

Assessment of onboard processing algorithms for cislunar space domain awareness

Kyle Merry

Sandia National Laboratories

Matthew Dykstra, Brian Hacsí

Sandia National Laboratories

ABSTRACT

Cislunar space presents a unique challenge for space domain awareness. Cislunar objects can inhabit a wide array of orbits and present a variety of sensor-to-object geometries. To adequately understand these challenges, we evaluated a hypothetical electro-optical sensor's performance across a range of sensor and object orbits including L1/L2 halos, distant retrograde, and near-rectilinear lunar.

To process these data, we simulated multiple onboard processing chains, with either traditional image processing or neural network-based detection algorithms. We studied the performance of these algorithms over a range of sensor-to-object geometries and sensor pointing configurations. For the neural networks, we also explored the effects of computing in reduced-precision quantized arithmetic, which enables low-SWaP onboard processing.

For a 5° field of view, 10 cm aperture system, we found that a feasible detection solution based on a deep convolutional neural network produced excellent coverage of cislunar objects at ranges up to 80,000 km under varied solar conditions and long-range detection performance beyond 120,000 km under favorable solar conditions.

1. INTRODUCTION

As nations accelerate their activities in lunar space, they have also recognized a need for enhanced Space Domain Awareness (SDA) capability in cislunar space [6]. Monitoring of space debris and management of traffic will be essential to the safe and successful operation of lunar missions and to the tracking of objects on lunar-assist trajectories [18]. The challenge is substantial; cislunar space is characterized by longer distances, longer timescales, and larger search volumes than the GEO domain [6]. The distance from the Earth to objects in the lunar sphere of influence complicates the application of terrestrial [5] or GEO/HEO [21] detection solutions, often requiring *a priori* knowledge of a target's state vector to assist detection by rate-tracking the sensor [14]. In this work, we assess a hypothetical electro-optical (EO) system in a cislunar orbit that can search cislunar space for unknown, previously undetected resident space objects (RSOs).

The communication challenges for such a system make it impractical to downlink imagery for ground-based processing. The diversity of cislunar orbits require large search volumes, while the distances and timescales involved require a sensitive imaging system and flexible, autonomous operation. This system would require sophisticated onboard processing in order to detect and extract RSOs from star background scenes in a robust and autonomous fashion.

This is a difficult image processing challenge, and is exacerbated by the need to perform the computation onboard. For large scale image processing in a space radiation environment, rad-tolerant or rad-hard Field Programmable Gate Array (FPGA) devices are often the most practical computational accelerators. However, implementing complex algorithms on FPGAs has traditionally required engineers with a highly specialized skillset and significant investment of development time. FPGAs are also not suited to accelerating all algorithms; they favor highly parallel, structurally simple algorithms that can be carried out in reduced precision fixed-point arithmetic¹. To make matters worse, modifying an onboard algorithm (e.g. to address an issue encountered in operations) will frequently require modifications to the FPGA bitstream. This bitstream is large and may be risky to update in an operational on-orbit sensor.

¹Modern System on a Chip (SoC) parts such as the Xilinx Versal ACAP have an improved ability to apply algorithms with extensive branching or random-access memory patterns, but reach peak performance with streamed algorithms and burst-memory-access patterns. ACAP devices also include AI/ML acceleration engines that are most useful for the large scale reduced-precision linear algebra operations common in AI/ML algorithms.

Fortunately, the proliferation of deep learning-based solutions to image processing problems has reached the FPGA toolchains as well. Multiple solutions exist for semi-automated mapping of sophisticated neural networks into FPGA accelerators [23, 3]. Deep learning algorithms such as Deep Convolutional Neural Networks (DCNNs) are structurally simple but mathematically dense and have been shown to be robust to implementation in reduced-precision arithmetic, or “quantization” [7]. After the network accelerator has been implemented into the FPGA, the network’s learned parameters are often read from a file system and can be updated at will. In practice, this means that a FPGA-accelerated neural network solution can be easily modified to address any problems that arise in operations, or to provide continuous improvements to the onboard processing algorithms.

In this paper, we will discuss many of the challenges inherent to processing EO imagery for cislunar space domain awareness. We consider a few potential onboard processing chains with feasible performance for a variety of cislunar RSO detection scenarios.

2. RELATED WORK

2.1 Space domain awareness

Holzinger *et al.* provide a primer on cislunar space and the challenges of cislunar SDA in [6], and [18] discusses cislunar orbital dynamics in detail. A system that successfully tracked a cislunar body with a terrestrial sensor can be found in [5]. Thornton *et al.* assess the capability of GEO and HEO-based sensors to detect cislunar objects in [21], and provide an excellent treatment of the SNR and integration time considerations for a 2° , 40 cm system. The NGSatSentry system in [2] includes a classical image processing solution that can detect RSOs in onboard processing, which informed some of our experiments with classical image processing techniques. The deep learning approach to RSO detection described by Woodward *et al.* [24] is suitable to onboard processing, and was included in our experiments.

2.2 Deep learning

We attempt to incorporate best practices from the deep learning literature when designing and training neural networks. Smith [20] provides a practical discussion of hyper-parameter configuration and training schemes. Focal loss [12] improves the performance of pixelwise segmentation networks on datasets with class imbalance, such as SDA imagery. Model ensembling [15] and Test Time Augmentation (TTA) [10] are established methods for improving a machine learning model’s prediction performance. There are a variety of methods for constructing neural networks with reduced computational requirements, but in this work we use a bottleneck structure derived from the MobileNetv2 architecture [19].

2.3 Quantized neural networks

The idea that neural networks are innately robust to evaluation in reduced-precision arithmetic (“quantization”) has been studied extensively [7, 23, 3]. In our work, we rely on frameworks developed by Xilinx Research for training and deploying highly quantized networks to FPGA devices. Namely, we use the Brevitas [16] library for Quantization-Aware Training. We use the FINN toolchain [23, 3] to transform and compile networks into an FPGA bitstream.

3. CHARACTERIZATION OF CISLUNAR DETECTION SCENARIOS

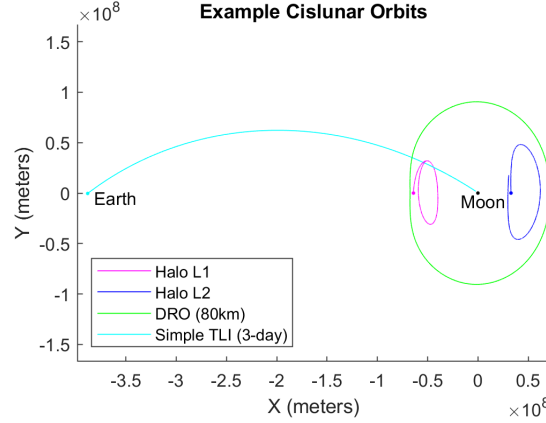


Fig. 1: Cislunar trajectories under consideration, projected on to the lunar orbital plane

Table 1: Representative cislunar detection scenarios. Angular rates are relative to the sidereal frame.

Scenario	Engagement range (km)	Angular rate ($\mu\text{rad}/\text{sec}$)	Integration time (sec)
Uncued Near	40,000	13	60
Uncued Mid	80,000	5	60
Uncued Far	120,000	3	60
Translunar Short Range	80,000	20	60
Translunar Long Range	300,000	2.6	600

To represent a cislunar RSO, we use a 1m diameter sphere with a 20% reflectance Lambertian surface under varied solar conditions. The Sun-Target-Observer-Angle (STOA) uniquely determines the amount of light reflected toward the observer by the Lambertian sphere. When the STOA is combined with the detection range and apparent motion, we can fully describe the expected signal on our simulated focal plane [21].

To characterize the sensor-to-RSO geometries, we constructed the exemplar orbital trajectories depicted in Fig. 1. We simulated sensors in the Distant Retrograde (DRO) and L_2 Halo orbits and measured their sensing geometries with respect to RSOs in the L_1 Halo, L_2 Halo, Distant Retrograde, or Trans-Lunar Insertion (TLI) trajectories for a number of orbital phasings. The result of this study was a series of figures such as Fig. 2, which depict the engagement range and the speed of angular motion between the two objects. These are our two variables of interest when assessing detection performance, since the range dictates the amount of signal received from the target, while the angular motion determines the number of sensor pixels transited by the RSO during an integration time.

Based on an analysis of these range/motion pairs, we can identify a few detection regimes. When observing RSOs that are in orbit near the moon, the sensor frequently encounters detection ranges from 20,000 km to 160,000 km. The angular motions of these targets are frequently less than $5 \mu\text{rad}/\text{sec}$, but some relatively close range ($<60,000$ km) objects will have angular motions of up to $25 \mu\text{rad}/\text{sec}$. RSOs that are transferring to a lunar orbit are characterized by much longer ranges and higher apparent angular motions than objects in orbit. We distilled this space of range/angular rate pairs into five exemplar cases for further study, shown in Table 1.

Three of these cases are examples of uncued cislunar SDA detections, where the sensor is expected to detect an RSO without *a priori* state vector knowledge. The ranges and geometries are representative of a DRO or L_2 -orbiting sensor that is observing an RSO in different phases of a L_1 Halo, L_2 Halo, or DRO. The *Near* scenario represents a relatively close-range detection between a sensor and an RSO in different orbits (e.g. DRO to L_1 Halo). At longer detection distances, the apparent motions between RSO and sensor decrease regardless of orbit. The *Mid* scenario has moderate

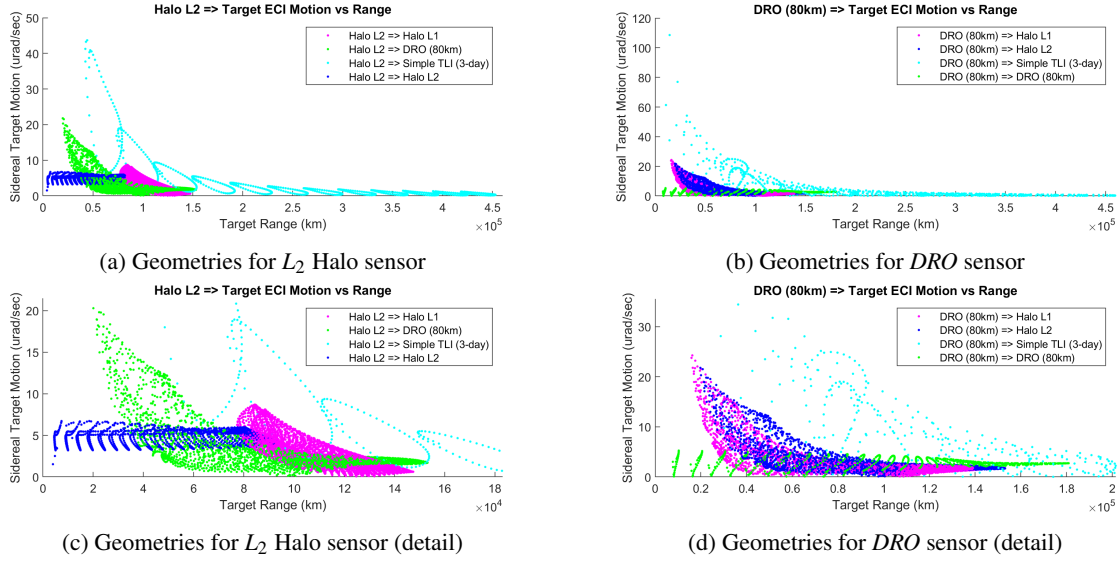


Fig. 2: Examples from sensor geometry study for a sensor in a distant retrograde or L_2 Halo orbit observing cislunar RSOs.

range but the RSO still moves during a 60-second integration period. In the *Far* case, the RSO's apparent motion is almost entirely due to the rotation between the sidereal and Earth-Moon rotating frame.

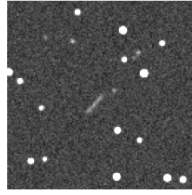
The remaining two cases are representative of a sensor in DRO or L_2 responding to a cue to re-acquire an RSO in a lunar transfer. These RSOs are characterized by very high angular rates and extremely long engagement ranges. We found that the sensor had limited capability to detect these RSOs in an uncued configuration, so we approach these cases with a cued rate-tracking configuration.

4. EXEMPLAR SENSOR CONCEPT

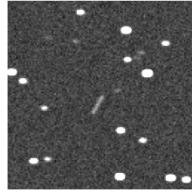
Table 2: Exemplar sensor specifications

Parameter	Value
Aperture Diameter	10 cm
Field of view	5°
Imager dimensions	2048 x 2048
Pixel IFOV	43 μ rad
Transmission	80%
Dark current	50 e-/s
Readout noise	5e-
Integration time	60 seconds
Summed scans	1 or 10
Ensquared energy	20%
Spectral band	0.4 – 0.9 μ m

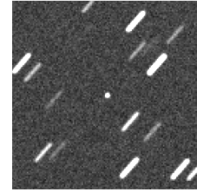
The principles of detection described in this paper apply over a range of sensor designs, but in order to study the effect of varying pointing configurations and processing algorithms we selected an exemplar sensor concept with the specifications shown in Table 2. This sensor is assumed to be hosted on a spacecraft in a DRO or L_2 Halo orbit in cislunar space. The modest aperture size and 5° field of view are intended to represent a lower-cost sensor for large



(a) Sidereal



(b) BRRF



(c) Rate-track

Fig. 3: An illustration of sensed imagery characteristics for three sensor pointing configurations. Colormap has been clipped to show contrast. See Section 5 for a description of the configurations.

volume surveillance with relatively frequent revisits, rather than a sensor intended for exquisite detection performance at long range. The ensquared energy value is intended to account for sensor jitter over the integration period. We assume that the sensor is sufficiently agile to rate-track moving objects.

5. POINTING CONFIGURATIONS

The method used to point the sensor in space is a primary contributor to overall sensor performance. We studied three pointing configurations: sidereal, rate-tracking, and Barycentric Rotating Reference Frame-fixed (BRRF). Note that these configurations refer to the behavior of the sensor during a collection (one or more consecutive integrations). The sensor will re-point between collections to observe a different region of its surveillance volume.

5.1 Sidereal stare

In a sidereal configuration, the sensor stares at a fixed point on the celestial sphere as depicted in Fig. 3a. This reduces the number of pixels influenced by stars, simplifying the background. However, since RSOs must move relative to the background in order to be detected, their signal will necessarily be spread over multiple pixels, reducing the signal-to-noise-ratio (SNR). Longer integration times lead to more spreading of the target energy, limiting the maximum SNR that can be achieved with long integration.

5.2 Rate-track stare

In a rate-track stare, the sensor is commanded to point along a known or hypothesized RSO trajectory as in Fig. 3c. If the hypothesis is correct, the RSO will resolve as a point or short streak with most of its energy focused into a few pixels. This configuration produces the highest theoretical SNR but requires a trajectory hypothesis. It also reduces the sensitivity of the sensor to any RSOs that do not match the same trajectory hypothesis. For these reasons, it is less ideal for general space domain awareness than it is for re-acquisition of a particular RSO.

5.3 Barycentric Rotating Reference Frame stare

As a compromise between these staring configurations, we propose a staring configuration that is unique to the cislunar space domain awareness problem. By staring in a fixed direction in the Barycentric Rotating Reference Frame (BRRF) defined by the Earth-Moon system, we can minimize the apparent motion of cislunar RSOs relative to our sensor. This is analogous to staring at the GEO belt with a terrestrial sensor: we can approximately rate-track many targets with one field of view, since they are slow-moving in a frame that is rotating with our own. The BRRF frame rotates at approximately $2.6 \mu\text{rad}/\text{sec}$ on a fixed inclination relative to the celestial background.

Unlike a sidereal stare, the apparent motion of a cislunar RSO in the BRRF frame approaches zero as range increases. This focuses the RSO's signal in a few pixels and allows BRRF to approach the SNR performance of rate-tracking against many long-range targets. In contrast with a rate-track stare, the BRRF star background has consistent appearance for a given integration time and pointing declination. The detection algorithms can take advantage of this fact to improve discrimination of RSOs from the star background.

For the results presented in Section 8, the sensor uses a BRRF pointing configuration for the uncued scenarios and rate-tracking for the cued, trans-lunar scenarios.

6. DETECTION METHODS

Many strategies for detecting RSOs fall in to one of two categories: by process of elimination, or by detection of anomalous motion.

Process of elimination In the process of elimination, all significant signals in the imager's field of view are compared to a star catalog. Signals that do not match with a star in the catalog are assumed to be RSOs. These algorithms have the advantage of detecting RSOs on short-integration imagery, where motion against the star background is not observable. Their disadvantages include a need for a relatively high-precision knowledge of the sensor's attitude and optical distortion. Some methods also require photometric calibration of the sensor. In non-sidereal staring configurations, the processing must become more sophisticated to handle the ambiguity resulting from computing a centroid on a star that is streaking across many pixels, and potentially overlapping other stars during an integration.

Anomalous motion When detecting anomalous motion, the algorithm must measure the shape of a signal in a single frame or the motion of a signal over several frames. For the purposes of this report, we are primarily considering algorithms which can detect anomalous motion of RSOs in a single frame of imagery. These algorithms are simple, do not require high-fidelity calibration of the sensor, and are easily applied to long-integration-time imagery and non-sidereal imagery.

6.1 Temporal algorithms

Classical and machine learning detection pipelines can use multiple frames of information to improve their detection performance. Temporal information is used to improve location and motion estimates for discriminating between stars and RSOs, and is an important tool for discarding false detections due to noise.

We assume that our single-frame detector is part of an onboard processing system that includes a temporal persistence check for detected clusters. A cluster will only be reported as a detection if a nearby cluster was detected on the previous frame. This filter is simple to implement is highly effective at eliminating uncorrelated false detections due to noise. In our experience, it can efficiently filter false detection rates at densities of 10^{-4} false positives per focal plane pixel without significant impact to the true positive detection quality, and substantially reduces the false detection rate. We use this target of 10^{-4} false positives per pixel to set our detection thresholds for all algorithms, with the assumption that the false detection rate reported out of the system will be manageable for storage and transmission to a ground-station.

A study of the performance of more sophisticated temporal algorithms for cislunar SDA would need to consider trades between single-frame integration time, number of frames collected per stare, onboard implementation complexity, and the impact on the system's revisit time. We anticipate that a multi-frame system would detect at longer ranges but with increased revisit period compared to the algorithms considered in this report.

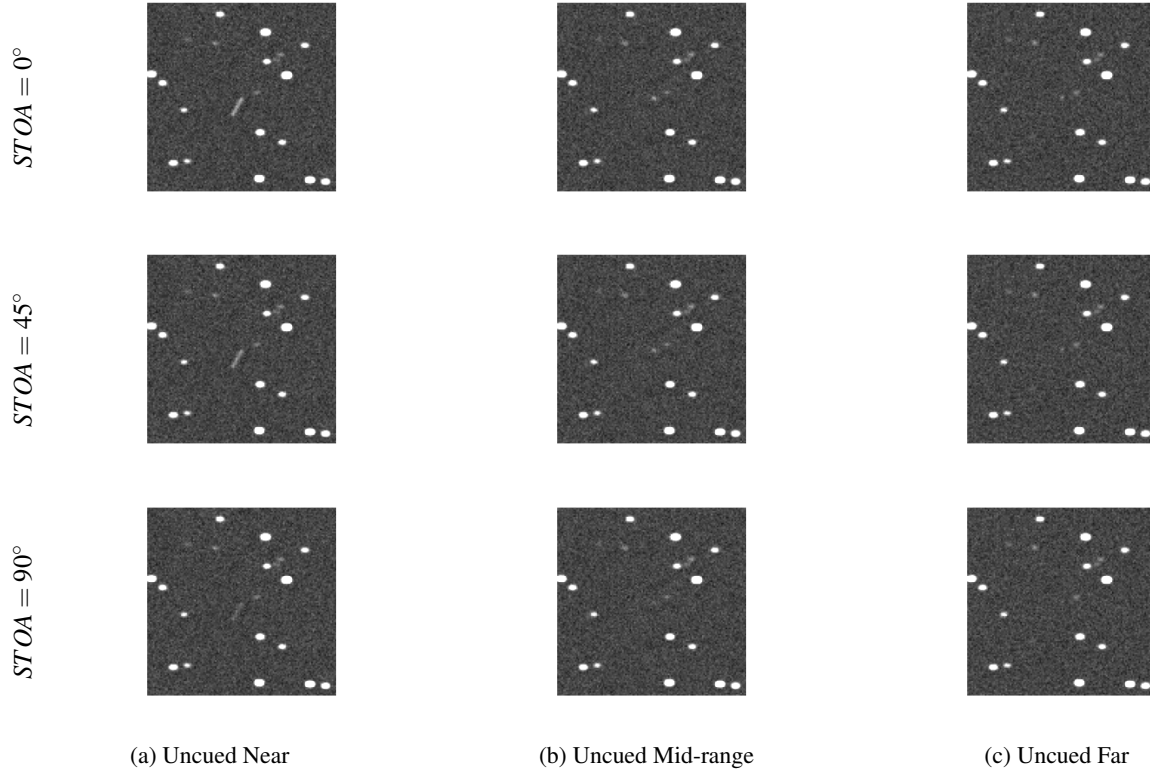


Fig. 4: A simulated RSO target with varied solar angle, target range, and motion characteristics. Collected in the BRRF pointing configuration with 60 sec integration. Colormap has been clipped to show contrast.

7. ALGORITHMS FOR ONBOARD DETECTION

In deciding which algorithms to consider in our study, we made judgments based on an assessment of suitability to onboard processing and focused our analysis on algorithm components that function with a single frame of data.

The first criterion, suitability to onboard processing, is difficult to objectively measure due to the variety of onboard processing solutions. An algorithm that is suitable on a System-on-a-Chip or a heterogeneous single-board computer with a dedicated application processor may not be feasible on a system with no general-purpose application processing unit. We focused our assessment on simpler algorithms that are reliant on parallelizable image processing operations which depend only on the availability of an FPGA accelerator.

The second criterion, single-frame-detection, is intended to allow us to focus on the effects of range, apparent motion, solar angle, and pointing configuration on the detection problem. We believe that a holistic study of the full-system performance including temporal preprocessing, temporal filtering, and multi-frame detection (e.g. spatio-temporal DCNN or velocity-matched filtering) will be essential to understand the full capabilities of an onboard processing system for cislunar SDA. These techniques will almost certainly further improve performance beyond the baseline capability discussed here.

7.1 Classical image processing

There are a number of methods for detecting RSOs in star field imagery with classical image processing techniques [4], but for our detection configuration we found that the performance hinges on an algorithm's shape discrimination ability. In the *Mid* and *Far* scenarios of Fig. 4, the algorithm must distinguish a point-like or anomalously moving streak from a field of very short, wide, elliptical stars. Similarly, rate-tracked collections of distant RSOs over 60 sec integration frequently result in very small motion of the star background. The two classical processing components that we tested for this task were *local normalization* and *form factor classification*. To estimate the theoretical maximum performance of a classical image processing algorithm, we also evaluated the *ideal single-frame* case.

Local normalization We construct an annulus kernel with inner radius r_1 and outer radius r_2 . A pixel at the center of the kernel is normalized by the mean and standard deviation of the pixels in the annulus range. By tuning r_1 and r_2 , we can construct a filter that suppresses short streaks while passing point-like signals.

Form factor Detailed in [4], the form factor calculation measures the elongation of a cluster to attempt to discriminate between circular and elliptical shapes.

Ideal single-frame To estimate what is possible for a classical image processing algorithm with perfect star removal, we simulated scenes without stars. We processed these scenes with an ideal matched filter and detected all clusters above a given SNR threshold.

7.2 UNet DCNN

The *UNet* architecture utilized by Woodward *et al.* [24] is based on a popular architecture for pixel-wise segmentation [17], with an up-sampling and down-sampling structure that allows the network to achieve a large receptive field with modest computational cost. By allowing the network to perceive large areas of the input image, the network can potentially use spatial context in its detection decisions. Woodward *et al.* demonstrated that this network architecture significantly outperformed the classical image processing baseline as represented by *Source-Extractor*.

7.3 Full resolution DCNN

We also propose a customized model architecture with no up- or down-sampling layers, which we refer to as the *fullres* or full-resolution architecture. We have observed that large receptive fields are not always necessary for SDA data, and we hypothesize that they may be detrimental. When evaluating the DCNN in an onboard accelerator, the image must often be broken in to tiles (e.g. 256x256) to evaluate properly with limited working memory. A DCNN with a large receptive field can exhibit reduced performance near the edges of these tiles, requiring a large overlap between adjacent tiles and considerable redundant computation for ideal performance. The *fullres* architecture intentionally limits its receptive field to 11x11 pixels, using 1x1 convolutions to provide more depth without increasing receptive field.

Convolutional block Our standard convolutional block is based on the MobileNetV2 [19], with 1x1 expansion convolution, 3x3 grouped convolution, and 1x1 bottleneck convolution on the output. We use a ReLU6 [11] activation after each of the first two convolutions, and a linear activation on the output bottleneck. The 3x3 grouped convolution provides a means to scale the computational requirements of the network against the memory bandwidth, as a convolution layer with more groups will apply fewer calculations for a given input feature map than a non-grouped convolution. We experimented with varying the grouping from $\#groups = 1$ (ungrouped, maximum computation) to $\#groups = \#channels$ (depthwise, minimal computation). The ungrouped convolution produces the best detection performance, but even a fully depthwise configuration produced acceptable detection performance.

Architecture experiments Before each activation layer we have a normalization layer, implemented as either Batch Normalization [9] or Instance Normalization [22]. Instance Normalization significantly improved our network's performance, and is preferred as long as it is supported by the inference acceleration framework.

We experimented with following each convolutional block with a Squeeze and Excitation module [8]. The intent of this module was to allow the network to account for global context despite its small receptive field. We did not observe any benefit from including this module in our architecture, so it has been excluded.

Fullres model The *fullres* model is intended to be as small as possible without severely impacting detection performance. It consists of one 2D 3x3 convolutional layer followed by 4 of the convolutional blocks described above. Finally, an output 2D 1x1 convolution with sigmoid activation produces the single-channel pixel-wise segmentation frame. Our bottleneck channel size (between convolutional blocks) is 16 channels, and the expansion factor within the blocks is 4.

7.4 DCNN training procedure

All DCNN models were trained using data from a low-fidelity simulator. This simulator produced 512 x 512 images with randomly generated stars, RSOs, dark current, and gradients used to represent stray light. The distributions of apparent motion of the stars and RSOs were configured to match our exemplar sensor with a given integration time and pointing configuration. A model was trained for a specific sensor, pointing configuration, and integration time.

We used focal loss [12] as the training criterion and updated the weights with an AdamW [13] optimizer under a one-cycle learning rate schedule [20]. We observed that a one-cycle learning rate schedule stabilizes and accelerates the training process, allowing us to train to higher final accuracy with fewer epochs. Focal loss was also critical to ensure that the network appropriately prioritizes the few positive RSO pixels relative to the vast majority of background pixels.

For quantization-aware training of the 8-bit and 4-bit variants, we used the Brevitas library [16] to define a quantized neural network that could be trained with the same procedure as the floating-point networks.

The networks were trained to take raw 16-bit sensor images that contain bad pixels, fixed pattern noise, and non-uniformity artifacts. No onboard imagery correction or normalization was assumed.

7.5 Ensembling (x4)

We trained multiple copies of the same model architecture using different initial weights and different samples of the training dataset. We formed a model ensemble by taking the per-pixel mean of each model’s output on a given input image [15]. Our experiments with several reduction functions – mean, median, min, and max – found that mean reduction produced the best F1 score performance.

7.6 Test time augmentation (TTA)

The weights of a trained DCNN are rarely symmetrical, so a trained DCNN will often respond differently to vertically- and horizontally-flipped permutations of the input image. We take advantage of this empirical observation and apply test time augmentation techniques [10], wherein the DCNN is shown horizontally- and vertically-flipped copies of the input image and the outputs are flipped back to the original orientation. Taking the mean of these outputs is equivalent to ensembling a model with horizontally- and vertically-flipped permutations of itself.

7.7 Computational considerations

Table 3: Comparison of DCNN model sizes and computational requirements.

Model identifier	Model size	Operations per megapixel	Weights Datatype	Activations Datatype	Est. Processing Time ² (2k x 2k)
<i>UNet32b (Woodward et al)</i>	8MB	30G	float32	float32	~
<i>UNet4b (Modified)</i>	2MB	30G	int4	uint4	6s
<i>UNet4b + TTA (Modified)</i>	2MB	120G	int4	uint4	24s
<i>Fullres32b (Ours)</i>	800kB	14G	float32	float32	~
<i>Fullres8b (Ours)</i>	200kB	14G	int8	uint8	6s
<i>Fullres4b (Ours)</i>	100kB	14G	int4	uint4	2s
<i>Fullres4b + TTA (Ours)</i>	100kB	56G	int4	uint4	8s
<i>Fullres4b x4 + TTA (Ours)</i>	400kB	224G	int4	uint4	32s

For cislunar SDA with our 10cm aperture and 5° field of view, we frequently require integration times of 60 seconds or more to accumulate enough signal for detection and observe anomalous motion against the stars. A processing solution which completes in less than 60 seconds can be used to reduce average power consumption or to allow a single FPGA accelerator to process data from multiple sensors.

²Runtime estimates for the 4-bit *fullres* network are derived from FINN RTL estimates for a Xilinx *Ultrascale+* FPGA. Runtime estimates for 8-bit and *UNet* networks are derived from scaling approximations [3] and benchmarks of comparable networks. Floating-point networks of this size are not considered to be viable on current generation FPGA accelerators. Runtime can vary wildly depending on embedded processor architecture and the amount of resources dedicated to the DCNN accelerator on the FPGA.

8. RESULTS

Table 4: Observed probability of detection for the DCNN-based detection algorithms on simulated scenes. Each scene consists of a simulated sensor collection against a star background with 100 synthetic targets injected at uniformly random focal plane locations and a fixed Sun-Target-Observer-Angle (STOA). The *ideal single-frame* benchmark is discussed in Section 8.1

STOA:	Uncued Near (BRRF) 40,000 km, 13 μ rad/sec				Uncued Mid (BRRF) 80,000 km, 5 μ rad/sec			
	0°	30°	60°	90°	0°	30°	60°	90°
<i>Ideal single-frame</i>	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.49
<i>UNet32b</i>	0.91	0.93	0.93	0.80	0.75	0.64	0.64	0.10
<i>UNet4b</i>	0.93	0.97	0.95	0.79	0.90	0.74	0.43	0.07
<i>UNet4b + TTA</i>	0.89	0.93	0.95	0.83	0.89	0.77	0.45	0.08
<i>Fullres32b</i>	0.95	0.99	0.99	0.95	0.95	0.96	0.82	0.08
<i>Fullres8b</i>	0.95	0.99	0.99	0.92	0.97	0.82	0.52	0.02
<i>Fullres4b</i>	0.95	0.99	0.99	0.92	0.95	0.91	0.75	0.07
<i>Fullres4b + TTA</i>	0.95	0.99	0.99	0.94	0.97	0.99	0.84	0.12
<i>Fullres4b x4 + TTA</i>	0.99	0.99	0.99	0.98	0.95	0.97	0.87	0.15
STOA:	Uncued Far (BRRF) 120,000 km, 3 μ rad/sec				Translunar Near (RT) 80,000 km, 20 μ rad/sec			
	0°	30°	60°	90°	0°	30°	60°	90°
<i>Ideal single-frame</i>	1.00	0.97	0.69	0.05	1.00	1.00	0.98	0.82
<i>UNet32b</i>	0.52	0.40	0.26	0.05	0.80	0.86	0.72	0.13
<i>UNet4b</i>	0.66	0.53	0.18	0.02	0.83	0.94	0.80	0.25
<i>UNet4b + TTA</i>	0.69	0.44	0.12	0.00	0.85	0.94	0.80	0.27
<i>Fullres32b</i>	0.86	0.69	0.10	0.00	0.96	0.98	0.94	0.27
<i>Fullres8b</i>	0.60	0.24	0.02	0.00	0.96	0.97	0.94	0.26
<i>Fullres4b</i>	0.76	0.55	0.15	0.00	0.96	0.97	0.94	0.58
<i>Fullres4b + TTA</i>	0.86	0.74	0.17	0.00	0.99	0.98	0.94	0.70
<i>Fullres4b x4 + TTA</i>	0.90	0.76	0.18	0.00	0.99	0.99	0.96	0.72

To evaluate the performance of our detection algorithms, we used imagery generated by a separate simulator with higher fidelity than the one used for training. The decision to use simulated data is anchored in two observations: first, the cislunar SDA data under consideration have simpler phenomenology than data from terrestrial telescopes. The absence of atmospheric effects such as seeing, airglow, or cloud obscuration simplify accurate simulation of these data. We have also observed that models trained on synthetic data were able to generalize to data from real sensors as long as the simulation parameters accurately represented the data from the test sensor.

In Table 4 we show the performance of several DCNN variants against the simulated test data. The 32-bit variants of the networks are likely not deployable in an onboard FPGA accelerator, but are useful as a benchmark. Our estimates in Section 7.7 suggest that the 8-bit and 4-bit variants of both DCNNs would be capable of running onboard in Ultrascale-class hardware given a 60-second integration time. The TTA and ensemble modes of the 4-bit models are also likely feasible, depending on hardware implementation details and the number of sensors handled by a single accelerator.

The *fullres* models provided the best performance in most cases, with the *UNet32b* model providing better detection probability in the SNR-limited *Uncued Far* cases. Of the onboard-feasible models, the *Fullres4b x4 + TTA* performed consistently well, with the *Fullres4b + TTA* model providing nearly identical detection performance at a fraction of the computation time. The interactions between quantization and ensembling for the *UNet* and *fullres* models are discussed in Sections 8.2 and 8.3, respectively.

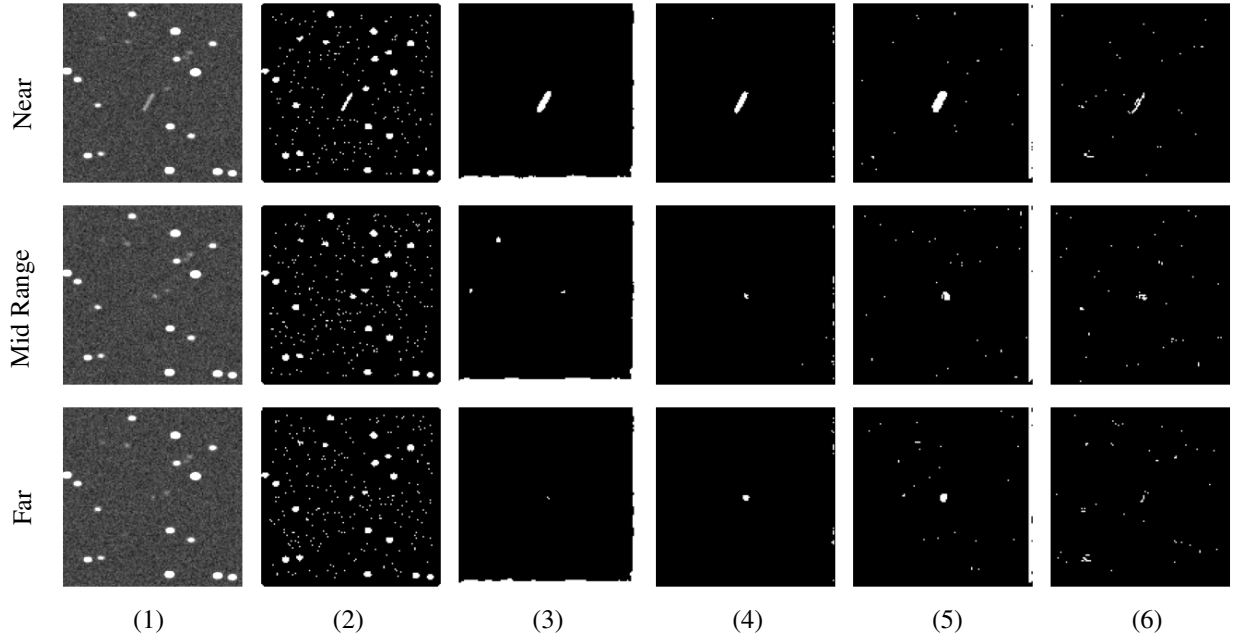


Fig. 5: Examples of detection algorithm performance for the Uncued Near, Mid, and Far detection scenarios. From left to right, the columns are: 1) raw input image, 2) local normalization, 3) UNet32b, 4) Fullres32b, 5) Fullres8b, 6) Fullres4b

8.1 Classical image processing

We were unable to produce satisfactory performance from the classical image processing algorithms over the full range of simulated scenes. They excel at identifying RSOs near the noise floor and can confidently reject stars when there is significant motion difference between the stars and RSO. However, they struggle to discriminate between the RSOs and stars in difficult cases. For example, the DCNNs in Fig. 6 are able to detect a point-like RSO against a field of very slightly horizontally-elongated stars. The traditional shape discrimination algorithms were unable to reject the stars in these collections, resulting in poor detection performance.

Similarly, traditional algorithms performed well in rate-track collections where the target moves significantly relative to the stars, but were unable to detect the slow-moving long-range objects in a 60 second integration time image.

We can't rule out the possibility that a sophisticated classical image processing system could be constructed that could operate over a range of pointing configurations and star/RSO motion, nor did we evaluate the performance of multi-frame or process-of-elimination approaches. To provide an upper bound on what is feasible for a single-frame detector, we simulated a perfect image processing pipeline capable of detecting RSOs with idealized noise-limited performance and with no influence from star confusors. To simulate this pipeline, we generated scenes without stars and processed them with an ideal matched filter detector. These results are listed as *ideal single-frame* in Table 4.

These results demonstrate the strength of classical image processing algorithm with respect to detecting low-SNR targets in noise. In future work, we hope to investigate architectures for combining classical and machine learning algorithm components that can provide the low-SNR capability of classical algorithms with the shape-discrimination power of a DCNN.

8.2 Effect of quantization

For the *fullres* architecture, we observe a slight decrease in detection performance in the most stressing cases when restricting the network to 4-bit quantization for weights and biases. For the *UNet* architecture, 4-bit quantization appears to act as a form of regularization, improving test performance. This could be an indication that our data are too simple for the *UNet* model, and it has been overfitting to our training set. It may require training with stronger regularization to improve performance.

We tested 2-bit quantization of our *fullres* network, but those results are not shown here as the performance was

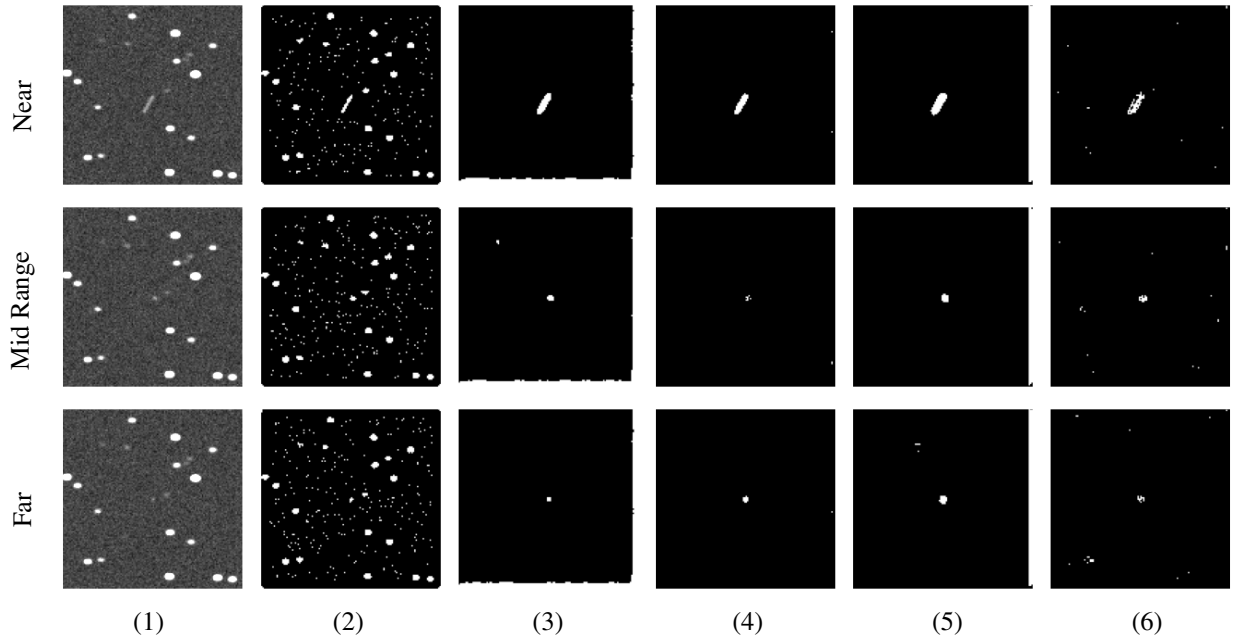


Fig. 6: Examples of detection algorithm performance for the Uncued Near, Mid, and Far detection scenarios after test time augmentation. From left to right, the columns are: 1) raw input image, 2) local normalization, 3) UNet32b + TTA, 4) Fullres32b + TTA, 5) Fullres8b + TTA, 6) Fullres4b + TTA

significantly degraded. This is not unexpected, as we know from [23] that special care must be taken to train a network with extremely low quantization values (e.g. 2- or 1-bit). We believe that there is more investigation to be done with respect to extremely low quantization values.

8.3 Ensembling and Test Time Augmentation

We observed comparable performance improvements for 4x ensembling and 4x test-time augmentation on the *fullres* network. Test-time augmentation will be slightly more efficient in practice due to re-use of a model's weights, and is recommended if one method must be selected. However, we noted that TTA and ensembling are not mutually exclusive, and can be combined to provide substantial detection performance at the cost of 16 total model evaluations per frame.

TTA appears to mildly degrade performance for the *UNet* architecture. This is a surprising result, but we were not able to ascertain a cause by inspection of the detection masks output by the model. This result certainly merits further investigation.

8.4 Trans-lunar insertion detection performance

The two trans-lunar scenarios in Table 1 were representative of our cislunar SDA sensor attempting to maintain custody of an RSO in the late stages of trans-lunar insertion. For the short range (80,000 km) scenario, the sensor demonstrated excellent performance in a rate-tracking configuration shown in Table 4, but was unable to detect the RSO in an uncued configuration. For the long range (300,000 km) scenario, we simulated collection with an extended 600 second integration time, formed by co-addition of 10 60-second rate-tracked frames. The sensor was unable to detect this RSO even in optimal solar conditions. A system with a larger aperture and/or narrower field of view would likely be required to maintain custody of trans-lunar RSOs with a reasonable integration period.

9. CONCLUSION

We evaluated an exemplar 5°, 10 cm electro-optical sensor system in a variety of cislunar space domain awareness (SDA) collection scenarios and pointing configurations. We asserted multiple onboard processing chains and measured the probability of detecting a resident space object (RSO) given an acceptable system-wide false detection rate. Deep convolutional neural network (DCNN)-based detection solutions were easily applied to this problem and presented excellent performance. Classical image processing approaches have the potential for superior low-SNR detection performance but require additional processing to reject the star background. We believe that more investigation into hybrid classical/machine learning architectures is warranted.

Our experiments suggest that SDA DCNNs are suited to evaluation in reduced-precision arithmetic, and that the system was very capable with the weights and activations quantized to 4-bit. We evaluated multiple methods for improving detection performance at the cost of additional computation, and found that model ensembling and test time augmentation can be highly effective for some DCNNs. This provides a designer with options for trading detection performance against the number of sensors that can be handled by a single FPGA accelerator.

These results demonstrate that such a system would have utility for uncued cislunar SDA, with excellent coverage of cislunar RSOs at ranges up to 80,000 km under varied solar conditions and long-range detection performance beyond 120,000 km under favorable solar conditions. It is significantly more difficult to detect objects in lunar transfer with this system. A cue is required to detect trans-lunar insertions of RSOs at 80,000 km, and the system is not sensitive enough to detect trans-lunar RSOs at 300,000 km.

In future work, we intend to expand our trade space to consider multi-frame detection algorithms and assess their effects on detection performance and orbit estimation.

10. ACKNOWLEDGMENTS

We thank Zach Mekus, John Ossorgin, and Jared Leyba for their contributions to designing, training, and assessing the performance of quantized neural networks for space domain awareness. We thank Taylor Clonts for his expertise in deploying quantized neural networks to FPGA accelerators. We thank Dave Cunningham and William Ledbetter for their work in simulating and visualizing cislunar object trajectories.

11. REFERENCES

- [1] *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, September 2021.
- [2] N. Bertrand, D. Cheung, J. Gu, and E. Zenker C. McNamara, C. Saidnawey. NGSatSentry: On-orbit detection system for space domain awareness. In *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference* [1].
- [3] M. Blott, T. Preußner, N. Fraser, G. Gambardella, K. O’Brien, Y. Umuroglu, M. Leeser, and K. Vissers. FINN-R: An end-to-end deep-learning framework for fast exploration of quantized neural networks. *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, 11(3):1–23, 2018.
- [4] G. Fitzgerald, Z. Funke, A. Cabello, V. Asari, and J. Fletcher. Toward deep-space object detection in persistent wide field of view camera arrays. In *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference* [1].
- [5] R. Furfaro, V. Reddy, T. Campbell, and B. Gray. Tracking objects in cislunar space: The Chang’e 5 case. In *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference* [1].
- [6] M. Holzinger, C. Chow, and P. Garretson. A primer on cislunar space. 2021.
- [7] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio. Quantized neural networks: Training neural networks with low precision weights and activations. *CoRR*, abs/1609.07061, 2016.
- [8] F. Iandola, S. Han, M. Moskewicz, K. Ashraf, W. Dally, and K. Keutzer. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size, 2016.
- [9] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.
- [10] M. Kimura. Understanding test-time augmentation. In *Neural Information Processing*, pages 558–569, Cham, 2021. Springer International Publishing.

- [11] A. Krizhevsky. Convolutional deep belief networks on CIFAR-10. 2010.
- [12] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2999–3007, 2017.
- [13] I. Loshchilov and F. Hutter. Decoupled weight decay regularization, 2017.
- [14] M. Lévesque. Detection of artificial satellites in images acquired in track rate mode. In *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, September 2011.
- [15] U. Naftaly, N. Intrator, and D. Horn. Optimal ensemble averaging of neural networks. *Network*, 8:283–296, 1997.
- [16] A. Pappalardo. Xilinx/brevitas, 2021.
- [17] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
- [18] A. Rosengren. Cislunar multiscale dynamics and implications for SSA. In *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference* [1].
- [19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen. MobileNetV2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.
- [20] L. Smith. A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay, 2018.
- [21] D. Thornton, B. Little, B. Steward, and R. Cobb. Detecting dim targets in cislunar space using GEO/HEO-based optical sensors. In *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference* [1].
- [22] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Instance normalization: The missing ingredient for fast stylization, 2016.
- [23] Y. Umuroglu, N. Fraser, G. Gambardella, M. Blott, P. Leong, M. Jahre, and K. Vissers. FINN: A framework for fast, scalable binarized neural network inference. In *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, FPGA ’17, pages 65–74. ACM, 2017.
- [24] D. Woodward, C. Manughian-Peter, T. Smith, and E. Davison. Pixelwise image segmentation with convolutional neural networks for detection of resident space objects. In *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference* [1].