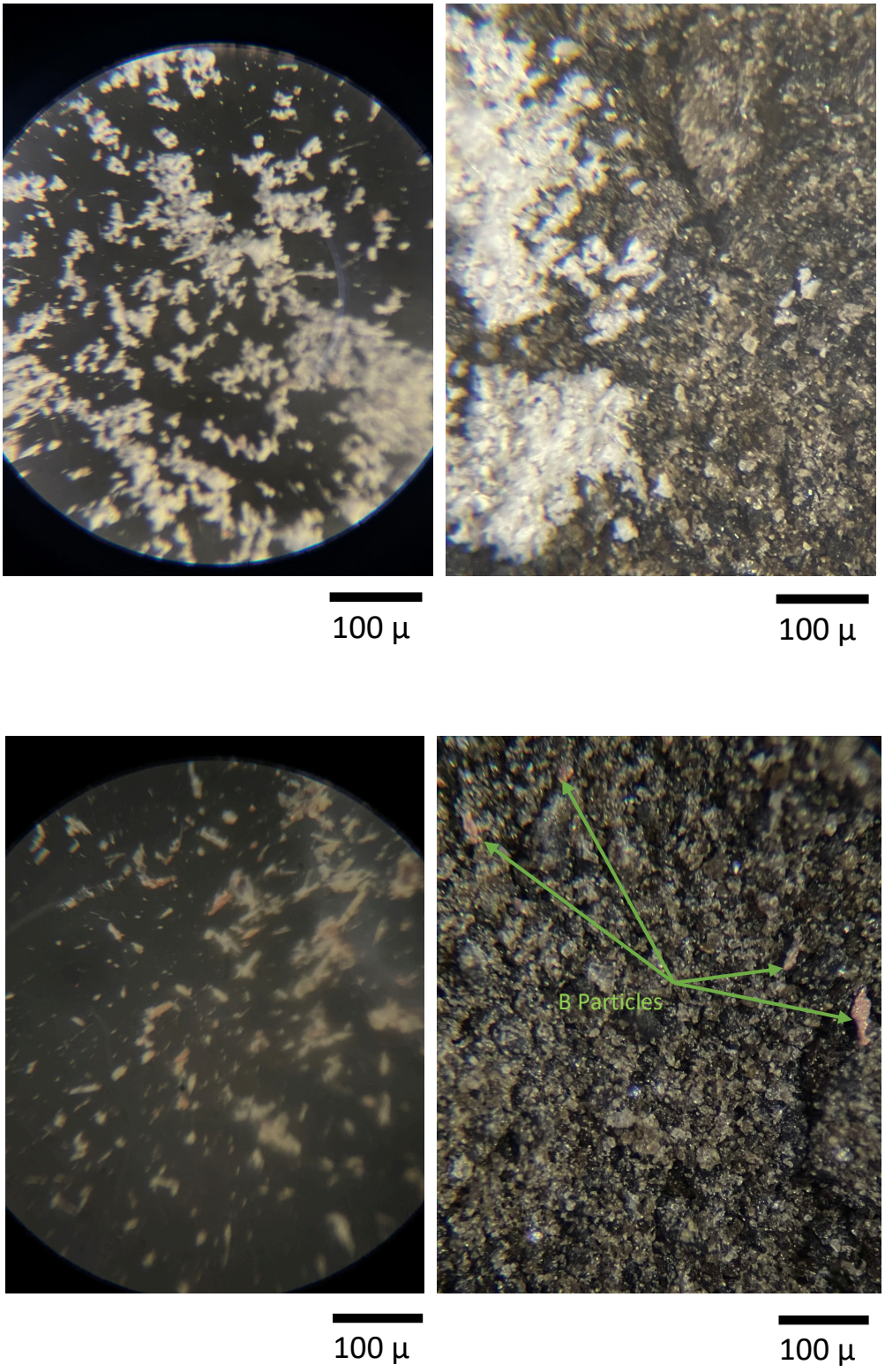


Figure 3.3: Samples of JSC-1 lunar simulant of 0.05 g each were prepared and evenly deposited into 2 cm diameter plastic vials. Test model portions of the CesiumAstro Nightingale satellite were particulated with a band saw, producing spiralled fragments that were then sieved to 80 µm or less to approximate average JSC-1 grain sizes, to set the challenge to distinguishing similar sized fragments amongst similar sized grains. Clockwise: CesiumAstro A particles from the antenna board; A particles on JSC-1; B particles from the Software Defined Radio board; B particles on JSC-1.

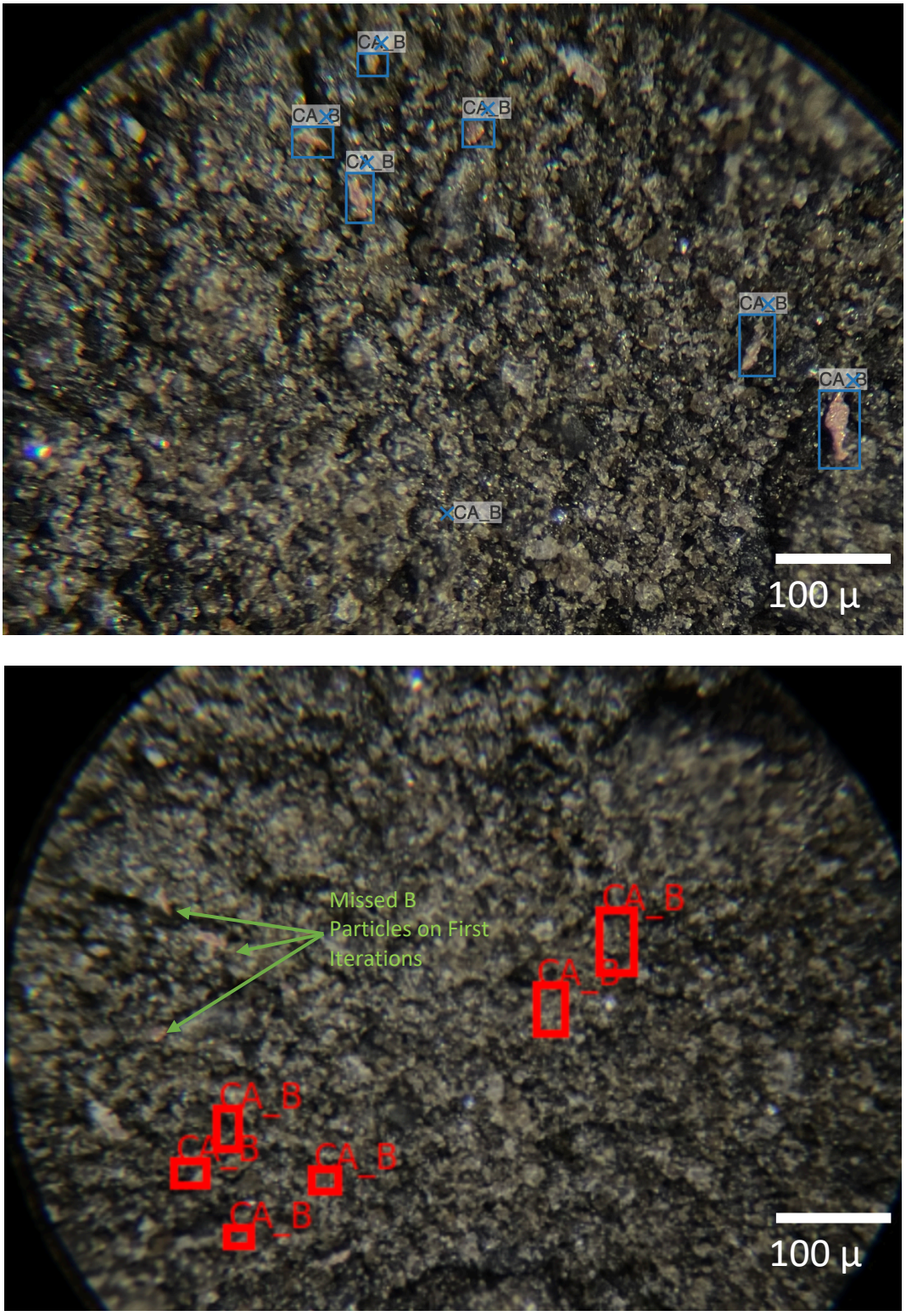


Figure 3.4: Early applications of YOLO-ET trained on known 'B' particles of CesiumAstro Nightingale Software Defined Radio Board automatically identified some unseen 'B' particles atop JSC-1 Lunar simulant but missed others. Top: BBox labelled particles; Bottom: Test result (different portions of the sample are shown).

similar to the early ground-truth training employed for the Tanpopo panel-surface objects as described in Chapter 2, despite the much more complex background of the regolith simulants. There is much room for improvement however, and further trials with more precise polygonal segmentation than provided by the BBox widget, as used e.g. in medical imaging using U-Net architectures, will be explored.

Such detections of spacecraft contaminants may bear promise for directly detecting Arkhipov Particles and remnants of Arkhipov-Bracewell Probes on the Moon. At least some portion of direct asteroid and cometary impactors on the Moon have been shown to likely survive (see Joy *et al.* 2012 [114] and Ong *et al.* 2009 [115]. Realistically however as discussed in Chapter 2 a larger portion of Arkhipov particles arriving at hypervelocity are likely to vaporise on impact. So to extend these experiments to distinguishing naturally occurring non-indigenous artefacts from manufactured debris, simulating the impacts of some of the materials from the above experiments is now in the planning phase. It is possible that the hypervelocity impacts of each may leave characteristic microcraters, with distinctive morphologies. Initial experiment designs are planned for the two-stage gas gun at University of Kent Canterbury (see Figures 3.5 and 3.6).

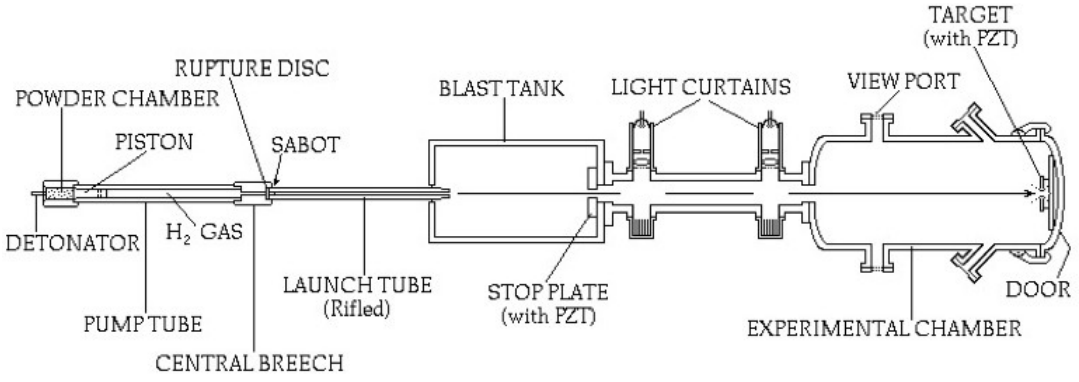


Figure 3.5: Two-stage gas gun schematic [116]

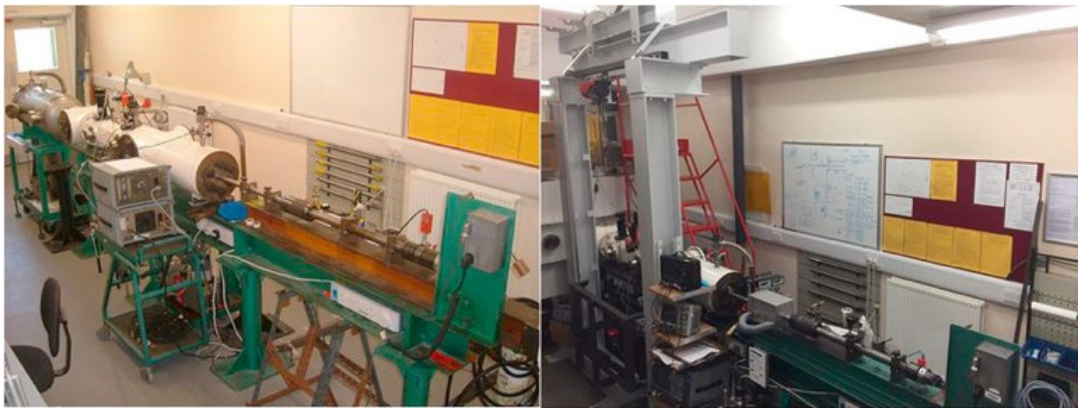


Figure 3.6: Two-stage gas gun set up at University of Kent Canterbury; Left, horizontal gun; Right, with vertical axis gun framework emplaced. Horizontal barrel length is c. 5 metres. (see Hibbert 2017 [116] and Burchell 1999 [117].

I procured a set of single-crystal Aluminium targets[30], so that shots of 80-100 μm natural projectiles (e.g. the glass beads I trained YOLO-ET on for the work on detecting particles atop lunar regolith simulant) and manufactured alloys and materials (e.g. starting with equal-sized titanium beads, and then leading to spacecraft materials). Projectile speeds from 2 to 7 km s^{-1} are planned for these future experiments. The use of the single-crystal Aluminium targets should allow for scaling what particle impacts of two orders of magnitude smaller projectiles might produce as characteristic craters. In general, the absence of grain boundary effects and other microstructural complexities in single crystals simplifies the interpretation of results, enabling more accurate modelling for extrapolation to smaller scales. Given the uniform isotropic mechanical properties of the single crystal Aluminium, its response to impact should be consistent in all directions, allowing for clearer analyses of the cratering process without the variability introduced by grain boundaries in polycrystalline materials. Varying the

[30] 10 x 10 x 1.0 mm Aluminium single crystal substrates in <1 1 1> orientation, one side polished, from MTI Corporation, Richmond CA, USA, product number mcALc101010S1.

orientation of the crystal lattice relative to the impact direction can also help determine how orientation affects the cratering process and scaling laws.

I anticipate that experimental data from impacts with ~100 μm impactors will provide insights into energy dissipation, deformation, and cratering mechanisms, which can then be correlated and scaled down to predict behaviours of Arkhipov-Particle-scale 1 μm impactors, both natural and manufactured. Data from these experiments can then help validate computational models of the impact cratering, helping to simulate cratering from the Arhkipov-Particle-scale projectiles.

YOLO-ET could then be used to identify, localise and distinguish characteristic microcrater morphologies. If microcraters produced by technological artefacts do produce characteristic craters, then YOLO-ET could be deployed for the detection of microcraters caused by hypervelocity Arkhipov Particles, with their spalled ejecta then examined for potential extraction of any remnant materials.

Potentially dust-sized grains of spacecraft hardware materials etc. could also be fired into cooled or frozen lunar simulant targets using Kent's two-stage gas gun, to simulate impact morphologies where water may be embedded in the lunar regolith. Previous experiments with the Kent two-stage gas gun typically used 3mm pellets (see e.g. [116]) – much larger than the dust grain scale planned for these experiments. Dust grains might however be agglomerated into a size that the gun can effectively accelerate. My plan

of work for the experimental set up includes the use of bulk shots of about 1000 grains each of 100 μm size, contained in 1mm diameter x 5 mm height cylindrical sabot interior. A variety of velocities would be tested, e.g. 2 km s^{-1} and 5 km s^{-1} and ultimately up to 7 km s^{-1} across the 5m horizontal gun barrel[31], to simulate both inter-planetary and interstellar velocity impacts. Given the significant artefact vaporisation that may be anticipated, spectral analyses of the impact flash across a range of wavelengths may be relevant, together with analyses of whatever is recovered in the target. Using the YOLO-ET algorithms developed and trained in Chapter 2, there is opportunity to improve the Convolutional Neural Network sequences done by Jaeger *et al.* [98] on automatic impact crater detection in the Stardust aluminium foils, beginning with adding their samples and images to the training, validation and test data sets.

In sum, I aim to train the YOLO-ET model I created and developed as highlighted in Chapter 2, to advance beyond its successes in detecting surface micro-objects on Tanpopo aerogel panels, to the detection and classification of non-indigenous micro-particles on the Moon, of both natural and artificial origin – both directly, and indirectly through their impacting micro-crater morphologies. As discussed above, first steps toward training the model on particles atop lunar simulant indicate this is a highly tractable challenge for YOLO-ET. My plan of work above for the two-stage light gas gun at University of Kent Canterbury will take particles from these experiments and other projectiles, and test, model and scale their impacts and remnants with the ultimate goal that the YOLO-ET

[31] Experiment's may also be conducted using the University of Kent Canterbury's vertical gun.

algorithm and its successive versions can reliably detect, localise and classify any Arkhipov Particles and or remnants of Arkhipov-Bracewell Probes on the Moon itself, directly or through their impacting crater morphologies.

3.3. Ryugu Asteroid Sample Experiments

Another avenue for potentially discovering Arkhipov Particles or remnants of Arkhipov-Bracewell Probes is as part of meteoritic materials that may have incorporated and processed them. It is possible that their parent body asteroids and the Earth-Moon system's own inventory of non-indigenous meteoritic materials may contain their traces. The Moon can offer both a larger quantity of samples unaltered by atmospheric re-entry, and more accessible sampling than the asteroids themselves, though micrometeorites raining through the Earth's atmosphere may land intact and unmelted. As a step toward testing the hypothesis that the YOLO-ET model might be trained to identify, localise and classify microstructures on the Moon, I trained and tested the YOLO-ET algorithm on the newly returned 'ground truth' of Scanning Electron Microprobe (SEM) images obtained from the Ryugu asteroid sample returned by Hayabusa2 [117], to establish potential correlations with SEM images from the suite of micrometeorites obtained from the TransAntarctic Mountains.

I participated in the initial inspection, sample handling and subsequent data interpretation and science discussions with a UK-Japanese team analysing samples of the Ryugu asteroid [118, 119, 120], and began experiments as part of this Thesis

to establish the potential for important time-and-manpower gains with YOLO-ET, as well to build toward detection capabilities for Arkhipov Particles and Arkhipov-Bracewell Probes.

Hundreds of nano-CT scans were conducted by the UK-Japanese team of the asteroid Ryugu sample A0108 to create segmented images whose cross-sections helped reveal microscale voids of particular interest [119, 120, 121]. Human-eye examination of void evidence in cross-section allowed re-integration of the images to reveal the voids in full dimension. This manual process of identifying and classifying evidence of voids in cross-section is ripe for ML identification, classification, and reintegration using the methods developed in this Thesis. I set out to test whether data from the Ryugu sample A0180 can now form a robust training data set for applying YOLO-ET to other more porous and aggregate samples from Ryugu, so that with time, the 3D optical images, nano-CT data, and external and internal SEM images from A0180 and other Ryugu samples can be archived to create further training data for searching the diversity amongst different groups of Ryugu samples and for practically comparing characteristic Ryugu micro-structures with for example structures from unmelted micrometeorites, which to the human eye seem to show strikingly similar characteristics and features (Figure 3.7).

Accordingly as a demonstration of these abilities YOLO-ET was also trained on SEM images of the Ryugu asteroid sample A0180, to establish ‘ground truth’ for characterising features of unmelted micrometeorites and unmelted portions of

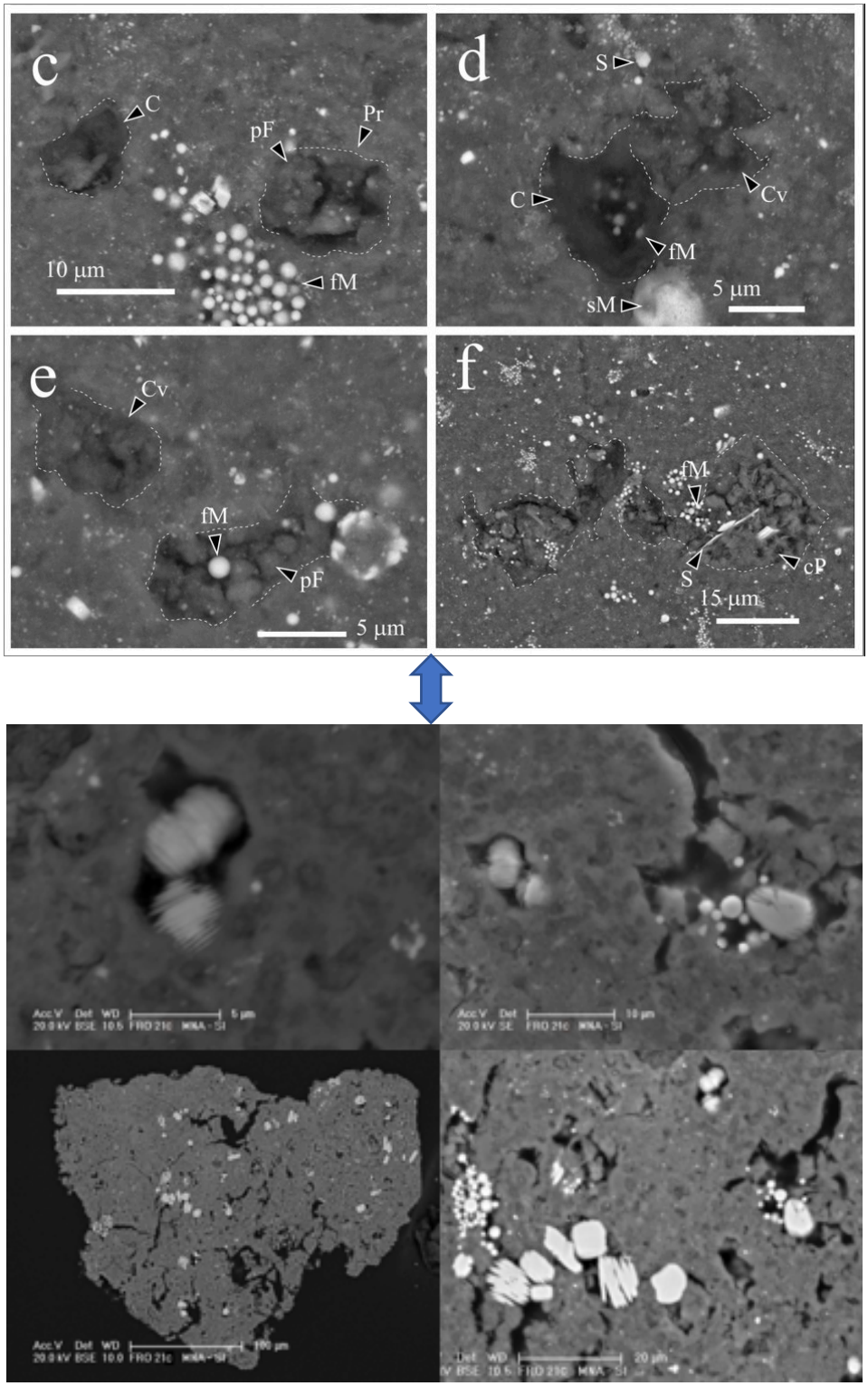


Figure 3.7: Exploring potential correlations in major features of indigenous asteroid samples Ryugu A0180, top, with those of unmelted micrometeorite samples from the TransAntarctic Mountain micrometeorites suite, below. *Image Credits:* Genge *et al.* [121] and van Ginneken *et al.* [122].

partially melted micrometeorites, using images of the TransAntarctic Mountains micrometeorite suite shared by the A0108 team (Figure 3.8).

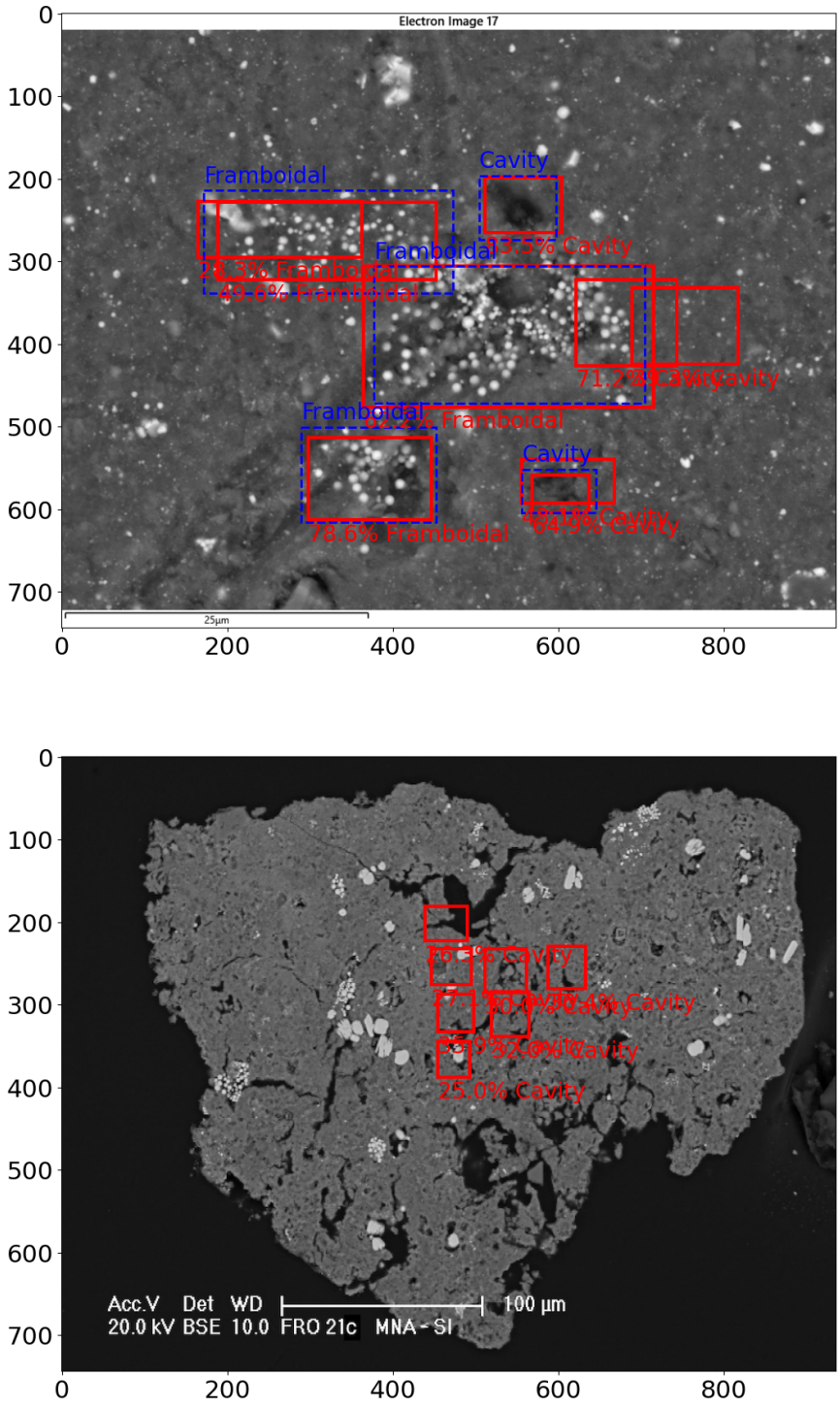


Figure 3.8: Early applications of my machine learning model YOLO-ET show that training on a Ryugu sample (top) automatically identifies similar features on unmelted TransAntarctic micrometeorites (bottom). *Image Credits:* Genge *et al.* [121] and van Ginneken *et al.* [122].

Unlike larger falls, micrometeorites are subject to less heating and alteration as they pass through Earth's atmosphere, and preserve important elements of their formation history and composition [122]. By comparing them for the first time to the ancient base line features of the indigenous asteroidal and cometary samples returned by spacecraft such as Stardust, Hayabusa, Hayabusa2 and OSIRIS-REx, it is possible to re-calibrate current assumptions about e.g. the proportional representation of CI chondrites amongst terrestrial meteorite collections. Many of the characteristic features of each sample image are notably subtle, and typically require considerable training and experience for research practitioners to deduce. In this Thesis YOLO-ET demonstrates capabilities for learning and establishing correlations amongst unseen micrometeoritic data sets. Distinguishable structural elements amongst images of micrometeorites with unmelted areas include: roughness and irregularities versus the smoother, glassy appearance of melted portions; differences in brightness reflecting compositional differences in backscattered electron images; micro-chondrules, mineral grains and inclusions indicative of parent body origins, otherwise obliterated in melted regions; and distinctive boundaries between melted and unmelted areas, with partial rims characteristically forming around unmelted areas.

Based on these types of parameters, as a proof of concept I trialled making selections of areas of interest on the Ryugu A0180 images, and using them to train and predict on a selection of the suite of TransAntarctic Mountain unmelted micrometeorites. My YOLO-ET model, trained on a limited dataset of 212 examples (without optimisation), detects features in unseen test data. While this showcases the model's potential, the current mean average precision (mAP) is only 10.1%, indicating significant room for improvement.

Based on my work undertaken in this Thesis with the Ryugu A0108 sample analysis team, Machine Learning Training using YOLO-ET has become part of the Ryugu sample analysis flow (Figure 3.9).

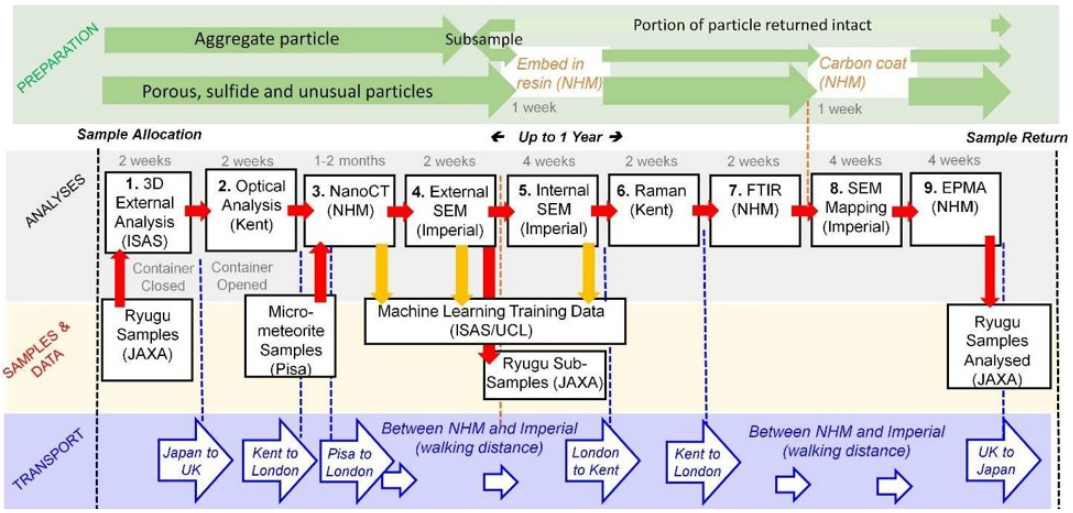


Figure 3.9: Based on the Proof of Concept work described in this Thesis, Machine Learning Training using YOLO-ET has been incorporated as part of the Ryugu sample analysis flow for subsequent proposals accepted by the JAXA Ryugu sample administration office. Source: JAXA Curation, The AO Administration Office.

3.4. Chapter 3 Conclusions

YOLO-ET demonstrates a 90% detection rate for surface contaminant microparticles on Tanpopo aerogels, and shows promising early results for detection of both microparticle contaminants on the Moon and for evaluating asteroid return samples. YOLO-ET's application to identifying spacecraft-derived microparticles in lunar regolith simulant samples and SEM images of Ryugu asteroid samples returned by Hayabusa2 indicate strong model performance and

transfer learning capabilities for future extraterrestrial applications. These are important first steps toward identifying: potential Arkhipov Particle impacts on collecting panels on the Moon and across the Solar System; characteristics of past Arkhipov Particle impacts on the Moon and other airless bodies, such as through their microcrater morphologies; and Arkhipov Particles or Arkhipov-Bracewell Probe remnants that may have been processed into meteorites. In the next Chapter 4, I look at the application of YOLO-ET at the macro-scale, for the potential detection of Arkhipov-Bracewell Probes.

Chapter 4

Experiments Detecting Spacecraft Hardware on the Moon

4.1. Overview

In this chapter I expand on Arkhipov's early ideas on the debris of other technological civilisations transiting the Galaxy as particles [123, 124], with particular emphasis on how such Arkhipov Particles might have been configured to develop into numerous self-replicating macro-scale probes, with the Moon as a natural catchment area for examination. I describe my early experiences with Moon Zoo in the course of this project, and its system for tagging human-operator-identified possible spacecraft hardware, and review my recent efforts searching for Apollo and Luna spacecraft hardware on the Moon using large-scale Machine Learning processes based on YOLO-ET.

I set out to focus on the application of YOLO-ET, the Machine Learning model I developed and optimised for mobile processing systems as described in Chapter 2, to the detection of spacecraft hardware on the Moon. My goal was to establish that YOLO-ET is well adapted to detecting, localising and classifying spacecraft hardware on the Moon from Lunar Reconnaissance Orbiter images. As highlighted in Chapter 1 of this Thesis, in 2013 Davies and Wagner had also identified LROC images as a potential data base for detecting alien artefacts on the Moon, and suggested 'computer automation' as a concept holding great promise, though limited by computer software that at that time could only search for specific features 'programmed in advance' [26]. In this Chapter I demonstrate that YOLO-ET, with its ability to use a very limited training data set to quickly establish robust detection confidence levels with light computing resource and high scalability, can now begin to meet the challenge. Trained on Apollo landing sites hardware, the model proves effective in detecting, localising and distinguishing Landing Modules, Lunar Roving Vehicles, and Other Spacecraft Hardware in unseen site images with robust levels of recall and confidence scores, and with minimal confusion with boulders, shadows and other natural artefacts.[32] As a means of establishing that the YOLO-ET model is not overly specific in its programming for these tasks, I also set out to test the Apollo-trained model on non-Apollo spacecraft hardware, and on LROC images of regions where spacecraft hardware might likely be found, but had not yet been identified.

[32] AI Machine Learning applications in planetary surface remote sensing have also made progress using YOLO models, e.g. for crater counting [125] and most recently very effective use of a U-Net CNN model, well suited to image segmentation tasks has also been applied to automated crater counting [126].

Accordingly I also applied my model to the general area where the Luna 9 spacecraft is believed to reside but whose precise coordinates have remained unknown, and the YOLO-ET model trained on Apollo spacecraft hardware successfully returns two candidate Luna 9 identifications from LROC imagery.

Given the lightweight integrated system and powerful on-board processing power demonstrated for mini-CLOXS microparticle-detection applications in Chapters 2 and 3, in this Chapter I also suggest how YOLO-ET and its system optimised for mobile processing could be incorporated into a comparatively inexpensive lunar satellite that could potentially scan the surface of the Moon in its entirety within a year, to complete a whole first catalogue of spacecraft hardware on the Moon. Finally, and back to the crowd-sourced inspiration of Moon Zoo, I conclude that by offering the YOLO-ET model for spacecraft hardware detection on the Moon to citizen scientists, it could complement such satellite deployments, and if the data proves homogenous enough, perhaps exhaustively mine the LROC archive to identify candidates for satellite remote sensing and *in situ* inspection.

4.2. Arkhipov-Bracewell Probes and the Moon

As evidence for an abundance of potentially life-bearing planets in the Galaxy significantly older than Earth becomes clearer, one increasingly compelling resolution of the Fermi Paradox is that the apparent absence of extraterrestrial technological civilisations owes to humans on Earth residing on a 'Zoo.' Maintained or observed or merely isolated by ExtraTerrestrial Intelligences (ETIs) that have kept themselves hidden from us, we may have been somehow

caged out from non-terrestrial civilisations. This 'Zoo Hypothesis,' as posited by astronomer John A. Ball in 1973 [127] may be the most logical alternative to ETIs otherwise being very rare or non-existent: it's the 'Zoo Hypothesis or nothing' per Crawford and Shulze-Makuch [36].[33]

The Moon is an attractive test bed to constrain each end of this hypothesis – allowing us to demonstrate either that ETIs are sufficiently rare or non-existent to leave not even microscopic traces of their past technologies on the Moon, or that if we are existing in a Zoo, allowing us to still find at least microscopic traces of their or others' earlier technological artefacts, or otherwise leave us to credit at least one ETI civilisation with astounding feats of Solar System shielding and/ or exhaustive clean-ups to sift out and remove even the tiniest traces of their evidence. As described in Chapter 1, Arkhipov Particles, the microscopic remnants that any early spacefaring civilisation is likely to have inadvertently dispersed across the Galaxy by now, would presumably require a fiendishly clever solution to preclude their detection on the Moon, perhaps via a self-replicating nanotechnology whose only task is to eat the trash then disappear. If an ETI developed space travel capabilities with velocities exceeding the dispersal waves of their own trash, then presumably they could set up a shield around the Solar System obscuring their penetration and existence (Crawford, personal communication 4 December, 2023) and avoid the clean-up task. Or if they were

[33] Others continue to follow Tipler's previously highlighted 1980 'Extraterrestrial Intelligent Beings Do Not Exist' [59], see e.g. Ellery's 2022 'The Prospect of Von Neumann Probes and the Implications for the Sagan-Tipler Debate' [128].

the dominant ETI per Bracewell [22] they thereby wouldn't have to worry about other civilisations' debris. But if there has been a plethora of earlier technological civilisations in the Galaxy, or even a few, and some number achieve early spacefaring but self-terminate before achieving interstellar travel or Solar System shielding or lunar housekeeping capabilities, there should still be Arkhipov Particles to find.

In Chapter 3, I described and set out the experiments conducted so far on images of surface contaminants collected on Tanpopo aerogels, spacecraft materials atop lunar regolith simulants, and microstructures detected in asteroid Ryugu and micrometeorites from the TransAntarctic Mountains; I also outlined my plan of work for experiments with a light two-stage gas gun, to evaluate possible characteristic micro-cratering that may distinguish natural and artefact impactors, all using variations of my YOLO-ET algorithm. Using these methods I aim to cover processes affecting Arkhipov Particles that are both endogenous and exogenous to the Solar System, including their littering the surface of the Moon as a result of impact gardening and then potentially further breaking down, or those arriving via high speed impacts from directly outside the Solar System.

If they exist on the Moon I believe it is now plausible that Arkhipov Particles can likely be found on its surface, by meticulous but now practical means. YOLO-ET video scans of large volumes of lunar material, as they are processed e.g. for any future lunar mining operations, would likely be able to detect Arkhipov Particles that arrived at non-destructive velocities, say from the trailing end of the Moon's orbital motions. And the experiments I've set out using a

hypervelocity gun and YOLO-ET may be able to determine whether the microcrater morphologies of particles that vaporise on impact offer characteristic clues about whether they are natural or artificial in origin, suggesting that spalled material around a candidate microcrater could be extracted for analyses to determine if it might be fragments of an Arkhipov Particle.

Discovering Arkhipov Particles on the Moon would offer immediate constraints on Crawford and Schulze-Makuch's 'Zoo or Nothing' hypothesis. 'Nothing' would be proven false. The Zoo hypothesis might also be weakened: even the confirmation of a single such particle might suggest that we have untidy zookeepers, or that they no longer care for or are no longer able to do the job, have limited powers, or that they are simply not there. If they are here with us in some form and anticipate such a discovery on the other hand, it may mark a particular milestone on our path to first ETI contact. The discovery of multiple Arkhipov Particles on the Moon, with isotope dating and composition indicating multiple sources and periods of impact, might be suggestive of a Galaxy that is fully capable of producing sentient life, early spacefaring technological civilisations, and their debris. But if debris is all we find, it may be that most ETIs are short-lived as spacefarers, and are unable or unmotivated to send intelligence-bearing artefacts – Arkhipov-Bracewell Probes – on interstellar pathways.

Even with the processing of massive amounts of lunar regolith material, aided by Computer Vision and Machine Learning Systems like YOLO-ET, finding Arkhipov Particles seems likely to mandate a scale of effort requiring major developments in human and robotic activities on the Moon's surface; it

will take time to industrialise as a process, and using YOLO-ET or successor programmes will have the added challenge of separating ETI debris from that of our own. It is possible that even relatively small volumes of lunar surface material could yield one or a small number of Arkhipov Particles, but even just a cubic metre of lunar material would likely require sifting 100's of billions of micron-sized particles to identify even one – not an impossible task using present technologies, but currently still a laborious one.[34] But the opposite is likely true for the discovery of Arkhipov-Bracewell Probes and their macro-scale products. Like Arkhipov Particles, if Arkhipov-Bracewell Probes exist, they should be here in our Solar System in some quantity, since even one early technological civilisation's use of them should have seen their Solar System arrival by now, per Tipler [59]. In a 'Zoo' scenario, such macro objects would be presumably easier to spot and clean up, but they should also be easier and quicker for us to find, as the experiments in this Chapter aim to demonstrate. As Crawford and Schulze-Makuch suggest, a determined search can discover the evidence.

YOLO-ET is well suited to the detection, localisation and classification of macro-scale objects, in both archived images and in real time. The core Model has been optimised so that objects of interest are detected with high recall over high precision rates, so that they are not as likely to be missed, and also fine-tuned

[34] Presently I am preparing an article as lead author together with Professor Crawford and researchers at the SETI Institute to estimate the volume of Arkhipov Particles that may have arrived in the Earth-Moon system from other parts of our Galaxy, based on a range of scenarios from a small number of technological civilisations with local asteroid mining operations, to multiples of Dyson-sphere scale construction activities. In our preliminary analyses even the low-end scenarios produce some detectable volume of Arkhipov Particles on or within the first metre of the Moon's surface, and the medium and higher-end scenarios suggest that a practical search may be conducted within a sample volume of the Moon's surface of a cubic metre or less.

in a way that real-time use in both Earth-based receiving laboratories and on the surface of the Moon and other space environments is practicable. The customised Model's capabilities are thus also well suited to prioritising detection of extraterrestrial macro-scale objects, both in archived images and in real-time orbital and fly-by missions, where the ability to discard non-detection data before transmission can prove valuable for optimising both bandwidth and time for further analyses.

Thus YOLO-ET can also be deployed for the detection of Arkhipov-Bracewell Probes or more likely, their remnants. As discussed in Chapter 1, just as Earth science and engineering draws ever closer to the creation of programmable matter at submicron scale, it is ever more conceivable that submicron particles of programmed matter could be designed to survive the space environment, and travel to the ends of the Galaxy via stellar winds and radiation pressure, without the aid of artificial propulsion systems. If they are designed to survive impacts with bodies encountered in space, and perhaps utilise impact energies to trigger and disperse their programmed elements, then much as DNA guides the growth of functioning organisms on Earth, such particles' components could be programmed to utilise materials they encounter to grow macro-scale elements, such as sensors and manipulators and processing, mobility and communications systems. Such Probes offer a relatively cheap means to explore the Galaxy, and could potentially be released in their many billions even from a short-lived technological civilisation.[35] If as part of their programmed function, when encountering

[35] I am also preparing a mission proposal with SETI Institute researcher Dr Peter Jenniskens to test dispersal patterns of inert micro-particles as a first step toward Arkhipov-Bracewell Probe production.

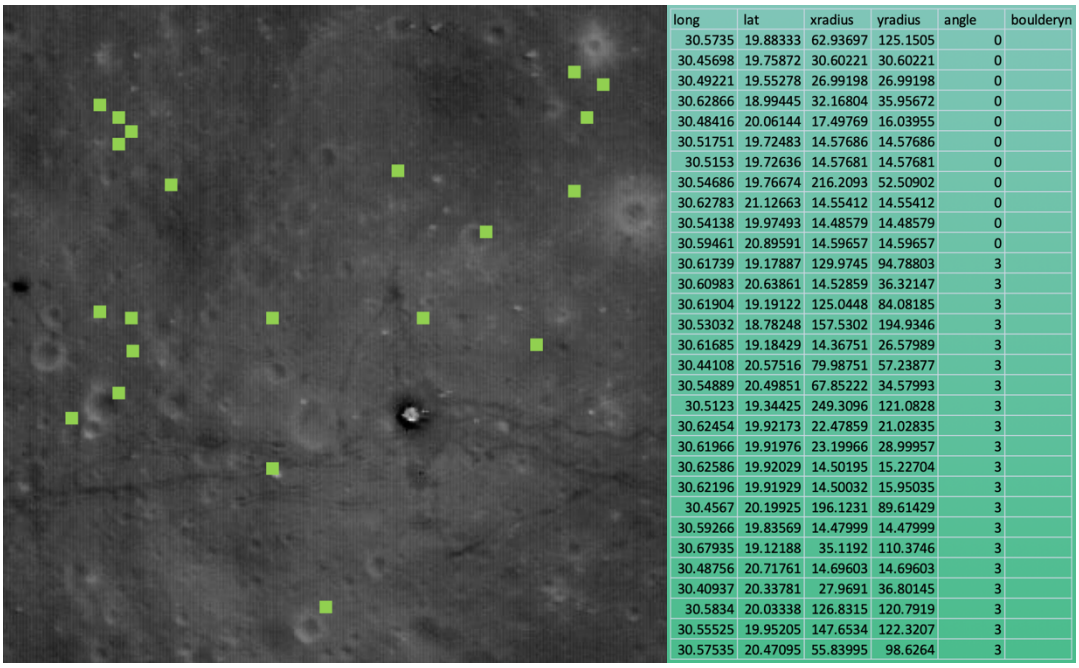
resource materials they create and disperse yet more Arkhipov-Bracewell Probes à la von Neumann machines, as established per Bracewell in 1960 [22], Tipler in 1980 [59] and others to the present day including Ellery [128], then the whole of the Galaxy could be covered efficiently in a few hundred Mega years, even if only one technological civilisation in the Galaxy were to undertake the exercise.

4.3. Examining Images of the Moon for Spacecraft Hardware: Prior Efforts

Thus as also discussed in Chapter 1, given that the Moon would be a natural collecting plate for Arkhipov-Bracewell Probes as well as Arkhipov Particles, over a span of more than 4 Giga years, it is attractive to deploy YOLO-ET to search for macro-artefacts on the Moon, since it is plausible that even if only one technological civilisation dispersed Arkhipov-Bracewell Probes, their remnants should now be detectable. Lunar Reconnaissance Orbiter and its Narrow Angle Cameras (LRO and NAC, and collectively, LROC) have been providing continuous imaging of the lunar surface since 2009, with millions of images capturing swathes of the surface at resolutions as fine as 0.5 to 1.0m, and using super-position techniques, as small as 0.25m per pixel [129]. Earth spacecraft and equipment are clearly visible in these images, and LROC featured images include detailed views of Apollo landing site artefacts, including descent engines, experimental packages, rovers and their tracks; Surveyor, some but not all Luna, and other soft landers; and numerous impacts of rocket stages and other de-orbited materiel [130, 131]. Over Mega years, these features and materials will be eroded

and churned into the surface regolith by micrometeorite bombardment, Solar wind and other influences [132]. But if an Arkhipov-Bracewell Probe arrived and developed on the Moon in the last c. 100 Mega years, and if it ever had features and edges distinguishing it as artificial, it might be similarly identified in LROC resolution images. There are however millions of LROC images under varying illumination, altitude and resolution conditions, and 40 billion square metres of the Moon covered, any one square metre of which might give evidence of technological artefacts. Distinguishing these and other features is thus a natural target for Computer Vision and Machine Learning, and the LROC team and others have begun to employ Machine Learning systems to classify features of interest on the Moon.

Moon Zoo was an early citizen science effort utilising LROC imagery and inspired by 'SETI at Home' [24]. Users were able to work with pre-selected LROC images to identify, trace and tag locations of interest on the Moon, with a pull down menu for classification that included craters and boulders but also spacecraft hardware. In my early work for this Thesis (see Figure 4.1) I correlated Moon Zoo user identifications of spacecraft hardware to actual locations using an ArcGIS package, and found that even when presented with a high resolution LROC image of an Apollo landing site, hardware features were largely missed by users, and other natural formations and shadows in the vicinity mis-identified as spacecraft hardware. I began to question the effectiveness of crowd-sourcing identification of alien artefacts in LROC images much as Wagner and Davies also saw the approach's likely limitations [26].



long	lat	xradius	yradius	angle	boulderyn
30.5735	19.88333	62.93697	125.1505	0	
30.45698	19.75872	30.60221	30.60221	0	
30.49221	19.55278	26.99198	26.99198	0	
30.62866	18.99445	32.16804	35.95672	0	
30.48416	20.06144	17.49769	16.03955	0	
30.51751	19.72483	14.57686	14.57686	0	
30.5153	19.72636	14.57681	14.57681	0	
30.54686	19.76674	216.2093	52.50902	0	
30.62783	21.12663	14.55412	14.55412	0	
30.54138	19.97493	14.48579	14.48579	0	
30.59461	20.89591	14.59657	14.59657	0	
30.61739	19.17887	129.9745	94.78803	3	
30.60983	20.63861	14.52859	36.32147	3	
30.61904	19.19122	125.0448	84.08185	3	
30.53032	18.78248	157.5302	194.9346	3	
30.61685	19.18429	14.36751	26.57989	3	
30.44108	20.57516	79.98751	57.23877	3	
30.54889	20.49851	67.85222	34.57993	3	
30.5123	19.34425	249.3096	121.0828	3	
30.62454	19.92173	22.47859	21.02835	3	
30.61966	19.91976	23.19966	28.99957	3	
30.62586	19.92029	14.50195	15.22704	3	
30.62196	19.91929	14.50032	15.95035	3	
30.4567	20.19925	196.1231	89.61429	3	
30.59266	19.83569	14.47999	14.47999	3	
30.67935	19.12188	35.1192	110.3746	3	
30.48756	20.71761	14.69603	14.69603	3	
30.40937	20.33781	27.9691	36.80145	3	
30.5834	20.03338	126.8315	120.7919	3	
30.55525	19.95205	147.6534	122.3207	3	
30.57535	20.47095	55.83995	98.6264	3	

Figure 4.1: Registered attempts to identify spacecraft hardware around Apollo Landing sites, using Moon Zoo. The preponderance of human-made identifications were false positives, potentially improving task performance with experience, a key advantage for future Machine Learning applications. *Right*: Moon Zoo Identifications of Other ('0' angle values) and Spacecraft Hardware ('3' angle values). *Left:* ArcGIS data overlaid; Apollo 17 landing site, with the Lunar Lander centre-lower right surrounded by a dark halo. Pinault 2015 [27].

Useful work was achieved however by Moon Zoo researchers, including a new method for dating the respective ages of features on the Moon through inferred crater erosion – less convergence by numerous human observers in defining crater edges has been seen to imply older and more eroded impact craters for example, and greater convergence on crater edges has been seen to imply sharper-edged fresher crater formations [24].

Through my own early work with Moon Zoo data sets, I set out to address the illumination, scale and coordinates challenges encountered distinguishing

spacecraft hardware remnants on the Moon from boulders, rilles, craters and associated shadows, based on early attempts using IntelCV. From these experiences I learnt that illumination and scale invariance are common issues also in standard (satellite) images processing, and became increasingly convinced that rapidly developing deep learning-based solutions could be applicable to Moon data. The deep learning approaches being applied to interpreting satellite remote sensing images have typically involved brute force applications requiring massive computing resource and run time, bespoke data libraries, and complex modelling tools. Commercial and government enterprises have been testing ideas for scaling up deep learning models to very large data sets, including unsupervised methods for natural images, using momentum contrast and unsupervised representation learning specifically for remote sensing data, using e.g. Tile2Vec [133]. The LROC data set itself has been highlighted [134] as an example presenting a host of ML remote sensing challenges: the datasets are not independently and identically distributed ('IID'); class imbalances are typical (e.g. many more craters than boulders); assessing generalization is difficult since objects are represented by multiple pixels; tiles are much larger than typical ML image sizes; the pre-processing/cleaning normally required is non-trivial; orbital tracks require interpolation and smoothing; co-registration is time-consuming; labelling often requires domain expertise; and differences in illumination conditions add to the weight of the task.

A notable effort to use unsupervised deep ML Virtual Auto Encoder techniques was introduced by Lesnikowski *et al.* [32] for LROC data sets explicitly

using Apollo landing site training sets. However as they were still engaging with whole LROC image tiles and their technique is limited to a similarity identification, their methods succeed for example in indicating that there is a probability of an object similar to an Apollo 15 landing module in a large tile known to contain another Apollo landing site, but without visualisation, localisation, or further classification. As a goal in applying YOLO-ET to LROC images, I aimed to at least cover known regions of interest within an unseen LROC image tile, and produce high-confidence identifications and localisations of spacecraft hardware, as a first step toward addressing whole image tile data sets and in future, real-time video coverage of the Moon.

4.4 LROC and YOLO-ET

The LROC team first established at University of Arizona offers a wide range of tools for accessing LROC data. Whilst the raw tiles can be a burden to process as described above, the Featured Sites section[36] offers roughly two dozen images per Apollo landing site, downloadable in several image formats. True to the nature of what the LRO narrow angle camera sees over its more than a decade of orbital passes, the images offer a wide array of phase angles, time of lunar day and illumination scenarios, as well differences in altitude resulting in differences in size and resolution. This can offer some robust and useful information for Machine Learning purposes however: elongated shadows of spacecraft at lunar dawn and dusk can produce characteristic sharp-edged rectilinear and complex shadows in

[36] https://www.lroc.asu.edu/featured_sites#ApolloLandingSites

different directions depending on the pass, for example. Low albedo regions caused by spacecraft regolith disturbances can provide contrast as dark halos or irregular shaped geometries to highly reflective glints of spacecraft material (Figure 4.2). Even without setting them as training targets for YOLO-ET, tracks of astronaut-dragged equipment carts or lunar rover tires can form characterising patterns pointing to spacecraft targets of interest (see also e.g. Figure 4.6).

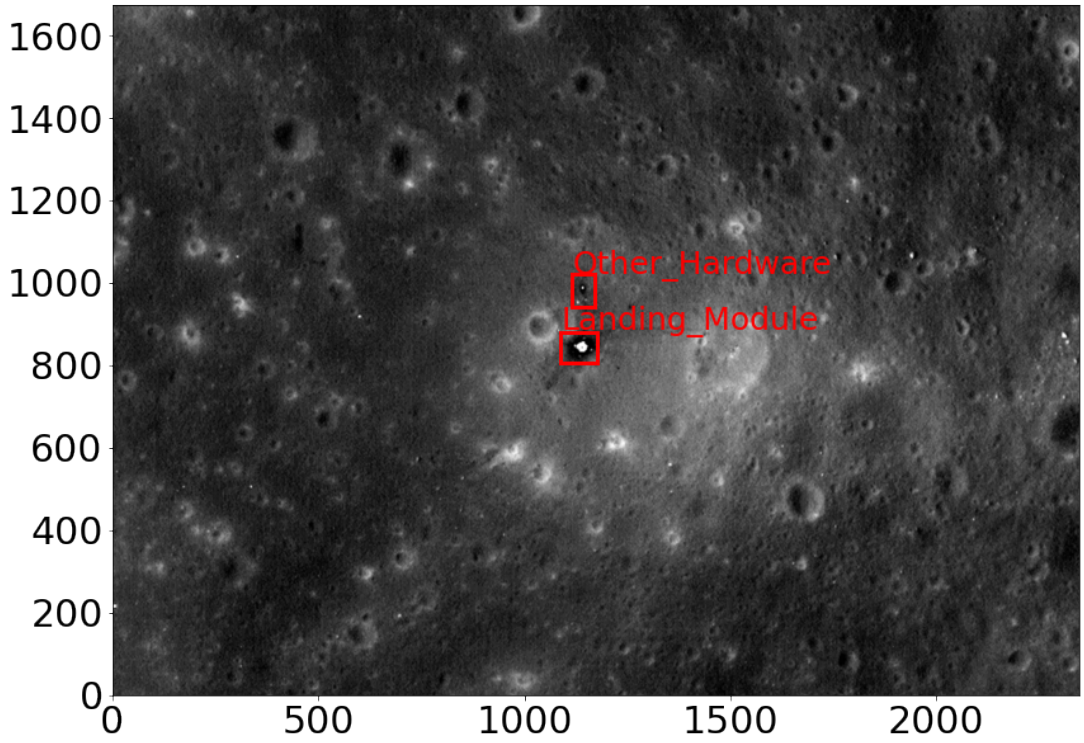


Figure 4.2: A plot of the Ground Truth and Annotations in YOLO-ET. Regolith disturbance by the Landing Module has created a characteristic dark halo around the spacecraft. Other Hardware has been identified by human researchers as the mission's seismometer deployment. Region centred around the Apollo 11 descent module. Coördinates are pixels, pixel scale approximately 1.18m/pixel.

4.5. Machine Learning dataset

This research focuses on Lunar Reconnaissance Orbiter Narrow-Angle Camera (LROC or LRO NAC) images of Apollo spacecraft and equipment. Millions of images are available on the LROC website hosted by the University of Arizona, but as described above the largest proportion are accessible as large tiles that make most Machine Learning analysis attempts intractable. Since the aim of this project is to demonstrate that a much lighter approach, using the YOLO-ET ML model and system for real-time object detection with on-device localisation and classification can ultimately be used for direct remote sensing applications, I have set the challenge to train my model to detect, localise and classify Apollo spacecraft material on the Moon as seen by the LRO NAC.

Rather than have it attempt to ingest a whole tile covering much of the Moon's surface from an orbital path segment, YOLO-ET was presented with high-resolution images from the Featured Sites selection from the LROC website. The Featured Sites pages offer downloads of some two-dozens of the best high resolution images of the six Apollo landing sites. These were used as a ready supply of data for the YOLO- ET model, similar to what might be observed by a mini-CLOXS type camera system as described in Chapter 3, were it orbiting the Moon. Since the combined data set of some 125 images is relatively small even for YOLO-ET, flips and rotations were made to expand the data set further, as with the Tanpopo data sets (Figure 4.3). The featured sites also tend to neatly centre Lunar Landing modules in the selected site scenes, so to reduce possible biases in the training (the algorithm might learn to focus on the centre of images)

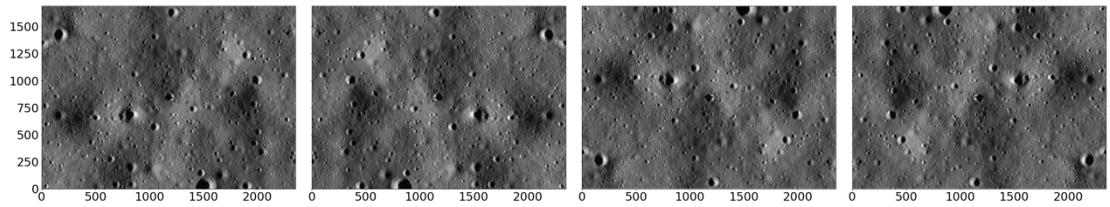


Figure 4.3: Still comprising a relatively small sized data set for Turi Create object detection, rotations and flips were performed. Then to avoid centring bias, off-centre cropping is done post-labelling.

a random 1000 x 1000 pixel cropping process was introduced alongside the flipping and rotations, helping create a larger data set. The random cropping also helps produce training and test images with no Apollo spacecraft material or activity traces in view. In practice, I might have best set up a manual K-fold cross-validation (CV) for the images as I did for the Tanpopo images in Chapter 2, but given the more limited number of images and no other model performance to directly compare them with, I elected to work without CV.

Ideally the YOLO-ET model should be optimised to expect that most of the Moon (for now at least) will be barren of spacecraft hardware objects, and to pick up the comparatively rare detections with high recall performance. Widget BBox was used to manually annotate each image with a bounding box around each object (as with the Tanpopo samples, some images contain multiple objects and others none). Initially per the LROC Featured Sites, class labelling of either ‘Landing Module’, ’Lunar Roving Vehicle’, ‘Flag’, ‘ALSEP’, and ‘Seismometer’ was used (Figure 4.5, 4.6). But distinguishing ALSEPs, Seismometers and Flags from each

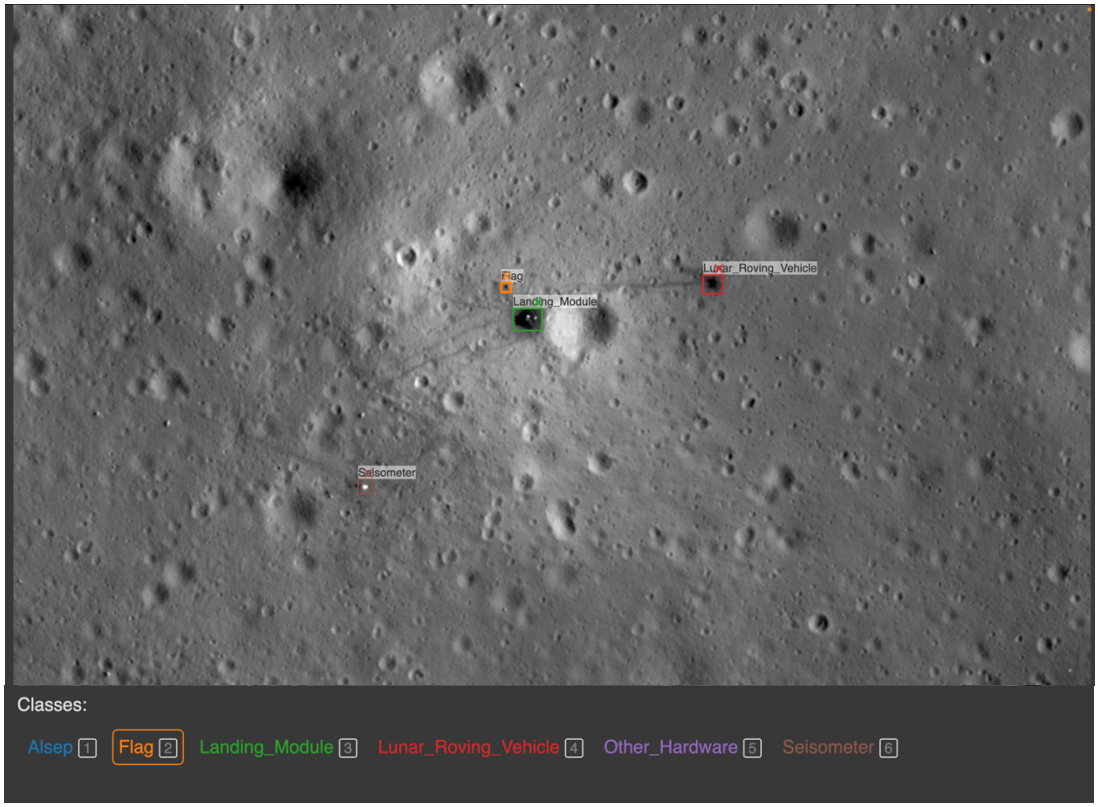


Figure 4.5: The 125 images obtained were hand labelled using the Python Bbox widget. The LROC website provides a useful guide to each respective landing site, with layers indicating key hardware elements, EVA traverses and stops, and nomenclature of surrounding geological features.

other in the images is actually only possible because exact coordinates are known to human researchers; from a Computer Vision perspective these are objects of only a few pixels that might be successfully distinguished by the algorithm from natural lunar features, but not from each other. So for the most recent YOLO-ET model runs classes of 'Landing Module,' 'Lunar Roving Vehicle,' and all 'Other Hardware' were used.

Even more so than the Tanpopo images analysed in Chapter 2 and 3 experiments, the LROC Apollo Featured Sites images have a variety of hues and

brightnesses (Figure 4.7), as they were captured under very different conditions of illumination and phase angle; however in the interest of both human time-saving and creating an independently operable on-device detector, the images have not been pre-processed to account for this.

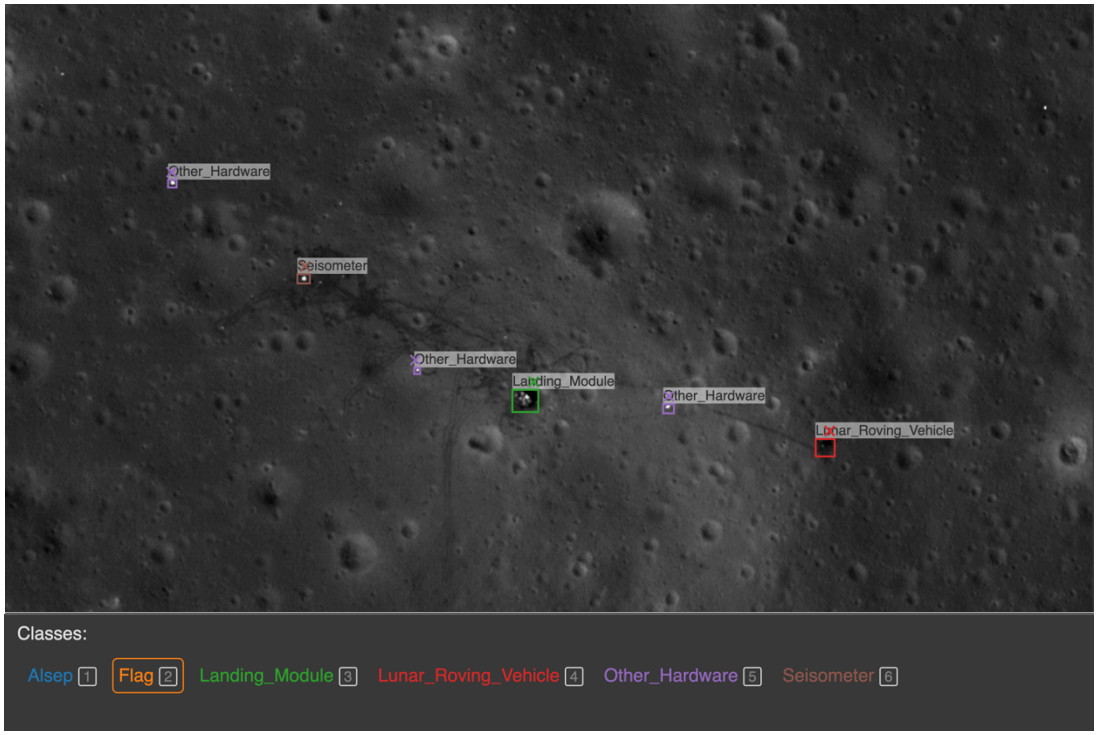


Figure 4.6: Initial BBox labelling was as detailed as 'Seismometer' and 'Flag;' for the initial model runs but these were consolidated to Landing Module, Lunar Roving Vehicle, and Other Hardware, to reflect what YOLO-ET could realistically discern without prior knowledge of what spacecraft hardware lies at which coördinates. Apollo 15 landing site; the labelled spacecraft hardware is distributed across an approximately 800m range.

The LROC data sample was split into 80% for training (1267 images) and 20% for testing (317 images); (Figure 4.8). As with Tanpopo these are randomly sampled whilst maintaining the baseline ratio between different classes; the training data are the images used to optimise the model weights during the

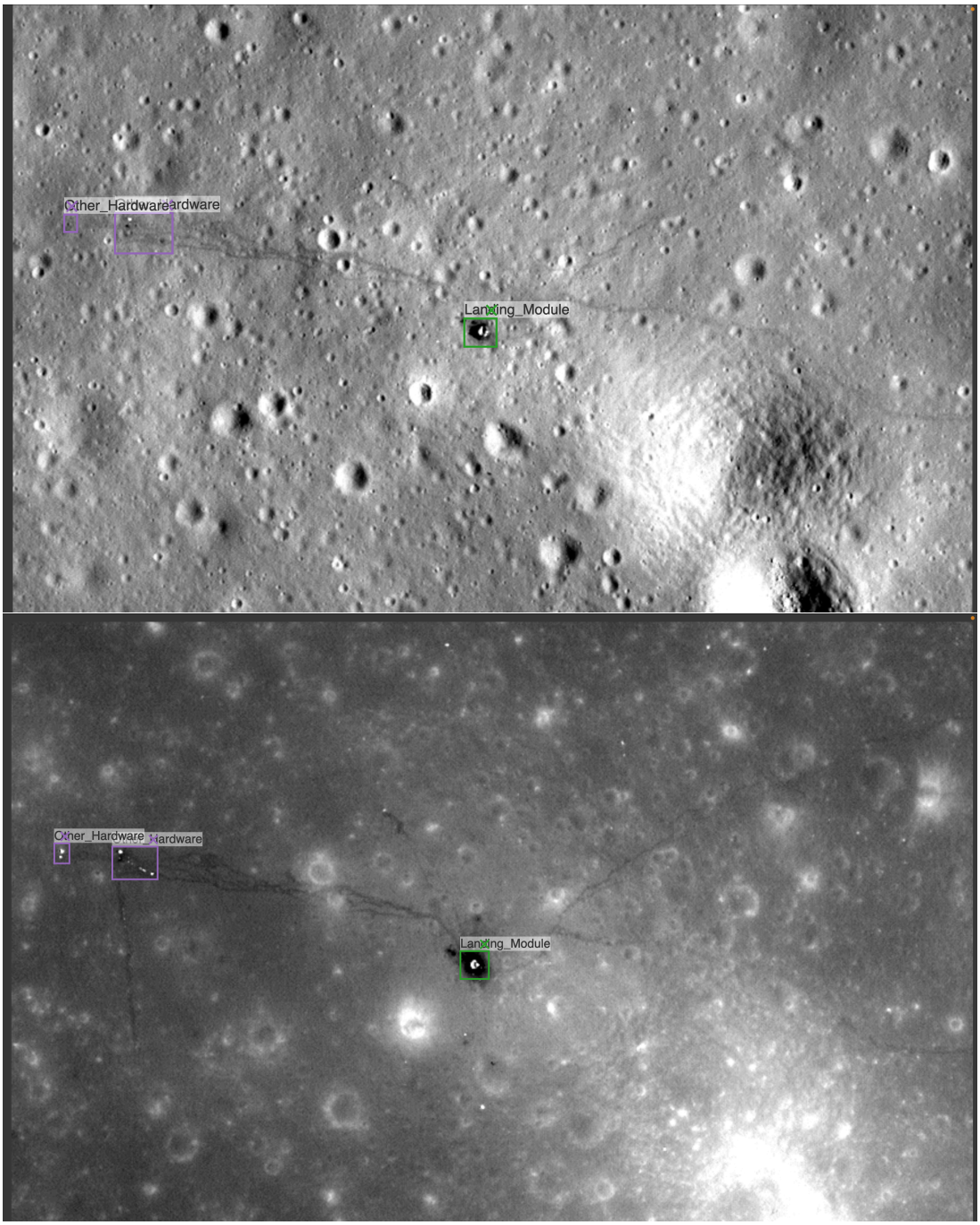


Figure 4.7 a, b : BBox labelled images of the Apollo 14 landing site, illustrating the variety of hues and brightnesses in the LROC images under different conditions of illumination, altitude and phase angle.

training phase of the model. The test data are not seen during the training of the model, and whilst it is common with larger CNN models to additionally set aside a validation dataset for informing when the model is sufficiently trained - i.e. not under- or over- fit, the still relatively limited data sample hampers the ability to reserve additional samples for validation, which could weaken the YOLO-ET model's performance. As noted in Chapters 2 and 3 Machine Learning methods are data greedy and perform better with more data, so to compensate for the relative data paucity, the rotations, flips and random crops were added.

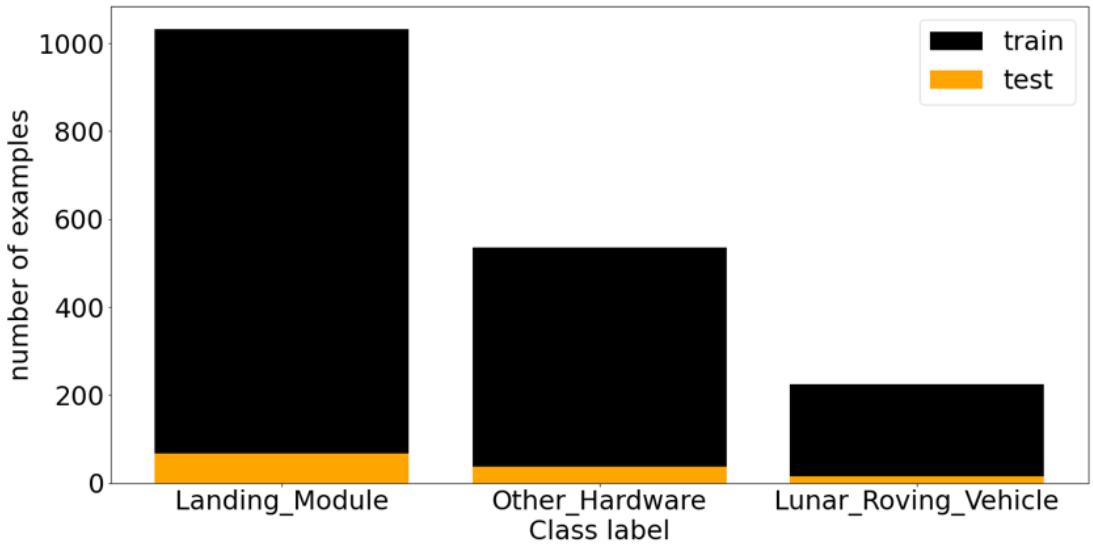


Figure 4.8: Distribution of training and test data over the different classes. Note that these do not directly correspond to the number of images, as images can contain multiple objects of different classes or no objects altogether. Additionally, there is significantly more training data as the augmentation is applied only to the training sample. With sparser data over a significant area of the Moon to cover, more flips and rotations as well random cropping and offsets have been added.

As described in Chapter 2, Machine Learning involves constructing layered architectures where each layer performs specific operations on data. During training the YOLO-ET algorithm processes input data through these layers, where each operation transforms the data based on the current weights. As with the Tanpopo samples, the goal with the LROC images is to optimise the weights to minimise a predefined cost or loss function which measures the difference between the algorithm's predictions and the actual outcomes.

Figure 4.9 re-illustrates these principles using LROC image examples.

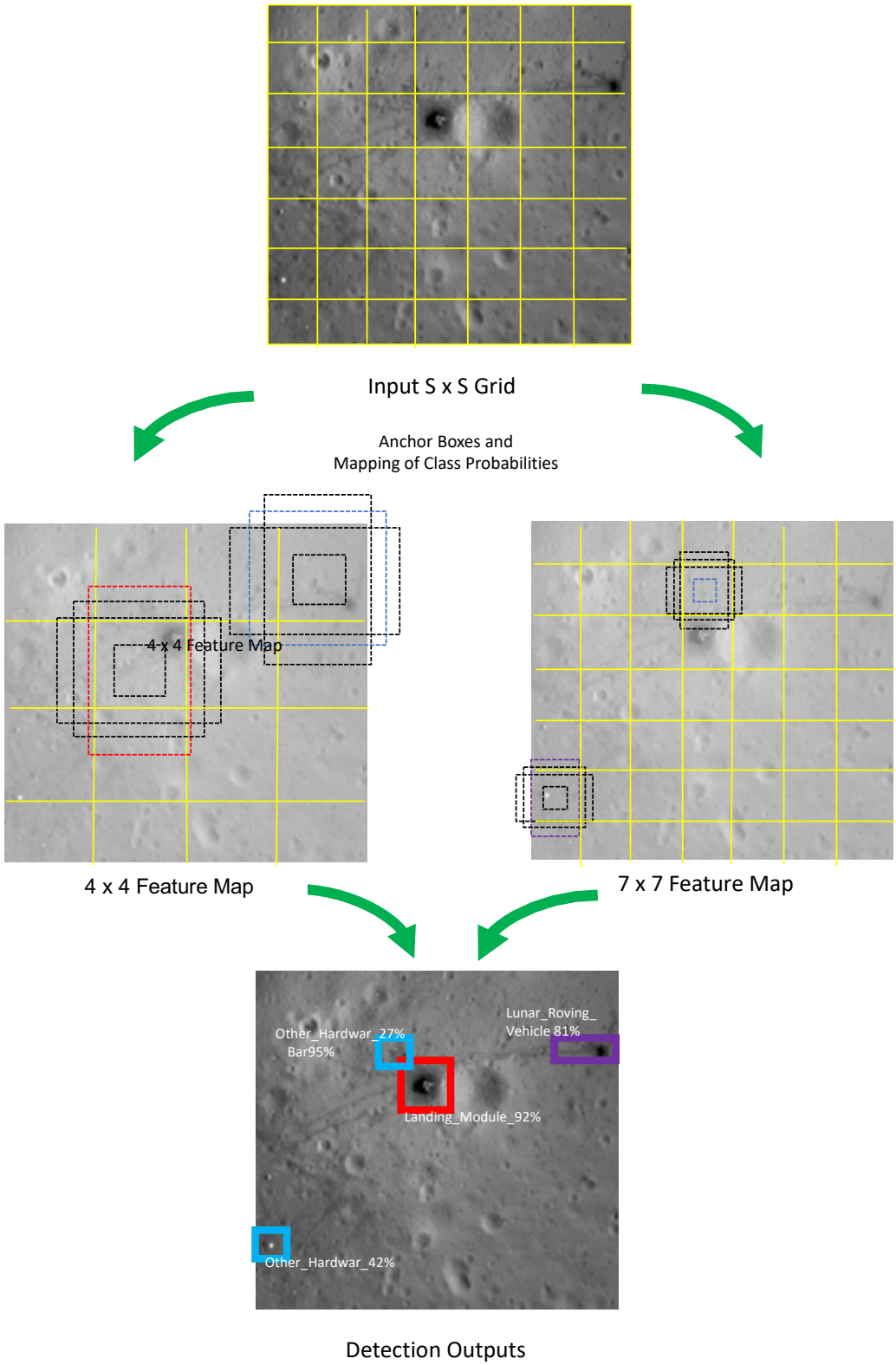


Figure 4.9: Diagram illustrating the concept of anchor boxes in YOLOv2, showcasing various predefined box shapes and sizes strategically positioned across an LROC image in YOLO-ET implementation. These anchor boxes enable the efficient and accurate prediction of object boundaries and classifications within a single pass of the network.

The work here on the LROC images applies supervised learning to object detection, training models to recognise and categorise Apollo spacecraft hardware on the Moon's surface, and as with Tanpopo the data consists of pairs of images and their corresponding labels that are the bounding box coordinates (x and y), height (h), width (w), and class.

The same principles and practices including Loss Equation as used and described in Chapter 2 are employed here on the LROC images; kindly refer to section 2.3 of this Thesis. Localisation is more important with the LROC Apollo images than with the Tanpopo images. On the Tanpopo images, particles and fragments stand out from the typically evenly hued blue and green backgrounds of the silica aerogels. Accurate predictions between classes prioritised the fourth and final terms, scaling by λnoobj to specifically penalise false detections and assessing the classification error for each class c across the objects detected, respectively, ensuring accurate class predictions.

For the LROC Apollo images work, the first two terms, weighted by λcoord, and penalising errors in the position (x,y) and size (w,h) of the predicted bounding boxes compared to the ground truth, are more crucial for precise localisation, in the face of the many craters, shadows and other features surrounding the hardware on the Moon. The third term penalise errors in object scores Ci, distinguishing between object presence and absence, relevant to both data set experiments [95].

During training, as with the Tanpopo work various batch sizes were experimented with. Given the larger LROC images, it was anticipated that larger batches demanding yet more memory due to the increased number of images

loaded simultaneously, might not be achievable and that resulting loss curve might not be as smooth, indicating a less stable model. But in fact with the same batch size of 32, a good convergence was achieved, with only a minor stutter in the curve and still expedited convergence to the global minimum of the loss function, without having to make further memory trade-offs. With the batch size of 32 again effectively balancing computational resource demands and learning stability, each Apollo LROC epoch consists of 1584 images/32= 49.5, approximately 50 iterations. Setting max iterations to 500 means the training process involved roughly 500/50 = 10 epochs. After experimenting with various numbers of max iterations, 500 was found to be optimal, striking a balance between model performance and computational efficiency (see Loss Curve Figure 4.10).

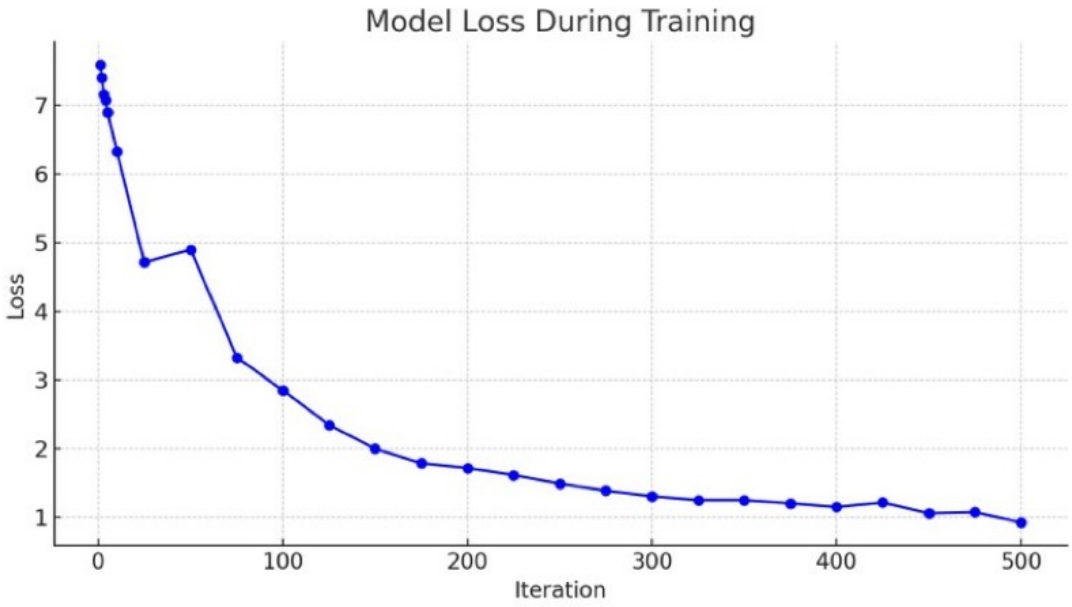


Figure 4.10: The YOLO-ET Apollo model's training loss over time. Halting the training process before overfitting - becoming excessively attuned to the training data -is particularly key for the Apollo site images, so that different types of spacecraft and spacecraft hardware can be recognised even in unseen scans of the lunar surface. In this case 500 iterations versus 2000 for Tanpopo proved a useful basis for optimal model convergence.

As with Tanpopo this choice was partly influenced by Turi Create's constraints on validation set usage, which limited the ability to employ traditional validation techniques to fine-tune the number of iterations. Again reliance was on trial and error in the creation of the model, along with runtime considerations, to determine the most effective training duration. Since the goal is automated on-device deployment for use in lunar orbit, the trial-and-error component in the model creation phase to achieve this goal is a practical trade off: the balance of convenience with a good modicum of configurability makes YOLO-ET's implementation a practical tool for rapid development, while recognising the limitations for more advanced experimentation and nuanced model optimisation. Evaluating the model on the test set that the model hasn't encountered during training, serves as a proxy for real-world, unseen data and provides a more accurate measure of how well it will perform in the real world in comparison to the training set (see Figure 4.11).

From the YOLO-ET Turi Create training log, the loss is decreasing over iterations, which a good sign that the model is learning from the training data. The loss starts at 7.5868 and has decreased to 0.920158 by iteration 500. The elapsed time per iteration also seems to be fairly consistent, indicating stable training performance (per Figure 4.10 above). As the training progresses the loss values' decreasing trend suggests that the model is improving. Since the loss does not start to increase or fluctuate significantly, it is less likely that it indicates overfitting or instability in the training process. Based on experiences with the Tanpopo data sets, a first next recourse would likely be to a more varied data set, not an adjustment to the learning rate or use of a different optimization algorithm.

With the training complete, the model's performance was then evaluated on a separate test set to ensure that it generalizes well to new, unseen lunar surface data. YOLO-ET's Turi Create implementation informs results qualitatively by visualizing the model's predictions on the test images, facilitating intuitive experiments with cropping, centring etc. for performance impact (see Figures 4.11, 4.12, 4.13, 4.14, 4.15).

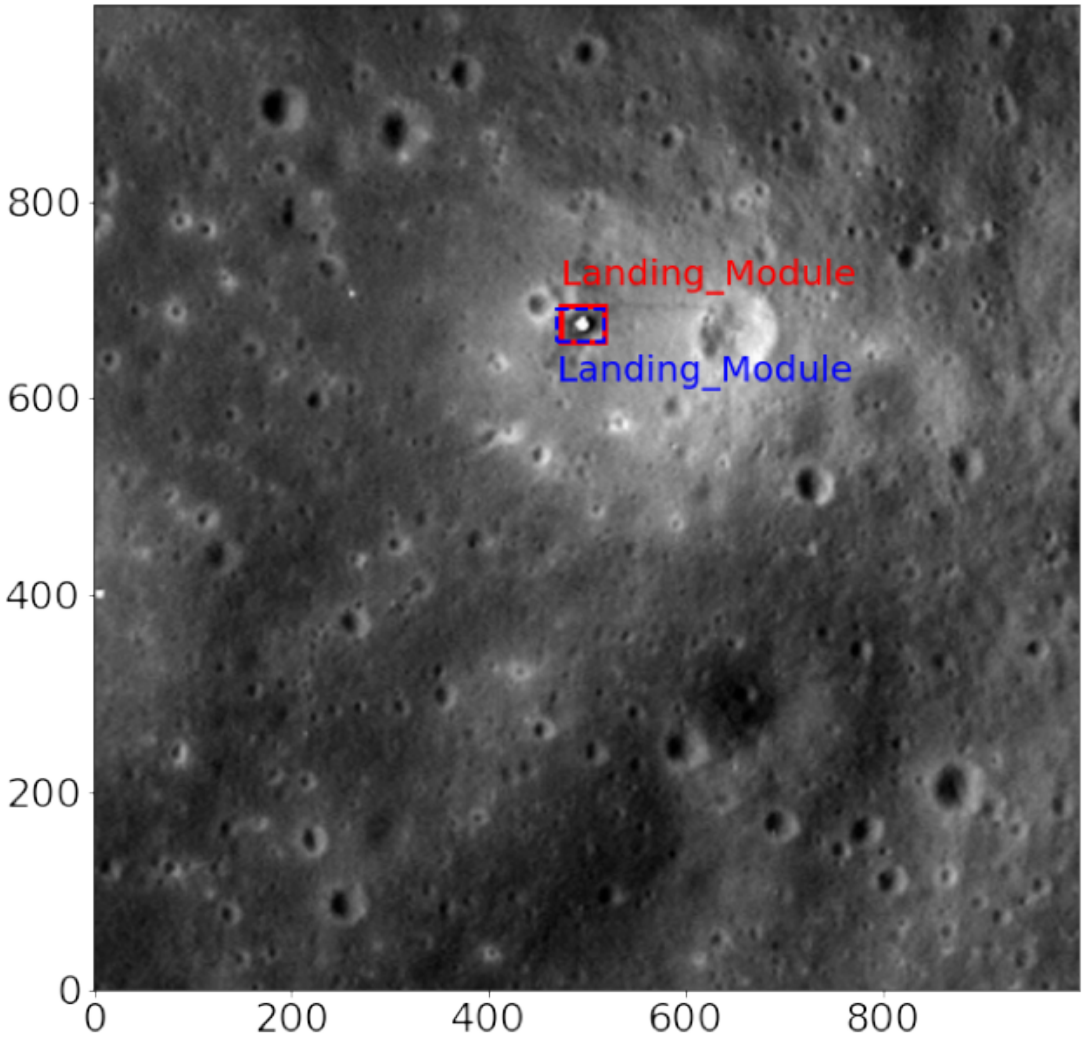


Figure 4.11: Training example image showing the Apollo 11 landing module with the ground truth bounding box in dashed blue and the prediction from the network in red. Units are pixel coördinates at approximately 1.18m/ pixel.

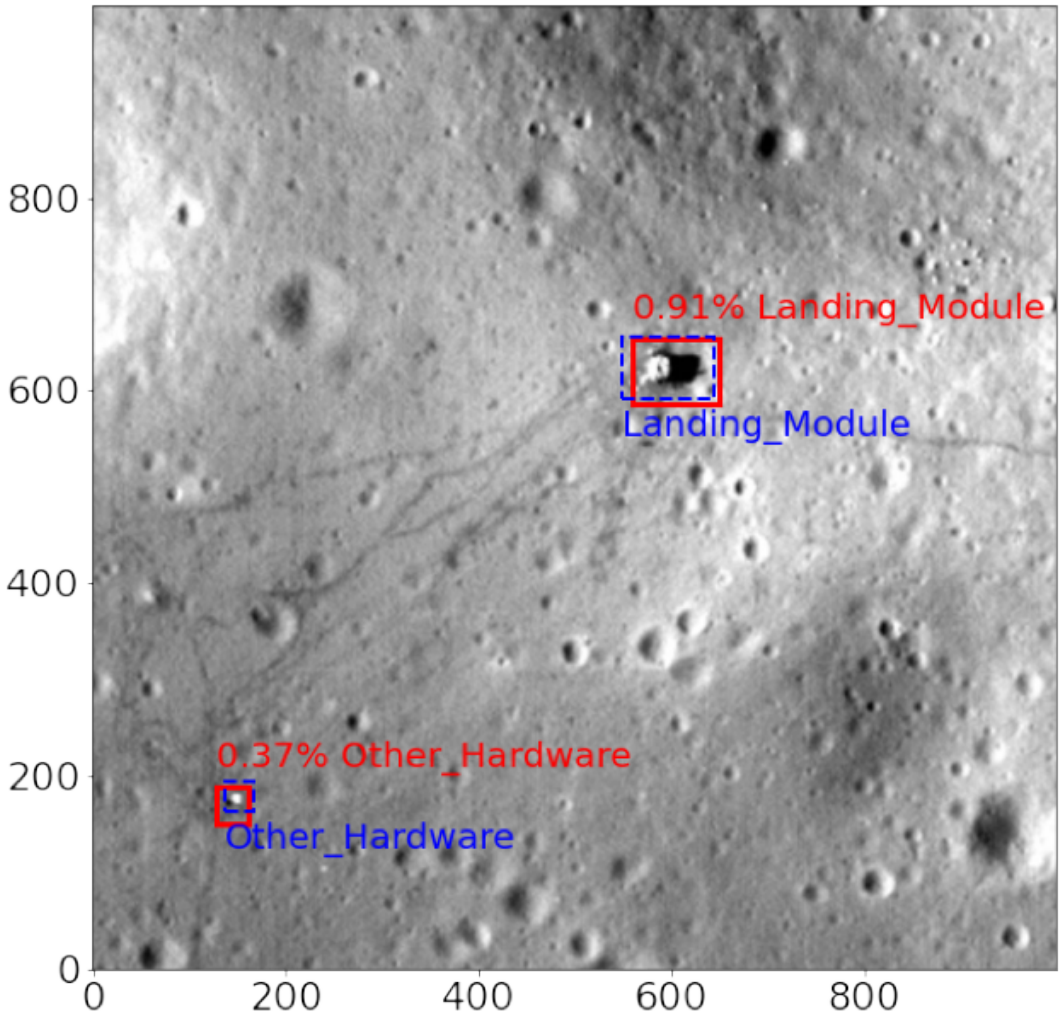


Figure 4.12: Landing Module detections outperform others in the YOLO-ET model. Ground truth in dashed blue. Confidence scores of the detected object are shown as percentages. Apollo 12 landing site, units in pixels, approximately 0.5m/ pixel.

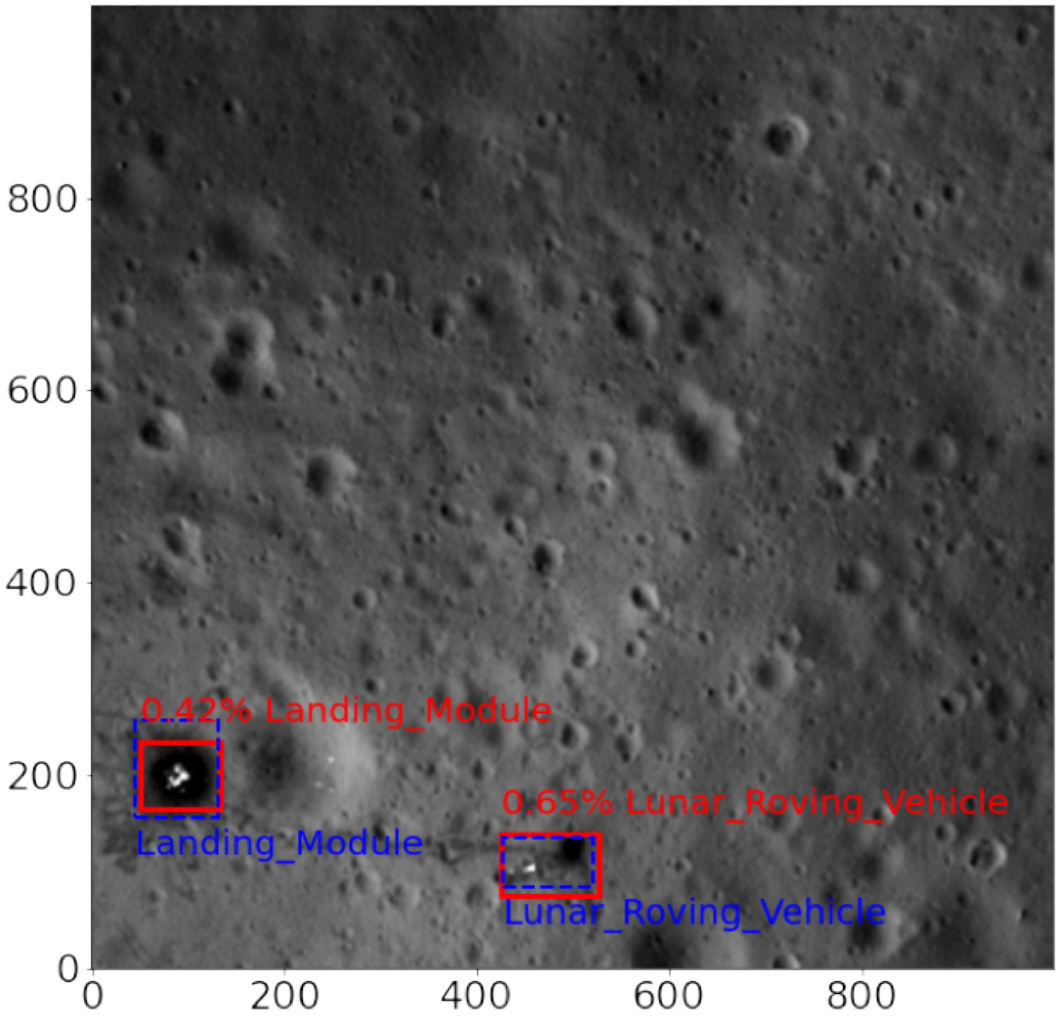


Figure 4.13: Test data examples depicting Apollo spacecraft hardware with ground truth bounding box in dashed blue and the network predictions in red. Confidence scores of the detected object are shown as percentages. Apollo 16 landing site, units in pixels, approximately 0.491m/ pixel.

.

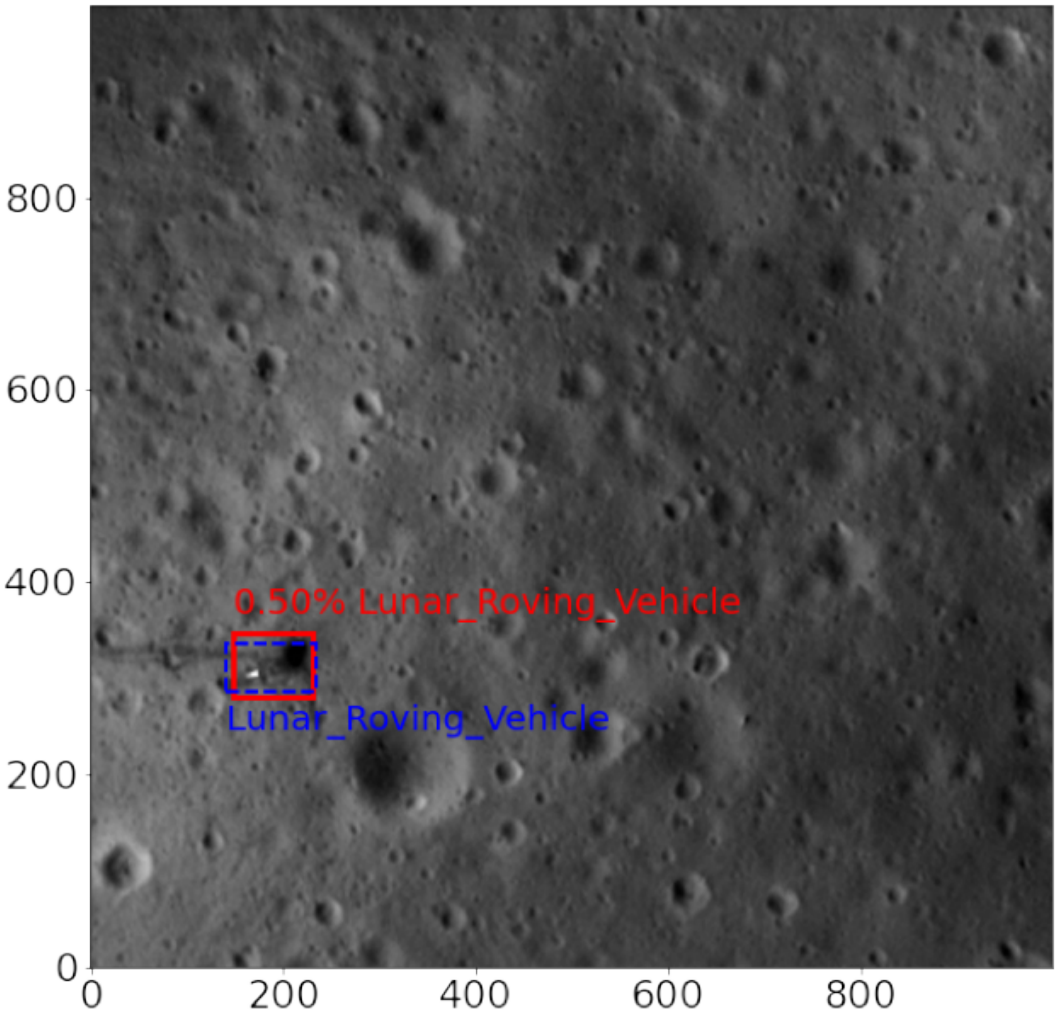


Figure 4.14: An image from the same file as Figure 4.15 featuring the Apollo 16 landing site, but here the Lunar Roving Vehicle is cropped away from other hardware elements, including the Landing Module. The model may have picked up better confidence levels when in proximity to other hardware in the image. Units in pixels, approximately 0.491m/ pixel.

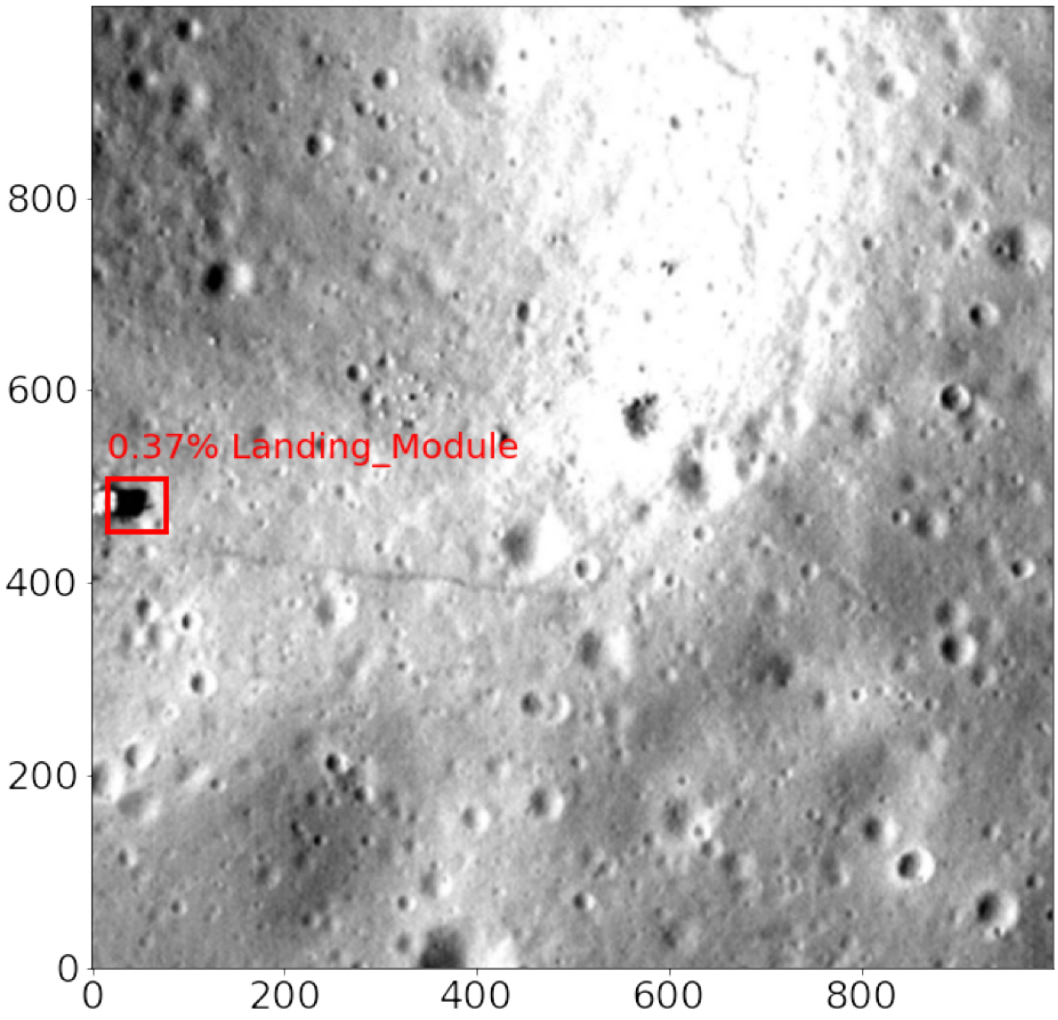


Figure 4.15: In this output prediction of YOLO-ET applied to test image, hardware cropped in isolation from other hardware and at the edge of the image seems to lower confidence scores. Apollo 12 site, units in pixels, approximately 0.486m/ pixel.

4.6. Evaluation Results

The final loss of the network is 0.9202. YOLO-ET correctly detects 42% of the Apollo spacecraft test data with over 50% overlap (IoU) with the ground truth box. Whilst this is lower than the 90% detection rate of surface microparticles on

Tanpopo aerogels, there two relevant factors to consider. Firstly, the lunar background itself is far more complex than the relatively smooth contexts for targets provided by the aerogels themselves. There is a natural risk that the Apollo YOLO-ET model could be overwhelmed by false positives of boulders, shadows, craters etc. Secondly, the results are somewhat 'dragged down' by the model's performance on Other Spacecraft Hardware, typically featured in a much smaller number of pixels. Results were 80% for Landing Modules, 32% for Lunar Roving Vehicles, and 14% for Other Hardware on the test data with over 50% overlap (IoU) with their ground truth boxes. Figure 4.16 shows the distribution of FPs and FNs over the respective class labels.

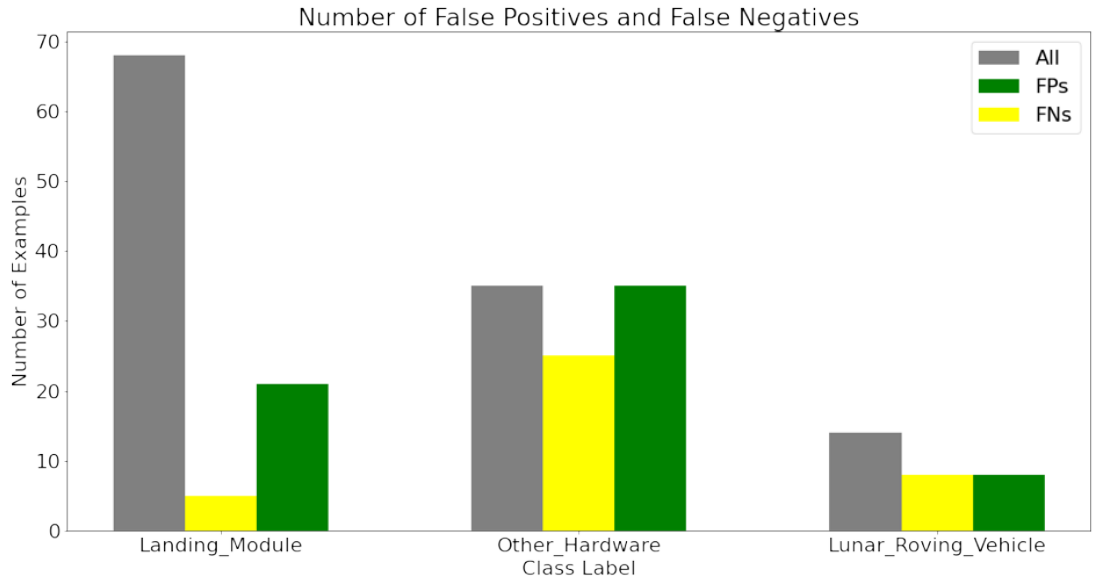


Figure 4.16: This chart shows the distribution of FPs and FNs over their respective class labels. 'All' comprises all detections of True Positives and False Positives. It's evident that the model performs best on detecting and classifying Landing Modules which has the least FNs and few FPs. From these values the evaluation metrics are computed, with the model attaining for the combined classes a precision, recall and F1 score of 57%, 64% and 60% respectively.

Whilst these results may seem low compared to human performance levels of 90% accuracy [106], unlike the 4000 Tanpopo image samples, there are millions of LROC images to potentially scan for Arkhipov-Bracewell Probes. Relatively uncluttered by false positives, the Apollo-trained YOLO-ET model works readily on unseen Apollo sites (Figure 4.17 a, b), offers practical opportunity to aid interested citizen scientists in examining areas of interest (as demonstrated by ready detection of an unseen Luna 9 candidate object as discussed below), and offers potential for automated examination of all LROC images, and thereby is well suited for deploying as an alert system on a lunar satellite continuously imaging the Moon's surface.

4.7. Non-Apollo Hardware Experiments: Luna 16, and Luna 9 Candidates

Test images of other, non-Apollo known spacecraft hardware unseen by the YOLO-ET model can also be returned with successful identifications. Here Luna 16 is identified as a landing module with 47% confidence (Figure 4.18). Experiments were then conducted to see if the YOLO-ET model could assist in detecting unseen hardware whose exact coordinates are unknown. Luna 9 was selected as a good example, having successfully landed and communicated with Earth, and with coordinates believed to be somewhere in the area of 7.08°N latitude and 295.63 °W longitude. Focusing on a search area of approximately 300 x 300m using the LROC QuickMap Image Search tool provided by Arizona State University, on first inspection a bright artefact is seen but rejected by the model (Figrue 4.19); to the seasoned eye it appears a fresh crater with no technosignatures.

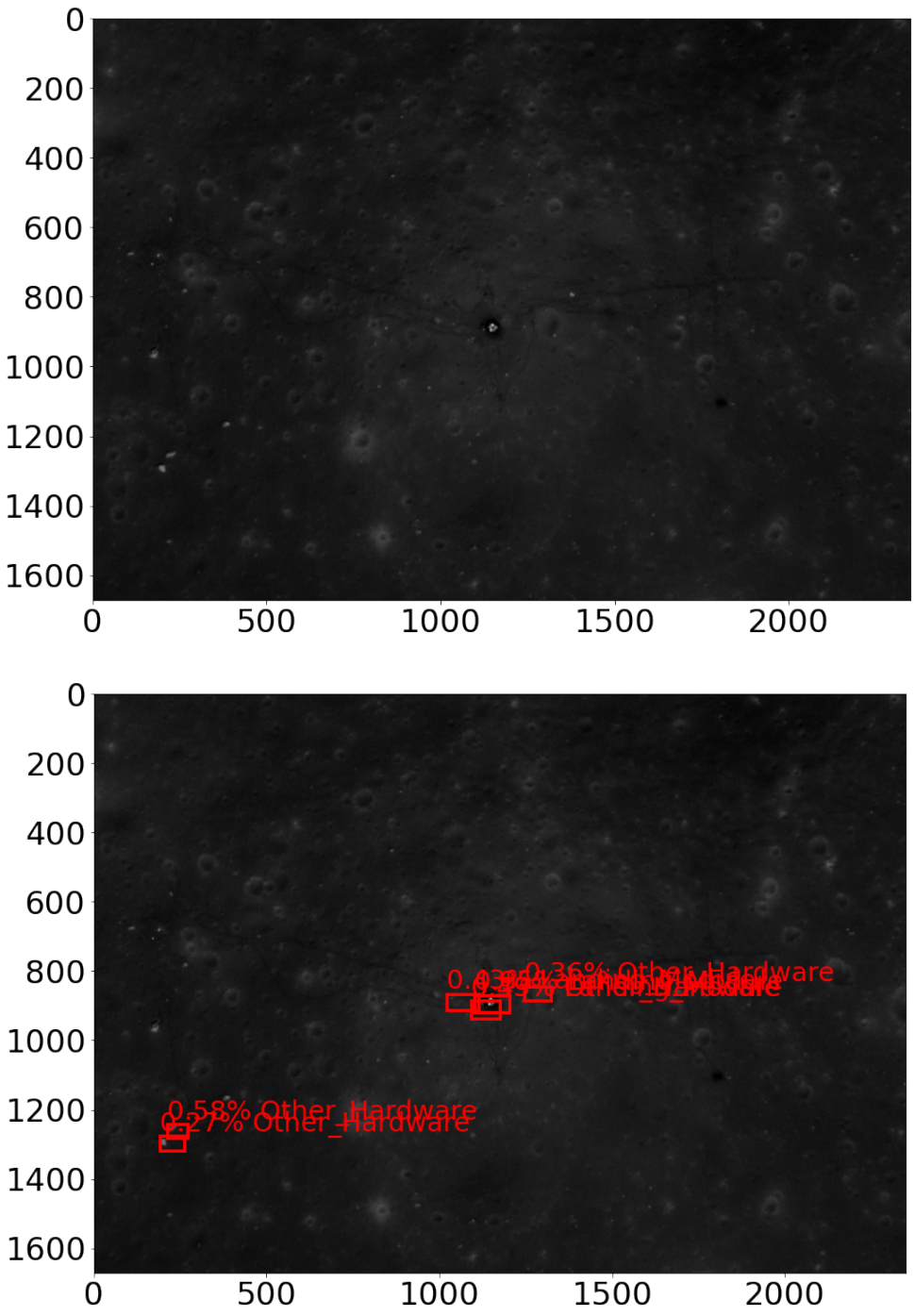


Figure 4.17 a, b: In deployment with the model fully trained and loaded, prediction tests are straightforward even in darker or more obscure images. Below are detections at the Apollo 17 landing site in an image unseen by the model. Units pixels, 0.515m/ pixel.

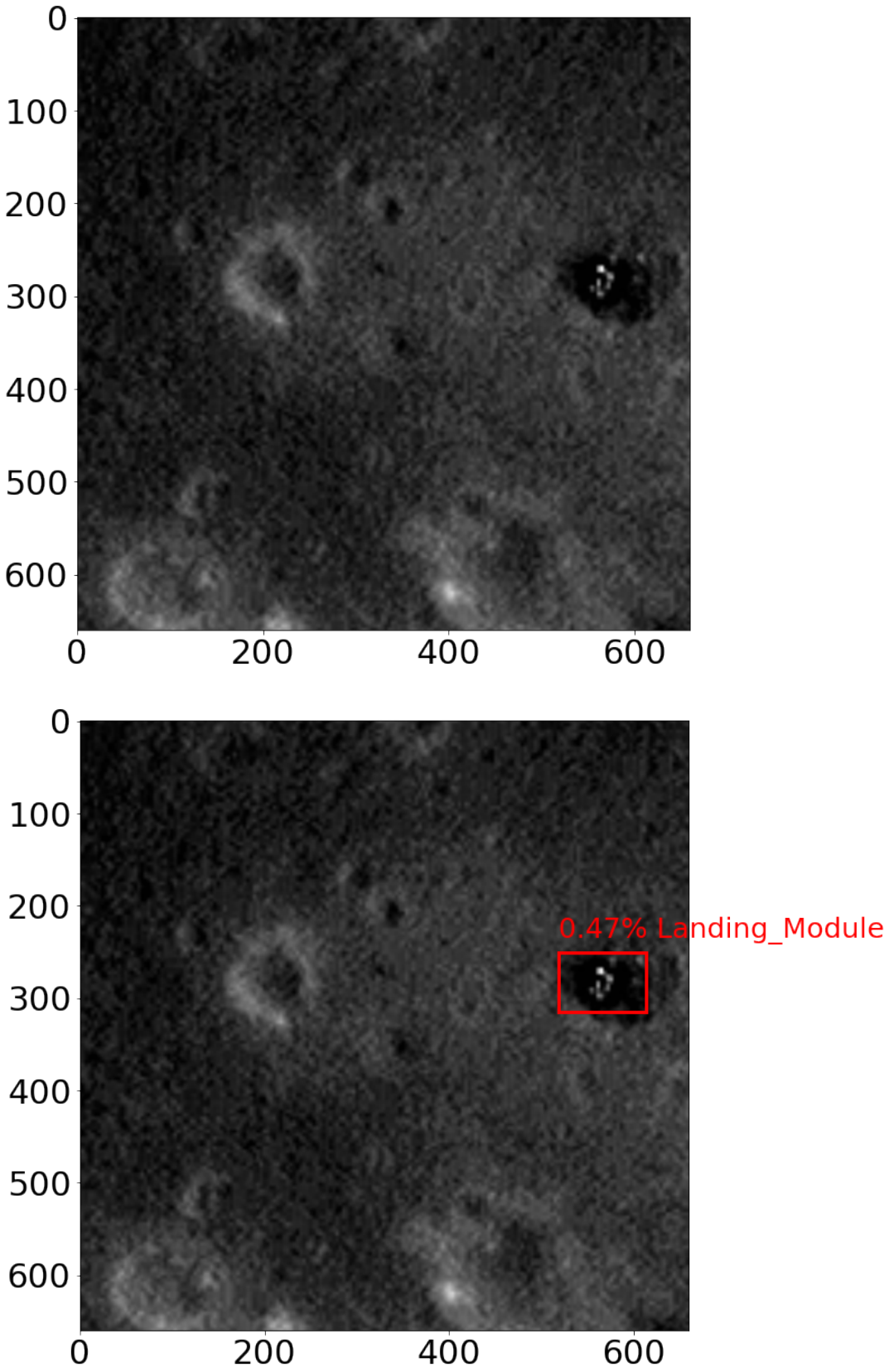


Figure 4.18 a, b: Luna 16 detected with 47% Confidence. Units pixels, 0.45m/ pixel.

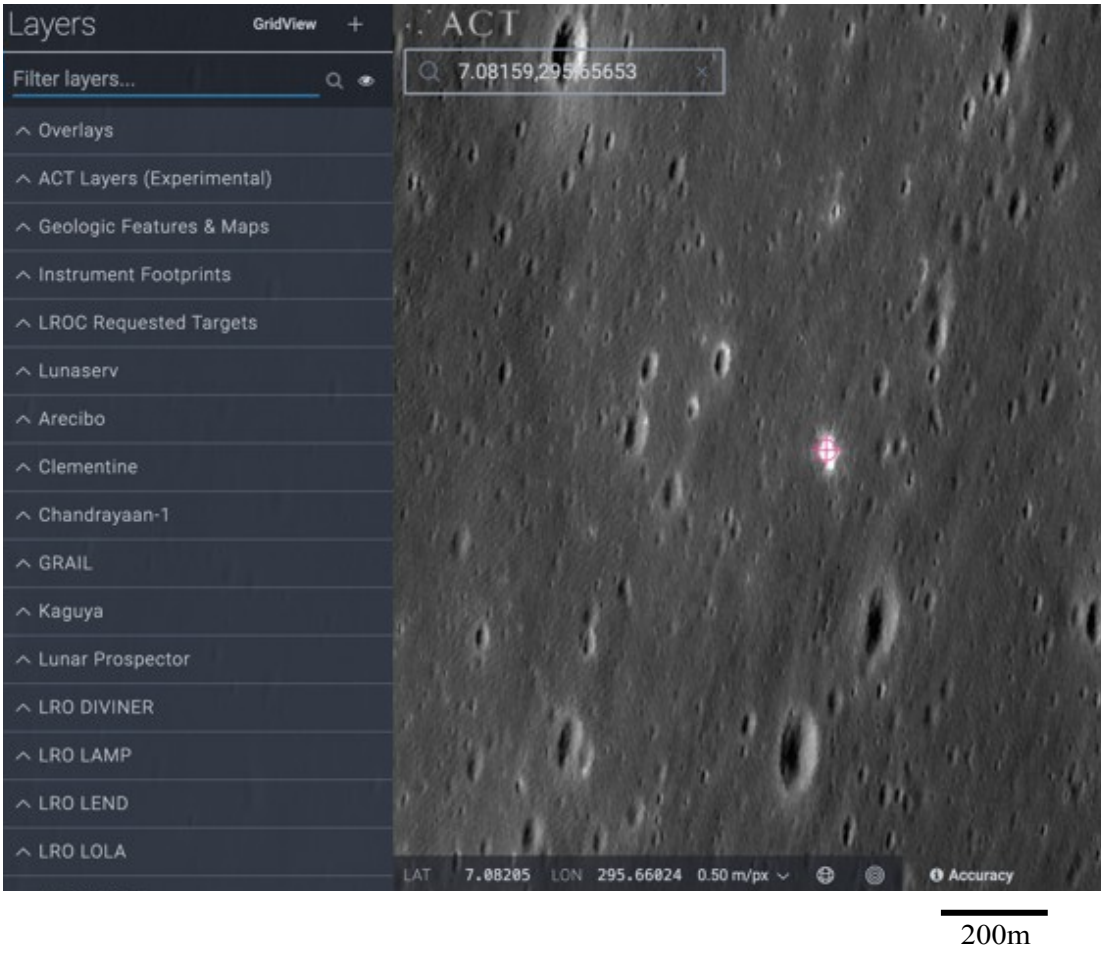


Figure 4.19: Flat reflection in the presumed landing area of Luna 9 catches human but not algorithm interest. LROC Quickmap search tool image.

Changing to an orthographic projection for a 'flatter' view and spanning out from the area with additional images however, the YOLO-ET algorithm made a first detection; a Landing Module (trained on Apollo spacecraft) is detected and localised at 7.02908 N. latitude and 295.67131 W. longitude with 32% confidence To the human eye a second similar object of potential interest can be seen in the upper right quadrant of the image, located at 7.03211 N. latitude and 295.6790 W. longitude (Figure 4.20).

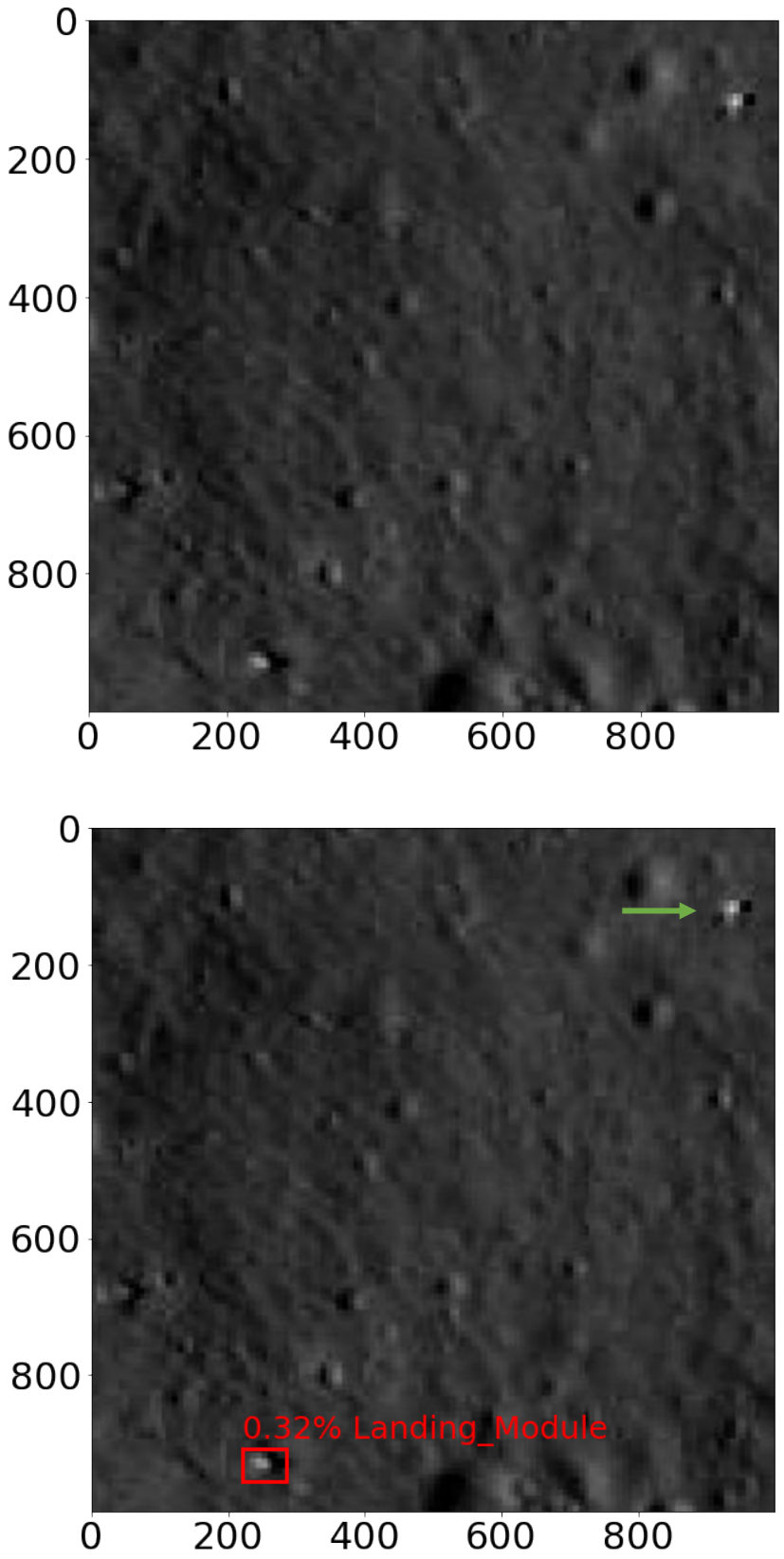


Figure 4.20 a, b: A first detection by YOLO-ET, trained only on Apollo spacecraft, of a Luna 9 candidate, at 32% Confidence of being a Landing Module. On (human) inspection another similar-looking artefact appears on the upper right of the image. Units Pixels, ~ 1m /pixel.

Since the YOLO-ET model, despite training on randomly cropped and flipped images may still have an image centre bias, both the original detected object in the presumed Luna 9 landing vicinity and the second object were cropped and centred separately. This improved the confidence score of the first object from 31% to 62% as a Landing module (Figure 4.21), and resulted in a first detection for the second object, as Other Spacecraft Hardware with a 43% confidence score (Figure 4.22).

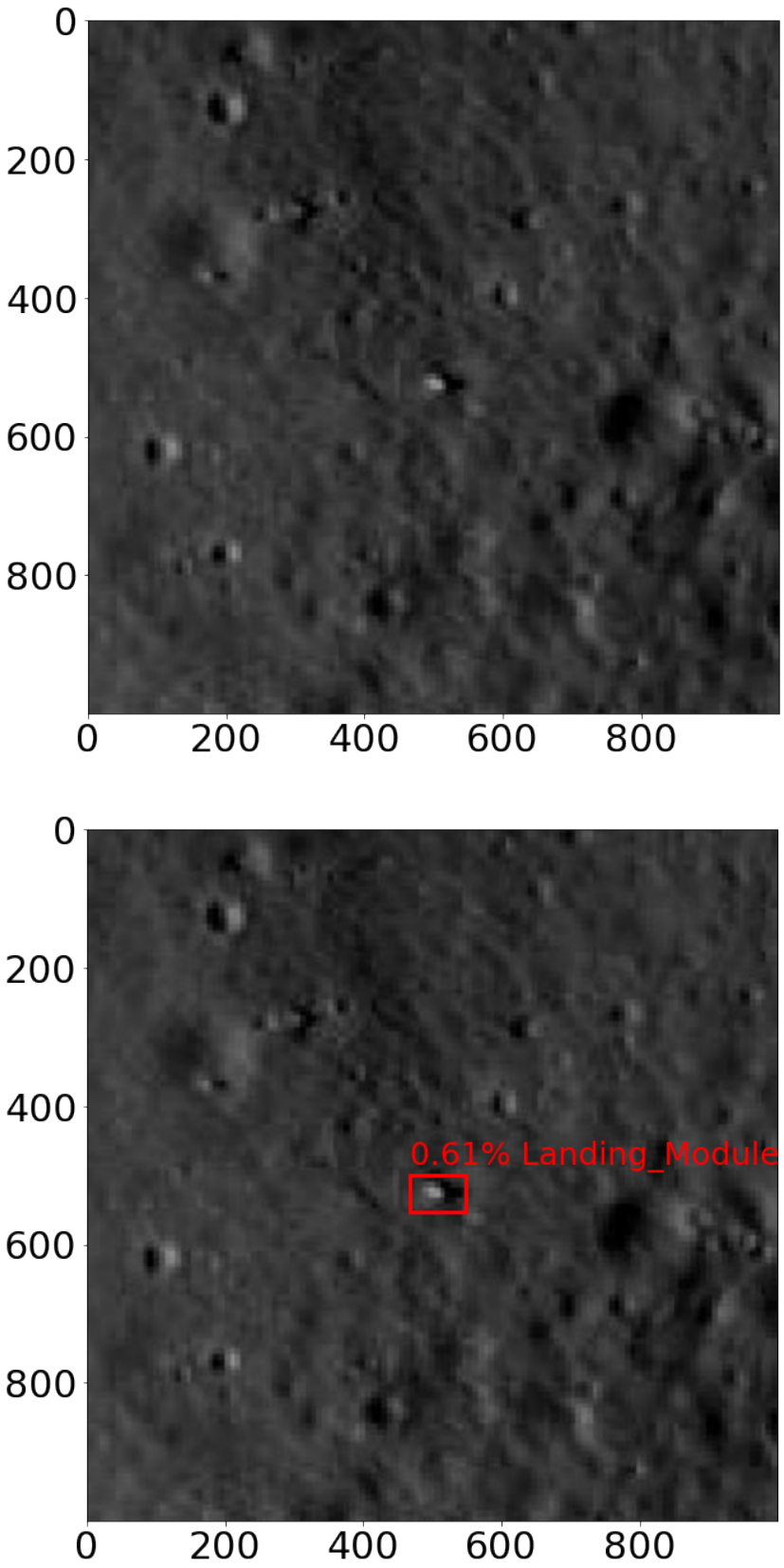


Figure 4.21: Centring the first artefact in an image, in light of possible algorithm bias, increases the Confidence score from 32% to 61%.

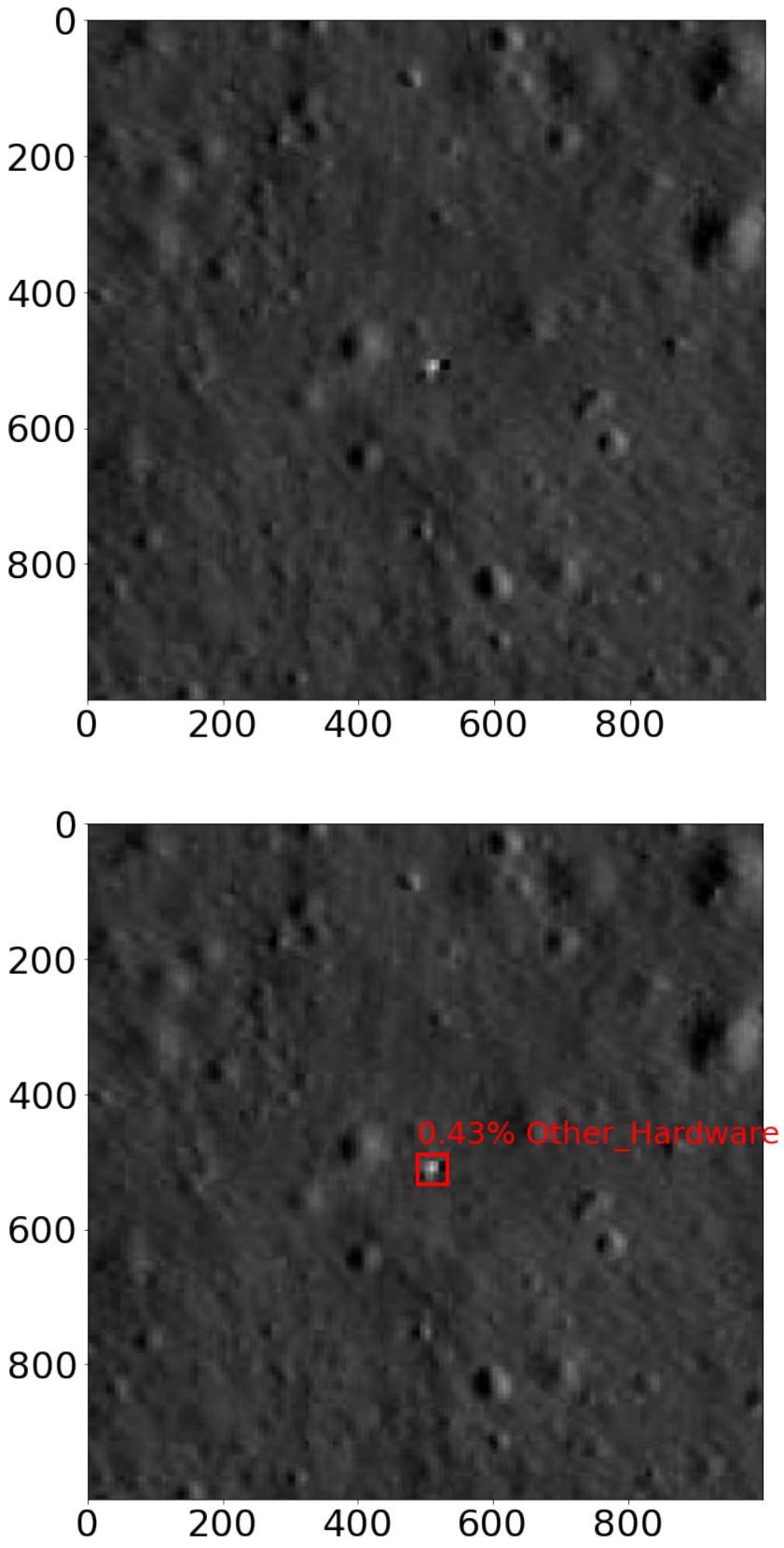


Figure 4.22: Centring on the second artefact, the YOLO-ET algorithm now picks up this one as well, with a 43% Confidence score.

One possibility for there being two objects in the single image that YOLO-ET has identified as Luna 9 hardware candidates is that they are different parts of the same spacecraft. The Luna 9 lander achieved the first robotic soft landing on the lunar surface inside an inflated cocoon with an impact speed estimated at 4 to 7 m s^{-1}. At 5 m above the surface, sensor probe contact commanded separation of the ball-like landing module, with its air bags programmed to jettison from the lander four minutes after touchdown, followed by a 10-second deployment of the lander and the unrolling of its antennas (Figure 4.23) [135].

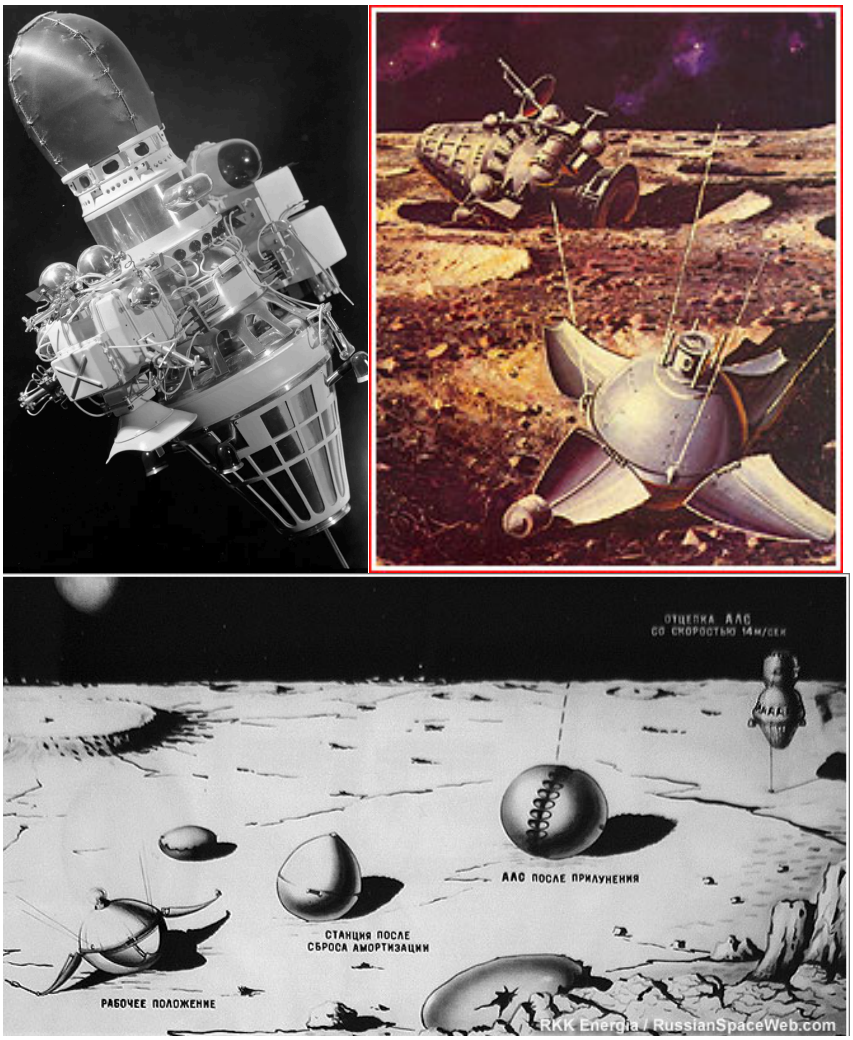


Figure 4.23: Top left, pre-launch configuration of Luna 9; top right, artist's conception of deployed module; below, landing sequence schematic. Image credits: RKK Energia and russianspaceweb.com

The distance between the coordinate points of the two artefacts detected by YOLO-ET is under 250m.[37] Surface images taken from the Luna 9 lander indicate at least one possible artefact between the Luna capsule and the horizon some 50m to 300m away, posited by Lockheed analysts shortly after Luna 9's landing to be a spherical element of the main spacecraft; their analyses of the size and distribution of craters in the landing capsule's panoramic site images also suggest a possible match for the terrain in the LROC image [136].

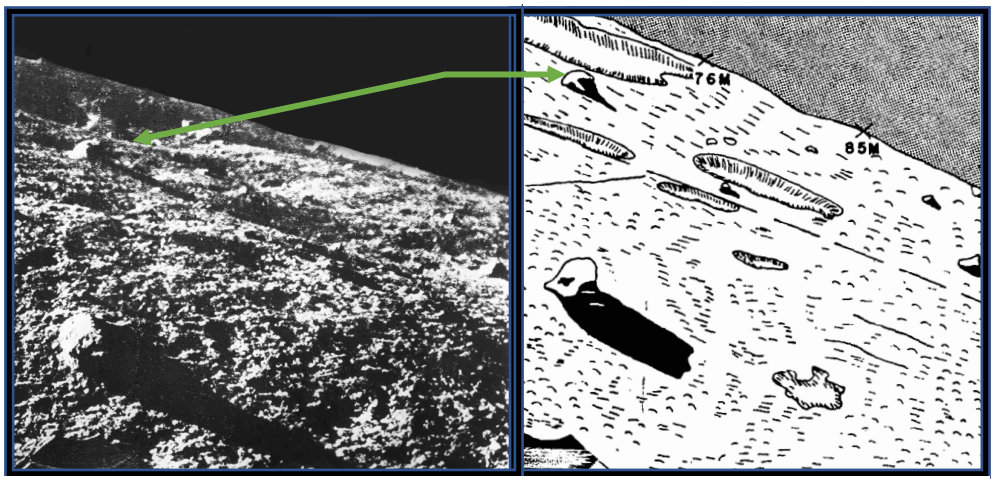


Figure 4.24: One of the first images transmitted by Luna 9 (and by any Earth spacecraft from the surface of another planetary body). Toward the top left of the image is a spherical object posited by Lockheed analysts to be part of the main spacecraft lander. Luna 9 capsule petal visible in the lower left of the image. Credits: Top, RIA Novosti; Bottom, NASA and Lockheed Electronics Company [135, 136].

Using the 'Layers' panel of the QuickMap tool to toggle between different datasets and images with 'Instrument Footprints' taken at different times by different instruments I am undertaking greater examination of these possible Luna 9 coordinates; based on my work here, which I'm now preparing for a short paper,

[37] Using the Lunar and Planetary Institute's distance calculator at https://www.lpi.usra.edu/lunar/tools/lunardistancecalc/

I plan to recommend additional imaging of the site by LROC, and the potential finding supports the case for deploying a dedicated detection satellite with YOLO-ET on board.

4.8. A YOLO-ET Lunar Satellite and a Citizen Science Approach to Searching for Arkhipov-Bracewell Probes on the Moon

Though it requires working with an additional application such as AV-Foundation,[38] using the CoreML framework described in Chapter 2 YOLO-ET can process video frames in real-time, feeding individual frames to the model at a rate that allows for real-time analysis in an iOS device. With this video capture and processing capability, much like the mini-CLOXS developments discussed in Chapter 3, it may be possible to put a relatively inexpensive system in place to continuously orbit the Moon to identify, localise and classify spacecraft hardware on the Moon, including potential Arkhipov-Bracewell Probes.

Using the optical capabilities of an iPhone 15 Pro Max as a gauge, it is possible to roughly calculate an optimal altitude to fly above the moon and take one square kilometre images that can be on-board processed in real time.[39] The focal length (f) and sensor size (S) of the iPhone camera are known quantities and can be used to help determine altitude.

---

[38] https://developer.apple.com/av-foundation/

[39] The iPhone is chosen for convenience based on the YOLO-ET work conducted with it so far; it is not spacecraft hardware, and not set up for other imaging capabilities (e.g. hyperspectral) that could ideally also be incorporated.

The iPhone 15 Pro Max camera Primary, Ultra-wide and Telephoto lens mode features are[40]:

- Primary: 48MP sensor, 2.44µm quad pixels, 24mm equivalent f/1.78-aperture lens, Dual Pixel AF, OIS;
- Ultra-wide: 12MP sensor, 13mm equivalent f/2.2-aperture lens, Dual Pixel AF; and
- Telephoto: 12MP sensor, 1.12µm pixels, 120 mm equivalent f/2.8-aperture lens, Dual Pixel AF, powered by a A17 Pro chipset.

With the focal length (f) of the iPhone camera lens and the sensor size (S) of the iPhone camera, a recommended altitude above the Moon's surface can be determined. Derived from the principles of similar triangles in optics and commonly used to calculate ground sampling distance (GSD) for aerial or satellite systems (reference), the following formula is typically used to ensure that the size of an image captured by a camera sensor is a projection of the desired width on the ground at the calculated altitude:

$$h = (1/S) \times f \times W \tag{4.1}$$

where W is the width of the area on the ground that the YOLO-ET satellite could practically image (in this case, the square root of 1 square kilometre, 1000 meters); f is the focal length of the iPhone 15 Pro max lens, S is the width of its camera sensor; and h is the resulting altitude to fly over the Moon's surface. This is then sufficient to capture surface resolution approximating YOLO-ET's spacecraft .

[40] https://www.apple.com/iphone-15-pro/specs/

object identification, based on the Apollo spacecraft training and identification experiments above, i.e. a pixel scale of about 0.5 to 2 metres per pixel.

Using the iPhone's telephoto lens with a 120 mm equivalent focal length and the telephoto sensor size approximating a standard 1/3 inch (0.85cm) sensor, about 4.8mm x 3.6mm, the $1km^2$ image area suggests that an altitude of some 50km like the LRO's would give image width of about 1km, or about 500 pixels, (1000m/ 2metres per pixel).

Whilst velocity and orbital dynamics must also be considered in the execution, this suggests that even before next generation iPhone developments and without seeking an orbital altitude lower than the LRO's (which would require higher velocity to sustain and challenge continuous image taking and processing), images well suited to high performance of the YOLO-ET model can be achieved.

With the above it is also possible to estimate the number of images the YOLO-ET satellite would make in one orbit, and how long it would take to cover the entirety of the lunar surface. Using the iPhone data above and the Lunar Reconnaissance Orbiter and its altitude and velocity as benchmarks, i.e. a Ground Sampling Distance (GSD) of approximately 2 meters per pixel at an altitude of 50 kilometres, and for an image width of 1 kilometre, or about 500 pixels (1000 metres / 2 metres per pixel), and with the LRO's orbital period at about 2 hours in one orbit, it could theoretically make about 11,228 one-kilometre-wide images.

This assumes continuous imaging, and does not account for manoeuvring, or breaks between imaging for data downlink and instrument calibration. To cover the entire surface area of the Moon, which is approximately 38 million square

kilometres, it would take about 3,384 orbits, assuming each orbit covers unique areas without overlap. In this scenario it would take approximately 6,768 hours or 282 days to cover the surface. Accounting for needed manoeuvres, data links, and optimised image overlaps, it reasonable to assume that the YOLO-ET satellite could complete a complete first cataloguing of existing and previously unidentified spacecraft hardware of 2m to 0.25m dimensions in under one year's time.

The cost for a YOLO-ET satellite would likely be comparatively in-expensive. The total cost of the Lunar Reconnaissance Orbiter (LRO) mission is reported to be approximately $583 million[41], including $504 million for the main LRO probe and an additional $79 million for the LCROSS satellite. This is an all-inclusive figure for design, planning, building, launching, and maintaining the orbiter. For comparison, startup companies like Planet, with their Earth Observation satellite cameras inspired by iPhone design, and ispace and CesiumAstro, two of the first companies attempting commercial landing and lunar orbital operations and communication satellites, are estimated to offer more attractive costing. Dove satellites are mass-produced CubeSats, building on standardised designs and off-the-shelf components. Based on complexity of the mission, they can cost approximately $50,000 to $300,000 per unit; additional testing can add approximately another $100,000 to the cost. ispace is offering its customers payload launches to lunar orbit at under $10 million US, and is developing an operations and communications platform together with

[41] See https://en.wikipedia.org/wiki/Lunar_Reconnaissance_Orbiter for further detail.

CesiumAstro at a likely subscription basis of less than $1 million per year. It is thus reasonable to assume that a pair of YOLO-ET satellites could be deployed to the Moon for a year of orbital operations at a cost of approximately $15 million or less, greater than an order of magnitude less than the LROC system. Such a mission would deliver highly efficient, purpose-driven scans of the lunar surface in real time, enabling edge processing to focus on high-confidence detections, conserving transmission bandwidth, and offering opportunity to navigate both orbital and surface resources to areas of key detections.

In advance of a YOLO-ET lunar satellite mission, citizen scientists could also use the YOLO-ET model to complement such satellite deployments, and if the data proves homogenous enough, perhaps exhaustively mine the LROC archive's millions of images to identify candidates for satellite remote sensing and *in situ* inspection. The YOLO-ET model I developed, tested and refined has been saved as a coreML model that can be used in any open-source Turi Create environment, and uploaded to any iOS system device for further experiments in remote sensing applications. With further advances in ML Computer Vision and implementing the latest versions of YOLO, citizen scientists can also contribute to strengthening the performance of future models for detecting, localising and classifying both Arkhipov Particles and Arkhipov-Bracewell Probes.

4.7. Chapter 4 Conclusions

The YOLO-ET model first developed and trained on images of particles on the surface of Tanpopo aerogel panels, and then used on more complex images of spacecraft particles on lunar regolith simulants and images of asteroid grains and micrometeorites, has now also demonstrated strong performance for detection of spacecraft hardware on the Moon in LROC images. Just as the work in the previous Chapters was a key step toward detection of Arkhipov Particles, YOLO-ET shows promise for the detection of Arkhipov-Bracewell Probes: it demonstrates an 80% detection rate for unseen Apollo lunar landing modules, and trained only on Apollo spacecraft, in a first test correctly identifies a known (to humans) Luna 16 as landing module in an unseen image. As a further demonstration of ability, the model detects two potential candidates for Luna 9, with confidence levels of 61% and 43%, a spacecraft whose exact location has thus far remained undetermined. The comparatively lightweight portability and processing demands of YOLO-ET suggest that is well suited to continuous video monitoring of the Moon's surface, with real-time object detection, localisation and classification ready to support the search for Arkhipov-Bracewell Probes.

Chapter 5

Future Work and Conclusions[42]

5.1. Future Work

Future space missions are expected to yield a much larger and more heterogeneous quantity of microscopic materials than treated in the experiments in Chapters 2 and 3. These missions include new asteroid interceptions, planetary expeditions, and most especially, robotic and human sample return missions to the Moon, facilitated by unprecedented cargo return capacities.[43] With the imminent rise of robotic and human activities on the Moon, the importance of *in situ* microscopic examination capabilities to distinguish these microparticles becomes increasingly important for (i) identifying and quantifying the flux of anthropogenic contaminants and lunar surface disturbances and (ii) for controlled experiments to better understand the flux of exogenous (IDPs, β-meteoroids, possible interstellar dust) and indigenous (secondary impact ejecta) microparticles, with important implications for characterising the quantity of volatiles held in micro-structures and their resource potential [133, 134] (Figure 5.1).

---

[42] As in Chapters 2 and 3 much of this Chapter is based on the full-length article I published April 2024 in *Astronomy and Computing* and attached here in this Thesis at the Special Appendix.

[43] Although sample returns from human missions are set to begin with Artemis III, including the aerogel exposure experiments I've proposed together with the Astrobiology Project Tanpopo team, the quantities of material will remain limited, making the *in situ* work all that much more important.

Figure 5.1: ispace lunar lander rendering. Startups like ispace are already adding to a rapid increase in robotic activities on the Moon, and are contracting for payloads that could accommodate a variety of experiments relevant to the work of this Thesis, including a mini-CLOXS. Credit: ispace

By moving from more conventional Machine Learning approaches to the YOLO-ET model specifically developed here for the detection, localisation and classification of microparticles from and in the space environment, this work has opened the door to rely more on compressed Machine Learning models, existing high-performance GPU code, and commercially available software libraries; this permits developing, training and testing algorithms on systems and hardware that would readily fit into a CubeSat class spacecraft, lunar rover, or planetary sampling missions (Figure 5.2).

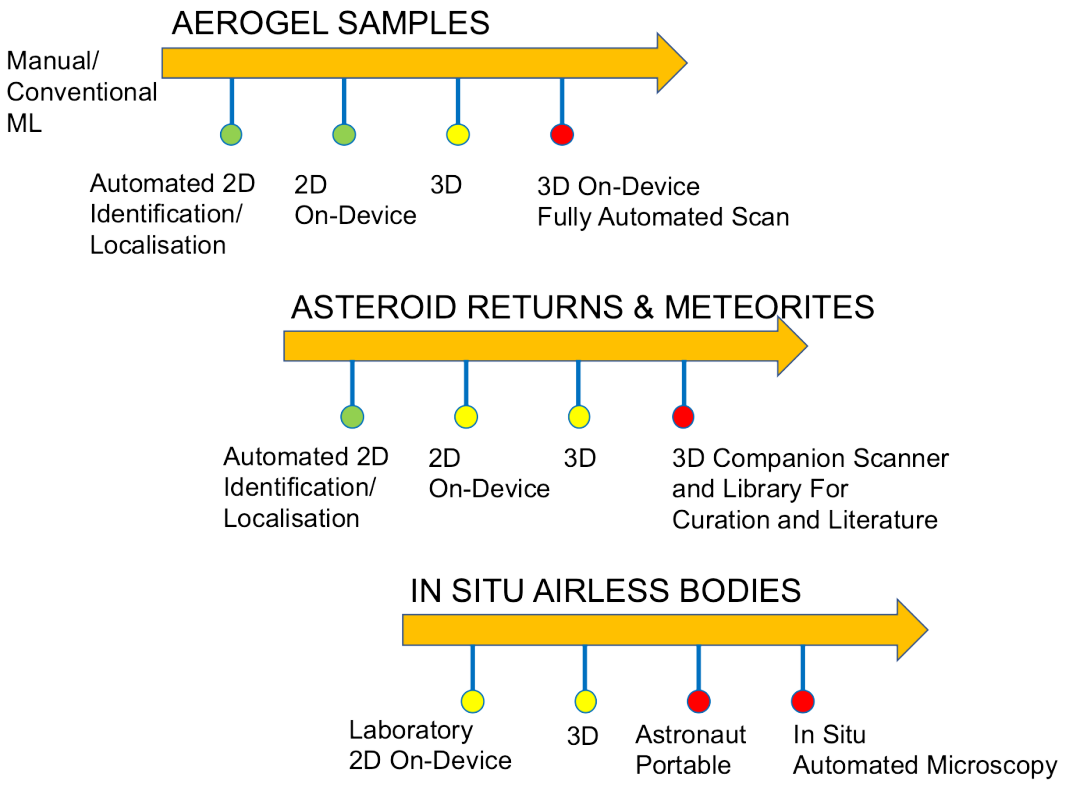


Figure 5.2: Current and Future Work: Progressing YOLO-ET on mobile devices from Earth receiving laboratories to *in situ* analyses on airless bodies. Green circles indicate experiments and proofs of concepts undertaken in this Thesis; yellow circles indicate next work in the planning phase; red circles indicate goals for each sequence.

For future work inspecting silica aerogels for captured microparticles, I am progressing to fully automated scans in real time in the receiving laboratory clean room environment, and from 2D surface scans to 3D inspection of hypervelocity impact candidates. Similarly, the work on 2D slice SEM images of the Ryugu samples correlated to images of micrometeorites as highlighted in Chapter 3 and other asteroid return samples is planned to be extended to 3D images. Moving from Ryugu to Bennu samples would be a natural next step [72].

This on-device, mini-CLOXS capability can then be made fully portable and ruggedised to perform scans *in situ* on airless bodies, obviating the need, risks, time and costs for sample return to Earth and better complementing Earth-based studies. These capabilities can also be extended from direct observation of micrometeorite particles on Earth to direct observation of microparticles on other planetary surfaces, without aerogel panel capture, by searching for the characteristic microcratering of impactors as discussed in Chapter 3, or e.g. by capturing low-velocity particles on carbon nanotube structures, as already deployed on witness plates on the International Space Station as part of Astrobiology Project Tanpopo [80].

Similarly moving from 2D to 3D scans, these devices can serve as companions for Earth-based space microparticle curators and assist in a more complete and integrated cataloguing from both newly collected samples and the extended prior literature. Ultimately, I aim to help create an astronaut-portable, 3D real-time scanning capability for microparticle detection, localisation and classification to assist automated microscopy on the Moon and other airless bodies. Having worked closely with Professor Yano and the Tanpopo and Ryugu teams at ISAS on applying YOLO-ET to samples and images in conjunction with this Project, I've had the exciting opportunity to help prepare a proposal for applying my model to Tanpopo-style panels to be placed on the lunar surface and returned by Artemis astronauts. As a direct product of our collaborations, our goal is to provide intact capture and post-

retrieval analyses of ß-meteoroids, Interstellar Dust Particles, and opportunistically, secondary impact ejecta from lunar regolith grains, using silica aerogel and carbon nanotube witness plate combinations as described above.

As discussed in Chapter 3, many of the particles of interest that might be distinguished on the surface of the Moon and other airless bodies, whether from cislunar space operations, asteroid and comet fragment intersections, or possible interstellar dust, including Arkhipov Particles, may arrive at hypervelocity impact speeds. Thus in future work I plan to have materials similar to those used in the YOLO-ET experiments distinguishing microparticles on lunar regolith simulants, fired at hypervelocities into monocrystalline target materials, to better scale and model characteristic microcrater morphologies as well as their patterns of vaporisation and spallation. It is envisaged that these tools and techniques can practically be packaged for use on board future mineral assay and mining devices deployed on the Moon and asteroids, to help identify micrometeoroids and other microscopic particles of interest in the surface regoliths.

Finally when contemplating spacecraft-borne YOLO-ET mobile device systems extended to both fly-bys and orbital missions it could be useful to adapt the image detection, localisation and classification processes from the microscopic-particle to the macro scale, as I detailed and proposed in Chapter 4. Real-time detection of features of interest on the Moon and asteroids to assist selection and navigation of surface sample operations, or to monitor and catalogue spacecraft debris and artefacts, could be amongst the next useful implementations. Thus the scope of future work also includes YOLO-ET analyses of Lunar

Reconnaissance Orbiter images for detection and classification of spacecraft hardware on the Moon, including for possible remnants of Arkhipov-Bracewell Probes, and commissioning and supporting citizen science efforts to use YOLO-ET on a publicly available basis, both on LROC images and as part of a dedicated YOLO-ET orbiter as described in Chapter 4. Future *in situ* missions could also include rovers, robots and astronauts equipped with similar detection systems deployed to some of the Moon's most pristine dust preserving environments, in ancient lava tubes exposed under pit craters.

Recent advances in materials science, sensing and computing can make construction of Arkhipov-Bracewell probes for our own exploration uses practical. And with appropriate space policy that can support using the Moon as a scientific and natural resource for surviving our planetary crises, in the course of our operations we may make potential discoveries of Arkhipov Particles or the remnants of Arkhipov-Bracewell Probes, and thereby also come to better understand and value our own uniqueness in the Galaxy.

5.2. Conclusions

The use of Computer Vision and Machine Learning with YOLO-ET in the search for Arkhipov Particles strongly complements how imminent robotic and human activities on the Moon and other planetary bodies would also benefit from these advanced capabilities *in situ* to identify and quantify microparticle terrestrial contaminants, lunar regolith disturbances, the flux of interplanetary dust particles, possible interstellar dust, β-meteoroids, and secondary impact ejecta.

In particular I have shown that:

1. Training on images from Tanpopo aerogel panels returned from Japan's Kibo module of the International Space Station, YOLO-ET demonstrates a 90% detection rate for all types of anthropogenic contaminants on aerogel surfaces and shows promising early results for detection of both microparticle contaminants on the Moon and for evaluating asteroid return samples.

2. YOLO-ET significantly improves on earlier ML processes in time savings, performance, and more efficient use of computing resources, and thereby can bring value to Tanpopo CLOXS machine processing by requiring fewer resources to more quickly and accurately identify samples of interest within the aerogel panels for extraction, while reducing contamination risk by more accurately and precisely selecting only those samples of interest, and allowing for the timeliest distribution of extracted samples to analysis groups around the world.

3. YOLO-ET's application to identifying spacecraft-derived microparticles in lunar regolith simulant samples and SEM images of Ryugu asteroid samples returned by Hayabusa2 also indicate strong model performance and transfer learning capabilities for future extraterrestrial applications. These are important first steps toward identifying: potential Arkhipov Particle impacts on collecting panels on the Moon and across the Solar System; characteristics of past Arkhipov Particle impacts on the Moon and other airless bodies, such as through their microcrater morphologies; and Arkhipov Particles that may have been processed into meteorites.

The application of YOLO-ET at the macro-scale, for the potential detection of Arkhipov-Bracewell Probes, also seems highly promising. On LROC images

it demonstrates an 80% detection rate for unseen Apollo lunar landing modules, and trained only on Apollo spacecraft, in a first test it correctly identifies a known (to humans) Luna 16 as landing module in an unseen image. As a further demonstration of ability, the model detects two potential candidates for Luna 9, with confidence levels of 61% and 43%, a spacecraft whose exact location has thus far remained undetermined. The comparatively lightweight portability and processing demands of YOLO-ET suggest that is well suited to continuous video monitoring of the Moon's surface, with real-time object detection, localisation and classification ready to support the search for Arkhipov-Bracewell Probes.

In sum, in this Thesis I have created a working model and search strategy for detecting Arkhipov Particles and Arkhipov-Bracewell Probes: for microparticles collected on clean surfaces, it does everything one could expect, tirelessly distinguishing localising and classifying as well or better than humans, and in portable fashion, ready to scale; it suggests real merit in putting collecting panels on everything in space. Where particles may vaporise, the model will be ready to learn from their impact morphologies. For complex microparticle and microstructure mixes, more data preparation for the training the model will likely be required, but the indications are good that it will work on these as well; I believe we should bring optical scanners to lunar and asteroid surfaces too.

The YOLO-ET model is performing remarkably well in spacecraft hardware detection. I expected the model might be overwhelmed by natural feature noise when moving it from one-shot, hyper-trained detections to full field random deployment. Instead it's making clean detections and appears very ready to learn more. I would be eager to engage both citizen scientists and automated systems

to inspect every LROC image in high resolution, and move quickly to get the scanner into lunar orbit and out to the asteroids.

My research supervisor Ian Crawford recently asked if we would now hypothetically be ready with YOLO-ET to find the 2001 monolith or its equivalent, perhaps *The Sentinel* pyramid that Sir Arthur based the story upon. At this point I believe I can assert the exhumed monolith would be identified by the model as Hardware with high confidence - its angularity, pattern of reflection and unique shadows in all illuminations would all be powerfully tell-tale signatures. For a still-buried monolith, if there were geometric patterns or tools left in the digging, burial and tidying then detection might also be probable, and it might be worth training the model to refine it in this direction – I believe the model is already learning on its own that sled and rover tracks point to artefacts. If the monoliths of our imagination were meant to house an extraterrestrial artificial intelligence, we now have our real-life AI companions to help us in the quest to discover them.

Chapter 6

Appendix and Bibliography

6.1. JSC-1 Lunar Simulant Components

JSC-1 Specifications https://ares.jsc.nasa.gov/projects/simulants/jsc-1-1a.html

6.2. Components Analyses Maana Electric Mare and Highlands Lunar Simulants

Lunar Mare Simulant Specifications https://maanaelectric.com/space_solutions/simulant_mare/ Lunar Highland Simulant Specifications https://maanaelectric.com/space_solutions/simulant_highland/

6.3. CesiumAstro Spacecraft Materials Ground and sieved to 80μm

SDR-1001 – the credit card-sized board with black solder mask CesiumAstro's Gen1 Software-Defined Radio (SDR) product for LEO and airborne applications with 100 MHz IBW, operating from 300 MHz to 6 GHz. The board construction: 16-layer PCB consisting of Arlon 85N (polyimide) and copper layers Prepreg layers are 85N with 106, 1080, and 2313 glass weaves. Plating is ENIG (gold over electroless nickel) Solder mask is Taiyo PSR-4000 MP. Black Solder applied to pads is SAC305 SAPA-1 – the board with four patches covered in reflective film S-band Antenna Patch Array for LEO applications operating in the 2.45 GHz region. The board construction: 6-layer PCB constructed of Rogers 4350B and copper layers Prepreg layers are RO4450F Plating is ENIG (gold over electroless nickel; latest revision of the antenna is in ImAg immersion silver finish). Solder mask is Taiyo PSR-4000 MP. The thin film radome on the front face of the antenna is Stamet-sputtered Kapton (Dunmore MO20295).

Bibliography

[1] L.C. Meiser, B.H. Nguyen, and Y.J. Chen. “Synthetic DNA applications in information technology.” *Nat Commun* 13 (2022), p. 352. doi:10.1038/s41467-021-27846-9.

[2] Dendra Systems. [online]. 2024 [Accessed 7 April 2024]. Available from: https://dendra.io/

[3] AtlasAI. [online]. 2024 [Accessed 7 April 2024]. Available from: https://www.atlasai.co/

[4] D.J. Kessler and B.G. Cour-Palais, "Collision Frequency of Artificial Satellites: The Creation of a Debris Belt", Journal of Geophysical Research, Vol. 83, No. A6, pp. 2637-2646, June 1, 1978

[5] D.J. Kessler and Johnson, N.L. Liou, J.-C., Matney, M. , “The Kessler Syndrome: Implications to Future Space Operations,” American Astronautical Society (AAS 10-016) In: Rocky Mountain Guidance and Control Conference ; 47-62 ; 2010

[6] LeoLabs. [online]. 2024 [Accessed 7 April 2024]. Available from: https://leolabs.space/

[7] A.C. Clarke. *The Sentinel in The Avon Science Fiction and Fantasy Reader*. Avon Periodicals: Inc, 1951. isbn: 0760701784.

[8] A.C. Clarke. *Introduction, The Sentinel, Masterworks of Science Fiction and Fantasy, Byron Preis Visual Publications*. NY, USA, 1983. isbn: 978-0425061831.

[9] D. Brin. *Existence*. A Tor book. Tom Doherty Associates, 2012. isbn: 9780765303615.

[10] Cixin Liu. *The Three-Body Problem*. Tor Books New York, NY, 2016. isbn: 9780765382030.

[11] Cixin Liu. *The Dark Forest*. Tor Books, New York, NY, 2016. isbn: 9780765386694.

[12] Cixin Liu. *Death’s End*. Tor Books, New York, NY, 2016. isbn: 9780765377104.

[13] Veerle, Sterken, A. Westphal, N. Altobelli, D. Malaspina, and F. Postberg (2019). “Interstellar Dust in the Solar System.” *Space Science Reviews*. 215. 43. 10.1007/s11214-019-0607-9.

[14] Sterken, Veerle J and Hunziker, S and Dialynas, K and Leitner, J and Sommer, M and Srama, R and Baalmann, L R and Li, A and Herbst, K and Galli, A and Brandt, P and Riebe, M and Baggaley, W J and Blanc, M and Czechowski, A and Effenberger, F and Fields, B and Frisch, P and Horanyi, M and Hsu, H-W and Khawaja, N and Krüger, H and Kurth, W S and Ligterink, N F W and Linsky, J L and Lisse, C and Malaspina, D and Miller, J A and Opher, M and Poppe, A R and Postberg, F and Provornikova, E and Redfield, S and Richardson, J and Rowan-Robinson, M and Scherer, K and Shen, M M and Slavin, J D and Sternovsky, Z and Stober, G and Strub, P and Szalay, J and Trieloff, M, "Synergies between interstellar dust and heliospheric science with an interstellar probe", RAS Techniques and Instruments, v2,1 pages 532-547, 2023. doi.org/10.1093/rasti/rzad034

[15] Tomonori Totani. Solid grains ejected from terrestrial exoplanets as a probe of the abundance of life in the Milky Way. International Journal of Astrobiology. 2023; 22(4):347-353. doi:10.1017/S147355042300006X

[16] A.V. Arkhipov. “The Moon as attractor of alien artifacts,” *Selenology, Journal of the American Lunar Society* 12.1 (1993), p. 6.

[17] A.V. Arkhipov. “Search for alien artifacts on the moon: a justification.” *RIAP Bulletin* 1.2 (1994), p. 9.

[18] A.V. Arkhipov. “Invasion effect on the moon. *Selenology, Journal of the American Lunar Society* 13.1 (1994), p. 9.

[19] Ian A. Crawford. “Back to the Moon?” *Astronomy Geo- physics: The Journal of the Royal Astronomical Society* 44.2 (2003), pp. 215–17.

[20] Ian A. Crawford. “The Astrobiological Case for Renewed Robotic and Human Exploration of the Moon.” *Internat. J. Astrobiology* 5 (2006), pp. 191–197.

[21] E. K. Jessberger, T. Stephan, D. Rost, P. Arndt, M. Maetz, F. J. Stadermann, D. E. Brownlee, J. P. Bradley, G. Kurat (2001). Properties of Interplanetary Dust: Information from Collected Samples, in Grün, E., Gustafson, B.A.S., Dermott, S.F., Fechtig, H. (Eds.) Interplanetary Dust, pp. 253–294, Springer-Verlag.

[22] R.N. Bracewell. “Communications from Superior Galactic Communities.” *Nature* 186.4726 (1960), pp. 670–671.

[23] Katherine H. Joy *et al.* “The Moon: An Archive of Small Body Migration in the Solar System.” *Earth, Moon, and Planets* 118.2-3 (2016), pp. 133–58.

[24] R. Bugiolacchi *et al. The Moon Zoo Citizen Science Project: Preliminary Results for the Apollo 17 Landing Site*. Icarus, 2016, 271:30–48.

[25] Robinson M, Brylow SM, Tschimmel M, Humm D, Lawrence SJ, Thomas PC et al. Lunar reconnaissance orbiter camera (LROC) instrument overview. Space Science Reviews. 2010 Jan; 150(1-4): 81-124. doi: 10.1007/s11214-010-9634-2

[26] W. Davies and R.V. Wagner. “Searching for alien artifacts on the moon.” *Acta Astronaut* 89 (2013), pp. 261–265.

[27] Lewis J. Pinault and Ian A. Crawford. “Advancing a Search for Non-Terrestrial Artefacts on the Moon.” *Poster Presentation. 3rd European Lunar Symposium*. Frascati Italy, 2015.

[28] L. Billings. *Search for Extraterrestrial Intelligence Nets Historic Cash Infusion*. Scientific American, 2015. [online]. Accessed July 2, 2024. Available from: https://www.scientificamerican.com/article/search-for-extraterrestrial-intelligence-nets-historic-cash-infusion/

[29] J.T. Wright. “Strategies and Advice for the Search for Extraterrestrial Intelligence.” *Acta Astronautica* 188 (2021), pp. 203–214.

[30] S.P. Worden *et al.* “Breakthrough Listen - A new search for life in the Universe.” *Acta Astronaut* 139 (2017), pp. 98–101.

[31] Robert A. Freitas, Francisco Valdes, The search for extraterrestrial artifacts (SETA), Acta Astronautica, Volume 12, Issue 12, 1985, Pages 1027-1034, ISSN 0094-5765, https://doi.org/10.1016/0094-5765(85)90031-1.

[32] A. Lesnikowski, V. Bickel, and D. Angerhausen. “Unsupervised distribution learning for lunar surface anomaly detection” arXiv 2001.04634. 2020.

[33] Totani T. Solid grains ejected from terrestrial exoplanets as a probe of the abundance of life in the Milky Way. *International Journal of Astrobiology*. 2023;22(4):347-353. doi:10.1017/S147355042300006X

[34] A.T. Basilevsky *et al.* “Survival times of Meter-sized Rock Boulders on the Surface of Airless Bodies.” *Planetary and Space Science* 117 (2015), pp. 312–28.

[35] James Benford, A Drake Equation for Alien Artifacts. Astrobiology Vol. 21, No. 6 (20210. https://doi.org/10.1089/ast.2020.2364

[36] Crawford, I.A., Schulze-Makuch, D. Is the apparent absence of extraterrestrial technological civilizations down to the zoo hypothesis or nothing? *Nat Astron* 8, 44–49 (2024). https://doi.org/10.1038/s41550-023-02134-2

[37] Dick, Steven J. “The Drake Equation in Context.” Cambridge University Press 2015.

[38] D.A. Vakoch and M.F. Dowd, eds. *The Drake Equation*. Cambridge University Press, Cambridge, UK. 2015.

[39] K. Denning. “Astrobiology, History and Society.” *Advances in Astrobiology and Biogeophysics*. Ed. by D.A. Vakoch. Springer Berlin, 2013, pp. 301–312.

[40] F. Drake and D. Sobel. *Is Anyone Out There? : The Scientific Search for Extraterrestrial Intelligence.* Delacorte Press, New York, NY, 1992. isbn: 9780385305327.

[41] C. Boettner, P. Dayal, M. Trebitsch, N. Libeskind, K. Rice, C. Cockell, B. I. Tieleman A&A Populating the Milky Way - Characterising planet demographics by combining galaxy formation simulations and planet population synthesis models 686 A167 (2024) DOI: 10.1051/0004-6361/202449557

[42] R.A. Freitas and F. Valdes. “A search for natural or artificial objects located at the earth-moon libration points.” *Icarus* 42 (1980), pp. 442–447.

[43] A.V. Arkhipov. “On the importance of nonclassical SETI.” *The Observatory*. V.113.- N1117.- P.306-307 (1993).

[44] Arkhipov A.V. Archeological elements of lunar exploration // Solar System Research.- 1994.- V.28.-N4-5.- P.467-469

[45] A.V. Arkhipov. “Problem of search for intelligent life on the moon.” *Institute of Radio Astronomy* (1994). preprint No. 70, Kharkov (in Russian).

[46] Crawford, Ian (2010). Astrobiological benefits of human space exploration. Astrobiology 10 (6), pp. 577-587. ISSN 1531-1074.

[47] I.A. Crawford and D. Schulze-Makuch. “Is the apparent absence of extraterrestrial technological civilizations down to the zoo hypothesis or nothing?” *Nat Astron* (2023). doi: 10.1038/s41550- 023-02134-2. url: https://doi.org/10.1038/s41550-023- 02134-2.

[48] Avi Loeb and E.L. Turner. “Detection technique for artificially illuminated objects in the outer Solar system and beyond.” *Astrobiology* 12.4 (2012), pp. 290–294.

[49] M. Lingam and Loeb A. *Searching the Moon for Extrasolar Material and the Building Blocks of Extraterrestrial Life*. arXiv:1907.05427v1 2019.

[50] G. Cocconi and P. Morrison. “Searching for Interstellar Communications”. *Nature* 184.(4690) 1959, pp. 844–846.

[51] R.N. Schwartz and C.H. Townes. “Interstellar and interplanetary communication by optical masers.” *Nature* 190.(4772) (1961), pp. 205–208.

[52] F.D. Drake. “Project Ozma.” *Physics Today* 14.(4) (1961), pp. 40–46.

[53] F.J. Dyson. “Search for Artificial Stellar Sources of Infrared Radiation.” *Science* 131.3414 (1960), pp. 1667–1668.

[54] J.C. Tarter. “The evolution of life in the Universe: Are we alone?” *Highlights of Astronomy* 14 (2007), pp. 14–29.

[55] M. Lingam and A. Loeb. *Life in the Cosmos. From Biosignatures to Technosignatures*. Harvard University Press, Cambridge Massachusetts, 2021.

[56] S.Z. Sheikh. “Nine axes of merit for technosignature searches.” *Int. J* Astrobiol.,19(3):237-243 (2020).

[57] S.Z. Sheikh *et al.* “The Breakthrough Listen search for intelligent life: A 3.95-8.00 GHz search for radio technosignatures in the restricted Earth Transit Zone.” *Astron. J* 160.1 (2020).

[58] Klein, M.J., Gulkis, S. (1991). The impact of technology on SETI. In: Heidmann, J., Klein, M.J. (eds) Bioastronomy. Lecture Notes in Physics, vol. 390. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-54752-5_216

[59] Tipler, F.J. “Extraterrestrial Intelligent Beings do not Exist” Royal Astronomical Society, Quarterly Journal, vol. 21, Sept. 1980, p. 267-281.

[60] von Neumann Theory of self-reproducing automata. by: Von Neumann, John, 1903-1957; Burks, Arthur W. (Arthur Walter), 1915-2008. Publication date: 1966.

[61] C. Maccone. “The statistical Drake Equation”. *Acta Astronaut* 67.11 (2010).

[62] Tanpopo Special Issue, Astrobiology vol 21 no 12 dec 2021.

[63] Dartois, E., Engrand, C., Brunetto, R., Duprat, J., Pino, T., Quirico, E., Remusat, L., Bardin, N., Briani, G., Mostefaoui, S., *et al*., 2013. Ultra Carbonaceous antarctic micrometeorites, probing the solar system beyond the nitrogen snow-line. Icarus 224 (1), 243–252.

[64] M.S. Prasad *et al.* “Characterisation, Sources and Flux of Unmelted Micrometeorites on Earth During the Last 50,000 Years.” *Nature Scientific Reports* 8887 (2018), pp.1–8.

[65] Rojas, J., Duprat, J., Engrand, C., Dartois, E., Delauche, L., Godard, M., Gounelle, M., Carrillo-Sánchez, J., Pokorny`, P., Plane, J., 2021. The micrometeorite flux at Dome C (Antarctica), monitoring the accretion of extraterrestrial dust on earth. Earth Planet. Sci. Lett. 560, 116794.

[66] Flynn, G.J., 1994. Interplanetary dust particles collected from the stratosphere: physical, chemical, and mineralogical properties and implications for their sources. Planet. Space Sci. 42 (12), 1151–1161.

[67] Kurat, G., Koeberl, C., Presper, T., Brandstätter, F., Maurette, M., 1994. Petrology and geochemistry of antarctic micrometeorites. Geochim. Cosmochim. Acta 58, 3879–3904.

[68] Taylor, A., Baggaley, W., Steel, D., 1996. Discovery of interstellar dust entering the earth's atmosphere. 380. 323-325. 10.1038/380323a0. Nature 380, 323–325.

[69] Brownlee *et al.* 2006 Science vol 314 pp 1711-1716

[70] Nakamura, T., Noguchi, T., Tanaka, M., Zolensky, M.E., Kimura, M., Tsuchiyama, A., Nakato, A., Ogami, T., Ishida, H., Uesugi, M., Yada, T., Shirai, K., Fujimura, A., Okazaki, R., Sandford, S.A., Ishibashi, Y., Abe, M., Okada, T., Ueno, M., Mukai, T., Kawaguchi, J., 2011. Itokawa dust particles: a direct link between S-type asteroids and ordinary chondrites. Science 333 (6046), 1113–1116.

[71] Yada, T., Abe, M., Okada, T., *et al.,* 2022. Preliminary analysis of the Hayabusa2 samples returned from C-type asteroid Ryugu. Nat. Astron. 6, 214–220.

[72] Goldwin, T., 2023. The importance of asteroid sample return. Nat. Geosci. 16, 833.

[73] Yano, H., Fitzgerald, H., Tanner, W., 1994. Chemical analysis of natural particulate impact residues on the long duration exposure facility. Planet. Space Sci. 42, 793–802.

[74] Yano, H., Kibe, S., Deshpande, S.P., Neish, M.J., 1997. The first results of meteoroid and debris impact analyses on the space flyer unit. Adv. Space Res. 20, 1489–1494.

[75] Yamagishi, A., Hashimoto, H., Yano, H., Imai, E., Tabata, M., Higashide, M., Okudaira, K., 2021. Four-year operation of Tanpopo: astrobiology exposure and micrometeoroid capture experiments on the JEM exposed facility of the international space station. Astrobiology 21 (12), 1461–1472.

[76] Genge, M.J., Van Ginneken, M., Suttle, M.D., 2020. Micrometeorites: Insights into the flux, sources and atmospheric entry of extraterrestrial dust at earth. Planet. Space Sci. 187, 104900.

[77] Sasaki, S., Imani, J.-Y., Yano, H., 2019. Design, fabrication and evaluation of an aerogel processor CLOXS for the astrobiology mission Tanpopo. Biol. Sci. Space 33, 7–11.

[78] Westphal et al. 2014 MAPS 49, pp 1509-1521

[79] Burchell M. J., Mann J. R. & Bunch A. W. *Monthly Notices of the Royal Astronomical Society* , 352. 1273 - 1278 (2004).

[80] Yamagishi, A., Yokobori, S.-I., Hashimoto, H., Yano, H., Higashide, M., Tabata, M., Imai, E., Yabuta, H., Kobayashi, K., Kawai, H., 2014. Tanpopo: astrobiology exposure and micrometeoroid capture experiments—proposed experiments at the exposure facility of ISS-JEM. Trans. Jpn. Soc. Aeronaut. Space Sci. Aerosp. Technol. Jpn. 12 (29), Tk_49–Tk_55.

[81] Tabata, M., Kawaguchi, Y., Yokobori, S.-I., Kawai, H., Takahashi, J.-i., Yano, H., Yamagishi, A., 2011. Tanpopo cosmic dust collector: silica aerogel production and bacterial DNA contamination analysis. Biol. Sci. Space 25 (1), 7–12.

[82] Kawaguchi, Y., Yokobori, S.-I., Hashimoto, H., Yano, H., Tabata, M., Kawai, H., Yamagishi, A., 2016. Investigation of the interplanetary transfer of microbes in the Tanpopo mission at the exposed facility of the international space station. Astrobiology 16 (5), 363–376.

[83] Sasaki, S., Imani, J.-Y., Yano, H., 2019. Design, fabrication and evaluation of an aerogel processor CLOXS for the astrobiology mission Tanpopo. Biol. Sci. Space 33, 7–11.

[84] Westphal *et al*. 2004 MAPS (39) 8, 1375 – 1386.

[85] Okudaira K., Yamagishi A., and Yano H. 2022 Database of Space-Captured Samples in Tanpopo: Astrobiology Exposure and Micrometeoroid Capture Experiments. darts.isas.jaxa.jp

[86] Yano, H., Yamagishi, A., Hashimoto, H., Yokobori, S., Kobayashi, K., Yabuta, H., Mita, H., Tabata, M., Kawai, H., Higashide, M., *et al.*, 2014. Tanpopo experiment for astrobiology exposure and micrometeoroid capture onboard the ISS-JEM exposed facility. 45th Annual Lunar and Planetary Science Conference. 2934.

[87] Tabata, M., Kawaguchi, Y., Yokobori, S.-I., Kawai, H., Takahashi, J.-i., Yano, H., Yamagishi, A., 2011. Tanpopo cosmic dust collector: silica aerogel production and bacterial DNA contamination analysis. Biol. Sci. Space 25 (1), 7–12.

[88] Krizhevsky, A., 2009. Learning Multiple Layers of Features from Tiny Images (Master's thesis). University of Toronto, Master's thesis. University of Toronto.

[89] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25.

[90] Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

[91] Mitchell, T.M., 1997. Machine Learning. McGraw Hill, New York.

[92] Huertas-Company, M., Lanusse, F., 2022. The DAWES review 10: The impact of deep learning for the analysis of galaxy surveys. doi:10.48550/arXiv.2210.01813, arXiv preprint arXiv:2210.01813.

[93] Jeffrey, N., Lanusse, F., Lahav, O., Starck, J.-L., 2020. Deep learning dark matter map reconstructions from DES SV weak lensing data. Mon. Not. R. Astron. Soc. 492 (4), 5023–5029.

[94] Grishin, K., Mei, S., Ilic, S., 2023. YOLO-CL: Galaxy cluster detection in the SDSS with deep machine learning. doi:10.48550/arXiv.2301.09657, arXiv preprint arXiv: 2301.09657

[95] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 779–788.

[96] Girshick, R., 2015. Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1440–1448.

[97] Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., Miao, Y., 2021. Review of image classification algorithms based on convolutional neural networks. Remote Sens. 13 (22), 4712.

[98] Jaeger, L., Butterworth, A.L., Gainsforth, Z., Lettieri, R., Zevin, D., Ardizzone, A., Capraro, M., Burchell, M., Wozniakiewicz, P., Ogliore, R.C., *et al.,* 2021. Automatic detection of impact craters on al foils from the stardust interstellar dust collector using convolutional neural networks. Meteorit. Planet. Sci. 56 (10), 1890–1904.

[99] Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. doi:10.48550/arXiv.1409.1556, arXiv preprint arXiv:1409.1556

[100] Redmon, J., Farhadi, A., 2017. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7263–7271.

[101] Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B., 2022. A review of YOLO algorithm developments. Procedia Comput. Sci. 199, 1066–1073.

[102] Zhou, H., Xiao, Y., Zheng, Z., Yang, B., 2022. YOLOv2-tiny target detection system based on FPGA platform. In: 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering. ICBAIE, pp. 289–292. doi:10.1109/ICBAIE56435.2022.9985817.

[103] Neubeck, A., Van Gool, L., 2006. Efficient non-maximum suppression. In: 18th International Conference on Pattern Recognition. ICPR'06, Vol. 3, IEEE, pp. 850–855.

[104] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A survey on deep transfer learning. In: Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27. Springer, pp. 270–279.

[105] Fei-Fei, L., Deng, J., Li, K., 2009. ImageNet: Constructing a large-scale image database. J. Vis. 9 (8), 1037–1037.

[106] van Dyck, L.E., Roland, K., Jochen, D.S., Roland, G.W., 2021. Comparing object recognition in humans and deep convolutional neural networks—An eye tracking study. Front. Neurosci. 15.

[107] Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2018. Focal loss for dense object detection. doi:10.48550/arXiv.1708.02002, arXiv:1708.02002v2

[108] Ateaque, S., 2022. Neurotrophin-3 Signalling in Neurons Derived from Human Embryonic Stem Cells (Ph.D. thesis). Cardiff University.

[109] Meng, T., Zheng, J., Chen, M., Zhao, Y., Sudarjat, H., MR, A.A., Kulkarni, V., Oh, Y., Xia, S., Ding, Z., et al., 2023. Six-month effective treatment of corneal graft rejection. Sci. Adv. 9 (12), eadf4608.

[110] Grün, E., Horanyi, M., Sternovsky, Z., 2011. The lunar dust environment. Planet. Space Sci. 59 (14), 1672–1680.

[111] Pokorny`, P., Janches, D., Sarantos, M., Szalay, J.R., Horányi, M., Nesvorny`, D., Kuchner, M.J., 2019. Meteoroids at the moon: orbital properties, surface vaporization, and impact ejecta production. J. Geophys. Res.: Planets 124 (3), 752–778.

[112] Szalay, J., Pokorny`, P., Horányi, M., 2020. Hyperbolic meteoroids impacting the moon. Astrophys. J. Lett. 890 (1), L11.

[113] Costello, E.S., Ghent, R.R., Lucey, P.G., 2021. Secondary impact burial and excavation gardening on the moon and the depth to ice in permanent shadow. J. Geophys. Res.: Planets 126 (9), e2021JE006933.

[114] Joy *et al.* 2012 Direct Detection of Projectile Relics from the End of the Lunar Basin–Forming Epoch**.** Science 336 1426-1429.

[115] Lissa Ong, Erik I. Asphaug, Donald Korycansky, Robert F. Coker, Volatile retention from cometary impacts on the Moon, Icarus, Volume 207, Issue 2, 2010, Pages 578-589, ISSN 0019-1035, https://doi.org/10.1016/j.icarus.2009.12.012.

[116] R. Hibbert *et al.* "The Hypervelocity Impact Facility at the University of Kent: Recent Upgrades and Specialized Capabilities." *Procedia Engineering* 204 (2017), pp. 208–14.

[117] M.J. Burchell *et al.* "Hypervelocity Impact Studies Using the 2 MV Van De Graaff Accelerator and Two-stage Light Gas Gun of the University of Kent at Canterbury." *Measurement Science Technology* 10.1 (1999), pp. 41–50.

[118] Lewis J. Pinault, Machine Learning Data Analyses for Asteroid and Micrometeorite Samples, Hayabusa Symposium 2023, S13-04, November 15 – 18, Institute of Space and Astronautical Sciences, JAXA, Sagamihara Japan.

[119] Genge, M.J., Almeida, N.V., van Ginneken, M., Pinault, L. J., Wozniakiewicz, P. J., Yano, H. Evidence from 162173 Ryugu for the influence of freeze–thaw on the hydration of asteroids. Nat Astron (2024). https://doi.org/10.1038/s41550-024-02369-7

[120] Genge, M.J., Almeida, N.V., van Ginneken, M., Pinault, L. J., Wozniakiewicz, P. J., Yano, H. The Discovery of a Large Porphyritic Chondrule in 162173 Ryugu. 86th Annual Meeting of the Meteoritical Society 2024 (LPI Contrib. No. 3036).

[121] Genge, M.J., Almeida, N., Van Ginneken, M., Pinault, L.J., Wozniakiewicz, P., Yano, H., 2023. Ice and liquid water in asteroid Ryugu – Constraints from sample A0180. In: The 14th Symposium on Polar Science, National Institute of Polar Research, November 14-17, 2023 Tachikawa, Tokyo, Japan. pp. 1–2.

[122] Matthias Van Ginneken *et al.* "Chondritic micrometeorites from the Transantarctic Mountains." *Meteoritics & Planetary Science* 47.2 (2012), pp. 228–247.

[123] Alexey V. Arkhipov. "Extraterrestrial Artefacts". *The Observatory* 116 (1996), pp. 175–176.

[124] A.V. Arkhipov. "Earth-Moon System as a Collector of Alien Artifacts". *J. Br. Interplanet. Soc* 51.5 (1998), pp. 181–184.

[125] La Grassa R, Cremonese G, Gallo I, Re C, Martellato E. YOLOLens: A Deep Learning Model Based on Super-Resolution to Enhance the Crater Detection of the Planetary Surfaces. *Remote Sensing*. 2023; 15(5):1171. https://doi.org/10.3390/rs15051171

[126] Sinha, M., Paul, S., Ghosh, M. *et al.* Automated Lunar Crater Identification with Chandrayaan-2 TMC-2 Images using Deep Convolutional Neural Networks. *Sci Rep* 14, 8231 (2024). https://doi.org/10.1038/s41598-024-58438-4

[127] John A. Ball, The zoo hypothesis, Icarus, Volume 19, Issue 3, 1973, Pages 347-349, ISSN 0019-1035, https://doi.org/10.1016/0019-1035(73)90111-5.

[128] Ellery A. The prospect of von neumann probes and the implications for the sagan - tipler debate. International Journal of Astrobiology, 2022; 21(4): 197-199. doi :10.1017/S1473550422000301.

[129] Haase, I., Oberst, J., Scholten, F., Wa¨hlisch, M., Glser, P., Karachevtseva, I., Robinson, M., 2012. Mapping the Apollo 17 landing site area based on lunar reconnaissance orbiter camera images and Apollo surface photography. Journal of Geophysical Research: Planets 117.

[130] R.V. Wagner, D.M. Nelson, J.B. Plescia, M.S. Robinson, E.J. Speyerer, E. Mazarico, Coordinates of anthropogenic features on the Moon, Icarus, Volume 283, 2017, Pages 92-103, ISSN 0019-1035, https://doi.org/10.1016/j.icarus.2016.05.011.

[131] P.J. Stooke, Identification of the SMART-1 spacecraft impact location on the Moon, Icarus, Volume 321, 2019, Pages 112-115, ISSN 0019-1035, https://doi.org/10.1016/j.icarus.2018.11.009.
(https://www.sciencedirect.com/science/article/pii/S0019103518303774)

[132] Burgess, R. (2016). "The Effects of Space Weathering on Lunar Materials." *Icarus*, 273, 211-220.

[133] Christopher Yeh *et al*. "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-being in Africa." *Nature Communications* 2583 (2020).

[134] Hannah Kerner, Machine Learning for Remote Sensing, a Talk at the Committee on Seismology and Geodynamics (COSG) Fall Meeting, National Academies of Science, Engineering, and Medicine, Washington, DC. 2019.

[135] Russian Space Web [online]. 2024. [Accessed February 28, 2024] Available from: russianspaceweb.com/luna9.html.

[136] Analysis of Luna 9 Photography. Prepared by Lockheed Electronics Company Houston, Texas for Manned Spacecraft Center NASA, Washington, D C. l February 1968.

Special Appendix:

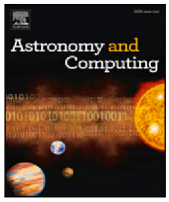
Astronomy and Computing YOLO-ET

Astronomy and Computing 47 (2024) 100828

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom



Full length article

YOLO-ET: A Machine Learning model for detecting, localising and classifying anthropogenic contaminants and extraterrestrial microparticles optimised for mobile processing systems

L.J. Pinault [a,b,*], H. Yano [c,d], K. Okudaira [c,e], I.A. Crawford [a,b]

[a] School of Natural Sciences Birkbeck College, Malet Street, London, WC1E 7HX, United Kingdom
[b] Centre for Planetary Sciences at University College London/Birkbeck, Gower Street, London, 61E 6BT, United Kingdom
[c] The Institute of Space and Astronautical Sciences, Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Chuo-ku, Sagamihara, Kanagawa, 252-5210, Japan
[d] Graduate University for Advanced Studies (SOKENDAI), 3-1-1 Yoshinodai, Chuo-ku, Sagamihara, Kanagawa, 252-5210, Japan
[e] Department of Computer Science and Engineering, University of Aizu, Aizu-Wakamatsu, Fukushima, 965-8580, Japan

ARTICLE INFO

Dataset link: http://github.com/apple/turicreate, https://github.com/LewisJPinault/YOLO-ET

Keywords:
Machine Learning
YOLO
Extraterrestrial materials
Lunar regolith
Aerogels
Anthropogenic contaminants

ABSTRACT

Imminent robotic and human activities on the Moon and other planetary bodies would benefit from advanced *in situ* Computer Vision and Machine Learning capabilities to identify and quantify microparticle terrestrial contaminants, lunar regolith disturbances, the flux of interplanetary dust particles, possible interstellar dust, β-meteoroids, and secondary impact ejecta. The YOLO-ET (ExtraTerrestrial) algorithm, an innovation in this field, fine-tunes Tiny-YOLO to specifically address these challenges. Designed for coreML model transference to mobile devices, the algorithm facilitates edge computing in space environment conditions. YOLO-ET is deployable as an app on an iPhone with LabCam® optical enhancement, ready for space application ruggedisation. Training on images from the Tanpopo aerogel panels returned from Japan's Kibo module of the International Space Station, YOLO-ET demonstrates a 90% detection rate for surface contaminant microparticles on the aerogels, and shows promising early results for detection of both microparticle contaminants on the Moon and for evaluating asteroid return samples. YOLO-ET's application to identifying spacecraft-derived microparticles in lunar regolith simulant samples and SEM images of asteroid Ryugu samples returned by Hayabusa2 and curated by JAXA's Institute of Space and Astronautical Sciences indicate strong model performance and transfer learning capabilities for future extraterrestrial applications.

1. Introduction

Extraterrestrial microparticles though millimetres or less in size, bear wide-ranging significance for understanding planetary system origins, delivery of water and life precursor materials to Earth and other planetary bodies, developing planetary protection measures, and identifying the distribution of potential resources in the Solar System. Missions to low Earth orbit, the Moon, asteroids and deep space destinations have already created a substantive inventory of these particles including:

1. Micrometeorites (MMs) — With their smaller mass, lower deceleration through the atmosphere and gentler Earth impact, some surviving micrometeorites are found to be relatively unaltered, with unmelted portions giving direct evidence of their precursor bodies and evolutionary sequence. MMs are generally categorised as meteoroids reaching the Earth's surface, and recovered like meteorites, with sizes in the 10s to 100s of μm (Dartois et al., 2013; Prasad et al., 2018; Rojas et al., 2021);
2. Interplanetary Dust Particles (IDPs) — Finer grained and captured in the stratosphere, with sizes up to 10 μm, IDPs are effectively a category of MMs, and are also presumed to be of asteroidal and cometary origin, like Antarctic Micrometeorites (AAMs) and Cosmic Spherules (CSs), fully melted and recondensated meteoroids recovered from the deepsea floor; (Flynn, 1994; Kurat et al., 1994);
3. Interstellar Dust Particles (ISPs) — Originating from outside our Solar system and owing to the Sun's Galactocentric orbit and other influences, these particles can travel at Earth encounter speeds of up to ~100 $\mathrm{km\,s^{-1}}$ or greater hypervelocities (Taylor et al., 1996);
4. Lunar and asteroidal regoliths — The Apollo and Luna missions of 1969–1976 and the Chang 'e 5 mission of 2020 returned dust

* Corresponding author at: School of Natural Sciences Birkbeck College, Malet Street, London, WC1E 7HX, United Kingdom.
E-mail address: l.pinault@ucl.ac.uk (L.J. Pinault).

https://doi.org/10.1016/j.ascom.2024.100828
Received 30 December 2023; Accepted 25 April 2024
Available online 3 May 2024

particles from the Moon to Earth, and the asteroid sample returns by Hayabusa in 2010 (Nakamura et al., 2011), Hayabusa2 in 2020 (Yada et al., 2022) and OSIRIS-REx in 2023 (Goldwin, 2023) also successfully added to Earth's inventory; and

5. Anthropogenic contaminants — Fragments from spacecraft exteriors, engines, spacesuit microfibres and outgassed materials from extravehicular activities are produced in both normal operations and as a result of material degradation and microparticle impacts in space. These are likely to accompany human and robotic activities on the Moon (e.g. Yano et al., 1994, 1997; Yamagishi et al., 2021).

Each of these categories of microparticles in and from the space environment has its own significance, but many are also interrelated, and in practice, on lunar and asteroid surfaces they may be mixed or amalgamated together. Micrometeorites may offer direct comparison to past asteroid and lunar sample returns for example, affording a recalibration of terrestrial micrometeorite collections by overcoming the selective biases of atmospheric entry, an important step toward better understanding Solar system formation processes (e.g. Genge et al., 2020).

Almost all microscopic analyses of extraterrestrial samples involve a detailed examination of their petrological features, textures, mineralogy, and chemical composition, drawing on a depth of research expertise, judgement and experience to offer classification suggestions and understand origins and implications for the early solar system and more. Besides a heavy experience requirement, the equipment required for these undertakings can be their own burden. Even the most recent automated micro-scanning systems in extraterrestrial sample labs, while increasingly powerful, are substantially sized and practically immobile (Sasaki et al., 2019). Their power-intensive requirements, slow speeds of operation and high consumption of computing resources can lead to lengthy processing times. Here we introduce a novel approach using Computer Vision and AI Machine Learning combined with advanced on-device optical and computing technologies, that can serve as an important complement to researchers' experience and a companion to their field efforts. These advancements can overcome the limitations of current systems to rapidly and accurately identify, localise, and classify microparticles making it a more robust and practical solution for *in situ* anthropogenic contaminant and extraterrestrial sample analyses.

In this study, we seek to harness the potential of AI Machine Learning to address a specific challenge in the planetary sciences: the identification and classification of micron to millimetre scale extraterrestrial particle impacts and features. The application of AI in image classification is a topic of fast-growing interest across various fields; our specific contributions are rooted in the novel application of these techniques to the unique domain of extraterrestrial particles, alongside the development of a specialised dataset and tailored training processes for their data handling.

We collect data from 'F (false)' samples which are not captured micrometeoroids but rather anthropogenic contaminants identified on the surfaces of the Tanpopo aerogel panels exposed outside of the JAXA Kibo module of the International Space Station. The data were prepared and imaged using a digital optical microscope, recording, and processing techniques designed for rapid and automated identification followed by initial morphological classification of the identified features by experienced space scientists (Section 2).

Following data acquisition and archiving, we introduce YOLO-ET (You Only Look Once ExtraTerrestrial), a modified YOLO deep learning algorithm trained on the aerogel panel images to provide an optimised pipeline for detecting these 'F' sample features (Section 3). We then analyse the performance of the trained model on unseen data in (Section 4.2) and conclude with discussions on the applications to anthropogenic contaminants introduced to lunar simulants, as well as micro-scaled features within asteroid Ryugu samples, and their potential correlations to micrometeorites from the TransAntarctic Mountains in (Section 6.1) and (Section 6.2) respectively.

2. Data acquisition and archiving

2.1. Astrobiology project Tanpopo

Tanpopo is Japan's first space experiment for astrobiology utilising the Exposure Facility of the Japan Experimental Module (JEM) of the International Space Station (ISS), designed for the exposure of extremophile microbes and astronomical organic analogues, and for the collection of potentially organic-bearing micrometeoroids impacting the ISS before entering the Earth's atmosphere, in order to explore the potential for and any evidence of two-way interplanetary transport of life precursors and life (Yamagishi et al., 2014). For impacting microparticles, including micrometeoroids, space debris, and possible terrestrial particles that might carry microbes as bioaerosols, the capture of these particles was achieved using silica aerogel capture panels (Tabata et al., 2011). These were first placed on the Exposed Experiment Handrail Attachment Mechanism (ExHAM) unit on the ISS in 2015–2019 for the Tanpopo mission and followed by the Tanpopo2 mission in 2019–2020 (Fig. 1).

The first set of silica aerogel panels was exposed for one year before being returned to Earth (Kawaguchi et al., 2016), and these aerogel panels were examined under the microscope in clean room facilities at the Institute of Space and Astronautical Sciences (ISAS) in Sagamihara Japan (Yamagishi et al., 2021). CLOXS, which stands for "Captured particles Locating Observation and eXtracting System" (Sasaki et al., 2019), is a specialised processing machine designed for the Tanpopo mission (Fig. 2). It processes the returned aerogel from space, placing them on an X-Y-Z coordinate stage, and autoscans and images them under the microscope to integrate a microscale map of the entire aerogel panel by moving the stage in micrometre increments (Fig. 3). When objects of interest are identified from the integrated mapping image by a scientist, the coordinates of the region of interest of the panel are recorded and the X-Y-Z stage can be automatically moved for revisiting the location for higher magnification investigation. The stack of the revisited images at different focal length depths may contain true penetration tracks and surface objects.

The Tanpopo mission's classification of surface impacts from hypervelocity impactors is pivotal to understanding not only the impact process, which can lead to vaporisation of the impactor, but also to glean information about the impactor's composition and origin from the remnants it leaves behind. Traditionally this involved laborious microscopic examination and imaging of 100s of samples, with inherent human errors, and earlier efforts by the authors focused on track types—carrot, pit crater, straight, and teardrop. The identified particles and particle impact tracks (Figs. 4 and 5) in the aerogel that are of interest are cut out into suitable-sized chips that contain impact tracks of particles captured in space, where a needle is then used to cut the aerogel without contamination; this is then distributed to research groups worldwide for detailed biological and chemical analyses of the captured microparticles (Yano et al., 2014).

As for microparticles collected on the surface of the Tanpopo aerogel panels, these are presumed not to be 'True' hypervelocity impactors associated with morphological features (i.e. carrot tracks, pit craters, straight tracks, and teardrop tracks), but rather 'False' incidentally collected particles impacted at much slower velocities such as material released by ISS docking and undocking activities, venting materials, secondary impacts from primary impact ejecta, possible spacecraft component fibres, and fragments of the aerogel itself. This study has prioritised accurately classifying surface residual effects of these 'F' samples, such as sputter, fibre, block, bar, and aerogel fragments. Semi-automated methods have been employed to enhance classification but until the work of this project, matching the expertise of human scientists has remained a challenge, requiring a series of manual resizings, whitening, and contrast adjustments to secure even modest levels of confidence. Given the abundance of samples, continuous improvement in automated techniques is essential to accurately assess the microparticle remnants.

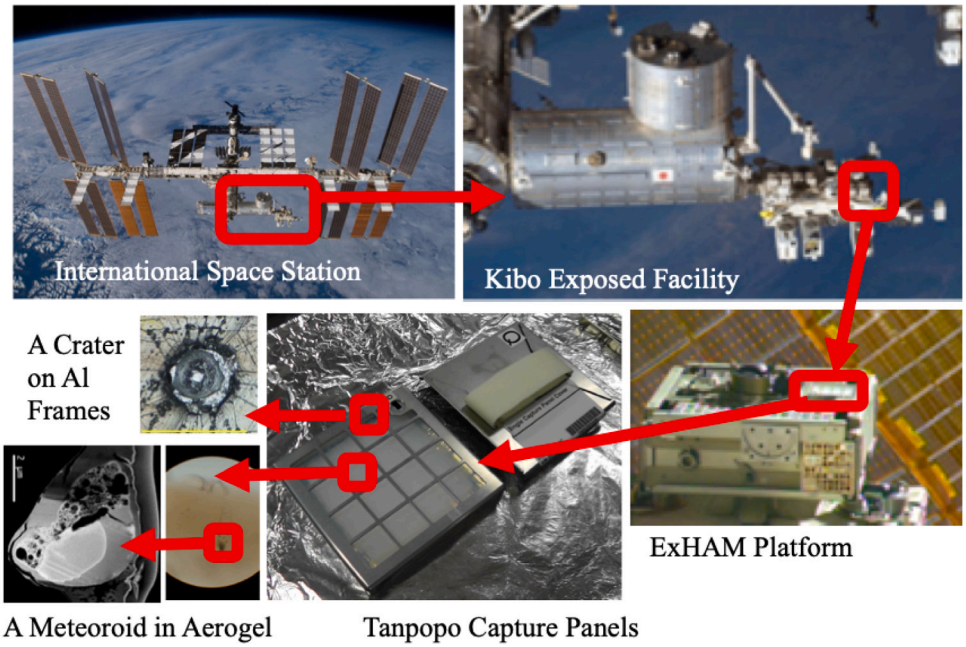


Fig. 1. Progressive zoom-in sequence of the Tanpopo astrobiology mission onboard the International Space Station (ISS), showing the placement of silica aerogel panels on the Kibo Exposed Facility for capturing impacting microparticles (Yamagishi et al., 2021). The sequence depicts the ISS with highlighted Kibo module, the ExHAM unit where aerogel panels are mounted, and close-up views of an impact crater on the aluminium frames of the panel and an aerogel panel in which an intact captured micrometeroid is discovered by subsequent analyses.

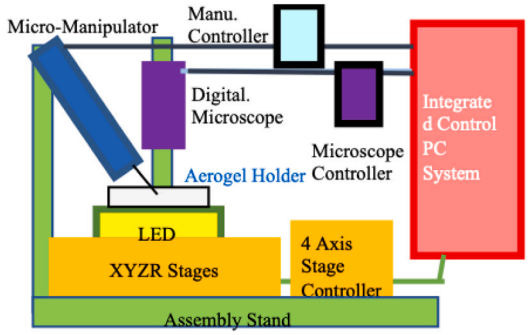


Fig. 2. Schematic representation of the CLOXS system, illustrating the precise arrangement of the micro-manipulator, aerogel holder, LED lighting, and XYZ stages mounted on an assembly stand, all coordinated by manipulator and microscope controllers, and integrated into a central control PC system for meticulous particle extraction and analysis.

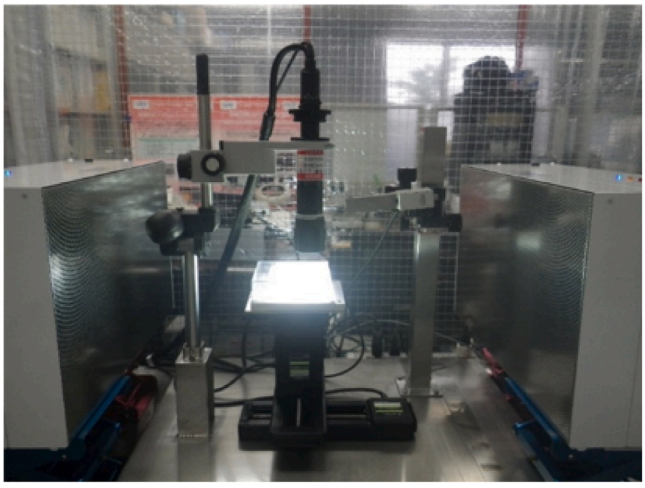


Fig. 3. The CLOXS system set-up in the ISAS clean room.

2.2. Machine learning dataset

This research focuses on the Tanpopo1-2 missions 2015–2020 collection of aerogel surface features larger than ~100 μm in the 10 cm × 10 cm aerogel panels captured at typically 100x to 245x magnifications. The total number of 'F' sample images in the collection is over 4000. In consideration for the computational power and memory limitations required to train machine learning models with large image input sizes, our data sample consists of 395 images, which is less than 10% of the total 'F' samples, in .jpeg format each 480 × 704 pixels in size. Our dataset is limited by the lack of annotated data (image-label pairs). With the open-source Python widget Bbox, we additionally manually annotate each image with a bounding box around each object (some images contain multiple particles and others none), and a class label of either 'sputter', 'block', 'fibre', 'bar', and 'aerogel fragment' (Fig. 6).

We note that the images have a variety of hues and brightnesses as they were captured under different lighting conditions, however in

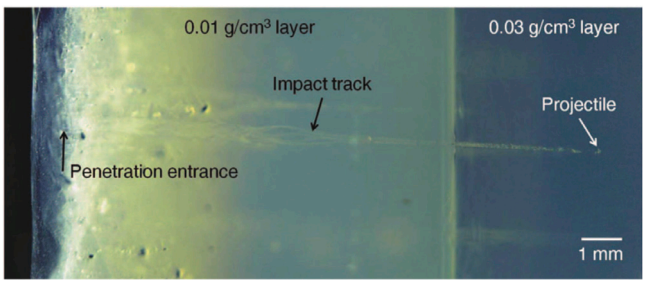


Fig. 4. Silica aerogel sample post-impact from a hypervelocity particle experiment, simulating the conditions for the Tanpopo project. This experiment conducted on Earth tests the resilience of the aerogel panels designed for microparticle collection onboard the ISS.
Credit: Tabata et al. (2011).

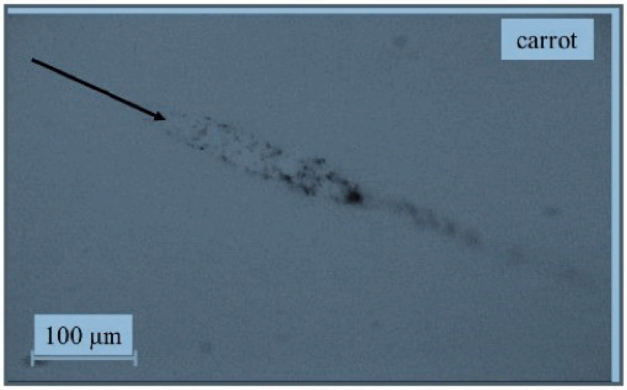


Fig. 5. A 'carrot' shaped track of a hypervelocity impactor in the Tanpopo silica aerogel panel returned from the International Space Station. The arrow indicates an impact direction.

Fig. 6. Examples of Tanpopo surface objects. From left to right, aerogel fragment, sputter, bar, block and fibre.

the interest of human time-saving we do not preprocess the images to account for this. Similarly we do not threshold, convert to grey scale or remove noise in the manner typically used to enhance the images for human inspection, as this would defeat the gains offered through the speed of Machine Learning. We are however inspired by the challenges and building on prior work using these more laborious methods (Krizhevsky, 2009; Krizhevsky et al., 2012).

We split our data sample into 80% for training (316 images) and 20% for testing (79 images). These are randomly sampled whilst maintaining the baseline ratio between different classes. The training data are the images used to optimise the model weights during the training phase of the model. The test data are not seen during the training of the model. We note that whilst it is common to additionally set aside a validation dataset for informing when the model is sufficiently trained — i.e. not under- or over- fit, our relatively limited data sample hampers our ability to reserve additional samples for validation, which could weaken the model's performance. Machine Learning methods are data greedy and perform better with more data, so to compensate for the relative data paucity, we apply automatic augmentation in the form of flipping the training images vertically, horizontally or both (corresponding to 180° rotation), resulting in a factor of 4 increase to the training sample (Fig. 7).

3. YOLO-ET: A highly efficient convolutional neural network for extraterrestrial microparticle detection and classification

3.1. Machine learning

Machine Learning (ML) (see Goodfellow et al., 2016, for a review) involves constructing layered architectures where each layer performs specific operations on data. These layers, particularly in neural networks, are composed of nodes or neurons with associated weights. During training, ML algorithms process input data through these layers, where each operation transforms the data based on the current weights.

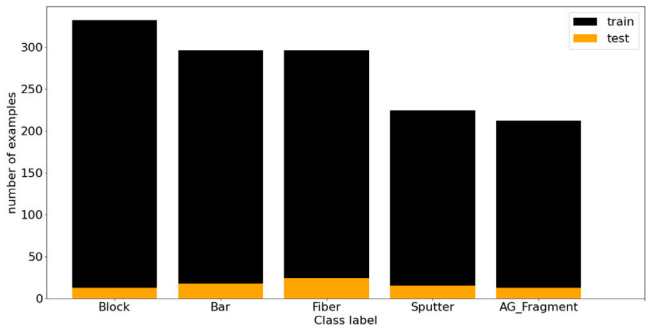


Fig. 7. Distribution of training and test data over the different classes. We note that these do not directly correspond to the number of images, as images can contain multiple objects of different classes or no objects altogether. Additionally there is significantly more training data as the augmentation is applied only to the training sample.

The goal is to optimise the weights to minimise a predefined cost or loss function (Section 3.3.2), which measures the difference between the algorithm's predictions and the actual outcomes. The optimisation is typically done using techniques like gradient descent, where the algorithm iteratively adjusts the weights based on the gradient of the loss, improving the model's predictions over time (Mitchell, 1997). Supervised learning, a subset of Machine Learning, involves algorithms that improve at tasks over time by learning from labelled data. Our project applies supervised learning to object detection, training models to recognise and categorise microscopic particles on aerogel panels. The data consists of pairs of images and their corresponding labels that are the bounding box coordinates (x and y), height (h), width (w), and class.

3.2. YOLO

Machine Learning (ML) algorithms employing Deep Learning techniques have been gaining traction in the astronomical sciences for nearly a decade, with applications ranging from galactic surveys (Huertas-Company and Lanusse, 2022), dark matter mapping (Jeffrey et al., 2020) and notably in regard to this work, galactic cluster detection (Grishin et al., 2023), using a streamlined YOLO technique. YOLO (Redmon et al., 2016), an acronym for "You Only Look Once", is a supervised learning approach to real-time object detection in computer vision. YOLO's novel architecture enables it to process images in a single pass, predicting both the bounding boxes and class probabilities (confidence scores) for objects within the image simultaneously. This contrasts with earlier two-step detection systems (e.g. Girshick, 2015), which would first propose regions and then classify them. The efficiency of YOLO allows it to detect objects rapidly with a high degree of accuracy, making it ideal for applications that require real-time processing.

Like more conventional techniques, YOLO is a type of convolutional neural network (CNN), consisting of a series of convolutional layers and pooling layers rather than neurons (Chen et al., 2021). Jaeger et al. (2021) use a 16-convolutional layer Visual Geometry Group CNN, VGG-16 (Simonyan and Zisserman, 2014) to classify impact craters on aluminium foils from the Stardust interstellar dust collector, which are typically less than one micrometre in size and sparse, making them difficult to find. While this method excels in accuracy for small objects, its deep architectures lacks YOLO's speed, limiting its use in real-time scenarios. Additionally, it primarily assesses the probability of crater presence without pinpointing exact locations, and is not optimised for images containing multiple objects of different classes.

3.3. YOLO-ET

YOLO-ET, is a modification of YOLO optimised for the detection of extraterrestrial microparticles. Specifically, we employ Tiny

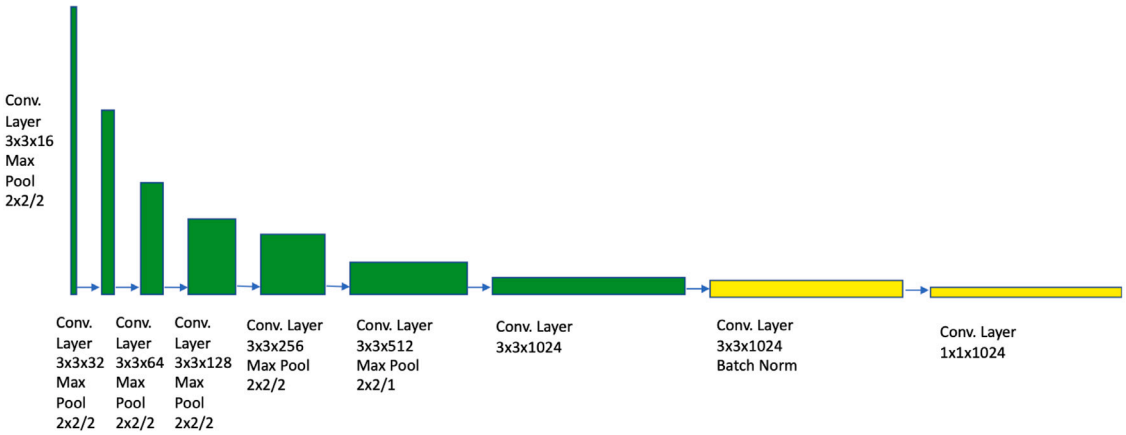


Fig. 8. TinyYOLOv2 architecture showing the series of convolutional and max pooling layers, with Batch Normalisation marked in yellow. The numbers represent the filter size and number of filters in each layer and the/represents the stride in the max pooling layers. Batch Normalisation is introduced after convolutional operations and before the activation functions, and leads to faster convergence during training by reducing internal covariate shift, i.e. the natural tendency to change the mean and variance of the inputs with each layer. Aside from helping to stabilise the training process by ensuring that the distribution of inputs to each layer remains more consistent during training, Batch Normalisation also helps regularise the model and reduce overfitting, so that the model can generalise better to unseen data.

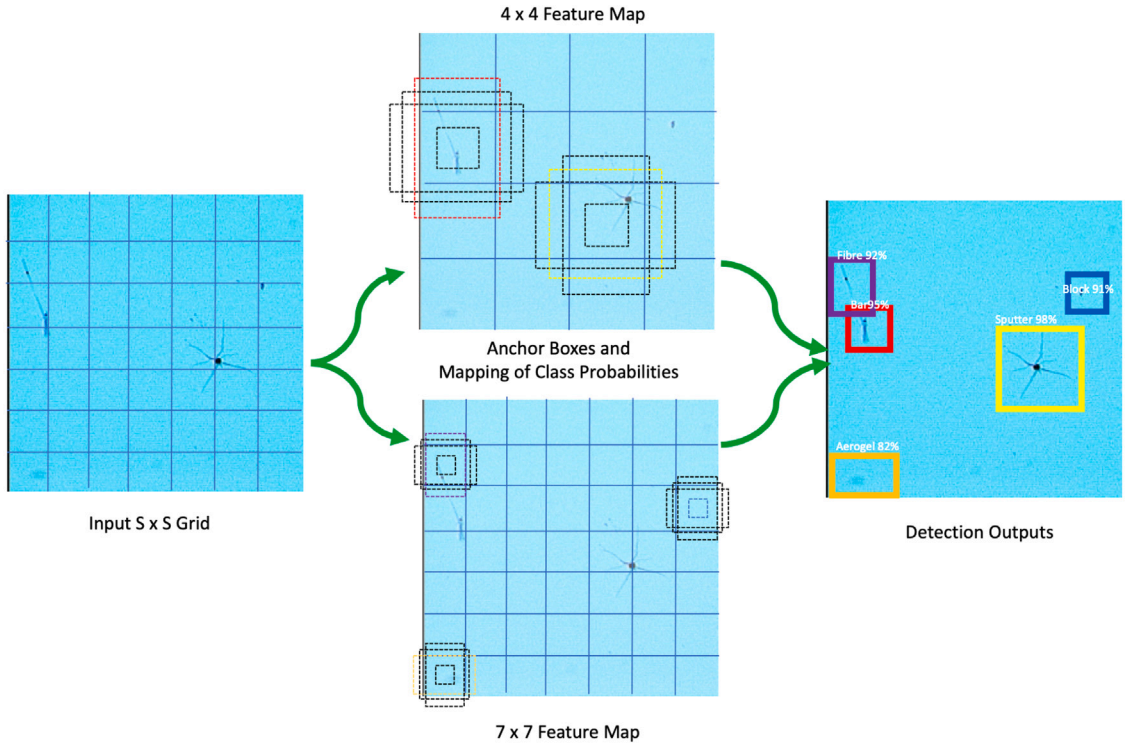


Fig. 9. Diagram illustrating the concept of anchor boxes in TinyYOLOv2, showcasing various predefined box shapes and sizes strategically positioned across a Tanpopo aerogel image in YOLO-ET implementation. These anchor boxes enable the efficient and accurate prediction of object boundaries and classifications within a single pass of the network.

YOLOv2 (Fig. 8), which is a smaller, simplified version of the original YOLOv2 (Redmon and Farhadi, 2017) with a Darknet-19 base network (a 19-layer network inspired by the VGG-16 model). For an overview on the development of YOLO see Jiang et al. (2022). YOLOv2 is designed to be more compact and faster than YOLO, making it suitable for applications with limited computational resources, such as mobile devices or real-time systems (see e.g. Zhou et al. (2022)). While maintaining the core principles of YOLO's single-pass detection, YOLOv2 simplifies the architecture with fewer convolutional layers and filters. In the original YOLO architecture, bounding box predictions were made relative to the dimensions of a grid cell; this approach had some limitations in terms of accuracy, particularly around predicting the correct size and location of objects. YOLOv2 improved upon this by predicting bounding box coordinates directly. Instead of the network learning offsets relative to a grid cell, YOLOv2 learns to predict bounding box coordinates relative to the location of the grid cell, along with anchor box dimensions, which makes predictions more precise (Fig. 9). This reduction in complexity results in faster processing speeds but typically at the cost of some detection accuracy compared to the full YOLO model. The "Tiny" version of YOLOv2 is specifically optimised to be more lightweight and faster, sacrificing some accuracy for the sake of speed and smaller model size. This makes Tiny YOLOv2 particularly well-suited for deployment in environments with limited computational resources, such as mobile devices, embedded systems, or applications where real-time performance is crucial. We underscore the suitability of the TinyYOLOv2 architecture for mobile use, which represents a significant advancement in deploying deep learning models on devices without the need for high-powered computing resources. YOLO-ET

thus embraces Tiny YOLOv2's trade-off between speed and precision, optimised for scenarios where real-time performance is crucial. For a deeper understanding of the efficiency and effectiveness of the TinyYOLOv2 architecture in mobile environments, we direct readers to the comprehensive study detailed in Zhou et al. (2022). This study provides empirical evidence supporting our choice of architecture, demonstrating its superior performance in scenarios demanding high efficiency and reliability on mobile devices.

3.3.1. Non-maximum suppression

Non-maximum Suppression (NMS) (Neubeck and Van Gool, 2006) is a post-processing technique commonly used in object detection algorithms and is a key feature in YOLO and YOLOv2, that ensures each detected object is only recognised once. When an object detection model predicts multiple bounding boxes for the same object, NMS helps in selecting the most probable bounding box and discarding the rest. It does this by comparing the overlap between two boxes (A and B) using a metric called Intersection Over Union (IOU),

$$IOU = \left| \frac{A \cap B}{A \cup B} \right|, \tag{1}$$

and retaining only the boxes below the defined IOU threshold, and above the defined confidence score threshold while suppressing the others. This reduces redundancy and increases both detection interpretability and accuracy. The default IOU threshold and confidence threshold in our work are 0.5 and 0.3, respectively, providing both a limited clutter of redundant bounding boxes and a practical level of accuracy for distinguishing microparticle types.

3.3.2. Loss function

The loss function quantifies the difference between the values predicted by the model and the actual values in the training data. A key goal in Machine Learning is to find the set of parameters, the weights and biases in the context of convolutional neural networks, that can optimise toward actual values by iteratively moving toward the minimum value of the loss function. The slope or derivative of the loss function with respect to its parameters is defined as its gradient, and moves in the direction of the steepest increase of the function. Moving in the opposite direction of the gradient, the algorithm iteratively adjusts the parameters to reduce loss, referred to as the gradient's descent. The learning rate is a hyperparameter that determines the size of the steps taken toward the minimum — too large, and the algorithm might overshoot the minimum, too small, and it will converge very slowly, consuming additional computing resource. Batch sizes determine the amount of data used to calculate the gradient at each step.

The YOLO loss function specifically combines terms for bounding box prediction accuracy, object presence confidence, and class prediction, ensuring the model is well-tuned across all aspects of object detection. In contrast to Grishin et al. (2023)'s work on galaxy clusters, YOLO-ET retains the comprehensive YOLO loss function, exploiting the full power of YOLO to simultaneously tackle the presence of multiple objects of different classes in an image. This is also useful for aerogel-captured particles, where multiple particles may overlap or appear at different depths in the aerogel panel, and it is essential for real world observational tasks in the planetary sciences like searching for microparticles and tell-tale microcraters *in situ* on the surface of the Moon.

Our loss function is defined as follows,

$$\begin{aligned} L = {} & \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\ & + \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\ & + \sum_{i=0}^{s^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} (C_i - \hat{C}_i)^2 \\ & + \lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{s^2} \mathbb{1}_{ij}^{obj} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2, \end{aligned} \tag{2}$$

where, the first two terms, weighted by λ_{coord}, penalise errors in the position (x, y) and size (w, h) of predicted bounding boxes compared to the ground truth. These are crucial for precise localisation. The third term penalise errors in object scores C_i, distinguishing between object presence and absence. The fourth term, scaled by λ_{noobj}, specifically penalise false detections and the final term assesses the classification error for each class c across the objects detected, ensuring accurate class predictions (Redmon et al., 2016).

3.3.3. Turi Create

We deploy YOLOv2 through Turi Create,[1] an open-source machine learning library developed by Apple. It provides a simplified approach to creating machine learning models, especially for developers interested in practical field application. Turi Create supports various types of models, including classifiers, recommender systems, and image classifiers, and is particularly known for its ease of use in creating models for iOS apps. The library is optimised for scalability and performance, enabling the development and deployment of models on both Macs and mobile iOS devices. Using Turi Create for object identification, localisation, and image classification is remarkably straightforward, allowing more user development time for focusing on the customisation of the learning model itself. AI Machine Learning is becoming increasingly accessible and user-friendly with applications such as Turi Create and Microsoft Lobe[2] providing highly accessible implementation of AI including in educational settings.[3]

This user-friendly entry point into object detection provides a streamlined experience, but at the cost of customisation depth. This abstraction means users are not able to fine-tune all model hyperparameters, i.e. the configuration settings of the network defined before training begins. In the case of YOLO-ET, these include: the learning rate — i.e. the magnitude by which the weights are updated during training, the anchor box dimensions, the NMS (confidence) threshold and the IOU threshold. It is also not trivial to employ a validation set directly within the framework. However, TuriCreate still affords some degree of control, allowing for the adjustment of hyperparameters such as batch size and maximum iterations, which can significantly influence model performance and training time.

During training, we experimented with various batch sizes. While larger batches demand more memory due to the increased number of images loaded simultaneously, they tend to smooth out the loss curve, leading to a more stable model. Conversely, smaller batches, although more memory-efficient, can result in a noisier gradient descent trajectory. High-resolution inputs restricted our batch capacity, thereby decelerating the training convergence. Nevertheless, larger batches expedited convergence toward the global minimum of the loss function. For this project, a batch size of 32 was identified as the most effective, balancing computational resource demands and learning stability.

In the training of our model, an epoch is defined as one complete pass through the entire dataset, whereas an iteration is one update of the model's weights, which occurs after processing a batch of samples. Recall our model uses 1264 training images with a batch size of 32, each epoch consists of 1264/32 = 39.5, approximately 40 iterations. Setting max_iterations to 2000 means the training process involved

[1] https://github.com/apple/turicreate.

[2] https://www.lobe.ai/.

[3] Lobe requires neither Machine Learning nor coding experience and should enable a wide range of user engagements. Currently however its templates are only set up for Image Classification tasks.

roughly 2000/40 = 50 epochs. After experimenting with various numbers of max_iterations, 2000 was found to be optimal, striking a balance between model performance and computational efficiency. This choice was partly influenced by Turi Create's constraints on validation set usage, which limited our ability to employ traditional validation techniques to fine-tune the number of iterations. Instead, we relied on trial and error in the creation of the model, along with runtime considerations, to determine the most effective training duration. Since our goal is automated on-device deployment for laboratory and *in situ* use in space environments, the trial-and-error component in the model creation phase to achieve this goal is a practical trade off: the balance of convenience with a good modicum of configurability makes Turi Create a practical tool for rapid development, while recognising the limitations for more advanced experimentation and nuanced model optimisation.

3.3.4. Transfer learning

Our dataset is relatively modest in size even with augmentation, but in order to train machine learning models effectively requires extensive datasets and prolonged computational training times. Transfer learning e.g. Tan et al. (2018) offers a practical solution to this challenge by utilising a pre-trained model – a model initially trained on a specific task and dataset – and adapting it to a different, yet related, task or dataset. This approach can take two forms:

- Direct Application: If the new task closely aligns with the original training task, the pre-trained model may be used as-is, leveraging its existing knowledge.
- Modification and Retraining: More commonly, the latter layers of the network are modified and retrained, while the initial layer weights are kept fixed. This tailors the model more closely to the new task and data.

Such a method is advantageous as it significantly reduces the volume of training data required and shortens the training time compared to training a model from scratch. This efficiency stems from the model's ability to build upon the knowledge already acquired during its initial training phase.

For further refinement, fine-tuning comes into play. This process involves making minor adjustments to the model's weights, already pre-trained on a large dataset, to achieve a more precise adaptation to the new task. Turi Create's implementation of TinyYOLO executes this process in a user-friendly manner. Initially, the model is pre-trained on the standard ImageNet dataset (Fei-Fei et al., 2009), which comprises over 1 million images of 224 × 224 resolution, spanning 1000 classes. This foundational training equips the model with a broad understanding of various visual features. Subsequently, it undergoes fine-tuning to adapt to higher-resolution images, specifically to a resolution of 448. This pre-trained model is further refined using our specialised Tanpopo dataset to create YOLO-ET. Using a desktop AMRadeon Pro Vega 64X 16 GB GPU, training time takes 0.02, 0.17 and 0.37 s per iteration for batch sizes of 1, 32 and 64 respectively. The runtime for prediction on 79 test images takes 1.14 s, demonstrating the efficient and practical application of transfer learning and fine-tuning in customising models for specific tasks in planetary and astronomical sciences, and opening the door to tasks that could be readily implemented on-device in laboratory and spacecraft environments.

4. Evaluation and results

Fig. 10 illustrates the model's training loss over time. While it may seem tempting to continue training until the loss approaches zero, it is critical to halt the training process beforehand to avoid overfitting. Overfitting occurs when a model becomes excessively attuned to the training data, to the extent that it perfectly predicts the classes and localisations. Such hyper-specific learning compromises the model's ability to generalise and perform accurately on new, unseen data. Therefore, identifying the right moment to stop training is essential for maintaining the model's effectiveness on diverse datasets. Evaluating the model on the test set that the model has not encountered during training, serves as a proxy for real-world, unseen data and provides a more accurate measure of how well it will perform in the real world in comparison to the training set (Fig. 11). Fig. 12 shows an example of the model applied to test data. Note that the NMS/IOU thresholds have failed at removing the duplicate detection, as both boxes are above the

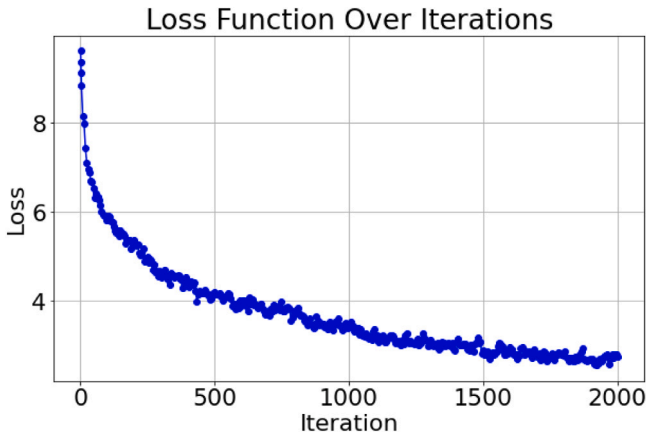


Fig. 10. YOLO-ET training loss of the network over time.

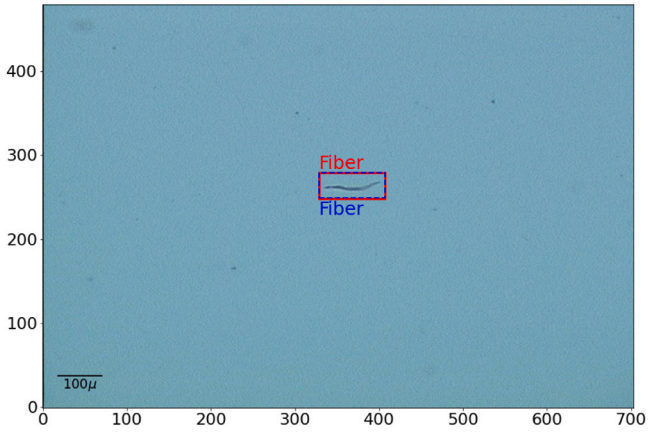


Fig. 11. Training example image depicting a fibre with the ground truth bounding box in dashed blue and the prediction from the network in red. Units are pixel coördinates.

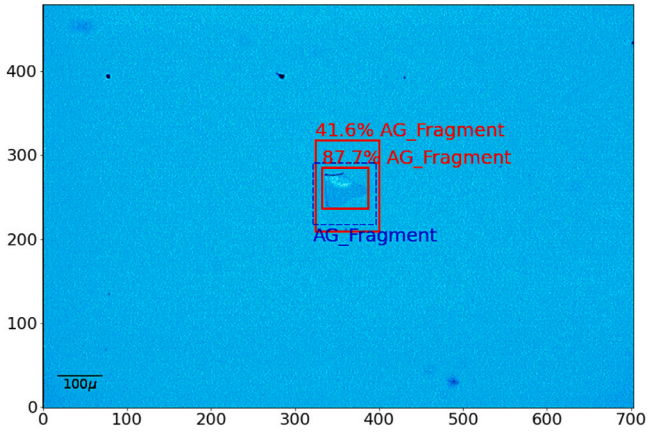


Fig. 12. Test data example depicting an aerogel fragment with ground truth bounding box in dashed blue and the network predictions in red. The confidence score of the detected object is also shown at 42% and 88%. Units are pixel coördinates.

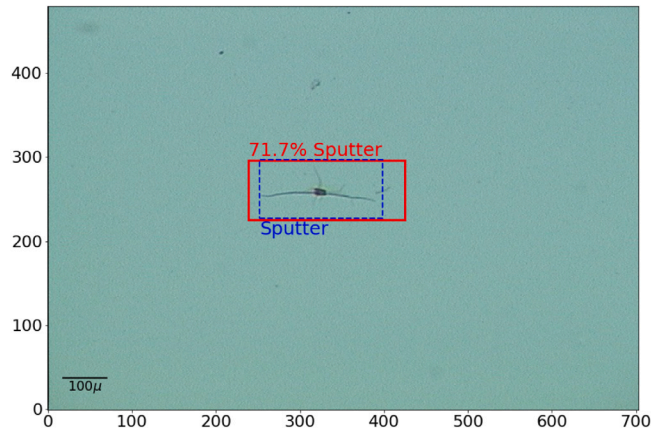


Fig. 13. Another output prediction of YOLO-ET applied to test image. The red box shows the predicted bounding box, the annotated classification and confidence probability at 72%. Units are pixel coördinates.

30% confidence level and the overlap of the boxes are less than 50%. Fig. 13 shows another test data example with a single detection. Note the variations in the hues, brightness and contrast of the images that typically make such classifications challenging.

4.1. Evaluation metrics

To accurately assess the performance of our network, we utilise a range of evaluation metrics, each chosen for its specific relevance and effectiveness in addressing the unique aspects of our problem. This diverse set of metrics ensures a comprehensive and nuanced understanding of the network's capabilities and weaknesses, allowing for a more targeted and effective optimisation.

4.1.1. Precision and recall

Precision and recall are two fundamental measures used in machine learning for evaluating the performance of classification models, especially in scenarios where the classes are imbalanced. Precision measures the accuracy of the positive predictions made by the model. It is the ratio of true positives (TP, correct detections) to the total predicted positives (both true positives and false positives FP).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{3}$$

High precision indicates a low false positive rate but does not consider false negatives (FN, missed detections).

Recall measures the ability of the model to find all the relevant cases within a dataset. It is defined as,

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

High recall indicates that the model is good at finding the positive instances but does not indicate how many negative instances were incorrectly labelled as positive.

The trade-off between precision and recall often depends on the specific requirements of the task. For example with the Tanpopo aerogel-captured surface samples above, if minimising the mis-identification of terrestrial debris particles as extraterrestrial in origin were paramount, then a high precision would help avoid us making incorrect categorisations. On the other hand high recall is essential when the goal is to ensure no potential extraterrestrial particle is missed. This might be prioritised in space environment studies where noting every possible particle is more critical than the occasional false identification. The priority, for Tanpopo surface samples, much like in ML applications for cell pathology, is not to miss anything.

Table 1
Confusion matrix based on the test data set.

		Predicted	
		Positive	Negative
Actual	Positive	71	15
	Negative	17	N/A

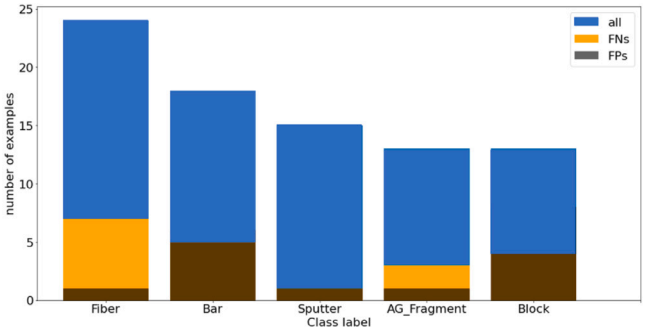


Fig. 14. Distribution of FPs and FNs over their classes.

In practice, ML modellers often look at both precision and recall together, sometimes combining them into a single measure called the F1 score, which is the harmonic mean of precision and recall:

$$\text{F1 Score} = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{5}$$

The F1 score provides a balance between precision and recall, considering both false positives and false negatives.

4.1.2. Average precision

We note that the definition of a TP also depends on the IOU threshold with respect to the ground truth box and confidence threshold. Average Precision quantifies the model's performance across different levels of precision and recall, which are typically varied by adjusting the threshold for classifying a detection as a true positive. It is calculated by plotting a Precision–Recall curve, which shows the trade-off between precision and recall for different thresholds. The area under this curve (AUC) represents the AP. Essentially, it is the average of precision values at different recall levels and is specified at a particular IOU threshold.

Mean Average Precision (mAP) is an extension of AP that is used when there are multiple classes to be detected. mAP is the mean of the APs calculated for each class individually. It is computed by first calculating the AP for each class independently. Then these AP values are averaged across all classes. This gives a single metric that summarises the performance across all classes. This is particularly important here where we need to detect multiple types of objects as it gives a holistic view of the model's performance across all these different classes, making it a more comprehensive and balanced metric.

4.2. Results

The final loss of our network is 0.8605. YOLO-ET correctly detects 90% of the test data with over 50% overlap (IoU) with the ground truth box. A summary of the results is shown in the confusion matrix (Table 1). Of the False positives, 47% are incorrectly identified as block, 35% as bar and only 1 each of fibre and sputter and aerogel fragment. The FPs are less of a concern as the confidence levels of all the detections are below 50% with the exception of the AG fragment with a confidence of 88%. On inspection this detection is a duplicate detection where 2 bounding boxes are picking up the same object with high confidence. It is also notable that blocks tend to be detected but mis-classified with 3 incorrectly classified as Fib and 2 incorrectly

Table 2
Average precision 50 (IOU > 0.5) and mean average precision from the Turi Create environment of the trained YOLO-ET applied to the train and test data from Tanpopo Project aerogel panels. The prioritised classes are grouped morphologically as aerogel fragments, bar, block, fibre and sputter, and likely represent surface residual effects of material released by ISS docking and undocking activities, venting materials, secondary impacts from primary impact ejecta, possible spacecraft component fibres, and fragments of the aerogel itself.

Class	AG fragment	Bar	Block	Fibre	Sputter	mAP
Train	0.852	0.789	0.938	0.943	0.996	0.903
Test	0.629	0.726	0.698	0.819	0.992	0.773

classified as AG fragment. Fig. 14 shows the distribution of FPs and FNs over their respective class labels. It is evident that the model performs best on detecting and classifying sputter which has the least FNs and FPs. From these values we compute the evaluation metrics, with model attaining a precision, recall and F1 score of 81%, 83% and 82% respectively. Additionally, the Turi Create environment provides tools to easily compute the average precision and mean average precision of the model. These are summarised in Table 2.

These results well surpass the targets set by the Tanpopo Astrobiology Project team of a relatively modest 70% or better recall for surface objects detected on the aerogel panels returned from the ISS, to help ensure that objects of interest and/or entry points to greater depths in the silica aerogel were not being missed. In practice human observers of inanimate objects outperform most convolutional neural networks unless fine-tuned, with overall accuracies of 90% (van Dyck et al., 2021), and YOLO-ET shows performance at these human-comparable levels. At least as notable however is the savings in human labour and computing resource with the implementation of YOLO-ET, and its on-device capabilities for ready real-time use in both laboratory and field environments.

Our F-sample dataset is proprietary to the Tanpopo Project team and currently no other published work exists for comparison, however, in prior work, before the development and implementation of the YOLO-ET system, a VGG-16 network was used but for object classification not detection, and Tanpopo aerogel surface object images at 245x magnification obtained by the CLOXS system were cropped from 704 × 480 pixels to 224 × 224 pixels and used to train a model with highest possible recall.

On the cropped images for most of the surface object categories the project goal of 70% or greater recall was met using the VGG-16 network: 93% for Blocks, 92% for Fibres, and 87% and 83% for Sputter and Bars respectively. Recall performance on Aerogel Fragments was poor however, at 29%. To improve performance a series of image pre-processing tasks were conducted: first a Zero Count Analysis (ZCA) whitening transform (Krizhevsky, 2009; Krizhevsky et al., 2012) was employed to accommodate the different colour hues of the aerogel panels, and next a thresholding sequence, where images were converted to grey scale, noise-reduced, and pixels re-filled against various thresholds, to help create more distinct object boundaries. Both ZCA and thresholding techniques brought Sputter recall to 95%, but there was a somewhat poorer recall performance with thresholding: in the 72%–79% range for Fibres, Blocks and Bars, and still under 70% for Aerogel Fragments, at 62%; ZCA for Aerogel Fragments achieved still only 33% recall.

The low recall rate using the VGG-16 model even with intensive pre-processing, for just one of the five Tanpopo surface object categories necessitated continuing need for human inspection of all panels. The YOLO-ET model and system was thus developed to achieve Tanpopo project recall performance across all categories, without the need for labour-intensive cropping and pre-processing techniques, implemented on a faster network and model that could allow real-time automatic capture on mobile devices. We compare the YOLO-ET model to the performance of VGG-16 on unprocessed data (i.e. no thresholding/ZCA) in Section 4.2 and Section 4.2. Turi Create does not natively

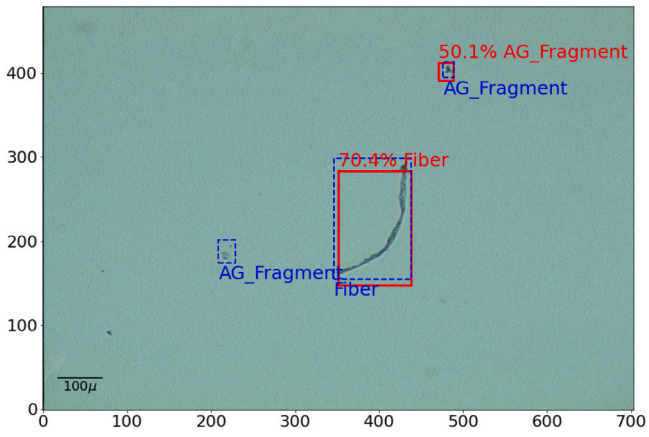


Fig. 15. Test image with multiple objects. True labels are shown as dashed blue boxes and model predicted by model CV-3 shown in red with confidence scores.

support K-fold cross-validation (CV) as a built-in feature however, but for purposes of direct comparison with the prior study, we manually created the folds for 4-fold CV and calculate the average Precision and Recall across all folds. In 4-fold CV, the dataset is divided into four equal segments. We then construct and evaluate four distinct models, each trained on a unique combination of three segments for training purposes, thereby ensuring that every segment is utilised once as a testing set. We note that this results in slightly less data available for training the model (75% versus 80%), but the implementation of CV allows us to check for robustness in the model which is important when no validation dataset is available. With no adjustment to Intersection over Union (bounding box overlap thresholds), YOLO-ET trained with this slightly smaller dataset demonstrates comparable if not better performance on Precision but slightly worse on Recall. We note however the models are not directly comparable as the VGG-16 study was an image classification task, where the objects were perfectly centred and cropped down. Our model introduces the additional complexity of object localisation, where images can contain more than 1 object and are not necessarily centred (Fig. 15). The performance accuracy of Image Classification-only models is generally higher than with Object Detection (see e.g. Lin et al. (2018)). But our augmented training dataset is more diverse and furthermore we note that the false negatives in the image classification model VGG-16 are defined as the number of objects incorrectly classified, whereas the false negatives in our model are both the number of objects that are incorrectly classified and those that are not detected. The balance between Precision and Recall is a trade-off, and setting for example a lower confidence threshold and IoU score would typically result in a higher Recall rate whilst reducing Precision. Despite the bigger challenges faced by object detection compared to image classification, by dropping the IoU threshold to 0.3 and the confidence score threshold to 0, the performance of YOLO-ET evaluated on Precision and Recall exceeds that of VGG-16 across all folds (Fig. 16).

Jaeger et al. (2021) also offers useful a useful contrast between Precision and Recall performance in Image Classification versus Object Detection, a study which explores the automatic detection of impact craters on aluminium foils utilising a Convolutional Neural Network (CNN), for Image Classification purposes. Their approach simplifies the problem to binary classification, focusing solely on distinguishing circular craters. Despite this simplification, Jaeger et al. employ synthetic data to train their model. This reliance on synthetically generated craters facilitates their model in achieving an impressive precision rate of 99.8%. However, it is crucial to note that the model's Recall rate stands at 66.7%. This discrepancy between high Precision and relatively lower Recall underscores the challenges inherent in balancing these metrics, particularly when training AI models on synthetic versus authentic datasets.

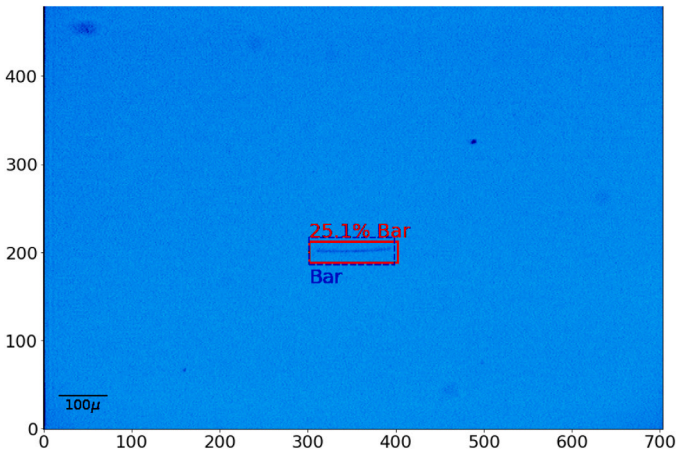


Fig. 16. Test image of an object with a low IoU threshold and confidence score of 25.1% that would not typically have made the threshold for positive detection. True labels are shown as dashed blue boxes and model prediction shown in red with confidence score.

Class	AG frag	Bar	Block	Fibre	Sputter
VGG-16	0.857	0.811	0.818	0.710	1.000
CV-0	0.929	0.938	1.000	0.944	1.000
CV-1	0.917	1.000	0.933	0.900	1.000
CV-2	1.000	1.000	1.000	1.000	1.000
CV-3	0.824	0.857	0.963	0.947	1.000
CV-avg	0.918	0.949	0.974	0.948	1.000
CV-std	0.063	0.059	0.028	0.035	0.000

Comparison of Precision using a VGG-16 image classification model and our 4-fold cross-validation runs. CV-avg and CV-std are the average and standard deviations of the cross-validation runs.

Class	AG frag	Bar	Block	Fibre	Sputter
VGG-16	0.286	0.833	0.931	0.917	0.872
CV-0	0.812	1.000	1.000	0.850	1.000
CV-1	0.957	1.000	0.933	0.818	1.000
CV-2	0.875	0.500	1.000	0.833	1.000
CV-3	0.933	1.000	0.963	1.000	0.950
CV-avg	0.894	0.875	0.974	0.875	0.988
CV-std	0.056	0.217	0.028	0.072	0.02

Comparison of Recall using a VGG-16 image classification model and our 4-fold cross-validation runs. CV-avg and CV-std are the average and standard deviations of the cross-validation runs.

5. Deploying the model

Using the methods developed for this project, we have demonstrated that applying AI Machine Learning to 2D aerogel images with YOLO-ET, greatly speeds up and simplifies the identification, localisation, and classification of Tanpopo aerogel-captured surface particles. These advances have been demonstrated to speed the object detection training process, improve accuracy, and consume fewer computing resources, all while taking advantage of the in-built optics and compact form factor of a mobile device. Our model requires no pre-processing of the data. However to truly realise the potential of this model for space missions, it is imperative to address the dependency on traditional microscopy. While the CLOXS system represents significant capabilities in that it can automatically relocate the coordinates of objects of interest and centre the stages accordingly, the YOLO-ET system can now greatly augment these capabilities by identifying, localising and classifying objects on the first pass in real time. The YOLO-ET core Machine Learning models can be readily translated to a mobile device, in this project as an App, allowing the iPhone's camera, enhanced by a LabCam® adaptor, to act as the object detector for untrained, real-world images, with the self-contained iPhone and App able to bound and classify new images based on the core Machine Learning models developed. We propose the integration of this model into a mobile application for both laboratory and space environments, harnessing the capabilities of widely accessible technology like smartphones. This integration marks a significant step toward edge computing, where data processing is performed at or near the source of data generation.

Table 3
iPhone Pro Max 12 specs.

Camera	iPhone Pro Max 12	
Ultra Wide	12 MP	f/2.4
Wide	12 MP	f/1.6
Telephoto	12 MP	f/2.2

5.1. Adapting and integrating LabCam® to CLOXS

Adapting from the most recent developments in field research (e.g. Ateaque, 2022; Meng et al., 2023), we selected the iPhone Pro Max 12 (Table 3) and the LabCam®,[4] a user-friendly combination, for our initial deployment (Fig. 17). The iPhone's advanced camera system, processing power and sophisticated autofocus technology make it an ideal choice for capturing high-quality images of microparticles. This autofocus feature is critical for our application, as it ensures that images are sharp and highly detailed, facilitating accurate Object Detection without the need for manual focus adjustments. Moreover, the convenience, portability and widespread availability of iPhones offer practical advantages for replicating our methodology across diverse settings, particularly fieldwork and applications in resource-constrained environments. While alternatives such as small PCs and specialised AI cameras exist, the iPhone's integrated ecosystem and the availability of sophisticated development tools in Turi Create make it an attractive choice for implementing advanced AI-driven object detection tasks directly on the device.

Meanwhile, the LabCam® attachment enhances the iPhone's capability to function as a makeshift microscope. It is a portable microscope that can be easily taken to the sample, rather than the other way around. This makes it ideal for real-world and real-time *in situ* detection of say micro-particle contaminants on the lunar surface. With the ability to capture images with up to 100x magnification in integration with an iPhone alone, the LabCam® provides an easy to use, ready system for microparticle detection. These capabilities are now being translated to real-time laboratory examination of aerogel panels, to identify, localise

[4] https://www.ilabcam.com/.

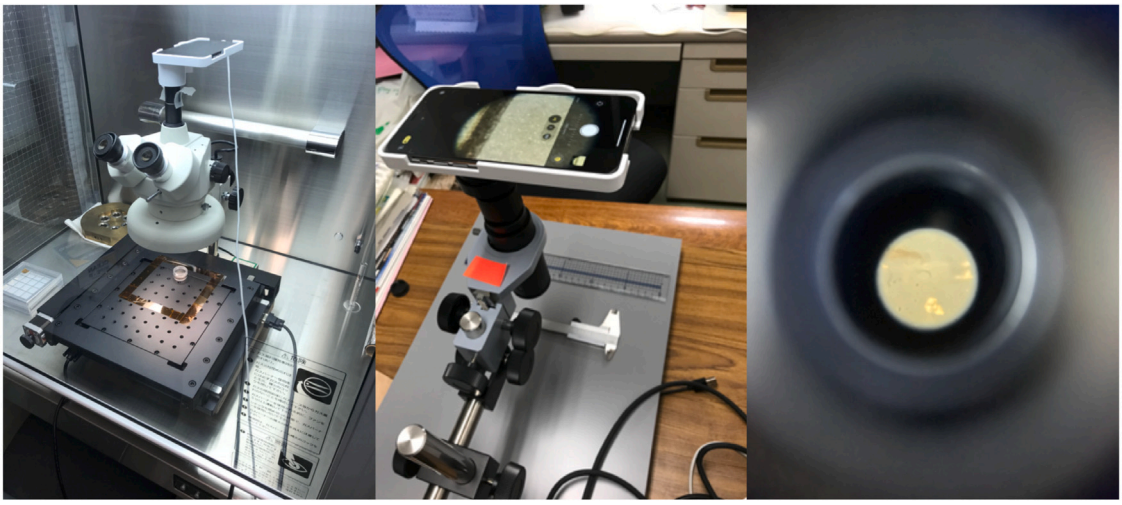


Fig. 17. Adapting the LabCam® and iPhone Pro Max 12 into the ISAS CLOXS system. *Left*: Clean bench set up to start single-operator imaging and analysis. Note the LabCam® mount, and 50 mm traverse motorised xy stage, 100 mm traverse stage and joystick controls. *Centre:* Hozon and LabCam® mounted with iPhone Pro Max 12 for calibration. *Right:* Looking remotely 'down the hole' of the iPhone Pro Max 12 at a magnified block particle fragment in the aerogel.

and classify 3D tracks and hypervelocity impact particles candidates, using mobile on-device Machine Learning models.

The combination of the iPhone Pro Max 12, LabCam® mount and integrated magnification, coupled to the CLOXS system at ISAS allows real-time processing to a GPU-equipped desktop running our YOLO-ET algorithms and core Machine Learning model – a full surrogate of what could be packed and space-hardened into *in situ* and in-spaceflight missions. With additional optics added to the system with LabCam®, the CoreML models developed and experiments described in this study demonstrate both real-world laboratory identification and classification of extraterrestrial microparticles and autonomous edge-computing capabilities for future spacecraft missions to detect, localise and classify them.

Thus a unique advantage of the methods employed in this study is that the YOLO-ET imaging and Machine Learning processing and classification is self-contained, with 'on-board' GPU processing whose form factor and computing power can be readily incorporated into small spacecraft. With these applications in mind, further experiments were conducted on identifying and classifying granulated microscopic spacecraft materials distributed atop lunar regolith simulants, as a surrogate for *in situ* detection of anthropogenic contaminants on the lunar surface. Finally, based on the newly returned 'ground truth' of Scanning Electron Microprobe (SEM) images obtained from the asteroid Ryugu sample returned by Hayabusa2, as a further demonstration the model was trained and tested to establish potential correlations with SEM images from the suite of micrometeorites obtained from the TransAntarctic Mountains.

6. Discussion

6.1. Experiments with spacecraft microparticles on lunar simulants

An anticipated future use of the CLOXS imaging system as adapted and coupled with YOLO-ET in this work is examination of Tanpopo-like aerogel panels deployed and retrieved from the lunar surface. We are developing a mission concept for aerogel panel deployment to the Moon as early as the Artemis lunar landing missions, with important opportunities to advance not only the core Tanpopo astrobiology objectives, but to collect more information and contribute to studies on microparticle anthropogenic contaminants, lunar regolith disturbances by human activities on the Moon, the flux of interplanetary, β-meteoroids, possible interstellar dust, and secondary impact ejecta (see e.g. Grün et al., 2011; Pokornỳ et al., 2019; Szalay et al., 2020; Costello et al., 2021).

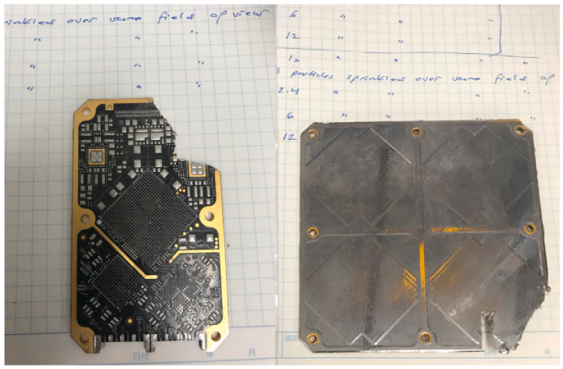


Fig. 18. *Left:* SDR board, Polyimide Arlon 85 with SAC 305 solder and copper layers, with impregnation material. *Right:* Nightingale Antenna board, Rogers material with copper and silver coating and STAMET radome.

Ultimately a 'Mini-CLOXS' could support both post-retrieval examination of aerogel panels returned from the lunar surface and other missions to Earth laboratories, as well as *in situ* examination on lunar and other planetary surfaces. A timely factor in bringing these capabilities to the Moon is to help establish a baseline for forward contamination caused by robotic and human activities there. A key element of potential forward contamination are particles of spacecraft, experimental packages, communications equipment etc. that may be deposited and distributed around the Moon by (i) routine operations, including outgassing of propellants and spacesuits, mechanical interfaces, vehicle track and wheel movements, etc.; (ii) natural material degradations from micrometeoroid bombardment, day/night temperature cycles, interaction with the solar wind etc. and (iii) larger scale de-orbited and hard-landed material.

With these in mind we aim to take the YOLO-ET algorithm developed and trained on surface particles captured on the Tanpopo aerogels, and used the same Turi Create, iPhone, and LabCam® system described above to test its capabilities for imaging and identifying spacecraft remnants mixed into lunar regolith simulants.

Samples of both JSC-1 (Appendix A) and Manna Electric lunar simulant (Appendix B) of 0.05 g each were prepared and evenly deposited into 2 cm diameter plastic vials. Test model portions of the CesiumAstro Nightingale satellite (Fig. 18) whose compositions are detailed in (Appendix C) were particulated with a band saw, producing

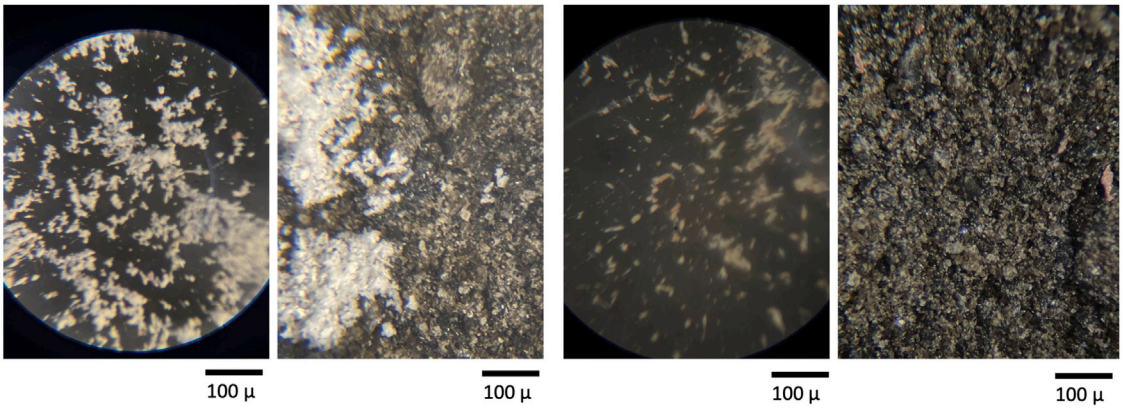


Fig. 19. Samples of JSC-1 (Appendix A) lunar simulant of 0.05 g each were prepared and evenly deposited into 2 cm diameter plastic vials. Test model portions of the CesiumAstro Nightingale satellite (Fig. 18) whose compositions are detailed in (Appendix C) were particulated with a band saw, producing spiralled fragments that were then sieved to 80 µm to approximate average JSC-1 grain sizes, to set the challenge to distinguishing similar sized fragments amongst similar sized grains. From left to right: CesiumAstro 'A' particles from the antenna board; A particles on JSC-1; 'B' particles from the Software Defined Radio board; B particles on JSC-1.

spiralled fragments that were then sieved to 80 µm to approximate average JSC-1 grain sizes, to set the challenge to distinguishing similar sized fragments amongst similar sized grains (Fig. 19). YOLO-ET first trained on a limited data set of images of 80 µm diameter glass beads, and then spacecraft grade Al, Ti, and CFRP in various lighting conditions and magnifications, set atop the lunar simulant deposits. Utilising the specially combined optics of a 4x Hozon lab microscope, a 10xLabCam®, and an iPhone Pro Max 12 with 1x-12x magnification, the overall range is 40x to 480x. Optimal experimental resolution for early training and ground truth experiments in YOLO-ET are 100x; with next-generation LabCam® and iPhone Pro Max 15, 225x magnification can be anticipated using purely analogue optics, highly suitable for spacecraft deployment.

At the optimum magnification of 100x, lighting conditions and depth, for which an *in situ* Mini-CLOXS would be designed, early YOLO-ET training and ground truth experiments have been demonstrated to show ready identification of Nightingale antenna particles and Software Defined Radio board microcircuitry particles (Fig. 20), and once optimised the resulting model can be directly exported to CoreML format, for streamlining integration into the on-device application.

6.2. Asteroid Ryugu sample experiments

As described above, the YOLO-ET convolutional network models developed and trained on aerogel-captured anthropogenic contaminant samples of the Tanpopo missions onboard the International Space Station using a mobile device camera and new processing techniques greatly speeds and automates their identification and classification. This technique moves beyond datasets that have already been laboriously centred, focused, scaled, photographed, and classed by human researchers, opening the door to automated transits by microscope across the Tanpopo aerogel panels at approximately 500 × 500 pixel increments, at different focal lengths, with images then fed directly into YOLO-ET — which then uses its object detection/localisation capabilities to automatically draw bounding boxes around the object or objects of interest in each image, and to automatically run a confidence prediction of which class of object it might be, displayed both on the image and as a searchable table.

Several of the authors have participated in analyses of samples of the asteroid Ryugu returned by the Hayabusa2 mission (Yada et al., 2022), and the potential for important time-and-manpower gains with YOLO-ET seems clear. Hundreds of nano-CT scans were conducted to create segmented images whose cross-sections helped reveal micro-scale voids of particular interest. Human-eye examination of void evidence in cross-section allowed re-integration of the images to reveal the voids in full dimension. This manual process of identifying and classifying evidence of voids in cross-section is ripe for ML identification, classification, and reintegration using the methods developed in this project. Data from the Ryugu sample A0180 can now form a robust training data set for applying YOLO-ET to other more porous and aggregate samples from Ryugu. With time, the 3D optical images, nano-CT data, and external and internal SEM images from A0180 and other Ryugu samples can be archived to create further training data for searching the diversity amongst different groups of Ryugu samples and for practically comparing characteristic Ryugu micro-structures with for example structures from unmelted micrometeorites, which to the human eye seem to show strikingly similar characteristics and features (Fig. 21).

Accordingly as a demonstration of these abilities YOLO-ET was also trained on SEM images of the Ryugu asteroid sample A0180, to establish 'ground truth' for characterising features of unmelted micrometeorites and unmelted portions of partially melted micrometeorites, using images of the TransAntarctic Mountains micrometeorite suite (Fig. 22). Unlike larger falls, micrometeorites are subject to less heating and alteration as they pass through Earth's atmosphere, and preserve important elements of their formation history and composition. By comparing them for the first time to the ancient base line features of the indigenous asteroidal and cometary samples returned by spacecraft such as Stardust, Hayabusa, Hayabusa2 and OSIRIS-REx, it is possible to re-calibrate current assumptions about e.g. the proportional representation of CI chondrites amongst terrestrial meteorite collections. Many of the characteristic features of each are notably subtle, and typically require considerable training and experience for research practitioners to deduce. In this work YOLO-ET demonstrates capabilities for learning and establishing correlations amongst unseen micrometeoritic data sets.

Distinguishable structural elements amongst images of micrometeorites with unmelted areas include: roughness and irregularities versus the smoother, glassy appearance of melted portions; differences in brightness reflecting compositional differences in backscattered electron images; micro-chondrules, mineral grains and inclusions indicative of parent body origins, otherwise obliterated in melted regions; and distinctive boundaries between melted and unmelted areas, with partial rims characteristically forming around unmelted areas. Based on these types of parameters we trialled, as a proof of concept, making selections of areas of interest on the Ryugu A0180 images, using them to train and predict on a selection of the suite of TransAntarctic Mountain unmelted micrometeorites. Our YOLO-ET model, trained on a limited dataset of 212 examples (without optimisation), detects features in unseen test data. While this showcases the model's potential, the current mean average precision (mAP) is only 10.1%, indicating significant room for improvement.

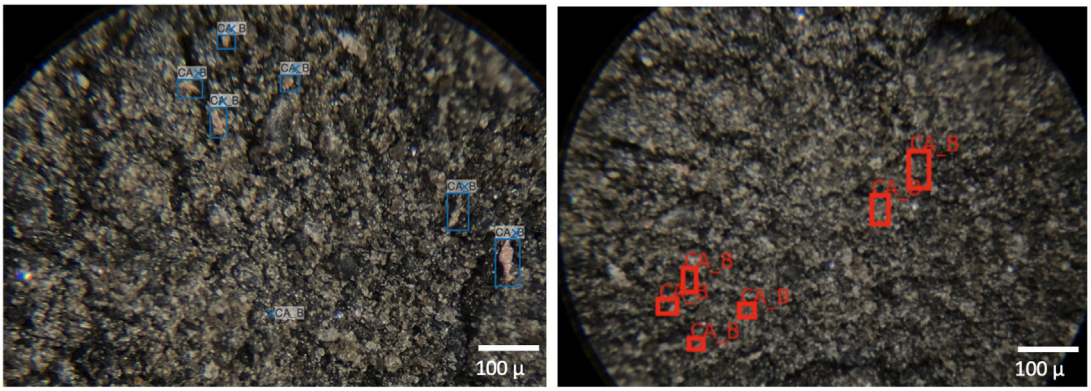


Fig. 20. Early applications of our machine learning model YOLO-ET demonstrated that they could be trained on known 'B' particles of CesiumAstro Nightingale Software Defined Radio Board to automatically identify unseen 'B' particles atop JSC-1 Lunar simulant.

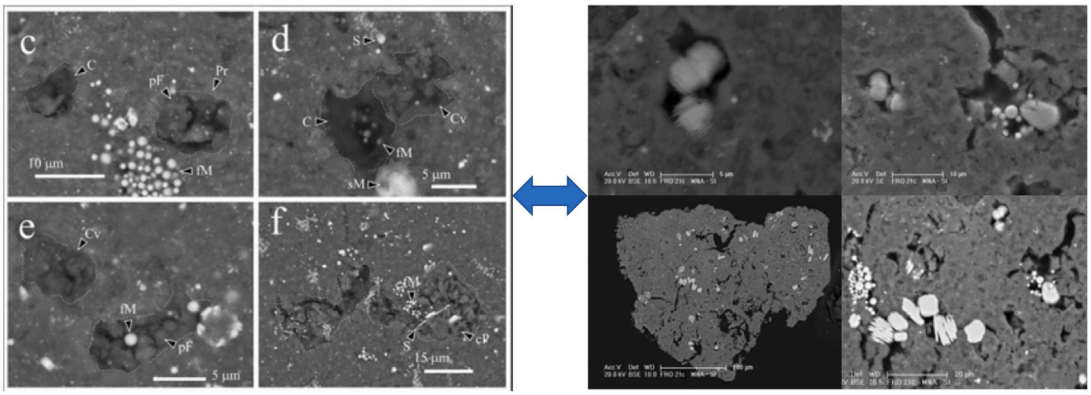


Fig. 21. Exploring potential correlations in major features of indigenous asteroid samples Ryugu A0180, left, with those of unmelted micrometeorite samples from the TransAntarctic Mountain micrometeorites suite.
Image Credits: Genge et al. (2023) and Van Ginneken et al. (2012).

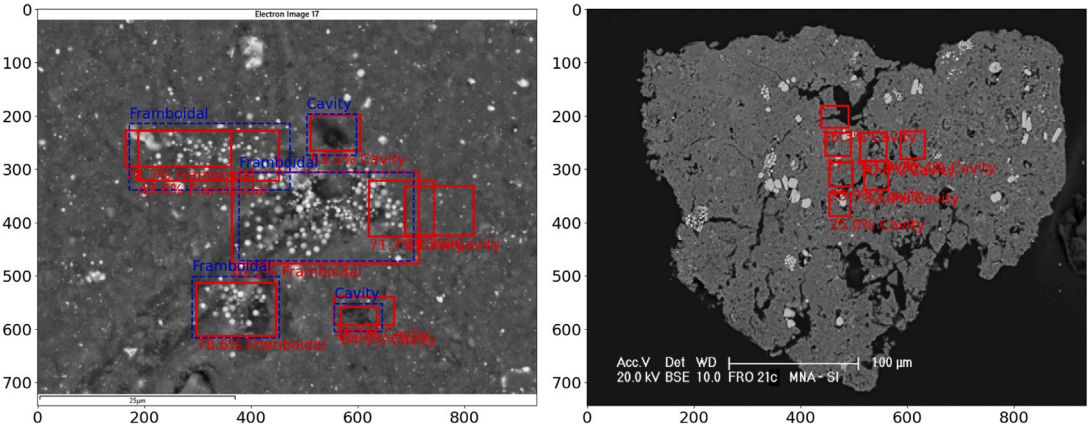


Fig. 22. Early applications of our machine learning model YOLO-ET show that training on a Ryugu sample automatically identifies similar features on unmelted TransAntarctic micrometeorites.
Image Credits: Genge et al. (2023) and Van Ginneken et al. (2012).

6.3. Future work

Future space missions are expected to yield a much larger and more heterogeneous quantity of microscopic materials than treated in these experiments. These include new asteroid interceptions, planetary expeditions, and most especially, robotic and human sample return missions to the Moon, facilitated by unprecedented cargo return capacities. With the imminent rise of robotic and human activities on the Moon, the importance of *in situ* microscopic examination capabilities to distinguish these microparticles becomes increasingly important for (i) identifying and quantifying the flux of anthropogenic contaminants and lunar surface disturbances and (ii) for controlled experiments to better understand the flux of exogenous (IDPs, β-meteoroids, possible interstellar dust) and indigenous (secondary impact ejecta) microparticles, with important implications for characterising the quantity of volatiles

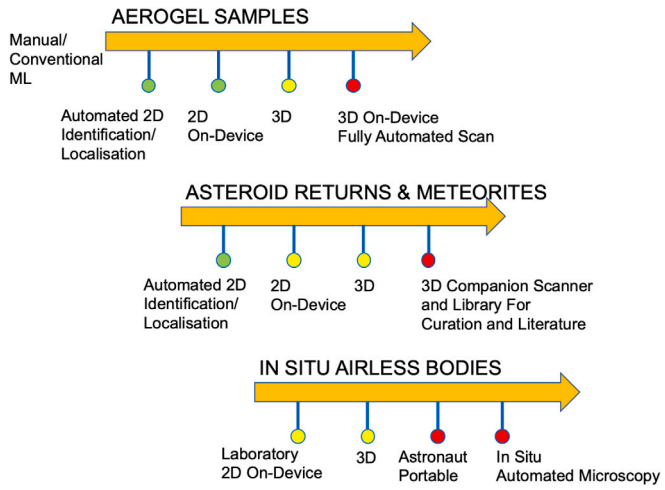


Fig. 23. Current and future work: Progressing YOLO-ET on mobile devices from Earth receiving laboratories to *in situ* analyses on airless bodies.

Fig. 24. Ispace lunar lander rendering. Startups like ispace are contracting for payloads that could accommodate a mini-CLOXS.
Credit: Ispace.

held in micro-structures and their resource potential (see e.g. He et al., 2023) for micro-structure potential.

In addition, researchers may also wish to investigate the possible panspermia delivery of pre-biotic and biotic materials within our Solar System (Napier, 2004). In the longer term, there is also the remoter possibility of discovering trace artefacts of other technological civilisations from across the Galaxy, pushed by radiation pressure and stellar winds (Arkhipov, 1996; Crawford, 2006). If such artefacts exist, they are most likely to be revealed at the microscopic level, upon the mass examination of many trillions of particles uncovered at the scale of industrialised lunar and asteroid mining.

By moving from more conventional Machine Learning approaches to the YOLO-ET model specifically developed here for the detection, localisation and classification of microparticles from and in the space environment, this work has opened the door to rely more on compressed Machine Learning models, existing high-performance GPU code, and commercially available software libraries; we are developing, training and testing algorithms on systems and hardware that would readily fit into a cubesat class spacecraft, lunar rover, or planetary sampling missions (Fig. 23).

For future work inspecting silica aerogels for captured microparticles, we are progressing to fully automated scans in real time in the receiving laboratory clean room environment, and from 2D surface scans to 3D inspection of hypervelocity impact candidates. Similarly, the work on 2D slice SEM images of the Ryugu samples correlated to images of micrometeorites and other asteroid return samples is planned to be extended to 3D images. Moving from Ryugu to Bennu samples may be a natural next step (Goldwin, 2023).

This on-device, mini-CLOXS capability can then be made fully portable and ruggedised to perform scans *in situ* on airless bodies, obviating the need, risks, time and costs for sample return to Earth and better complementing Earth-based studies (Fig. 24).

These capabilities can also be extended from direct observation of micrometeorite particles on Earth to direct observation of microparticles on other planetary surfaces, without aerogel panel capture. Similarly moving from 2D to 3D scans, these devices can serve as companions for Earth-based space microparticle curators and assist in a more complete and integrated cataloguing from both newly collected samples and the extended prior literature. Ultimately, we aim to help create an astronaut-portable, 3D real-time scanning capability for microparticle detection, localisation and classification to assist automated microscopy on the Moon and other airless bodies.

Many of the particles of interest that might be distinguished on the surface of the Moon and other airless bodies, whether from cislunar space operations, asteroid and comet fragment intersections, or possible interstellar dust, may arrive at hypervelocity impact speeds. Thus in future work we plan to have materials similar to those used in the YOLO-ET experiments distinguishing microparticles on lunar regolith simulants, fired at hyper-velocities into monocrystalline target materials, to better scale and model characteristic microcrater morphologies as well as their patterns of vaporisation and spallation. It is envisaged that these tools and techniques can practically be packaged for use on board future mineral assay and mining devices deployed on the Moon and asteroids, to help identify micrometeoroids and other microscopic particles of interest in the surface regoliths.

Finally when contemplating spacecraft-borne YOLO-ET mobile device systems extended to both fly-bys and orbital missions it could be useful to adapt the image detection, localisation and classification processes from the microscopic-particle to the macro scale. Real-time detection of features of interest on the Moon and asteroids to assist selection and navigation of surface sample operations, or to monitor and catalogue spacecraft debris and artefacts, could be amongst the next useful implementations. Thus the scope of future work also includes YOLO-ET analyses of Lunar Reconnaissance Orbiter images for detection and classification of spacecraft hardware on the Moon (Haase et al., 2012; Lesnikowski et al., 2020).

7. Conclusions

In this work, we have adapted Turi Create's Object Detection capabilities to identify and classify features in images of extraterrestrial microparticle impacts, microstructures, and anthropogenic debris.

1. Training on images from Tanpopo aerogel panels returned from Japan's Kibo module of the International Space Station, YOLO-ET demonstrates a 90% detection rate for all types of anthropogenic contaminants on aerogel surfaces and shows promising early results for detection of both microparticle contaminants on the Moon and for evaluating asteroid return samples.
2. YOLO-ET significantly improves on earlier ML processes in time savings, performance, and more efficient use of computing resources, and thereby can bring value to Tanpopo CLOXS machine processing by requiring fewer resources to more quickly and accurately identify samples of interest within the aerogel panels for extraction, while reducing contamination risk by more accurately and precisely selecting only those samples of interest, and allowing for the timeliest distribution of extracted samples to analysis groups around the world.
3. Preliminary tests of YOLO-ET's application to identifying spacecraft-derived microparticles in lunar regolith simulant samples and SEM images of Ryugu asteroid samples indicate strong model performance and transfer learning capabilities for future extraterrestrial applications.

Our research provides insights into the practical challenges and solutions associated with applying AI in planetary science, especially in environments where data characteristics differ markedly from those used to train general-purpose Image Classification models. We detail our methodology for dataset preparation, model training, and validation, offering a blueprint for other scientists looking to apply AI in specialised research areas. Our hope is that this work will inspire further research that leverages the power of AI to tackle domain-specific challenges.

CRediT authorship contribution statement

L.J. Pinault: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **H. Yano:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **K. Okudaira:** Resources, Methodology, Formal analysis, Data curation. **I.A. Crawford:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Lewis Pinault reports financial support was provided by Airbus Ventures. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that support the findings of this study are available on request from the corresponding author, l.pinault@ucl.ac.uk. The data are not publicly available as Tanpopo Astrobiology Project and Ryugu sample analyses are currently ongoing. The source code for TuriCreate is open source and developers can access, modify and contribute to its development on GitHub at http://github.com/apple/turicreate. Additionally our trained model can be found at https://github.com/LewisJPinault/YOLO-ET.

Acknowledgements

We thank the anonymous Reviewers for contributing to the improvement of our paper.

Airbus Ventures and Airbus Japan supported LJP and coordinated collaborations with HY at JAXA/ISAS.

Key thanks to Professors Yamagishi, H. Mita, and Dr. J. Imani for all their support to HY's work on the Tanpopo aerogel capture experiments and CLOXS development and operation.

Special thanks to Honaka Kiryu for her work supporting KO in the first applications of CLOXS images using VGG-16 models to tackle the Machine Learning approaches to Tanpopo 'F' sample classification: the images and results obtained were key to evaluating YOLO-ET performance.

CesiumAstro kindly provided test model materials from their Nightingale phased array radar unit.

Manna Electric of Luxembourg kindly provided samples of their lunar regolith simulant, used in their lunar *in situ* resource utilisation process preparations, including for their prospective manufacturing of solar panels on the Moon.

HY and LJP have been working with Matt Genge of Imperial College London, Natasha Almeida of the Natural History Museum in London, Matthias van Ginneken and Penny Wozniakiewicz at the University of Kent Canterbury, and Luigi Folco of Università di Pisa on the asteroid Ryugu A0180 samples returned by Hayabusa2. Their careful sample preparations imaging and analyses have greatly stimulated our thinking on present and future directions for YOLO-ET, and we thank them for their permission to use the images in this paper. The A0180 sample has been loaned to the proposal PI (HY) of this research consortium since 2022, by the JAXA/ISAS Astromaterial Science Research Group through the first international Ryugu sample AO. The authors are most grateful to the Hayabusa2 project team for having successfully returned the indigenous samples from Ryugu to the Earth in 2020.

We thank iDu Optics® (New York, NY, USA) for providing LabCam® adaptors aiding the capture of the images through the iPhones used in this project.

Appendix A. JSC-1 Lunar simulant components

JSC-1 Specifications https://ares.jsc.nasa.gov/projects/simulants/jsc-1-1a.html

Appendix B. Components analyses maana electric Mare and highlands Lunar simulants

Lunar Mare Simulant Specifications https://maanaelectric.com/space_solutions/simulant_mare/Lunar Highland Simulant Specifications https://maanaelectric.com/space_solutions/simulant_highland/

Appendix C. CesiumAstro spacecraft materials ground and sieved to 80 μm

- SDR-1001 – the credit card-sized board with black solder mask
  - CesisumAstro's Gen1 Software-Defined Radio (SDR) product for LEO and airborne applications with 100 MHz IBW, operating from 300 MHz to 6 GHz
  - The board construction:
    - 16-layer PCB constructed of Arlon 85N (polyimide) and copper layers
    - Prepreg layers are 85N with 106, 1080, and 2313 glass weaves
    - Plating is ENIG (gold over electroless nickel)
    - Solder mask is Taiyo PSR-4000 MP Black
  - Solder applied to pads is SAC305
- SAPA-1 – the board with four patches covered in reflective film
  - S-band Antenna Patch Array for LEO applications operating in the 2.45 GHz region
  - The board construction:
    - 6-layer PCB constructed of Rogers 4350B and copper layers
    - Prepreg layers are RO4450F
    - Plating is ENIG (gold over electroless nickel; latest revision of the antenna is in ImAg immersion silver finish)
    - Solder mask is Taiyo PSR-4000 MP
  - The thin film radome on the front face of the antenna is Stamet-sputtered Kapton (Dunmore MO20295)

References

Arkhipov, A.V., 1996. Extraterrestrial artefacts. Observatory 116, 175–176.

Ateaque, S., 2022. Neurotrophin-3 Signalling in Neurons Derived from Human Embryonic Stem Cells (Ph.D. thesis). Cardiff University.

Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., Miao, Y., 2021. Review of image classification algorithms based on convolutional neural networks. Remote Sens. 13 (22), 4712.

Costello, E.S., Ghent, R.R., Lucey, P.G., 2021. Secondary impact burial and excavation gardening on the moon and the depth to ice in permanent shadow. J. Geophys. Res.: Planets 126 (9), e2021JE006933.
Crawford, I.A., 2006. The astrobiological case for renewed robotic and human exploration of the moon. Int. J. Astrobiol. 5, 191–197.
Dartois, E., Engrand, C., Brunetto, R., Duprat, J., Pino, T., Quirico, E., Remusat, L., Bardin, N., Briani, G., Mostefaoui, S., et al., 2013. UltraCarbonaceous antarctic micrometeorites, probing the solar system beyond the nitrogen snow-line. Icarus 224 (1), 243–252.
Fei-Fei, L., Deng, J., Li, K., 2009. ImageNet: Constructing a large-scale image database. J. Vis. 9 (8), 1037–1037.
Flynn, G.J., 1994. Interplanetary dust particles collected from the stratosphere: physical, chemical, and mineralogical properties and implications for their sources. Planet. Space Sci. 42 (12), 1151–1161.
Genge, M.J., Almeida, N., Van Ginneken, M., Pinault, L.J., Wozniakiewicz, P., Yano, H., 2023. Ice and liquid water in asteroid Ryugu – Constraints from sample A0180. In: The 14th Symposium on Polar Science, National Institute of Polar Research, November 14-17, 2023 Tachikawa, Tokyo, Japan. pp. 1–2.
Genge, M.J., Van Ginneken, M., Suttle, M.D., 2020. Micrometeorites: Insights into the flux, sources and atmospheric entry of extraterrestrial dust at earth. Planet. Space Sci. 187, 104900.
Girshick, R., 2015. Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1440–1448.
Goldwin, T., 2023. The importance of asteroid sample return. Nat. Geosci. 16, 833.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
Grishin, K., Mei, S., Ilic, S., 2023. YOLO-CL: Galaxy cluster detection in the SDSS with deep machine learning. doi:10.48550/arXiv.2301.09657, arXiv preprint arXiv:2301.09657.
Grün, E., Horanyi, M., Sternovsky, Z., 2011. The lunar dust environment. Planet. Space Sci. 59 (14), 1672–1680.
Haase, I., Oberst, J., Scholten, F., Wählisch, M., Glser, P., Karachevtseva, I., Robinson, M., 2012. Mapping the apollo 17 landing site area based on lunar reconnaissance orbiter camera images and apollo surface photography. J. Geophys. Res.: Planets 117.
He, H., Ji, J., Zhang, Y., 2023. A solar wind-derived water reservoir on the moon hosted by impact glass beads. Nat. Geosci. 16, 294–300.
Huertas-Company, M., Lanusse, F., 2022. The DAWES review 10: The impact of deep learning for the analysis of galaxy surveys. doi:10.48550/arXiv.2210.01813, arXiv preprint arXiv:2210.01813.
Jaeger, L., Butterworth, A.L., Gainsforth, Z., Lettieri, R., Zevin, D., Ardizzone, A., Capraro, M., Burchell, M., Wozniakiewicz, P., Ogliore, R.C., et al., 2021. Automatic detection of impact craters on al foils from the stardust interstellar dust collector using convolutional neural networks. Meteorit. Planet. Sci. 56 (10), 1890–1904.
Jeffrey, N., Lanusse, F., Lahav, O., Starck, J.-L., 2020. Deep learning dark matter map reconstructions from DES SV weak lensing data. Mon. Not. R. Astron. Soc. 492 (4), 5023–5029.
Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B., 2022. A review of YOLO algorithm developments. Procedia Comput. Sci. 199, 1066–1073.
Kawaguchi, Y., Yokobori, S.-I., Hashimoto, H., Yano, H., Tabata, M., Kawai, H., Yamagishi, A., 2016. Investigation of the interplanetary transfer of microbes in the Tanpopo mission at the exposed facility of the international space station. Astrobiology 16 (5), 363–376.
Krizhevsky, A., 2009. Learning Multiple Layers of Features from Tiny Images (Master's thesis). University of Toronto, Master's thesis. University of Toronto.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25.
Kurat, G., Koeberl, C., Presper, T., Brandstätter, F., Maurette, M., 1994. Petrology and geochemistry of antarctic micrometeorites. Geochim. Cosmochim. Acta 58, 3879–3904,.
Lesnikowski, A., Bickel, V.T., Angerhausen, D., 2020. Unsupervised distribution learning for lunar surface anomaly detection. doi:10.48550/arXiv.2001.04634, arXiv preprint arXiv:2001.04634.
Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2018. Focal loss for dense object detection. doi:10.48550/arXiv.1708.02002, arXiv:1708.02002v2.
Meng, T., Zheng, J., Chen, M., Zhao, Y., Sudarjat, H., MR, A.A., Kulkarni, V., Oh, Y., Xia, S., Ding, Z., et al., 2023. Six-month effective treatment of corneal graft rejection. Sci. Adv. 9 (12), eadf4608.
Mitchell, T.M., 1997. Machine Learning. McGraw Hill.
Nakamura, T., Noguchi, T., Tanaka, M., Zolensky, M.E., Kimura, M., Tsuchiyama, A., Nakato, A., Ogami, T., Ishida, H., Uesugi, M., Yada, T., Shirai, K., Fujimura, A., Okazaki, R., Sandford, S.A., Ishibashi, Y., Abe, M., Okada, T., Ueno, M., Mukai, T., Kawaguchi, J., 2011. Itokawa dust particles: a direct link between S-type asteroids and ordinary chondrites. Science 333 (6046), 1113–1116.
Napier, W., 2004. A mechanism for interstellar panspermia. Mon. Not. R. Astron. Soc. 348, 46–51.
Neubeck, A., Van Gool, L., 2006. Efficient non-maximum suppression. In: 18th International Conference on Pattern Recognition. ICPR'06, Vol. 3, IEEE, pp. 850–855.
Pokornỳ, P., Janches, D., Sarantos, M., Szalay, J.R., Horányi, M., Nesvornỳ, D., Kuchner, M.J., 2019. Meteoroids at the moon: orbital properties, surface vaporization, and impact ejecta production. J. Geophys. Res.: Planets 124 (3), 752–778.
Prasad, M., Rudraswami, N., de Araujo, A., Khedekar, V., 2018. Characterisation, sources and flux of unmelted micrometeorites on earth during the last 50,000 years.. Nat. Sci. Rep. 8887, 1–8.
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 779–788.
Redmon, J., Farhadi, A., 2017. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7263–7271.
Rojas, J., Duprat, J., Engrand, C., Dartois, E., Delauche, L., Godard, M., Gounelle, M., Carrillo-Sánchez, J., Pokornỳ, P., Plane, J., 2021. The micrometeorite flux at Dome C (Antarctica), monitoring the accretion of extraterrestrial dust on earth. Earth Planet. Sci. Lett. 560, 116794.
Sasaki, S., Imani, J.-Y., Yano, H., 2019. Design, fabrication and evaluation of an aerogel processor CLOXS for the astrobiology mission Tanpopo. Biol. Sci. Space 33, 7–11.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. doi:10.48550/arXiv.1409.1556, arXiv preprint arXiv:1409.1556.
Szalay, J., Pokornỳ, P., Horányi, M., 2020. Hyperbolic meteoroids impacting the moon. Astrophys. J. Lett. 890 (1), L11.
Tabata, M., Kawaguchi, Y., Yokobori, S.-I., Kawai, H., Takahashi, J.-i., Yano, H., Yamagishi, A., 2011. Tanpopo cosmic dust collector: silica aerogel production and bacterial DNA contamination analysis. Biol. Sci. Space 25 (1), 7–12.
Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A survey on deep transfer learning. In: Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27. Springer, pp. 270–279.
Taylor, A., Baggaley, W., Steel, D., 1996. Discovery of interstellar dust entering the earth's atmosphere. 380. 323-325. 10.1038/380323a0. Nature 380, 323–325.
van Dyck, L.E., Roland, K., Jochen, D.S., Roland, G.W., 2021. Comparing object recognition in humans and deep convolutional neural networks—An eye tracking study. Front. Neurosci. 15.
Van Ginneken, M., Folco, L., Cordier, C., Rochette, P., 2012. Chondritic micrometeorites from the transantarctic mountains. Meteorit. Planet. Sci. 47 (2), 228–247.
Yada, T., Abe, M., Okada, T., et al., 2022. Preliminary analysis of the Hayabusa2 samples returned from C-type asteroid Ryugu. Nat. Astron. 6, 214–220.
Yamagishi, A., Hashimoto, H., Yano, H., Imai, E., Tabata, M., Higashide, M., Okudaira, K., 2021. Four-year operation of Tanpopo: astrobiology exposure and micrometeoroid capture experiments on the JEM exposed facility of the international space station. Astrobiology 21 (12), 1461–1472.
Yamagishi, A., Yokobori, S.-I., Hashimoto, H., Yano, H., Higashide, M., Tabata, M., Imai, E., Yabuta, H., Kobayashi, K., Kawai, H., 2014. Tanpopo: astrobiology exposure and micrometeoroid capture experiments—proposed experiments at the exposure facility of ISS-JEM. Trans. Jpn. Soc. Aeronaut. Space Sci. Aerosp. Technol. Jpn. 12 (ists29), Tk_49–Tk_55.
Yano, H., Fitzgerald, H., Tanner, W., 1994. Chemical analysis of natural particulate impact residues on the long duration exposure facility. Planet. Space Sci. 42, 793–802.
Yano, H., Kibe, S., Deshpande, S.P., Neish, M.J., 1997. The first results of meteoroid and debris impact analyses on the space flyer unit. Adv. Space Res. 20, 1489–1494.
Yano, H., Yamagishi, A., Hashimoto, H., Yokobori, S., Kobayashi, K., Yabuta, H., Mita, H., Tabata, M., Kawai, H., Higashide, M., et al., 2014. Tanpopo experiment for astrobiology exposure and micrometeoroid capture onboard the ISS-JEM exposed facility. In: 45th Annual Lunar and Planetary Science Conference. p. 2934.
Zhou, H., Xiao, Y., Zheng, Z., Yang, B., 2022. YOLOv2-tiny target detection system based on FPGA platform. In: 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering. ICBAIE, pp. 289–292. doi:10.1109/ICBAIE56435.2022.9985817.