

AI based Sensor Fusion for Robust Feature Extraction for Autonomous Navigation of Spacecraft Missions to Asteroids with application to ESA Hera mission

Iain Hall*, Jinglang Feng[†], Hao Peng[‡], and Massimiliano Vasile[§]

Missions to visit asteroids depend on autonomous navigation to carry out operations. The estimation of the relative position and attitude (pose) of the spacecraft to the target asteroid is a key step but can be challenging in the poor illumination conditions which can occur for asteroids. We explore how sensor fusion using deep learning can allow for robust estimation of pose. Visible images, thermal images, and different levels of fusing visible images and thermal images are tested using synthetic images of Didymos, the target of ESA's Hera mission. Pose is estimated through the extraction of centroid and keypoint features using Convolutional Neural Networks which are also used for sensor fusion. We find that sensor fusion has little effect in centroid estimation when compared to using just visible or thermal images. Sensor fusion improves keypoint estimation over using a single image type with feature fusion outperforming source fusion. This leads to a more accurate estimate of relative pose.

I. Nomenclature

θ	=	Sun phase angle
e_p	=	Pixel error
r	=	Spacecraft asteroid range
f	=	Focal length
μ	=	Pixel pitch
e_d	=	Distance error
e_{pos}	=	Position error
e_q	=	Orientation (quaternion) error
\mathbf{P}^{Pred}	=	Predicted keypoint position
\mathbf{P}^{GT}	=	Ground Truth keypoint position

II. Introduction

DEEP space missions to asteroids depend significantly on autonomous navigation systems to perform proximity operations around the target asteroids [1], which requires the perception of the target and its relative pose (position and attitude). The dynamical environment around asteroids is challenging due to poor illumination conditions and large uncertainties in the target's shape and motion. Developing robust pose estimation methods is essential to a mission's success.

Visual relative navigation has been a key component of recent missions to visit asteroids. Hayabusa-1, Hayabusa-2, and OSIRIS-REx all operated vision based navigation allowing them to autonomously operate[2–4]. This allows the missions to carry out operations that wouldn't be possible with the long delays when communicating with Earth.

All these missions used monocular visual navigation, but the Hera mission - ESA's mission to visit Didymos and Dimorphos, has a range of sensors which can be used for navigation. The sensors include a visual camera (AFC), a thermal camera (TIRI), a multi-spectral camera, and a laser altimeter (PALT). The GNC system on the mission will use sensor fusion to take advantage of these different sensors [5]. Sensor fusion will allow for visual images to be combined

*PhD. Student, Mechanical and Aerospace Engineering, The University of Strathclyde. Corresponding author, email: iain.hall@strath.ac.uk

[†]Associate Professor, Mechanical and Aerospace Engineering, The University of Strathclyde.

[‡]Assistant Professor, Aerospace Engineering Department, Embry-Riddle Aeronautical University

[§]Professor, Mechanical and Aerospace Engineering, The University of Strathclyde.

with thermal images to reduce the effect of sun phase angle and illumination conditions on the navigation. The fusion of visual images with altimeter data will allow for more accurate estimation of range (perpendicular to the camera) which is a weakness of visual cameras.

Sensor data fusion can also be done at different levels of data processing: source fusion (low-level) – fusing the data directly output by the sensors, feature fusion (mid-level) – fusing features identified partway through the data processing, and decision fusion (high-level) – fusing the end results of the data processing from different sensors [6]. Hera will use Decision level fusion for its GNC.

For navigation of Hera in the 30-8.5km range from Didymos, centroid based position estimation will be used. The centroid will be extracted using a Lambertian sphere correlation [7]. For navigation at closer than 8.5km features must be extracted for relative pose estimation. This will be done using Kanade-Lucas-Tomasi (KLT) feature tracker [8]. These are both conventional image processing techniques and do not take advantage of Deep learning (DL).

DL has been used for a range of different image processing applications [9–11], with Convolutional Neural Network (CNN) architectures being common and very successful. DL methods have been explored for other pose estimation and pose estimation related tasks. In autonomous ground vehicles DL feature extraction methods are used to carry out simultaneous localisation and mapping (SLAM)[12]. For artificial satellite pose estimation DL methods have been explored in the literature [13], with the satellite pose estimation challenge (SPEC) competitions in 2019 [14] and 2021 [15] pushing its development. In the literature for artificial satellite pose estimation the extraction of known keypoints and estimation of pose using Perspective-n-Point (PnP) solvers have been the most successful in the SPEC competitions. The extraction of keypoints is done using DL methods while the PnP problem is solved by non-DL methods. The PnP problem is the challenge of estimating a camera’s relative pose from a set of points in the camera image, with knowledge of the relative 3D physical positions of the points to each other [16].

DL’s success in these tasks has also attracted research in asteroid relative navigation. In Kaluthantrige et al.[17] the HRNet architecture is used to extract the centroid, sub solar point, and keypoints on the edge of the Didymos asteroid. These are used for estimating the relative position of the asteroid. In Pugliatti et al.[18] a range of DL architectures are developed, primarily CNNs for extraction of features and semantic segmentation of asteroids. These works both focus on how DL can be applied to the navigation of ESA’s Hera mission. Both rely on synthetic images allowing for labels to be easily defined, in Kaluthantrige et al. ESA’s PANGU[19] tool is used, while Pugliatti et al. develop a tool, CORTO [20], for synthetic image generation based on Blender[21].

DL slam feature extractors have been explored in Knuuttila et al. [22] and Driver et al. [23]. In Knuuttila et al. they adapt the R2D2 feature extractor architecture [24] training it on a real image dataset made using images from NASA’s Planetary Data System [25] and ESA’s Planetary Science Archive [26]. Synthetic image augmentation is used to create image pairs for training feature extraction and tracking. In Driver et al. they train the SuperPoint[27], R2D2[24], and ASLFeat[28] architectures on a dataset of real asteroid images. The dataset is also made using images from NASA’s Planetary Data System, using stereophotogramy to develop pixel correspondences between images. In Letizia et al. [29] thermal and visible images for asteroid SLAM are explored. They use synthetically generated images and use ORB [30] and SURF [31] feature extractors. The visible images are obtained using PANGU [19] while they develop thermal models for producing thermal images.

Through this work we will explore how source fusion and feature fusion can be combined with DL for asteroid navigation. The developed architectures will be applied to synthetically generated images of Didymos, the target of ESA’s Hera Mission.

This work aims to evaluate how sensor data fusion can improve the robustness of pose estimation, which allows for the shortcoming of an individual sensor to be addressed by combining it with complimentary data from other sensors. The fusion of visible camera data and thermal camera data is explored in this study as they are common sensors onboard spacecraft around asteroids. This allows the limitations of visible images in shadow to be addressed by thermal images.

The key contribution of this work is to assess how an existing CNN architecture can be used to fuse visible and thermal images for extracting features from asteroids. The different fusion architectures are then assessed for accuracy and robustness of feature extraction. The features being extracted are the centroid and a set of known keypoints which can be used in a PnP solver.

This paper is arranged as follows, Section III describes the method. Section IV shows the results for the models on the test dataset. Section V discusses the results, Section VI contains the conclusion, and Section VII outlines future work.

III. Method

This section outlines four key parts of the method: the methods of fusing sensor data, the CNN architecture used for the identification of features, the PnP method, and the synthetic image generation pipeline.

A. Sensor Data Fusion

In this work we explore two types of sensor fusion, source level sensor fusion and feature level sensor fusion. For source level fusion the visible (1 channel) and thermal (1 channel) images are input as a single image with 2 channels made from concatenating the two individual image tensors. This 2 channel image is used as an input to a CNN which is trained to extract either centroid or keypoint features. This approach minimizes computational cost and network complexity for a given network size. This fusion architecture is shown in Figure 1. The input images must be modified because the cameras have different focal lengths, sensors sizes, and resolutions. The thermal image is modified so that the angular resolution is the same as for the visible image, so that the pixels in the thermal image and visible image correspond to the same physical location on the asteroid.

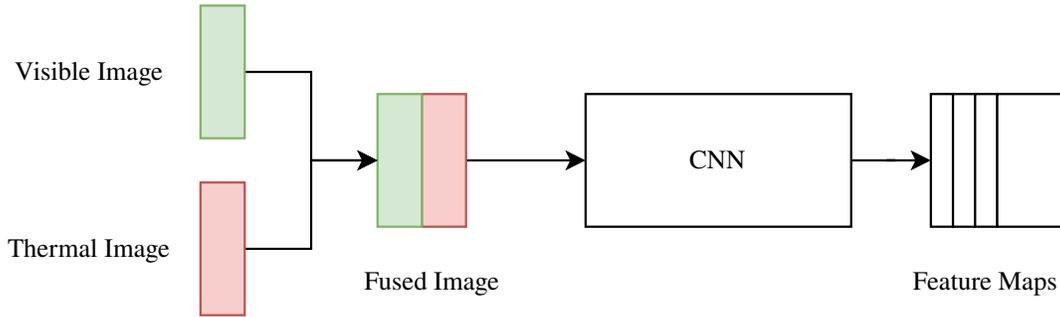


Fig. 1 Source level sensor fusion architecture.

In feature level fusion the data from the sensors are input individually into separate CNN's, processing them into feature space. These features can then be fused through concatenation before further processing through a joint CNN. This allows sensor specific features to be identified individually and is slightly more computationally complex than source fusion. This architecture is shown in Figure 2. The thermal image is also modified before entry into the CNN so its angular resolution matches the visible image. This is done so that the features extracted from it correspond in location to features extracted from the visible image. Both sensor fusion architectures are established architectures of fusing visible and infra-red images in other fields [32].

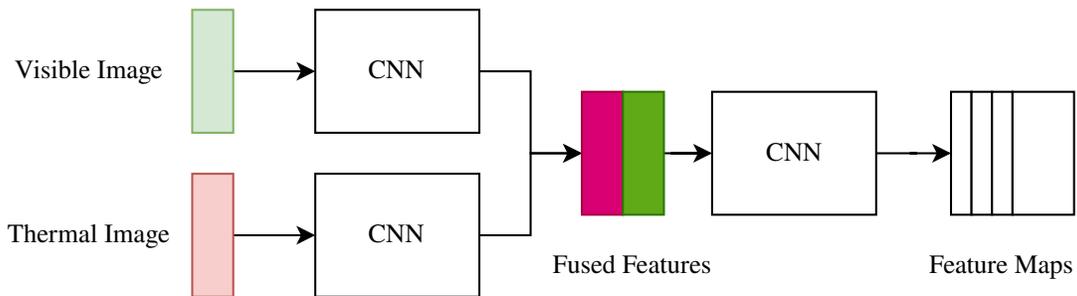


Fig. 2 Feature level sensor fusion architecture.

B. Deep Learning Model

The CNN architecture used is the High Resolution Network (HRNet) [33] which was originally developed for human pose estimation. It has also been successfully used for satellite pose estimation and for asteroid centroid and edge detection[17, 34]. The base HRNet architecture is shown in Figure 3, the parallel structure of feature extraction at different resolutions allows it to extract features at both large and small scales and recombine them. The input to the network is an image and it outputs a set of feature maps representing the estimated position of each keypoint. Each

feature map only represents one keypoint and the location of the key point is taken from the peak pixel location. To improve the training a Gaussian distribution around the key point is actually used as the label. The network is modified to match the fusion methods described, with the network being split in half for feature fusion.

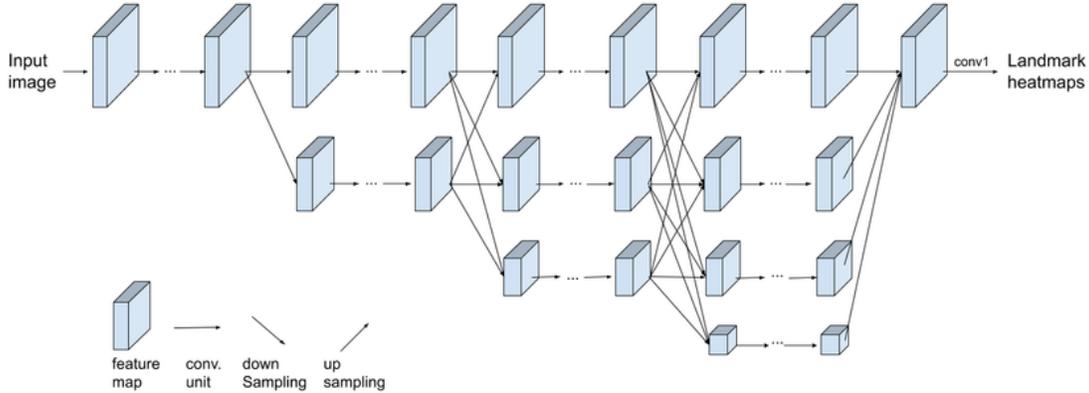


Fig. 3 HRNet architecture[34].

A set of 4 models were produced for centroid estimation using the two sensor fusion architectures described and the base HRNet architecture. The base HRNet architecture was used to do visible image to centroid and thermal image to centroid estimation. The Fusion architectures were then used to make networks for source fusion to centroid and feature fusion to centroid respectively. A further 4 models were produced similarly but for estimating the full 16 keypoints instead of just the centroid. This led to a total set of 8 models to be trained and tested. The loss function used for training is the Pixel-wise Mean Square Error (MSE) between the output feature maps and the labels. The error refers to the difference between the activation of a specific pixel in the output and the desired output activation as specified by the label.

The parameters used in training are shown in Table 1, and they were trained on an NVIDIA L4 GPU [35].

Table 1 Model training parameters

Training Parameter	Value
Learning rate	10^{-3}
Number of Epochs	30
Batch Size	64
Input image Dimensions	256x256
Output feature map Dimensions	64x64

C. Perspective-n-Point

The PnP problem is solving the pose of a camera relative to a reference frame with knowledge of a set of keypoints. In Figure 4 the keypoints identified by the camera are a 2D projection of the 3D points on the asteroid, and their 2D projection is a result of the camera's pose and intrinsic properties (focal length and resolution). This means that it is possible to solve the camera's pose from the keypoints if the position of the keypoints is known in the asteroid reference frame. The 16 keypoints are used to estimate the pose of the spacecraft using the MATLAB implementation [36] of a PnP solver by Gao et al. [16].

D. Synthetic Data Generation

For training and testing DL models large datasets are needed. Due to the difficulty in labeling asteroid datasets containing visible and thermal images synthetically generated images are used. This is done using Blender, an artistic rendering and modelling tool, taking advantage of the physically based Cycles render engine [21]. Blender's open-source

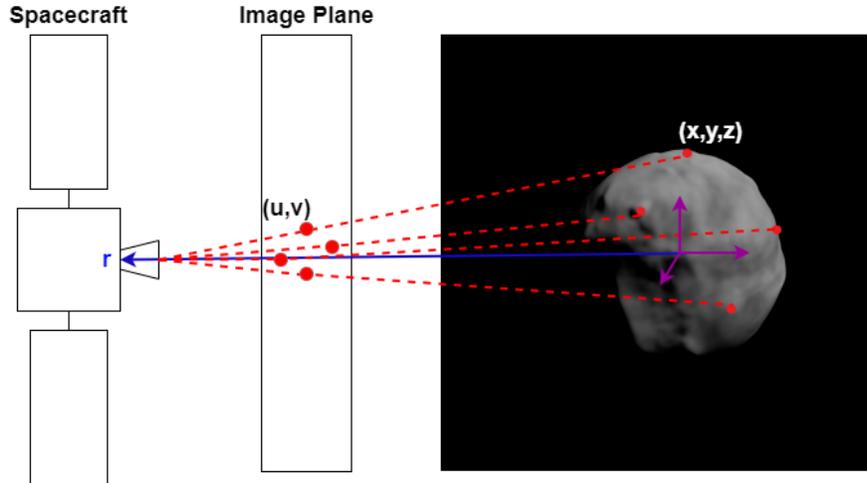


Fig. 4 Perspective-n-Point for relative pose estimation to Didymos

nature and Python API has led to its use for synthetic training data generation for other DL based satellite pose estimation methods and for synthetic image generation of asteroids[20, 37].

A pipeline was developed for generating the synthetic image dataset and labels, which is shown in Figure 5. There were 3 main tasks - generation of relative positions and labels, thermal modelling, and rendering. For the generation of relative positions and labels a set of random satellite position vectors within the range, of 8.5km to 30km are generated to provide a large number of different views of the asteroid. The positions of the keypoints in the camera image were also extracted at this point to provide labels. The selection of 8.5km to 30km was done to represent Early Characterisation Phase (ECP) and Detailed Characterisation Phase (DCP) in the Hera mission with a range of sun phase angles. The range, r , and sun phase angle, θ , are shown in Figure 6.

Thermal modelling is carried using the MATLAB PDE toolkit [38]. The temperature model of the asteroid is run for 1 month with the asteroid rotating. This was to provide a more representative quasi steady state temperature, as using a purely steady state model would result in the sun facing side of the asteroid being much hotter than the dark side, which is not accurate due to Didymos's rotation. The asteroids thermal parameters are shown in Table 2, with 3 different versions being run at different distances from the sun: 1AU, 1.5AU, 2AU. These distances were selected to reflect the perihelion, a mid-point, and the aphelion in Didymos's orbit. A temperature map was then produced by interpolating the temperature from the model over each surface of the asteroid's blender model. This could then be applied as a texture in Blender to be used when rendering thermal images.

Table 2 Thermal Model properties [39]

Property	Value	Unit
Didymos Density	2550	kg/m^3
Surface emissivity (thermal)	0.9	n/a
Albedo (visible)	0.07	n/a
Thermal inertia	320	$Jm^{-2}K^{-1}s^{-1/2}$
Rotation period	2.26	hours

The relative positions and the temperature maps were then used in the rendering step. An image pair containing a visible image and thermal image was generated for each relative position with the camera properties of the thermal and visible camera shown in Table 3. 5,000 images were rendered for each temperature map, for a total of 15,000 images. For each of these images the centroid of Didymos was in the centre of the frame, which lacks diversity and encourages the model to learn to identify the middle pixel rather than the centre of Didymos. To address this a second image pair was generated for each rendered image pair, in which the original images undergo an affine transform rotating and translating them. The affine transform was a random rotation and a random translation, within limits that ensure all of

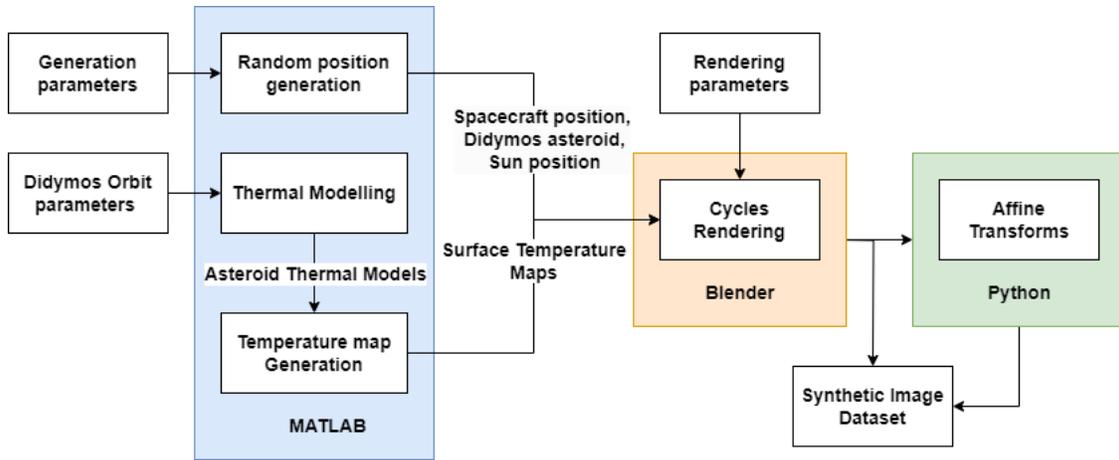


Fig. 5 Pipeline for producing synthetic data for training and testing of the network.

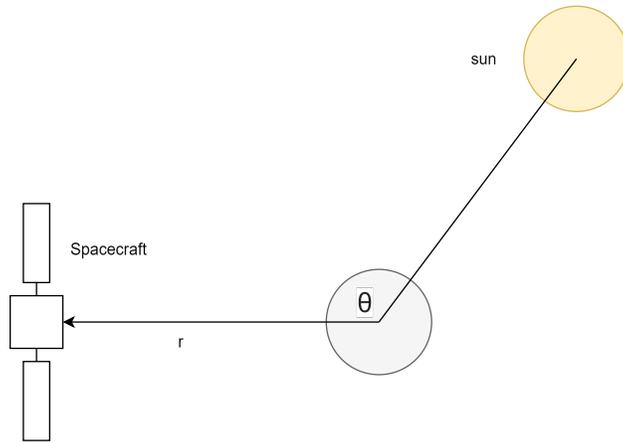


Fig. 6 Diagram showing range (r) and sun phase angle (θ).

Didymos remained in frame. This extended the dataset to 30,000 images. A final step was undertaken to make sure the thermal and visible images had the same angular pixel pitch ($^{\circ}/pixel$) so that pixels mapped to the same physical points. This was done by cropping and resizing the thermal image to match the angular pitch of the visible image. The dataset of 30,000 image pairs were split into a training set of 21,000 pairs and a testing set of 9,000 pairs. An example image pair with 2 of 16 labels is shown in Figure 7. For each image pair two sets of labels were produced for training the CNNs. One was a single feature map per image pair which contained a Gaussian around the centroid of Didymos. This was used for training the identification of the centroid. The second set was a group of 16 feature maps per image pair where each feature map represents a single keypoint, manually chosen as salient features on the surface of Didymos. These can be used to estimate the relative pose by solving the PnP problem.

Table 3 Thermal camera and Visible camera properties

Camera Property	Visible	Thermal	Unit
Focal Length	106	100	mm
Sensor Width	10.2	17.5	mm
Image Dimensions	1020x1020	768x768	pixels

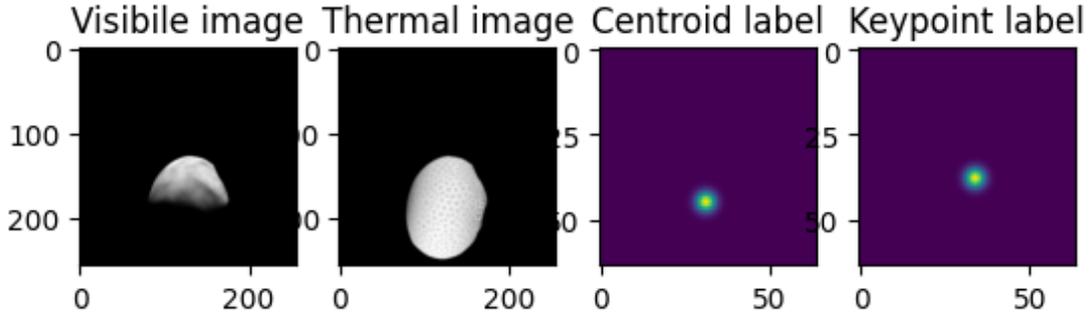


Fig. 7 Example image pair with the centroid label and label for Keypoint 2.

IV. Results

The results for the different models performances on the 9,000 image pair test set is presented in this section. A summary of the results are shown in Table 4 where the performance of each model is measured in the MSE, the pixel error, and the distance error. The pixel error, e_p , is the distance between the peak pixel of the output feature map and the position of the ground truth keypoint in that feature map. The pixel error is shown in Equation (2) where each Position vector, \mathbf{P} , is a 2 element vector representing its horizontal, u , and vertical, v , positions in the image. The position vectors are the ground truth position vector, \mathbf{P}^{GT} , and the predicted position vector, \mathbf{P}^{Pred} .

$$\mathbf{P}_{pixel} = (u, v) \quad (1)$$

$$e_p = \left| \mathbf{P}_{pixel}^{Pred} - \mathbf{P}_{pixel}^{GT} \right| \quad (2)$$

This can be converted into a distance at the range of the asteroid using the ground truth range, r . The conversion from pixels to metres at the asteroid is shown in Equation (3), which is then used to find the distance error, e_d , defined in Equation (4).

$$\mathbf{P}_{asteroid} = \mu \mathbf{P}_{pixel} \times \frac{r}{f} \quad (3)$$

$$e_d = \frac{\left| \mathbf{P}_{asteroid}^{Pred} - \mathbf{P}_{asteroid}^{GT} \right|}{r} \quad (4)$$

An example result for the extraction of a keypoint is shown in Figure 8 where an image pair is used to predict keypoints. The predicted labels are shown on the right with red crosses at the point the keypoint is identified, the predicted keypoint from the visible only model is shown on the top right while the predicted keypoint using the feature fusion model is shown in the bottom right. It can be seen that while the prediction by the visible only model appears cleaner, the prediction by the feature fusion model is actually closer to the ground truth. Similar results are found over the full test set, and this is reflected in the average performances.

A. Centroid Estimation

In Figure 9 the distance error, e_d , is shown against sun phase angle (a) and range (b). For sun phase angle e_d is averaged over 5° bins where it can be seen that the centroid error does not appear to be influenced by sun phase angle. For range e_d is averaged over $1km$ bins and it can be seen that the models all perform very similarly, with the distance error showing no clear trends with range. This similarity in performance can also be seen in their overall average error shown in Table 4.

B. Key point Estimation

The keypoints extracted from the feature maps are compared to the ground truth locations to provide e_d for the predicted keypoint position. The mean e_d over the 16 keypoints is shown in Figure 10 plotted over the sun phase angle, with e_d averaged over 5° bins. The keypoint e_d over range is averaged over $1km$ bins. There is significantly more

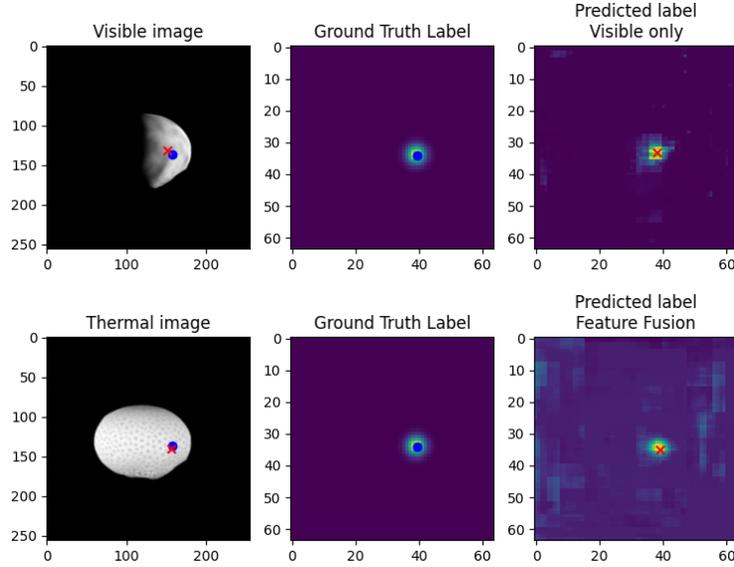


Fig. 8 Example result of keypoint prediction with Visible only and Feature fusion models.

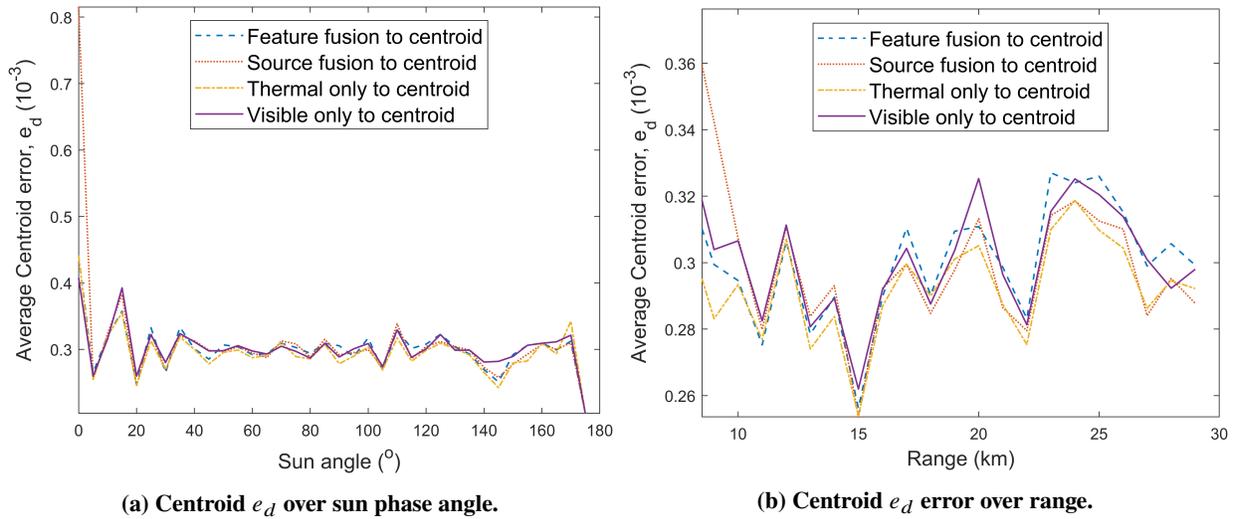


Fig. 9 Centroid distance error, e_d

difference in the performance of the different models. Feature fusion performs best followed by Source fusion, then Thermal only, with Visible only performing the worst by a significant margin. The relative performances can also be seen in Table 4. There are clear trends with sun phase angle and range in keypoint estimation. It can be seen in Figure 10 that the visible model performs worse at higher sun phase angles while the other models don't vary with sun phase angles. For all four models e_d decreases as the range increases.

C. Pose Results

The relative pose of the spacecraft to the asteroid is found from the keypoints by solving the PnP problem. A sample result showing the ground truth orientation and the predicted orientation is shown in Figure 11. The quaternion error, e_q , is used to define the error in the orientation estimation aspect of pose estimation. It is defined in Equation (5) where \mathbf{q}_{Pred} is the predicted rotation quaternion of the satellite to Didymos and \mathbf{q}_{GT} is the ground truth rotation quaternion. The position error, e_{pos} , is defined in Equation (6) and is the distance between the ground truth satellite position, \mathbf{r}_{GT} ,

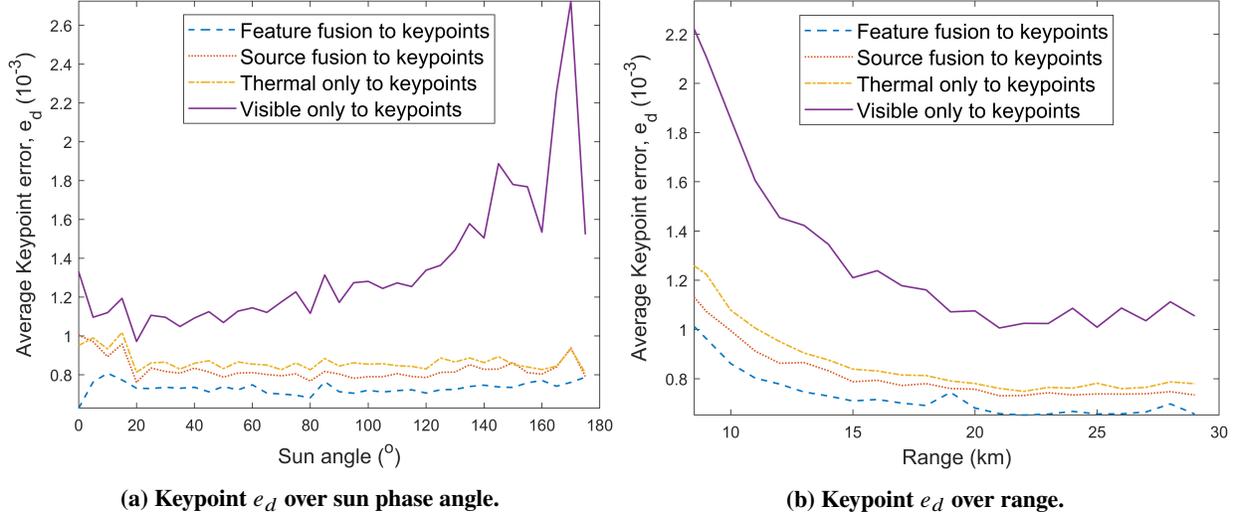


Fig. 10 Keypoint distance error, e_d

Table 4 Mean errors for pixel position (e_p), feature map MSE, and distance error e_d with their standard deviations (Sd.)

Model	e_p	Sd. (e_p)	MSE (10^{-5})	Sd. (MSE)	e_d (10^{-3})	Sd. (e_d)
Visible only to centroid	0.20	0.24	0.95	3.17	0.300	0.354
Thermal only to centroid	0.19	0.23	0.92	2.88	0.293	0.338
Source fusion to centroid	0.20	0.23	0.68	2.88	0.299	0.352
Feature fusion to centroid	0.20	0.23	1.16	3.38	0.300	0.352
Visible only to keypoints	0.84	0.75	15.1	18.0	1.27	1.13
Thermal only to keypoints	0.57	0.20	8.09	5.80	0.858	0.307
Source fusion to keypoints	0.54	0.18	7.10	2.88	0.809	0.277
Feature fusion to keypoints	0.48	0.22	5.25	4.36	0.724	0.332

and the predicted satellite position, \mathbf{r}_{Pred} , normalised to the ground truth range, r .

$$e_q = 2\arccos(\mathbf{q}_{Pred} \cdot \mathbf{q}_{GT}) \quad (5)$$

$$e_{pos} = \frac{|\mathbf{r}_{Pred} - \mathbf{r}_{GT}|}{r} \quad (6)$$

The results for e_q over sun phase angle and range are shown in Figure 12. The feature fusion, source fusion, and thermal only models show no variation with sun angle. The visible only model increases in error at sun phase angles over 60° . It can also be seen that there is no clear trend with range for all 4 models. The results for e_{pos} over sun angle and range are shown in Figure 13. A clear increase in e_{pos} with sun angle for the visible only network while the other networks show no clear trend. As range varies we see relatively little influence on e_{pos} for the visible network. For the thermal network the error slightly increases with range, while e_{pos} for the source and feature fusion significantly increases with range, with a higher error than the visible and thermal networks at ranges greater than 26km. The mean performance of the models for pose estimation are shown in Table 5.

V. Discussion

The results in centroid estimation show that there is almost no effect on error as a result of sun phase angle, range, or fusion method. The models are able to perform very effectively on the task of estimating the centroid keypoint with an

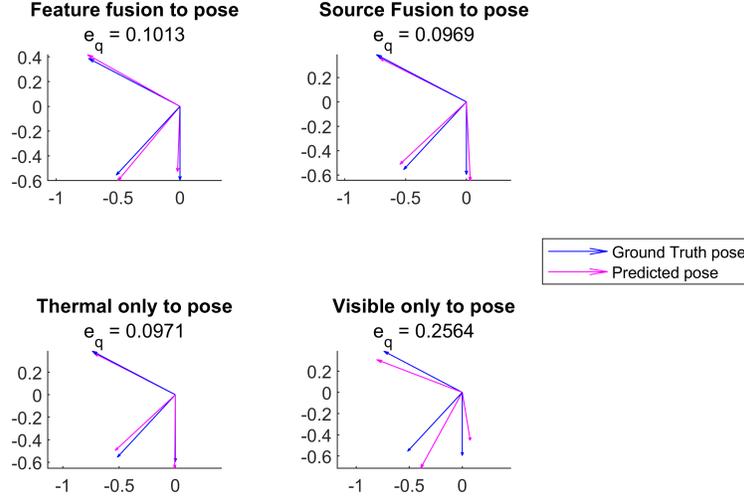


Fig. 11 Orientation example result

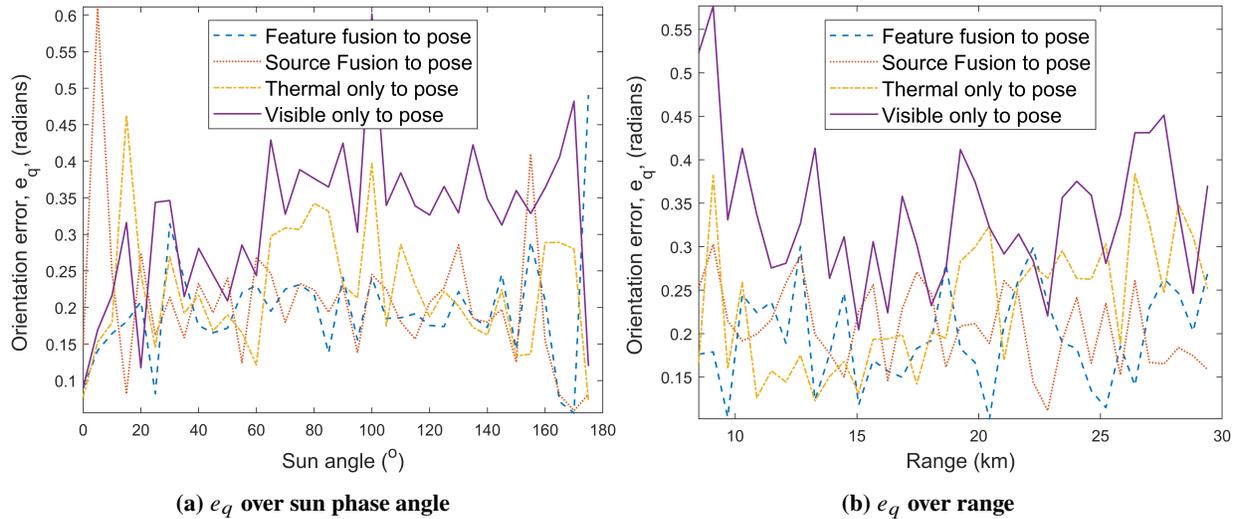


Fig. 12 Orientation errors (e_q) for the pose estimated by solving the PnP problem

Table 5 Mean Errors for position and orientation during pose estimation with their standard deviations (sd.)

Model	e_{pos}	Sd. (e_{pos})	e_q	Sd. (e_q)
Visible only to pose	0.172	0.278	0.335	0.951
Thermal only to pose	0.124	0.229	0.229	0.798
Source fusion to pose	0.114	0.207	0.207	0.757
Feature fusion to pose	0.104	0.206	0.196	0.759

average error 0.2 pixels. This means that the use of fusion methods is unnecessary for this task as the model is able to achieve high performance without this. The average error of 0.2 also likely arises from the difference in the nature of the ground truth keypoints and the predicted keypoints. The predicted keypoints are always limited to integer pixel positions from the feature map while the ground truth keypoints are a floating point values based on the camera properties. To try and reduce the error resulting from this a few strategies could be pursued.

If the network were to downsize the image less the difference between the predicted keypoint and the output pixel keypoint would be smaller. The models currently downsize from 256 to 64 pixels, but if they instead go from 256 to

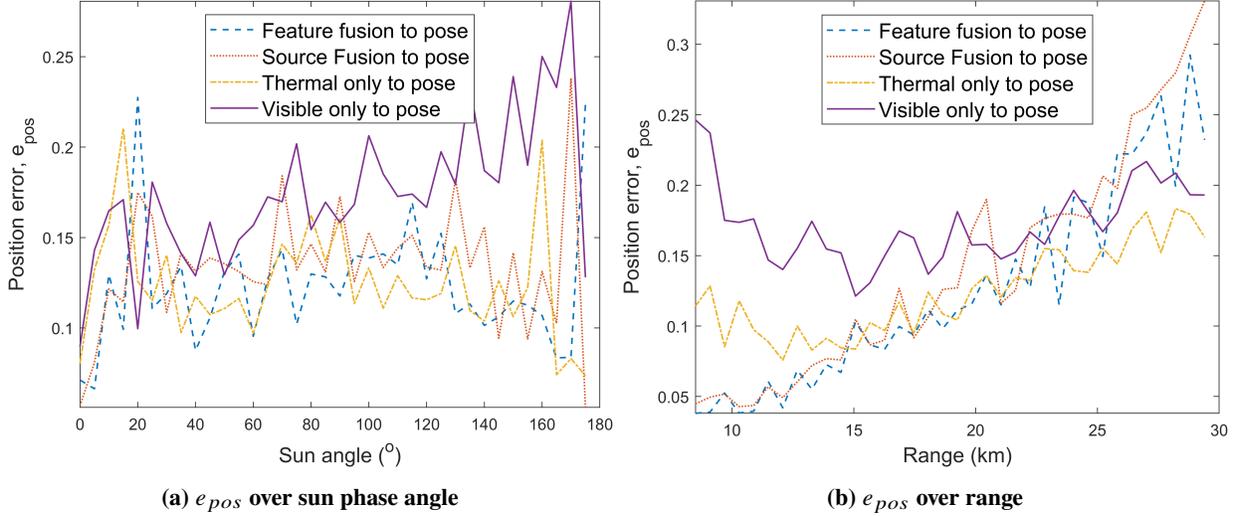


Fig. 13 Position errors (e_{pos}) for the pose estimated by solving the PnP problem

256 pixels this would reduce the error to a quarter of its current value. This would however lead to a 16 times larger memory requirement making it impractical. Instead the network could be modified to have up sampling layers at the final stages to take the image back to 256 by 256 pixels without having to process it at that size for the entire network. This method would lose information during the downsizing and up-sampling which would likely make it less accurate than not downsizing at all. A final potential method would be to use another method for selecting the centroid/keypoint such as the weighted mean of the feature map. This could allow the network to estimate the position of keypoints and more accurately, reducing the final pose error.

From the results for the full 16 keypoint estimation we can see that the sun phase angle has a significant effect on the estimation of the keypoint positions for the visible only model, with increasing sun phase angle leading to a higher error in keypoint prediction. This is because as the sun phase angle increases a greater portion of Didymos will be in shadow and their will be less asteroid visible for features to be extracted from. This does not effect the thermal or fusion methods as the thermal image is not significantly affected by the sun phase angle. On a slower rotating asteroid their may be an effect as the dark side would be able to cool down, reducing the quality of the thermal images at high sun phase angles.

For distance error with range we see a similar trend for all four keypoint extraction models. As the range increases all the models tend to perform better, with feature fusion performing best across all distances. This trend is only true for shorter distances with the distance error not significantly varying at ranges greater than 20km. The absolute distance between the keypoint ground truth and prediction will be increasing, but not at the same rate as the range does, which leads to a reduction in e_d .

The relative performance gap between visible only model for keypoint estimation and the other models shows that the use of thermal images, not data fusion, has a more significant effect on performance. This would indicate that using thermal data is more important, however this is only tested against synthetic data. It is of note that the model of Didymos used has relatively few surface features compared to real images of asteroids while the thermal image appears to have a higher level of relative detail. It would be useful for understanding the effect of data fusion to try and vary the quality of the different image types. This could be done by using a more detailed visible model and less detailed thermal model and assess the effect on the different models.

The improvement in performance by the feature fusion network compared to the source fusion network could be due to the feature fusions networks ability to learn data type specific features in the early stages instead of having to learn to extract features which are common to both image types. It could also be explained simply by the larger number of parameters in the feature fusion model, and further investigation would be necessary to determine which was the more important factor.

The large variation in the models performance on different image pairs indicated by the large standard deviations indicates the possibility of an unaccounted for factor effecting them. Because there is no indication if the model is confident in a predicted keypoint position and has large variation in performance, there would be insufficient trust in the system to use it in GNC.

The pose results show that the performance on keypoint extraction doesn't exactly indicate the performance on pose estimation. Feature fusion still achieves the lowest errors for both position and orientation estimation while visible only still performs worst. However, the size of the standard deviation relative to the average pose error is larger indicating a significant sensitivity to errors in the keypoint estimation. The trend of improving performance with range is also not clear in orientation error, while position error actually increases at higher ranges for all models but visible. The nature of using a PnP solver requires known keypoints, for this to be used in a mission an observation period at longer distances before pose estimation is necessary would be required to allow for the keypoints to be identified and trained on.

VI. Conclusion

In this work we explored how thermal images, visible images, and fusion of those images can be used to extract features for estimating satellite pose using DL methods. We developed a dataset for training and testing DL models, use the HRNet architecture to carry out data fusion and tested the effect of data fusion using our dataset. In the results we show that using only visible images results in poor performance at high sun phase angles, with thermal images allowing for robustness to illumination conditions. We also found that using feature fusion outperformed other models, demonstrating the advantage it can provide for robust and accurate feature extraction. We used the extracted features to estimate the relative pose of the spacecraft by using a PnP solver, with features extracted by fusion methods performing best.

VII. Future work

The high standard deviation in the pose estimation shows that the method is highly sensitive to some unknown factors, so further analysis should be done to try and identify these factors. Further experimentation with different DL architectures for the models should be done to assess the effect of output heat map size, network size, and different feature extraction methods on the pose estimation. Similarly the models should be tested against additional datasets which provide a more varied quality of visible and thermal images to assess their effect. Future work should explore how the models developed can be validated against real data. This could be done using publicly available image data from the Hayabusa-2 mission which operated both visible and thermal cameras. The challenge for this would be in the generation of labels for the data.

Acknowledgments

The authors like to acknowledge the Strathclyde Doctoral Training Centre for AI based multi-sensor data fusion for space applications for co-funding this work. The authors would also like to acknowledge the ESA Open Space Innovation Platform for co-funding this work.

References

- [1] Johnson, A. E., Cheng, Y., and Matthies, L. H., "Machine vision for autonomous small body navigation," *IEEE Aerospace Conference Proceedings*, Vol. 7, 2000, p. 661 – 671. <https://doi.org/10.1109/AERO.2000.879333>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0034432456&doi=10.1109%2fAERO.2000.879333&partnerID=40&md5=efcc5b779df30c9cf59f9ecc8ac7259a>, cited by: 66.
- [2] Hashimoto, T., Kubota, T., Kawaguchi, J., Uo, M., Shirakawa, K., Kominato, T., and Morita, H., "Vision-based guidance, navigation, and control of Hayabusa spacecraft - Lessons learned from real operation -," *IFAC Proceedings Volumes*, Vol. 43, No. 15, 2010, pp. 259–264. <https://doi.org/https://doi.org/10.3182/20100906-5-JP-2022.00045>, URL <https://www.sciencedirect.com/science/article/pii/S1474667015318504>, 18th IFAC Symposium on Automatic Control in Aerospace.
- [3] ONO, G., TERUI, F., Ogawa, N., MIMASU, Y., Yoshikawa, K., YASUDA, S., MATSUSHIMA, K., Takei, Y., SAIKI, T., and Tsuda, Y., "GNC Design and Evaluation of Hayabusa2 Descent Operations," *TRANSACTIONS OF THE JAPAN SOCIETY FOR AERONAUTICAL AND SPACE SCIENCES, AEROSPACE TECHNOLOGY JAPAN*, Vol. 19, 2021, pp. 259–265. <https://doi.org/10.2322/tastj.19.259>.
- [4] Lorenz, D. A., Olds, R., May, A., Mario, C., Perry, M. E., Palmer, E. E., and Daly, M., "Lessons learned from OSIRIS-REx autonomous navigation using natural feature tracking," *2017 IEEE Aerospace Conference*, 2017, pp. 1–12. <https://doi.org/10.1109/AERO.2017.7943684>.

- [5] Pellacani, A., Graziano, M., Fittock, M., Gil, J., and Carnelli, I., “HERA vision based GNC and autonomy,” *Proceedings of the 8th European Conference for Aeronautics and Space Sciences. Madrid, Spain, 1-4 July 2019*, 2019. <https://doi.org/10.13009/EUCASS2019-39>.
- [6] Castanedo, F., “A review of data fusion techniques,” *The Scientific World Journal*, Vol. 2013, 2013. <https://doi.org/10.1155/2013/704504>, publisher: Hindawi Publishing Corporation.
- [7] Buonagura, C., Pugliatti, M., and Topputo, F., “Image Processing Robustness Assessment of Small-Body Shapes,” *The Journal of the Astronautical Sciences*, Vol. 69, 2022. <https://doi.org/10.1007/s40295-022-00348-6>.
- [8] Tomasi, C., “Detection and Tracking of Point Features,” 1991. URL <https://api.semanticscholar.org/CorpusID:238434334>.
- [9] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., and Sánchez, C. I., “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, Vol. 42, 2017, p. 60 – 88. <https://doi.org/10.1016/j.media.2017.07.005>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85026529300&doi=10.1016%2fj.media.2017.07.005&partnerID=40&md5=7a8f8ab63ff9d2cdaab2acbba2e9c2f>, cited by: 8494; All Open Access, Green Open Access.
- [10] Kamilaris, A., and Prenafeta-Boldú, F. X., “Deep learning in agriculture: A survey,” *Computers and Electronics in Agriculture*, Vol. 147, 2018, p. 70 – 90. <https://doi.org/10.1016/j.compag.2018.02.016>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85042262881&doi=10.1016%2fj.compag.2018.02.016&partnerID=40&md5=d26168382cc20d035ef8467ad305d6a7>, cited by: 2449; All Open Access, Green Open Access.
- [11] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D., “Image Segmentation Using Deep Learning: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 7, 2022, p. 3523 – 3542. <https://doi.org/10.1109/TPAMI.2021.3059968>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85100948197&doi=10.1109%2fTPAMI.2021.3059968&partnerID=40&md5=115a97279a77fb75cf00d20ca1578417>, cited by: 1384; All Open Access, Green Open Access.
- [12] Fayyad, J., Jaradat, M. A., Gruyer, D., and Najjaran, H., “Deep Learning Sensor Fusion for Autonomous Vehicle Perception and Localization: A Review,” *Sensors*, Vol. 20, No. 15, 2020. <https://doi.org/10.3390/s20154220>, URL <https://www.mdpi.com/1424-8220/20/15/4220>.
- [13] Song, J., Rondao, D., and Aouf, N., “Deep learning-based spacecraft relative navigation methods: A survey,” *Acta Astronautica*, Vol. 191, 2022, pp. 22–40. <https://doi.org/10.1016/j.actaastro.2021.10.025>, arXiv: 2108.08876 Publisher: Elsevier Ltd.
- [14] Kisantal, M., Sharma, S., Park, T. H., Izzo, D., Martens, M., and D’Amico, S., “Satellite Pose Estimation Challenge: Dataset, Competition Design, and Results,” *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 56, No. 5, 2020, p. 4083–4098. <https://doi.org/10.1109/taes.2020.2989063>, URL <http://dx.doi.org/10.1109/TAES.2020.2989063>.
- [15] Park, T. H., Märten, M., Jawaid, M., Wang, Z., Chen, B., Chin, T.-J., Izzo, D., and D’Amico, S., “Satellite Pose Estimation Competition 2021: Results and Analyses,” *Acta Astronautica*, Vol. 204, 2023, pp. 640–665. <https://doi.org/https://doi.org/10.1016/j.actaastro.2023.01.002>, URL <https://www.sciencedirect.com/science/article/pii/S0094576523000048>.
- [16] Gao, X.-S., Hou, X.-R., Tang, J., and Cheng, H.-F., “Complete solution classification for the perspective-three-point problem,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 8, 2003, pp. 930–943. <https://doi.org/10.1109/TPAMI.2003.1217599>.
- [17] Kaluthantrige, A., Feng, J., and Gil-Fernández, J., “CNN-based Image Processing algorithm for autonomous optical navigation of Hera mission to the binary asteroid Didymos,” *Acta Astronautica*, Vol. 211, 2023, pp. 60–75. <https://doi.org/https://doi.org/10.1016/j.actaastro.2023.05.029>, URL <https://www.sciencedirect.com/science/article/pii/S0094576523002709>.
- [18] Pugliatti, M., Franzese, V., and Topputo, F., “Data-Driven Image Processing for Onboard Optical Navigation Around a Binary Asteroid,” *Journal of Spacecraft and Rockets*, Vol. 59, No. 3, 2022, p. 943 – 959. <https://doi.org/10.2514/1.A35213>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85130852100&doi=10.2514%2f1.A35213&partnerID=40&md5=aee7fc316728c0cd1f590994a038ebcd>, cited by: 21; All Open Access, Green Open Access.
- [19] STAR-Dundee, “Planet and Asteroid Natural Scene Generation Utility,” , 2024. URL <https://pangu.software/>.
- [20] Pugliatti, M., Buonagura, C., and Topputo, F., “CORTO: The Celestial Object Rendering TOol at DART Lab,” *Sensors*, Vol. 23, No. 23, 2023. <https://doi.org/10.3390/s23239595>, URL <https://www.mdpi.com/1424-8220/23/23/9595>.
- [21] “Blender 4.0,” , 2023. URL <https://www.blender.org/>.

- [22] Knuutila, O., Kestila, A., and Kallio, E., “CNN-Based Local Features for Navigation Near an Asteroid,” *IEEE Access*, Vol. 12, 2024, pp. 16652–16672. <https://doi.org/10.1109/ACCESS.2024.3358021>, publisher: Institute of Electrical and Electronics Engineers Inc.
- [23] Driver, T., Skinner, K., Dor, M., and Tsiotras, P., “AstroVision: Towards Autonomous Feature Detection and Description for Missions to Small Bodies Using Deep Learning,” *Acta Astronautica*, Vol. 210, 2023, pp. 393–410. <https://doi.org/10.1016/j.actaastro.2023.01.009>, URL <http://arxiv.org/abs/2208.02053>, arXiv:2208.02053 [astro-ph].
- [24] Revaud, J., Weinzaepfel, P., de Souza, C. R., and Humenberger, M., “R2D2: Repeatable and Reliable Detector and Descriptor,” *NeurIPS*, 2019.
- [25] ESA, “NASA Planetary Data System (PDS),” , 2022. URL <https://pds.nasa.gov/>.
- [26] NASA, “ESA Planetary Science Archive (PSA),” , 2018. URL <https://psa.esa.int/psa/#/pages/home>.
- [27] DeTone, D., Malisiewicz, T., and Rabinovich, A., “SuperPoint: Self-Supervised Interest Point Detection and Description,” *CoRR*, Vol. abs/1712.07629, 2017. URL <http://arxiv.org/abs/1712.07629>.
- [28] Luo, Z., Zhou, L., Bai, X., Chen, H., Zhang, J., Yao, Y., Li, S., Fang, T., and Quan, L., “ASLFeat: Learning Local Features of Accurate Shape and Localization,” *CoRR*, Vol. abs/2003.10071, 2020. URL <https://arxiv.org/abs/2003.10071>.
- [29] Letizia, G., Politecnico, C., Milano, D., Piccinin, M., Di Milano, P., Civardi, G. L., Piccinin, M., and Lavagna, M., “Small Bodies IR Imaging for Relative Navigation and Mapping Enhancement SMALL BODIES IR IMAGING FOR VISION BASED RELATIVE NAVIGATION AND MAPPING ENHANCEMENT,” Tech. rep., 2021. URL <https://www.researchgate.net/publication/356641732>.
- [30] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G., “ORB: An efficient alternative to SIFT or SURF,” *2011 International Conference on Computer Vision*, 2011, pp. 2564–2571. <https://doi.org/10.1109/ICCV.2011.6126544>.
- [31] Bay, H., Tuytelaars, T., and Van Gool, L., “SURF: Speeded Up Robust Features,” *Computer Vision – ECCV 2006*, edited by A. Leonardis, H. Bischof, and A. Pinz, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 404–417.
- [32] Ma, J., Ma, Y., and Li, C., “Infrared and visible image fusion methods and applications: A survey,” *Information Fusion*, Vol. 45, 2019, pp. 153–178. <https://doi.org/10.1016/j.inffus.2018.02.004>.
- [33] Sun, K., Xiao, B., Liu, D., and Wang, J., “Deep high-resolution representation learning for human pose estimation,” 2019, p. 5686 – 5696. <https://doi.org/10.1109/CVPR.2019.00584>, cited by: 3086; All Open Access, Green Open Access.
- [34] Chen, B., Cao, J., Parra, A., and Chin, T., “Satellite Pose Estimation with Deep Landmark Regression and Nonlinear Pose Refinement,” *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, IEEE Computer Society, Los Alamitos, CA, USA, 2019, pp. 2816–2824. <https://doi.org/10.1109/ICCVW.2019.00343>, URL <https://doi.ieeecomputersociety.org/10.1109/ICCVW.2019.00343>.
- [35] NVIDIA, “NVIDIA L4 Tensor Core GPU,” , 2023. URL <https://www.nvidia.com/en-gb/data-center/l4/>.
- [36] MathWorks, “estworldpose,” , 2022. URL <https://uk.mathworks.com/help/vision/ref/estworldpose.html>.
- [37] Guthrie, B., Kim, M., Urrutxua, H., and Hare, J., “Image-based attitude determination of co-orbiting satellites using deep learning technologies,” *Aerospace Science and Technology*, Vol. 120, 2022, p. 107232. <https://doi.org/https://doi.org/10.1016/j.ast.2021.107232>, URL <https://www.sciencedirect.com/science/article/pii/S1270963821007422>.
- [38] MathWorks, “MATLAB Partial Differential Equation Toolbox,” , 2021. URL <https://uk.mathworks.com/products/pde.html>.
- [39] Rozitis, B., Green, S. F., Jackson, S. L., Snodgrass, C., Opitom, C., Müller, T. G., Kolb, U. C., Chesley, S. R., Daly, R. T., Thomas, C. A., and Rivkin, A. S., “Pre-impact Thermophysical Properties and the Yarkovsky Effect of NASA DART Target (65803) Didymos,” *The Planetary Science Journal*, Vol. 5, No. 3, 2024, p. 66. <https://doi.org/10.3847/PSJ/ad23eb>, URL <https://iopscience.iop.org/article/10.3847/PSJ/ad23eb>.