

# DEEP MONOCULAR HAZARD DETECTION FOR SAFE SMALL BODY LANDING

Travis Driver<sup>\*,\*</sup>, Kento Tomita<sup>\*,†</sup>, Koki Ho<sup>‡</sup>, and Panagiotis Tsiotras<sup>§</sup>

Hazard detection and avoidance is a key technology for future robotic small body sample return and lander missions. Current state-of-the-practice methods rely on high-fidelity *a priori* terrain maps, which require extensive human-in-the-loop verification and expensive reconnaissance campaigns to resolve mapping uncertainties. We propose a novel safety mapping paradigm that leverages deep semantic segmentation techniques to predict landing safety directly from a single monocular image, thus reducing reliance on high-fidelity *a priori* data products. We demonstrate precise and accurate safety mapping performance on real *in-situ* imagery of a prospective sample site from the OSIRIS-REx mission.

## INTRODUCTION

Hazard detection and avoidance (HD&A) is a key technology for future robotic small body sample return and lander missions. Current approaches rely on high-fidelity digital elevation maps (DEMs) derived from digital terrain models (DTMs), local topography and albedo maps, generated on the ground.<sup>1</sup> However, DTM construction involves extensive human-in-the-loop verification, carefully designed image acquisition plans, and expensive reconnaissance campaigns to resolve mapping uncertainties.<sup>2,3</sup> We instead propose a novel safety mapping paradigm that leverages Bayesian deep learning techniques to accurately predict landing safety maps *directly from monocular images* in order to reduce reliance on expensive high-fidelity, *a priori* data products (i.e., DTMs).

Safety mapping methodologies that leverage deep learning have demonstrated potential to improve the accuracy of onboard hazard detection. Previous works<sup>4,5</sup> have leveraged deep semantic segmentation to classify safe and unsafe landing locations from digital elevation maps (DEMs) from simulated LiDAR scans. However, generating reliable DEMs from LiDAR scans is non-trivial and requires accurate state estimates and range measurements. Moreover, LiDARs typically feature a relatively small effective operating range<sup>6</sup> and increased size, weight, and power (SWaP) requirements relative to passive sensors such as monocular cameras. Thus, we propose to derive landing safety directly from monocular images without assuming any *a priori* data or relying on the fidelity of the current state estimate. Deep semantic segmentation has been previously employed for surface characterization of small bodies from simulated monocular images, primarily focusing on boulder detection.<sup>7,8</sup> Conversely, we apply our models to real images and directly predict safety maps that conform to realistic landing parameters and constraints.

---

<sup>\*</sup>These authors contributed equally to this work.

<sup>\*</sup>PhD Student, Institute for Robotics and Intelligent Machines, School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30313, USA.

<sup>†</sup>PhD Student, School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30313, USA.

<sup>‡</sup>Associate Professor, School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30313, USA.

<sup>§</sup>David & Andrew Lewis Chair, Professor, Institute for Robotics and Intelligent Machines, School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30313, USA.

The contributions of this paper are as follows: (i) we develop a novel safety mapping paradigm that leverages Bayesian deep learning techniques to accurately landing safety *directly from monocular images*; (ii) we construct a dataset of *real* monocular images and corresponding landing safety maps that conform to realistic landing parameters for training and testing our models; (iii) we demonstrate precise safety mapping performance on real imagery of prospective sample sites from the OSIRIS-REx mission.

## RELATED WORK

Current hazard detection methodologies for small body missions rely on high-fidelity digital elevation maps (DEMs) derived from digital terrain models (DTMs), local topography and albedo maps.<sup>1</sup> However, DTM construction typically involves extensive human-in-the-loop verification and carefully designed image acquisition plans to achieve optimal results.<sup>2,3</sup> Consequently, autonomous hazard detection and avoidance (HD&A) has been identified as a high-priority technology<sup>9</sup> to promote and enable new mission concepts to near-earth asteroids, comets, the Moon, Mars, and beyond.

The Autonomous Landing Hazard Avoidance Technology (ALHAT)<sup>10,11</sup> program was launched in 2005, followed by the Safe & Precise Landing—Integrated Capabilities Evolution (SPLICE) program<sup>12</sup> in 2018, in order to develop autonomous landing technologies. These programs have focused on developing HD&A algorithms that operate on DEMs generated from range measurements acquired by active sensors such as flash and scanning LiDARs. However, these methods are constrained by the relatively small effective operating range and the increased size, weight, and power (SWaP) requirements of LiDARs relative to passive sensors such as monocular cameras. Indeed, the OSIRIS-REx Guidance, Navigation, and Control (GNC) flash LiDAR had a maximum operational range of approximately 1 km,<sup>13,14</sup> while preliminary testing of the Hazard Detection LiDAR by the SPLICE program demonstrated a 5 cm ground sample distance at 500 meters and near-nadir pointing.<sup>12</sup> Conversely, the OSIRIS-REx Camera Suite (OCAMS)<sup>15</sup> was able to acquire 5 cm GSD images at almost 4 km, providing higher-resolution measurements earlier in the mission than the active sensors onboard and allowing for detailed surface characterization during early phases of the mission.<sup>6</sup> Moreover, constructing a DEM from LiDAR scans is non-trivial and requires accurate range measurements and precise knowledge of the spacecraft’s relative pose to the landing plane. Instead, we focus on estimating landing safety directly from a single monocular image.

Safety mapping methodologies that leverage deep learning have demonstrated potential to improve hazard detection accuracy, and have also been shown to offer competitive runtimes on flight relevant hardware.<sup>16</sup> Previous works have leveraged deep semantic segmentation to classify safe and unsafe landing locations from high-resolution DEMs. Moghe and Zanetti<sup>4</sup> leverage a deep neural network architecture for predicting safety maps from a DEM and design a novel loss function specifically designed to decrease the false safe rate and encourage more precise safe predictions. Tomita et al.<sup>5</sup> employ a Bayesian SegNet architecture for segmentation of input DEMs into safe and unsafe landing locations. The Bayesian architecture implemented by Tomita et al.<sup>5</sup> enables uncertainty quantification of the predicted safety map through the predictive entropy of the model, allowing for more precise predictions through global thresholding with respect to this uncertainty measure. We build upon this work and demonstrate its efficacy on monocular imagery.

Methods based on deep learning have also been developed for surface segmentation from monocular imagery. Pugliatti and Maestrini<sup>7</sup> employ a custom U-Net architecture for classification of surface landmarks, i.e., boulders and crater rims. Caroselli et al.<sup>8</sup> apply deep semantic segmen-



tation to boulder detection on synthetic images of a fabricated small body model and post-process the network prediction to derive a landing safety map based on boulder density. Conversely, we apply our models to real images and directly predict safety maps that conform to realistic landing parameters and constraints.

## PROPOSED APPROACH

Our novel safety mapping paradigm leverages Bayesian deep learning techniques to develop a uncertainty-aware semantic segmentation model to predict safety maps from monocular imagery. We train and test our model on real *in-situ* imagery from the OSIRIS-REx mission to asteroid 101955 Bennu with corresponding ground truth safety maps generated using realistic landing parameters and constraints.

### Bayesian Deep Learning

Given training inputs  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  and corresponding labels  $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ , Bayesian deep learning employs *Bayesian inference* to maximise the posterior distribution of the network parameters  $\boldsymbol{\theta}$  given the training data  $\mathcal{X}, \mathcal{Y}$ :

$$p(\boldsymbol{\theta} | \mathcal{X}, \mathcal{Y}) = \frac{p(\mathcal{Y} | \mathcal{X}, \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathcal{Y} | \mathcal{X})}. \quad (1)$$

The distribution above can then be used to predict the output  $\mathbf{y}^*$  for a new input  $\mathbf{x}^*$  via

$$p(\mathbf{y}^* | \mathbf{x}^*, \mathcal{X}, \mathcal{Y}) = \mathbb{E}_{p(\boldsymbol{\theta} | \mathcal{X}, \mathcal{Y})}[p(\mathbf{y}^* | \mathbf{x}^*, \boldsymbol{\theta})], \quad (2)$$

where we may assume a softmax likelihood for  $p(\mathbf{y}^* | \mathbf{x}^*, \boldsymbol{\theta})$ . However, computing  $p(\boldsymbol{\theta} | \mathcal{X}, \mathcal{Y})$  is intractable and must be approximated using *variational inference*.

Gal and Ghahramani<sup>17,18</sup> showed that training a NN with dropout layers is equivalent to approximate variational inference with a variational distribution  $q(\boldsymbol{\theta})$  which imposes a Bernoulli distribution over the model weights. Specifically, consider a convolutional layer  $i$  with  $c_{i-1}$  input channels,  $c_i$  output channels, and weights  $W_i \in \mathbb{R}^{c_{i-1} \times k \times k \times c_i}$ . Then dropout can be viewed as imposing a distribution over the layer weights according to

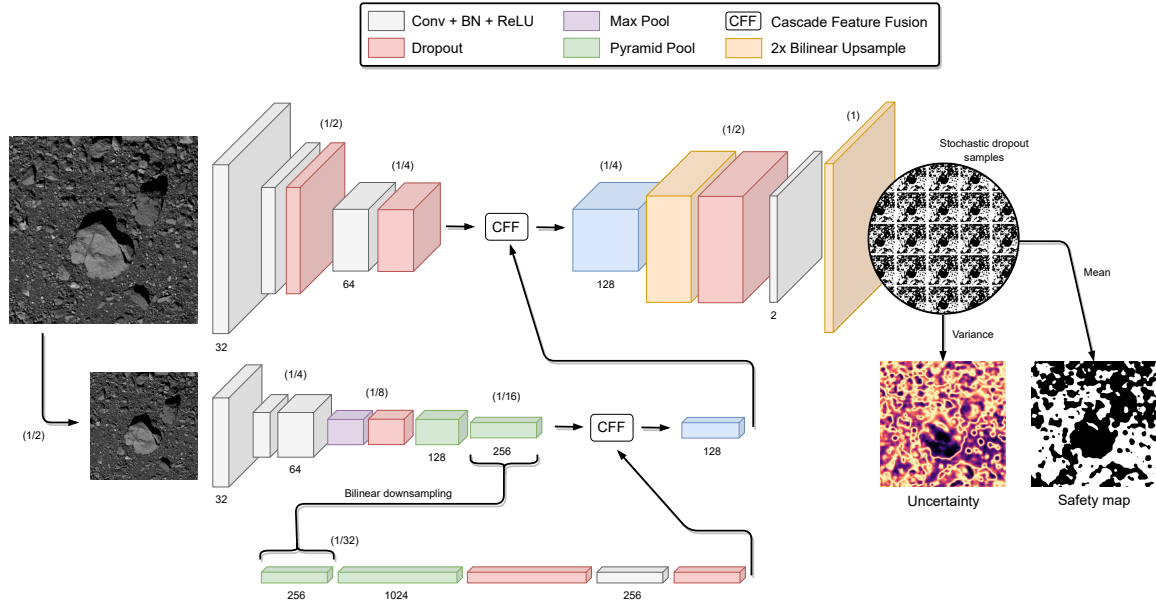
$$W_i = M_i \cdot \text{diag}([\epsilon_j]_{j=1}^{c_i}), \quad \epsilon_j \sim \text{Bern}(p_j), \quad (3)$$

where  $\epsilon_j$  are Bernoulli distributed random variables with parameter  $p_j$  (which we take to be 0.5), and  $M_i$  are the variational parameters optimized during training. Therefore, Equation (2) may be approximated by

$$p(\mathbf{y}^* | \mathbf{x}^*, \mathcal{X}, \mathcal{Y}) \approx \mathbb{E}_{q(\boldsymbol{\theta})}[p(\mathbf{y}^* | \mathbf{x}^*, \boldsymbol{\theta})]. \quad (4)$$

Finally, employing dropout at test time permits the use of the predictive entropy approximated through  $T$  *stochastic forward passes* through the network as a measure of uncertainty for each pixel-wise label  $y \in \mathcal{Y}$ :

$$\mathbb{H}[y | \mathbf{x}, \mathcal{X}, \mathcal{Y}] \approx - \sum_c \left( \frac{1}{T} \sum_{t=1}^T p(y = c | \mathbf{x}, \boldsymbol{\theta}_t) \right) \log \left( \frac{1}{T} \sum_{t=1}^T p(y = c | \mathbf{x}, \boldsymbol{\theta}_t) \right), \quad (5)$$



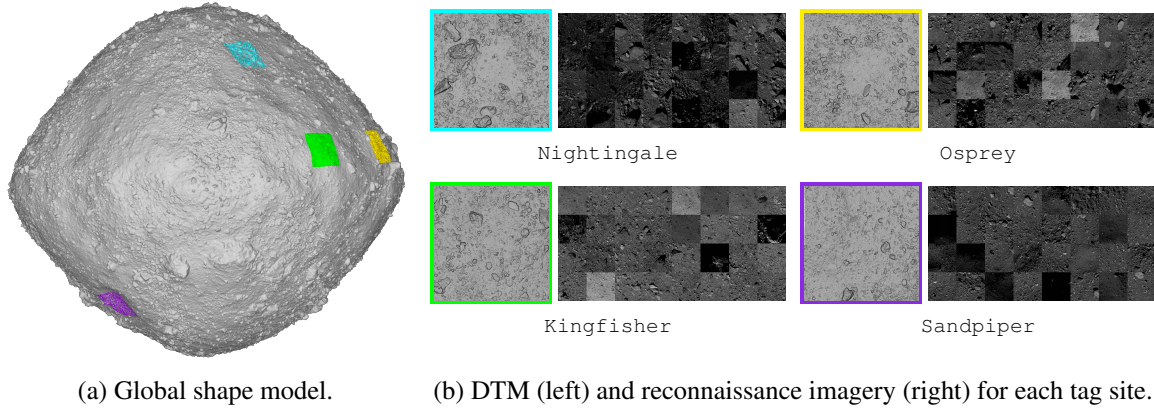
**Figure 1: Bayesian ICNet architecture.** Multiscale fusion is conducted within the Cascade Feature Fusion (CFF)<sup>19</sup> modules. The ratio in parentheses denotes the relative magnitude of the spatial dimensions with respect to the original image.

where  $\theta_t \sim q(\theta)$  corresponds to network parameters sampled during a stochastic forward pass through the network, and the average over forward passes is taken to be the final prediction probabilities. This process is referred to as *Monte Carlo (MC) dropout*. We will demonstrate that leveraging this uncertainty measure in our network predictions leads to increased precision and accuracy of safe landing locations.

### Uncertainty-Aware Semantic Segmentation

We leverage an uncertainty-aware semantic segmentation architecture based on the image cascade network (ICNet),<sup>19</sup> shown in Figure 1. ICNet is a highly efficient segmentation architecture that blends coarse prediction maps obtained from down-sampled inputs with high-resolution feature maps obtained from high throughput networks that operate on the full-resolution image, allowing for fast inference on high-resolution images while maintaining accuracy. Multiscale feature map fusion is conducted by the *cascade feature fusion (CFF)* modules, whereby a reduced-resolution segmentation map is computed from the two multiscale feature map inputs. The multiscale predictions are used to train the network via a weighted softmax cross entropy loss.<sup>19</sup>

We implement a Bayesian version of ICNet, which we denote as BICNet, where dropout layers are added to allow for stochastic sampling with respect to the model parameters using techniques from Bayesian deep learning, i.e., MC dropout, as described in the previous subsection. Ideally, a Bayesian NN would feature a dropout layer after every hidden layer of the network.<sup>17,18</sup> However, as was observed in previous works,<sup>20,21</sup> adding dropout layers after every convolutional layer in more complex networks is too strong of a regularizer and results in underfitting. Therefore, we follow the work of Mukhoti et al.<sup>20</sup> and Kendall et al.<sup>21</sup> and only insert dropout layers after the central encoder and decoder layers. At test time, we perform  $T = 8$  stochastic forward passes and



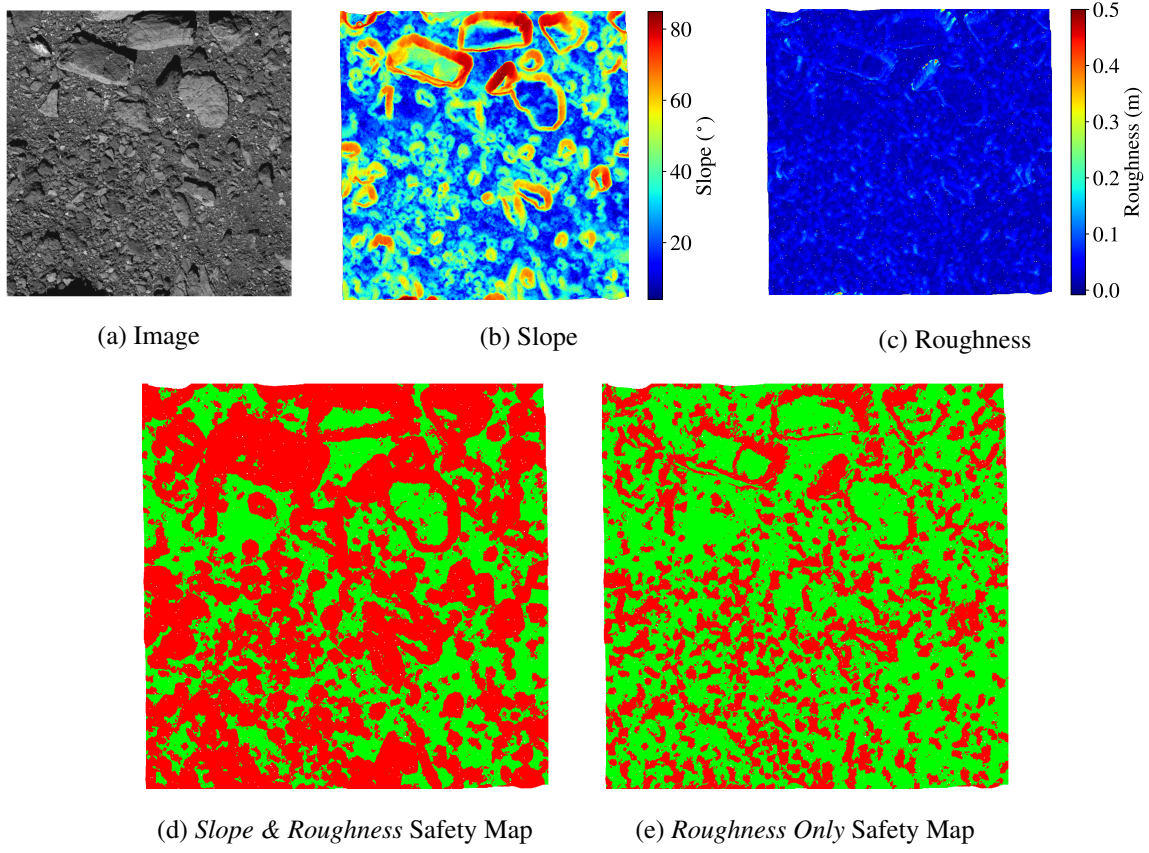
**Figure 2: OSIRIS-REx TAG site datasets.** TAG site locations are indicated by the corresponding color in the global shape model.

use the predictive entropy, defined in Equation (5), as a measure of uncertainty and use the average of this measure over all training instances to mask out high uncertainty regions in the image.

### Data Generation

High-fidelity DTMs (i.e., 5 cm ground sample distance) of the four prospective Touch-And-Go (TAG) sample sites developed as part of the OSIRIS-REx mission to asteroid 101955 Bennu were used to generate ground truth safety map labels for reconnaissance imagery from the mission. Specifically, we leverage *monocular* reconnaissance imagery and the corresponding camera pose labels, relative to a body fixed frame of the asteroid, provided through the AstroVision dataset.<sup>22</sup> For each image, DEMs are constructed by transforming the DTM into a local coordinate system in which the  $+z$ -axis points opposite the vector corresponding to the direction of the gravitational force due to the target body at the point on the surface closest to the center of the image. Safety mapping was conducted on the DEM and then projected back into the image to produce pixel-wise landing safety labels. Example reconnaissance images for each prospective TAG site are provided in Figure 2.

Safety was computed from the DEMs using the method developed by the Autonomous Landing Hazard Avoidance Technology (ALHAT) project.<sup>23</sup> The ALHAT method evaluates the lander contact locations for all pixels and for all orientations to assess the worst-case surface slope and roughness values with respect to the surface elevation data contained in the ground truth DEMs. Specifically, a landing plane is computed for each pixel by assessing the elevation of four evenly spaced contact points, emulating lander foot pads, on the perimeter of a circle specified by the diameter of the lander. Slope is defined as the largest angle between the landing plane and  $x$ - $y$  plane of the ground truth DEM for all orientations, and the roughness is the largest perpendicular distance to the terrain above the landing plane for all orientations. Any pixel with slope and roughness exceeding a given threshold is labeled as unsafe, where we chose a threshold of  $30^\circ$  for slope and 3.5 cm for roughness. We specify a lander with a 35 cm diameter, similar to the MASCOT (Mobile Asteroid surface SCOUT) lander that was deployed during the Hayabusa2 mission to asteroid 162173 Ryugu.<sup>24</sup> An example safety map along with its corresponding monocular image is provided in Figure 3.



**Figure 3: Ground truth safety map example.** Safe and unsafe regions are drawn in green and red, respectively, in the safety map.

Moreover, we provide the data distributions of our datasets with respect to the ground sample distance, imaging depth, viewing angle, and visibility ratio. *Ground sample distance* (GSD) measures the average distance on the surface spanned by a single pixel, which is a function of the distance to the surface and the camera intrinsics, as landing safety becomes increasingly difficult to observe as the relative of the lander in the image decreases. The *imaging depth* measures the average distance to the surface when the image was taken, and provides context for the GSD values. Specifically, the MapCam of the OSIRIS-REx Camera Suite (OCAMS),<sup>15</sup> with a focal length of  $\sim 125$  mm, can provide 5 cm GSD measurements of the surface at distances of approximately 1 km, while the PolyCam, with a focal length of  $\sim 620$  mm, provides the same resolution at distances of almost 4 km. *Viewing angle* measures the angle between the  $-z$ -axis of the ground truth DEM and the camera boresight. Finally, the *visibility ratio* is the ratio of visible (i.e., not occluded by shadows) pixels to total pixels in the image and provides a measure of the illumination conditions in the image.

## RESULTS

In this section, we first present our suite of metrics used to evaluate the performance of our approach. We then validate our approach on two different experiments using real images from the OSIRIS-REx mission to asteroid 101955 Bennu, including images captured during the actual TAG sample collection event.



**Figure 4: Data distributions with respect to imaging depth, GSD, viewing angle, and visibility ratio.** Our dataset features a total of 770 annotated images: 133 of Kingfisher, 342 of Nightingale, 162 of Osprey, and 91 of Sandpiper, and 42 from the TAG sample collection event at Nightingale.

### Metric Definitions

We measure the quality of the predicted per-pixel safety map labels of our model with respect to precision, sensitivity, accuracy, and mean intersection over union (mIoU):

$$\text{precision} = \frac{\text{true safe}}{\text{true safe} + \text{false safe}}, \quad (6)$$

$$\text{sensitivity} = \frac{\text{true safe}}{\text{true safe} + \text{false unsafe}}, \quad (7)$$

$$\text{accuracy} = \frac{\text{true safe} + \text{true unsafe}}{\text{valid pixels}}, \quad (8)$$

$$\text{mIoU} = \frac{1}{2} \left( \frac{\text{true safe}}{\text{valid pixels} - \text{true unsafe}} + \frac{\text{true unsafe}}{\text{valid pixels} - \text{true safe}} \right). \quad (9)$$

**Table 1: Overall performance for the Sandpiper experiment.** The values in parentheses are the metrics with shadowed pixels ignored. All reported values are percentages.

| METHOD                       | PRECISION            | SENSITIVITY          | ACCURACY             | mIoU                 |
|------------------------------|----------------------|----------------------|----------------------|----------------------|
| <b>SLOPE &amp; ROUGHNESS</b> |                      |                      |                      |                      |
| WITHOUT UNCERTAINTY          | 60.66 (62.91)        | <b>67.21 (70.05)</b> | 69.53 (69.41)        | 52.05 (52.27)        |
| WITH UNCERTAINTY             | <b>76.98 (77.67)</b> | 20.09 (21.86)        | <b>82.29 (82.01)</b> | <b>65.76 (65.93)</b> |
| <b>ROUGHNESS ONLY</b>        |                      |                      |                      |                      |
| WITHOUT UNCERTAINTY          | 77.24 (78.92)        | <b>61.61 (63.66)</b> | 63.53 (64.41)        | 44.87 (45.31)        |
| WITH UNCERTAINTY             | <b>85.71 (86.32)</b> | 28.78 (31.63)        | <b>73.77 (73.93)</b> | <b>55.11 (54.98)</b> |

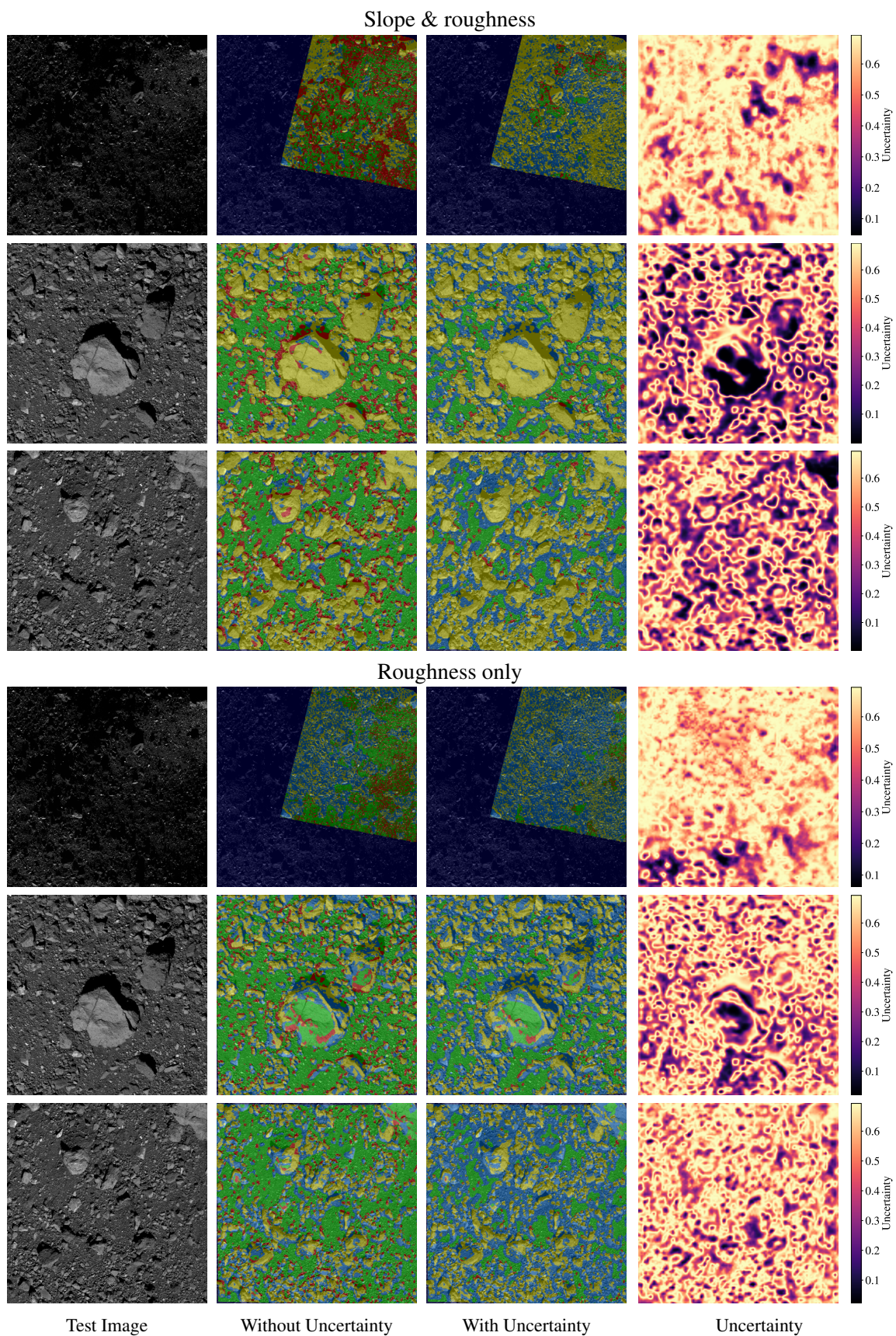
*True safe (false safe)* include pixels predicted to be safe by our models that are safe (unsafe) in the ground truth labels, and *true unsafe (false unsafe)* includes pixels predicted to be unsafe that are unsafe (safe) in the ground truth labels. Note that false unsafe includes the safe pixels that are ignored and not labeled safe due to high uncertainty. For our application, we can interpret *precision* as the reliability of the pixels predicted to be safe, and *sensitivity* as detection rate of true safe sites, respectively. *Accuracy* and *mIoU* are the metrics evaluated for the valid pixels, which are the pixels with smaller uncertainty than the threshold. In other words, valid pixels corresponds to the predictions that the network is 'certain' about. For the results without uncertainty thresholding, accuracy and mIoU are evaluated for all the pixels with valid safety labels. In the following analysis, we refer to the ratio of pixels that fall above our uncertainty threshold, and consequently marked as unsafe, as the *screening rate*.

### Experiment 1: Prospective Landing Site Sandpiper

We train our model on images from three of the prospective landing sites from the OSIRIS-Rex mission, i.e., Nightingale, Osprey, and Kingfisher, and test our model on images of the remaining sample site, Sandpiper. These results are detailed in Table 1, and qualitative examples are provided in Figure 5. Our models are able to predict safety maps from just a single monocular image of the prospective landing site, which is completely unseen during training, with accuracy over 69% for the slope and roughness hazard detection case, and over 63% for the roughness only hazard detection case, even without uncertainty thresholding. Moreover, our Bayesian ICNet architecture enables uncertainty thresholding in order to further boost performance by ignoring regions in which the model's prediction has high entropy. With the uncertainty threshold, accuracy increases to 82.29% and 73.77% for the slope and roughness and roughness only cases, respectively, at the cost of decreased sensitivity. Importantly, we are able to achieve 76.98% and 85.71% precision for the slope and roughness and roughness only cases, respectively. We also have a slight increase in all the metrics by ignoring shadowed pixels, which are reported in parentheses in Table 1.

Comparing the two different tasks of hazard detection, which consists of the slope and roughness case and the roughness only case, the roughness only case has lower values of accuracy and mIoU, but higher values of precision and sensitivity. This means that the roughness only case is a harder task in terms of precisely labeling safe and unsafe pixels on average, resulting in lower accuracy and mIoU, but is an easier task in terms of identifying only safe pixels, thus resulting in higher precision and sensitivity. This is partially due to the fact that roughness only hazard detection





**Figure 5: Qualitative monocular safety mapping results for the Sandpiper experiment.** Green, yellow, blue, and red labels represent true safe, true unsafe, false unsafe, and false safe, respectively.

requires attention to more detailed features, but also provided with more safe pixels than the slope and roughness case, as illustrated in Figure 3.

Additionally, we analyze the per-image metrics with respect to GSD, viewing angle, and visibility ratio in Figures 6 and 7. As a general trend, we can observe that a higher uncertainty results in lower performance metrics of precision, sensitivity, accuracy, and mIoU. As a cause of high uncertainty, the low visibility is a common factor that results in a higher uncertainty in our model for both of slope and roughness case (Figure 6) and roughness only case (Figure 7). Our models assign a higher uncertainty to images with larger GSD for slope and roughness case (Figure 6), and larger viewing angle for roughness only case (Figure 7). Note that increased uncertainty for images at higher GSDs may also be due to these instances being less represented in the training data as shown in Figure 4. In either case, we demonstrate that the uncertainty threshold serves as a powerful tool for detecting and accounting for difficult or out-of-distribution input conditions.

## Experiment 2: OSIRIS-REx TAG Sequence

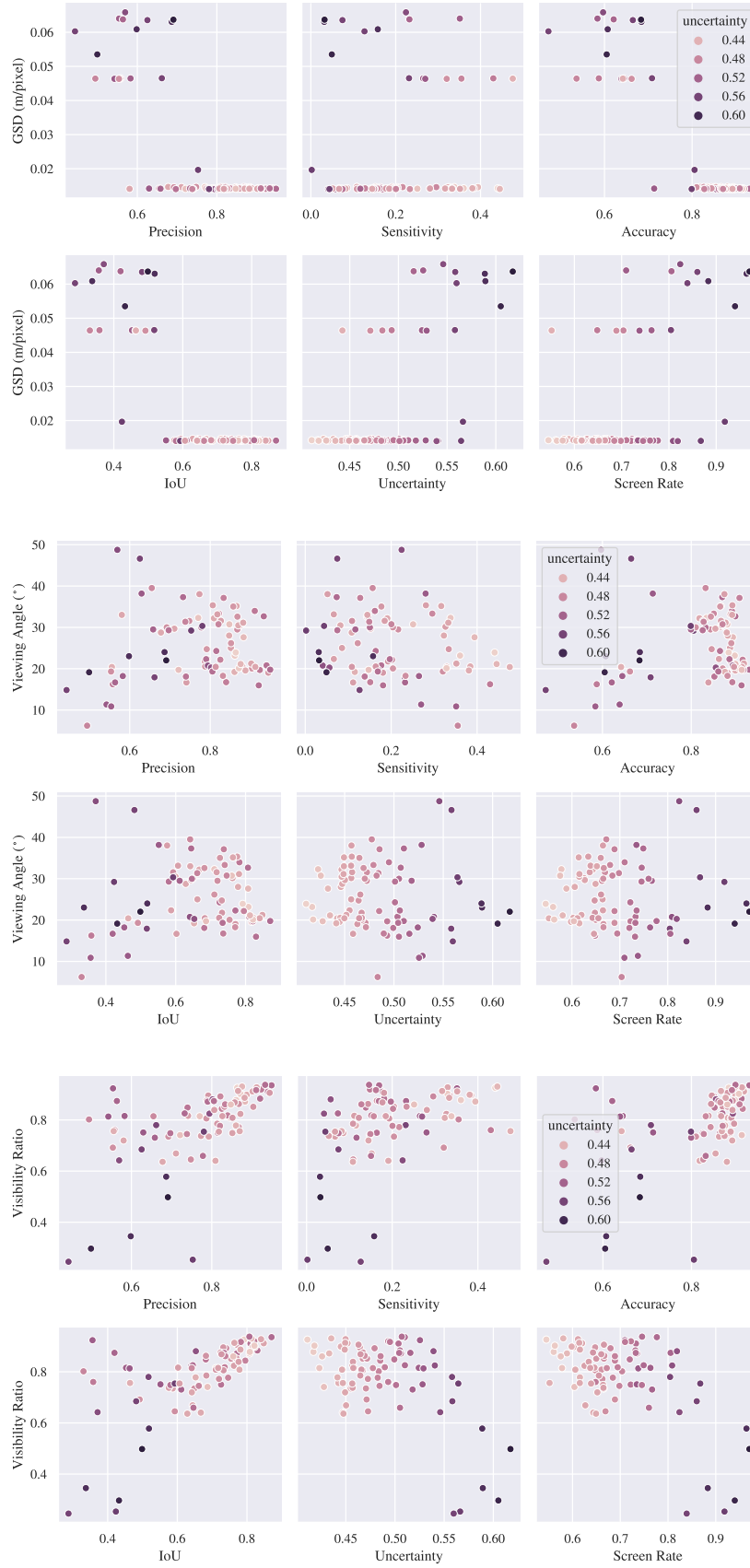
For the second experiment, we train two models using different combinations of images from the OSIRIS-Rex mission: one model, denoted by BICNet-NKO, is trained on Nightingale, Kingfisher, and Osprey, and the other model, denoted by BICNet-KOS, is trained on Kingfisher, Osprey, and Sandpiper. We test the two models on images captured during the TAG sample collection event at the Nightingale sample site. We present both models to illustrate the effect of the training distribution on the test results. A subset of the 42 image sequence is shown in Figure 8. Note that the 42 test images are not included in the training set even for the Nightingale images.

Table 2 and Figure 10 show quantitative results and qualitative examples, respectively. Comparing BICNet-NKO and BICNet-KOS in Table 2, we can see sensitivity of BICNet-KOS is significantly lower than that of BICNet-NKO after uncertainty thresholding. Indeed, BICNet-KOS assigns a high uncertainty to almost all safe regions of the input images and are thus overwritten as unsafe after uncertainty thresholding, as shown in Figure 10, which is not the case for BICNet-NKO. These differences are explained by the training and testing data distribution relationship; as illustrated in Figure 4, the TAG images have less overlap with the images of the prospective landing sites except for Nightingale with respect to GSD, viewing angle, and visibility ratio, and so the exclusion of Nightingale from training data increases the predictive uncertainty for TAG images at test time for BICNet-KOS. This effect of training data distributions on the uncertainty level and the accompanying predictive performance is consistent with the per-image metrics with respect to GSD that is shown in Figure 9; for BICNet-NKO, as GSD of testing images gets closer to the peak of the training data, the predictive performance increases and uncertainty decreases. On the other hand, for BICNet-KOS, the testing images have a high screening rate and a low sensitivity due to the high uncertainty. These results illustrate the effect of training data distribution and the ability of the uncertainty measure to identify out-of-distribution data for uncertainty-aware segmentation networks.

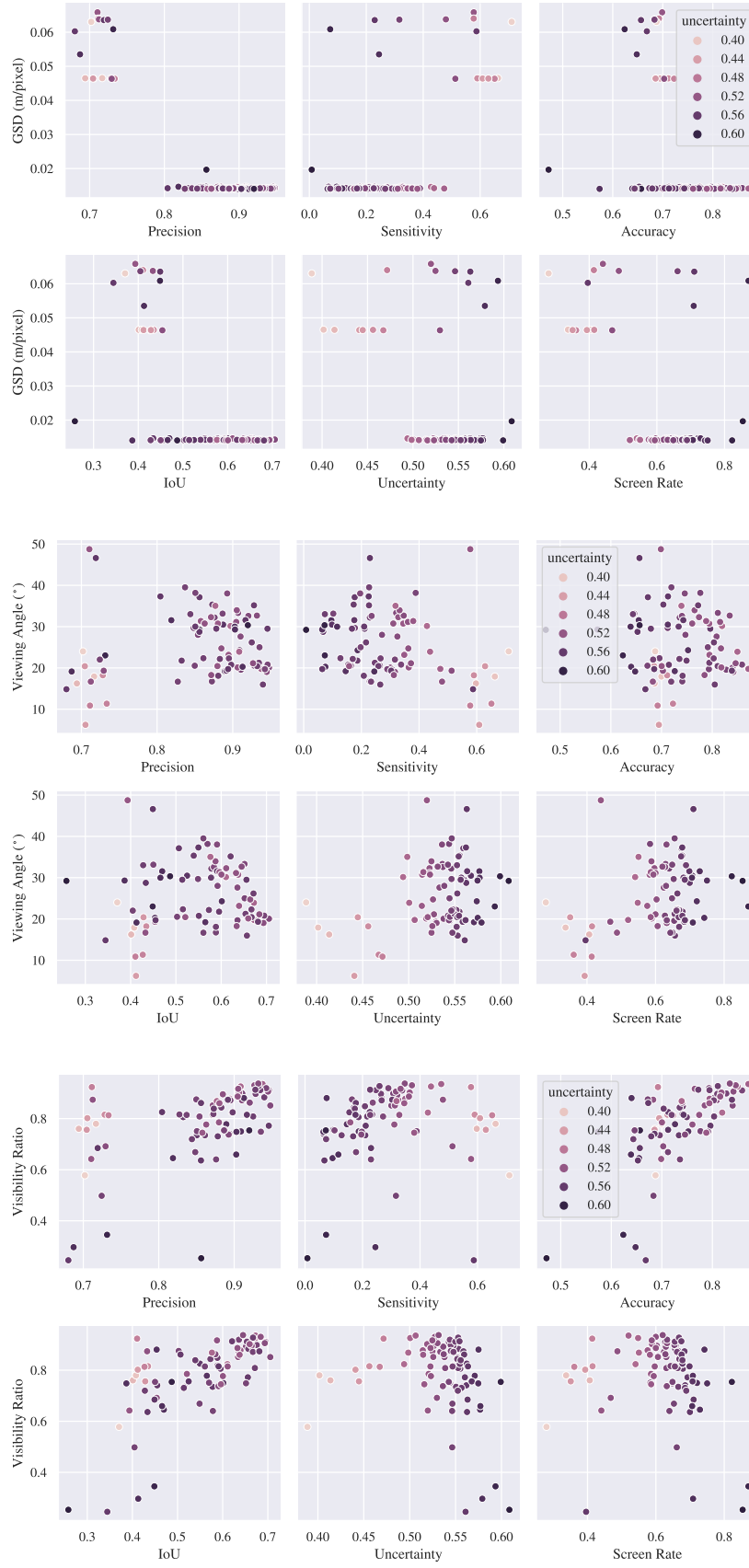
## CONCLUSION

In this paper we presented a novel landing hazard detection approach for small body missions that predicts safety maps directly from monocular imagery. We implemented an efficient, uncertainty-aware segmentation network that demonstrated impressive hazard detection performance on challenging test instances from real imagery obtained from the OSIRIS-REx mission. Future work involves identifying the causes of uncertainty to increase the reliability of proposed approach.

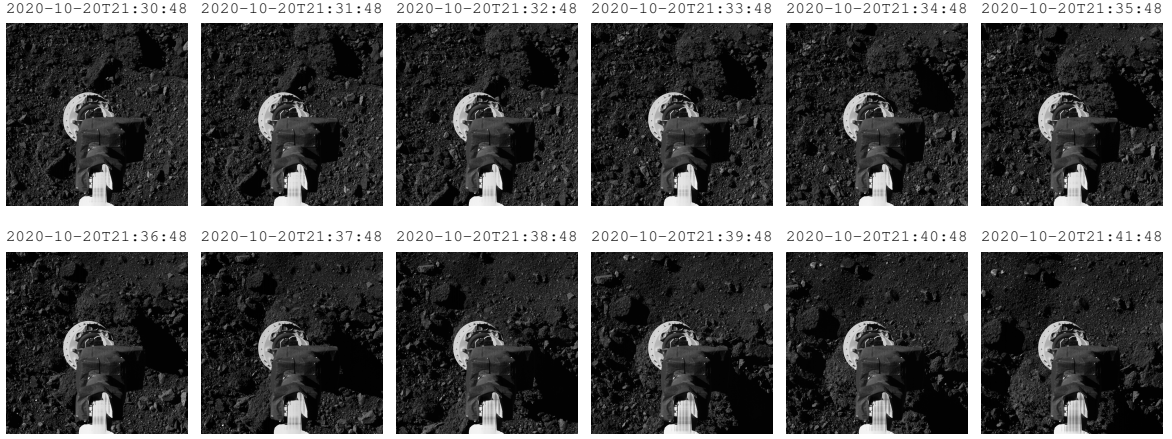




**Figure 6: Per-image metrics for *slope* & *roughness* on the Sandpiper experiment with respect to GSD, viewing angle, and visibility ratio.**



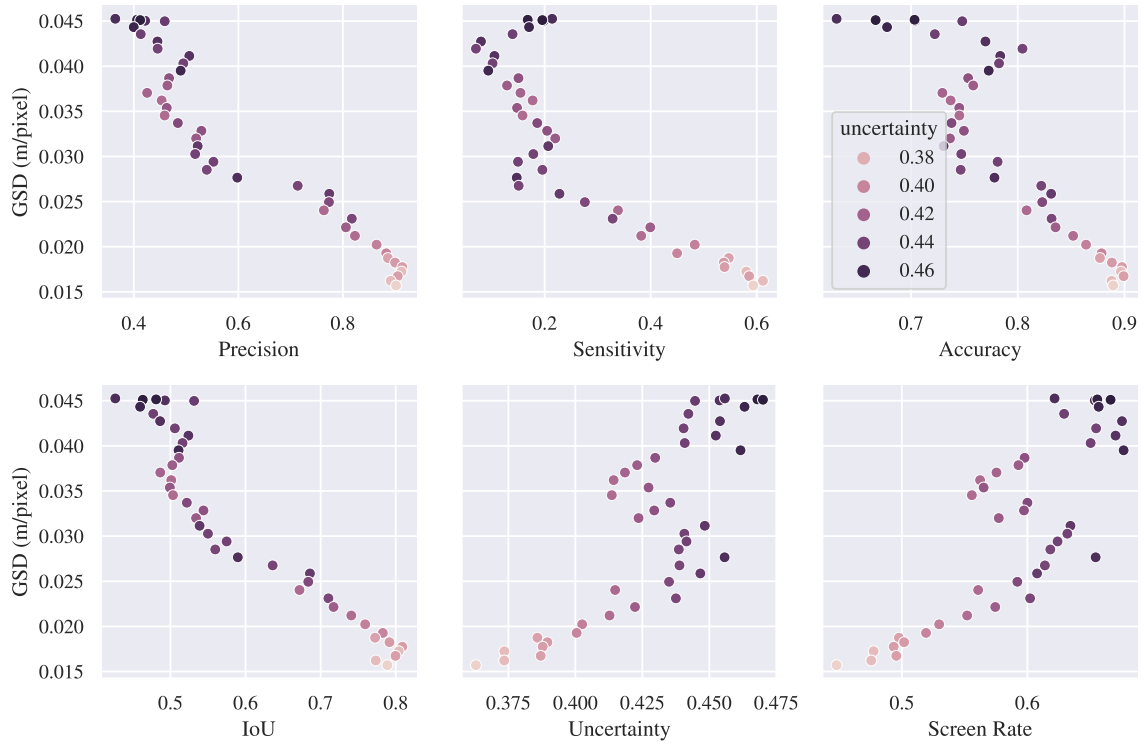
**Figure 7: Per-image metrics for *roughness only* on the Sandpiper experiment with respect to GSD, viewing angle, and visibility ratio.**



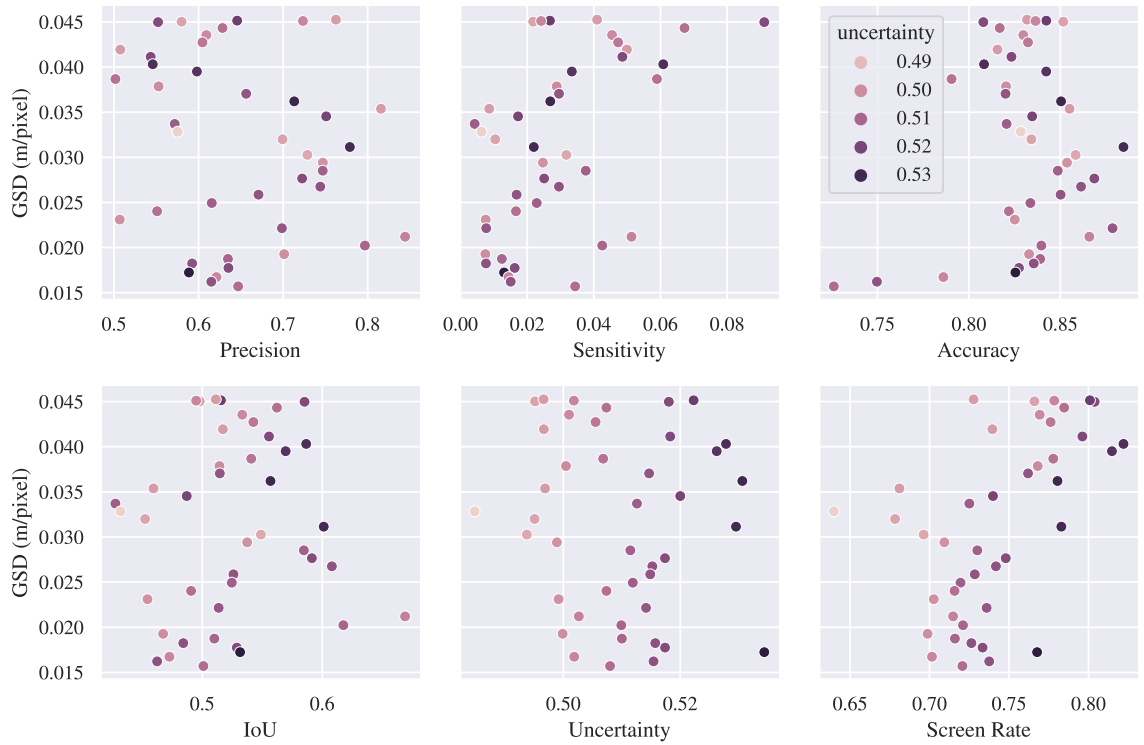
**Figure 8: Frames from the OSIRIS-Rex Tag sequence..**

**Table 2: Overall performance for the TAG experiment for slope & roughness safety.** BICNet-NKO is our Bayesian ICNet model trained on Nightingale, Kingfisher, and Osprey, and BICNet-KOS is trained on Kingfisher, Osprey, and Sandpiper. All reported values are percentages.

| METHOD     |                     | PRECISION            | SENSITIVITY          | ACCURACY             | MIoU                 |
|------------|---------------------|----------------------|----------------------|----------------------|----------------------|
| BICNET-NKO | WITHOUT UNCERTAINTY | 49.02 (50.21)        | <b>61.16 (66.53)</b> | 67.44 (67.09)        | 48.65 (48.66)        |
|            | WITH UNCERTAINTY    | 61.38 (61.66)        | 26.60 (30.80)        | 78.62 (77.09)        | <b>60.04 (59.23)</b> |
| BICNET-KOS | WITHOUT UNCERTAINTY | 49.73 (50.06)        | 42.83 (50.97)        | 67.30 (65.27)        | 45.64 (45.21)        |
|            | WITH UNCERTAINTY    | <b>62.72 (62.46)</b> | 3.06 (4.29)          | <b>83.07 (82.33)</b> | 52.96 (55.96)        |

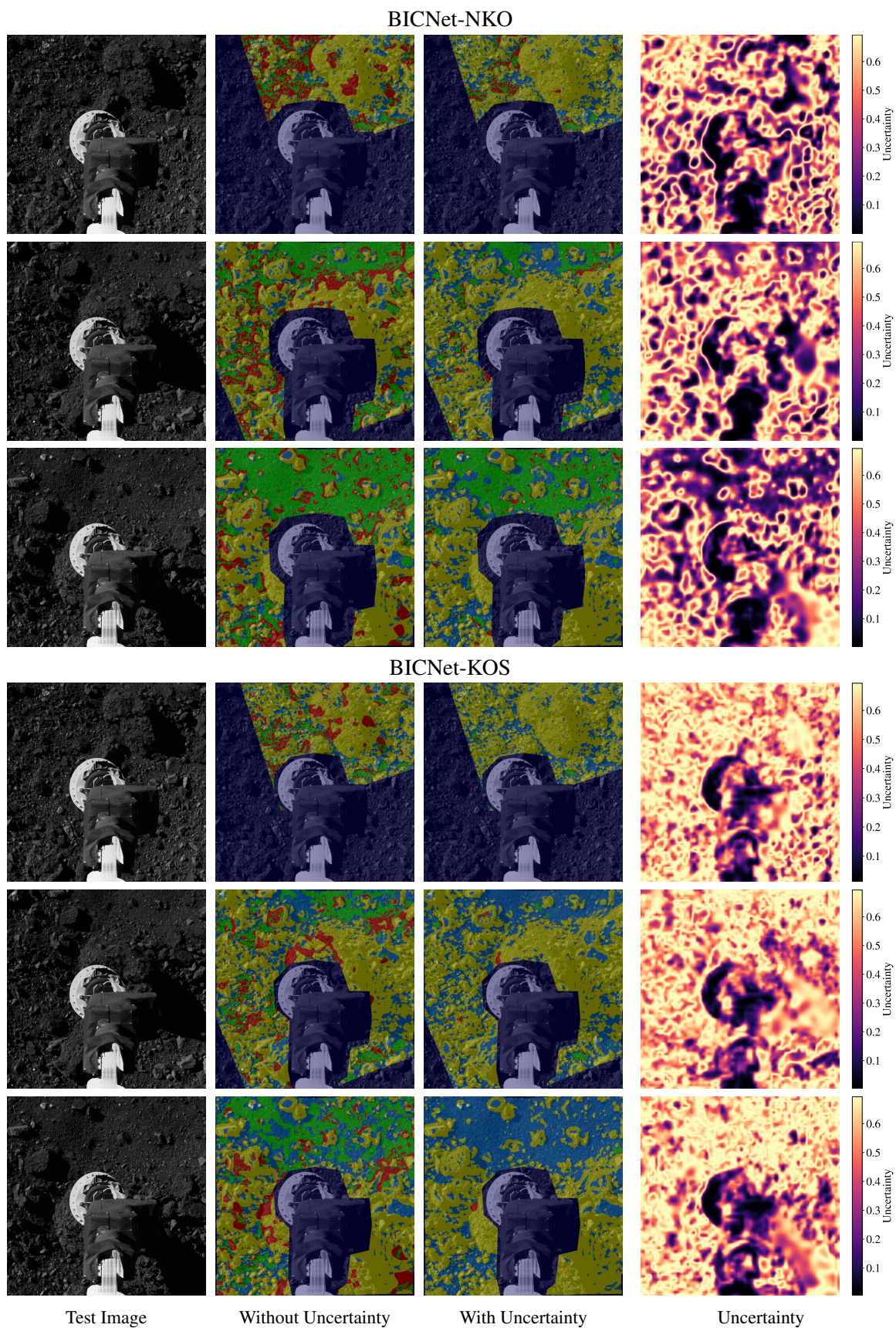


(a) BICNet-NKO



(b) BICNet-KOS

**Figure 9: Per-image metrics for the TAG experiment with respect to the GSD for slope & roughness safety.**



**Figure 10: Qualitative monocular safety mapping results for the TAG experiment.** Green, yellow, blue, and red labels represent true safe, true unsafe, false unsafe, and false safe, respectively.



## REFERENCES

- [1] K. Berry, K. M. Getzandanner, M. C. Moreau, S. M. Rieger, P. G. Antreasian, C. D. Adam, D. Wibben, J. M. Leonard, A. H. Levine, J. Geeraert, *et al.*, “Contact with Bennu! Flight Performance Versus Prediction of OSIRIS-REx “TAG” Sample Collection,” *AIAA SciTech Forum*, 2022, p. 2521.
- [2] O. Barnouin, M. Daly, E. Palmer, C. Johnson, R. Gaskell, M. Al Asad, E. Bierhaus, K. Craft, C. Ernst, R. Espiritu, H. Nair, G. Neumann, L. Nguyen, M. Nolan, E. Mazarico, M. Perry, L. Philpott, J. Roberts, R. Steele, J. Seabrook, H. Susorney, J. Weirich, and D. Lauretta, “Digital terrain mapping by the OSIRIS-REx mission,” *Planetary and Space Science*, Vol. 180, 2020, p. 104764.
- [3] E. E. Palmer, R. Gaskell, M. G. Daly, O. S. Barnouin, C. D. Adam, and D. S. Lauretta, “Practical Stereophotoclinometry for Modeling Shape and Topography on Planetary Missions,” *Planetary Science*, Vol. 3, No. 102, 2022, pp. 1–16.
- [4] R. Moghe and R. Zanetti, “A Deep learning approach to Hazard detection for Autonomous Lunar landing,” *J. of the Astronautical Sciences*, Vol. 67, No. 4, 2020, pp. 1811–1830.
- [5] K. Tomita, A. K. Skinner, and K. Ho, “Bayesian Deep Learning for Segmentation for Autonomous Safe Planetary Landing,” *J. of Spacecraft and Rockets*, 2022, <https://doi.org/10.2514/1.A35104>.
- [6] D. A. Lorenz, R. Olds, A. May, C. Mario, M. E. Perry, E. E. Palmer, and M. Daly, “Lessons learned from OSIRIS-REx autonomous navigation using natural feature tracking,” *IEEE Aerospace Conf.*, 2017, pp. 1–12.
- [7] M. Pugliatti and M. Maestrini, “Small-Body Segmentation Based on Morphological Features with a U-Shaped Network Architecture,” *J. of Spacecraft and Rockets*, 2022, pp. 1–15.
- [8] E. Caroselli, F. Belien, A. Falke, F. Curti, and R. Förstner, “Deep Learning-Based Passive Hazard Detection for Asteroid Landing in Unexplored Environment,” *AAS GN&C Conf.*, 2022, pp. 1–16.
- [9] “NASA Technology Taxonomy,” tech. rep., National Aeronautics and Space Administration (NASA), 2020.
- [10] C. D. Epp, E. A. Robertson, and T. Brady, “Autonomous landing and hazard avoidance technology (ALHAT),” *IEEE Aerospace Conf.*, IEEE, 2008, pp. 1–7.
- [11] J. M. Carson, N. Trawny, E. Robertson, V. E. Roback, D. Pierrottet, J. Devolites, J. Hart, and J. N. Estes, “Preparation and integration of ALHAT precision landing technology for Morpheus flight testing,” *AIAA SPACE Conf.*, 2014, pp. 1–16.
- [12] R. R. Sostaric, S. Pedrotty, J. M. Carson, J. N. Estes, F. Amzajerjian, A. M. Dwyer-Cianciolo, and J. B. Blair, “The SPLICE Project: Safe and Precise Landing Technology Development and Testing,” *AIAA Scitech Forum*, 2021, pp. 1–9.
- [13] E. Church, T. Bourbeau, J. Curriden, A. Deguzman, F. Jaen, H. Ma, K. Mahoney, C. Miller, B. Short, K. Waldorff, *et al.*, “Flash Lidar On-Orbit Performance at Asteroid Bennu,” *AAS Guidance, Navigation and Control (GN&C) Conference*, 2020.
- [14] J. M. Leonard, M. C. Moreau, P. G. Antreasian, K. M. Getzandanner, E. Church, C. Miller, M. G. Daly, O. S. Barnouin, and D. S. Lauretta, “Cross-Calibration of GNC and OLA LIDAR Systems Onboard OSIRIS-REx,” *AAS Guidance, Navigation and Control (GN&C) Conference*, No. 22-166, 2022.
- [15] B. Rizk, C. D. d’Aubigny, D. Golish, C. Fellows, C. Merrill, P. Smith, M. Walker, J. Hendershot, J. Hancock, S. Bailey, D. DellaGiustina, D. Lauretta, R. Tanner, M. Williams, K. Harshman, M. Fitzgibbon, W. Verts, J. Chen, T. Connors, D. Hamara, A. Dowd, A. Lowman, M. Dubin, R. Burt, M. Whiteley, M. Watson, T. McMahon, M. Ward, D. Booher, M. Read, B. Williams, M. Hunten, E. Little, T. Saltzman, D. Alfred, S. O’Dougherty, M. Walthall, K. Kenagy, S. Peterson, B. Crowther, M. Perry, C. See, S. Selznick, C. Sauve, M. Beiser, W. Black, R. Pfisterer, A. Lancaster, S. Oliver, C. Oquest, D. Crowley, C. Morgan, C. Castle, R. Dominguez, and M. Sullivan, “OCAMS: The OSIRIS-REx Camera Suite,” *Space Science Reviews*, Vol. 214, No. 26, 2018, pp. 1–55.
- [16] T. Claudet, K. Tomita, and K. Ho, “Benchmark Analysis of Semantic Segmentation Algorithms for Safe Planetary Landing Site Selection,” *IEEE Access*, Vol. 10, 2022, pp. 41766–41775.
- [17] Y. Gal and Z. Ghahramani, “Bayesian convolutional neural networks with Bernoulli approximate variational inference,” *arXiv preprint arXiv:1506.02158*, 2015.
- [18] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian approximation: Representing model uncertainty in deep learning,” *international conference on machine learning*, PMLR, 2016, pp. 1050–1059.
- [19] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, “ICNet for real-time semantic segmentation on high-resolution images,” *European Conf. on Computer Vision (ECCV)*, 2018, pp. 405–420.
- [20] J. Mukhoti and Y. Gal, “Evaluating Bayesian deep learning methods for semantic segmentation,” *Int. Conf. on Learning Representations (ICLR)*, 2018.

- [21] V. B. Alex Kendall and R. Cipolla, “Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding,” *Proceedings of the British Machine Vision Conference (BMVC)* (G. B. Tae-Kyun Kim, Stefanos Zafeiriou and K. Mikolajczyk, eds.), BMVA Press, September 2017, pp. 57.1–57.12, 10.5244/C.31.57.
- [22] T. Driver, K. Skinner, M. Dor, and P. Tsiotras, “AstroVision: Towards Autonomous Feature Detection and Description for Missions to Small Bodies Using Deep Learning,” *preprint*, 2022. arXiv:2208.02053.
- [23] T. Ivanov, A. Huertas, and J. M. Carson, “Probabilistic hazard detection for autonomous safe landing,” *AIAA Guidance, Navigation, and Control (GNC) Conf.*, 2013, pp. 1–13.
- [24] T.-M. Ho, V. Baturkin, C. Grimm, J. T. Grundmann, C. Hobbie, E. Ksenik, C. Lange, K. Sasaki, M. Schlotterer, M. Talapina, *et al.*, “MASCOT—the mobile asteroid surface scout onboard the Hayabusa2 mission,” *Space Science Reviews*, Vol. 208, No. 1, 2017, pp. 339–374.